

Albrecht Iseler, Statistik II, Fassung SS 2005

Vorbemerkung:

Einstweilen sind aus technischen Gründen noch einige Details zu korrigieren, insbesondere die folgenden:

- Die Seiten, Abschnitte etc., auf die sich Querverweise beziehen, können sich bei der Überarbeitung geändert haben, was aber möglicherweise noch nicht überall auch in dem Querverweis berücksichtigt wurde. Insbesondere kennzeichnet das Zeichen @ zu überprüfende oder zu ergänzende Querverweise. Es kann aber sein, daß noch weitere Querverweise korrigiert werden müssen.
- Damit nicht jedes Semester das ganze Skript für wenige Korrekturen neu ausgedruckt werden muß, stehen Korrekturhinweise hinter dem Inhaltsverzeichnis (auf S. 350). Bitte übertragen!
- << - Einige Abschnitte "für Spezialisten" sind - wie dieser Hinweis - durch doppelte spitze Klammern auf dem Linksrand markiert. Außerdem sind solche Abschnitte mit geringerem Zeilenabstand gedruckt. Das ist aber noch nicht durchgehend erfolgt. Stattdessen gilt: "Obligatorisch" ist, was in den Merkposten steht, soweit nicht in der Vorlesung anders angegeben.

>>

3 Seitenzählungen: Hauptteil ohne Buchstaben (Inhaltsverzeichnis S. 345), dann M1, M2, ... (Merkposten) T1, T2 ... (Tabellen).

B. Wahrscheinlichkeitstheorie

Die Wahrscheinlichkeitstheorie ist für die Psychologie in verschiedener Weise von Bedeutung. In dieser Einleitung soll auf drei Anwendungsformen eingegangen werden: Wahrscheinlichkeitsaussagen über das Erleben und Verhalten von Menschen, ferner die sog. Inferenzstatistik und schließlich eine besondere Art der Modellierung von Kognitionen und Entscheidungen.

Bei den *Wahrscheinlichkeitsaussagen über das Erleben und Verhalten von Menschen* sind vor allem zwei Arten mit jeweils zwei Unterformen zu unterscheiden: Allgemeine Gesetzaussagen und Aussagen über einzelne Menschen in der Diagnostik. Ein Beispiel für eine erste Unterform von Gesetzaussagen in der Sprache von Wahrscheinlichkeiten wäre etwa die folgende Darstellung der Wirkung von positiver Verstärkung ("Belohnung"): Wenn ein Mensch in einer Situation für ein Verhalten "belohnt" worden ist, dann ist die Wahrscheinlichkeit, daß dieses Verhalten in ähnlichen Situationen wieder auftritt, größer, als wenn er nicht belohnt worden ist. Allgemein liegt bei dieser ersten Unterform von Gesetzaussagen das folgende Schema vor: Unter einer Bedingung a ist die Wahrscheinlichkeit eines bestimmten Erlebens oder Verhaltens (oder hoher Meßwerte auf einer bestimmten Skala) größer als unter Bedingung b .

<<

Manchmal werden auch mehr als zwei Abstufungen einer Bedingung unterschieden, so daß Je-desto-Formulierungen der folgenden Art entstehen: Je ausgeprägter Bedingung a , desto größer ist die Wahrscheinlichkeit des interessierenden Erlebens oder Verhaltens. Zu betonen ist, daß die in solchen Aussagen auftretenden Bedingungen (z.B. a und b oder abgestufte Ausprägungen von a) sowohl in der Situation als auch im einzelnen Menschen liegen können. Letzteres ist z.B. der Fall, wenn die Wahrscheinlichkeit von Erlebens- oder Verhaltensweisen in Bezug zu Persönlichkeitsmerkmalen gesetzt wird.

>>

Da diese erste Unterform von Gesetzesaussagen sich nur auf die Größer-kleiner-Beziehung von Wahrscheinlichkeiten beziehen, kann man auch von *Gesetzesaussagen mit ordinalen Wahrscheinlichkeitsvergleichen* sprechen, um diese Unterform zu kennzeichnen.

Es gibt auch Versuche, über solche ordinalen Vergleiche hinauszugehen und *mit präzisen Formeln zu beschreiben, wie die Wahrscheinlichkeit bestimmter Erlebens- und Verhaltensweisen von Bedingungen abhängt*.

<<

Das Verständnis eines Beispiels setzt Grundkenntnisse in Wahrscheinlichkeitstheorie voraus. Das folgende Beispiel ist deshalb möglicherweise auch "für Spezialisten" erst nachvollziehbar, nachdem Abschnitt @B.II.1 bearbeitet ist. Im Moment kommt es aber auch nur darauf an, den Unterschied zwischen ordinalen Wahrscheinlichkeitsvergleichen und Wahrscheinlichkeitsangaben in präziseren Formeln zu demonstrieren.

Die Vpn in einem Experiment sehen auf einem Bildschirm Gegenstände, die sie sich einprägen sollen, z.B. Hammer, Zange, Schraubenzieher, Gabel, Löffel, Tafelmesser, Nadel, Fingerhut und Garnrolle. Nach einer Pause (in der sie mit einer anderen Aufgabe beschäftigt waren) sollen sie sagen, um welche Gegenstände es sich handelte. Dabei spielt dann die Zugehörigkeit der Gegenstände zu den drei Kategorien Werkzeug, Besteck und Nähzeug eine Rolle: Auch wenn die Gegenstände auf dem Bildschirm nicht nach diesen Kategorien sortiert waren, werden beim Reproduzieren gehäuft Gegenstände derselben Kategorie unmittelbar hintereinander genannt. Diese meist als "clustering" (wörtlich: Traubenbildung) bezeichnete Tatsache gilt als Beleg für die Beteiligung von Begriffs-Strukturen an einfachen Gedächtnisleistungen. In der Sprache ordinaler Wahrscheinlichkeitsvergleiche könnte man dieses clustering folgendermaßen formulieren: Die Wahrscheinlichkeit, daß ein bestimmter Gegenstand als nächster genannt wird, ist größer, wenn unmittelbar vorher ein Gegenstand derselben Kategorie genannt wurde, als wenn der zuletzt genannte Gegenstand einer anderen Kategorie angehörte. Eine exaktere Angabe von Wahrscheinlichkeiten könnte beispielsweise folgendermaßen aussehen: Sind die ersten j genannten Gegenstände bekannt (wobei j kleiner als die Gesamtzahl der Gegenstände ist), dann beträgt die Wahrscheinlichkeit, daß als nächstes ein Gegenstand der Kategorie k auftritt,

$$c \cdot u_{jk} / (c \cdot u_{jk} + n - j - u_{jk}),$$

falls der zuletzt genannte Gegenstand zur Kategorie k gehört. Anderenfalls beträgt diese Wahrscheinlichkeit

$$u_{jk} / (c \cdot v_j + n - j - v_j).$$

Dabei ist n die Gesamtzahl der gezeigten Gegenstände und u_{jk} die Zahl der Gegenstände, die zur Kategorie k gehören und sich nicht unter den bereits genannten j Gegenständen befinden. In der zweiten Formel ist v_j die Zahl derjenigen noch nicht genannten Gegenstände, die zur gleichen Kategorie gehören, wie der zuletzt genannte. Schließlich ist c eine Zahl, die angibt, um

welchen Faktor die Nennungs-Wahrscheinlichkeit der zu einer Kategorie gehörenden Gegenstände erhöht wird, wenn zuvor ein Gegenstand dieser Kategorie genannt wurde. Diese Zahl drückt also aus, wie stark die Tendenz zum clustering ist. Daß in derartigen Formeln Zahlen auftauchen, deren Wert aus den Daten geschätzt werden muß, ist typisch für die hier demonstrierte Art der Formulierung von Gesetzesaussagen in Wahrscheinlichkeitssprache. Solche Zahlen werden auch als "Parameter" bezeichnet. Unser c könnte man also als den "clustering-Parameter" bezeichnen.

Derartige Formeln kann man nicht wahrscheinlichkeitstheoretisch beweisen. Man kann aber mit Wahrscheinlichkeitstheorie überprüfen, ob die in einer empirischen Untersuchung erzielten Ergebnisse mit der Annahme vereinbar sind, daß diese Formeln zutreffen.

Der wesentliche Schritt, durch den dieses Modell über ordinale Wahrscheinlichkeitsvergleiche hinausgeht, besteht in der Annahme, daß die Erhöhung der Auftretenswahrscheinlichkeit eines Gegenstandes durch das vorherige Erinnern eines Gegenstandes derselben Kategorie als Multiplikation mit einer für alle Gegenstände identischen Konstanten - dem clustering-Parameter c - dargestellt wird.

Typisch für derartige Modelle ist aber auch, daß meistens noch weitere vereinfachende Annahmen erforderlich sind, um zu exakten Formeln zu kommen. Bei den obigen Formeln wird z.B. nicht berücksichtigt, daß verschiedene Gegenstände unterschiedlich leicht zu merken sind bzw. unterschiedlich leicht wieder einfallen. Mit komplexeren Annahmen könnte man auf diese Vereinfachung verzichten; aber das zu demonstrieren, würde den hiesigen Rahmen sprengen.

>>

Bei den Wahrscheinlichkeitsaussagen über einzelne Personen in der Diagnostik kann man als Unterform die Diagnostik i.e.S von Prognosen unterscheiden. Bei der *Diagnostik i.e.S.* geht es dabei um die Wahrscheinlichkeit, daß ein "latenter" (d.h. nicht beobachtbarer) Zustand vorliegt, wenn bestimmte beobachtbare Tatsachen bekannt sind. Bei "hypermotorischen" Kindern (beobachtbare Tatsache) liegt z.B. häufig eine geringfügige Hirnschädigung (latenter Zustand) vor, und für die Diagnostik ist es hilfreich, zu wissen, wie groß die Wahrscheinlichkeit einer solchen geringfügigen Hirnschädigung ist, wenn ein Kind hypermotorisch ist. Bei *Prognosen* geht es dagegen um die Vorhersage von Zukünftigem aus Bekanntem. Ein Psychologe im Strafvollzug könnte z.B. gefragt werden, wie groß er die Wahrscheinlichkeit einschätzt, daß ein bestimmter, ihm bekannter Häftling eine zur Diskussion stehende Hafterleichterung (z.B. Freigang) mißbraucht, und für einen Therapeuten ist es wichtig, wie groß die Wahrscheinlichkeit einer deutlichen Besserung im Sinne der Therapieziele seines Klienten bei verschiedenen möglichen therapeutischen Interventionen ist. In beiden Fällen handelt es sich um Prognosen unter der Annahme bestimmter Interventionen ("Maßnahmen"), und diese sind natürlich auch von besonderer Bedeutung bei der Entscheidung über die bei einer bestimmten Person günstigsten Maßnahmen.

Die hier aufgeführten Formen von Wahrscheinlichkeitsaussagen über das Erleben und Verhalten von Menschen haben eines gemeinsam: Sie bringen ein Wahrscheinlichkeitsdenken ausdrücklich zur Sprache, das bei vielen Psychologen auf einer intuitiven Ebene unausgesprochen bleibt. Es ist auch unter Methodikern umstritten, ob es überhaupt Ziel der Psychologie sein soll, zu präziseren Wahrscheinlichkeitsaussagen zu kommen, oder ob man nicht eher suchen soll, aus welchen Bedingungen sich das Erleben und Verhalten nicht nur mit

Wahrscheinlichkeiten, sondern mit Sicherheit ("deterministisch" statt probabilistisch) vorhersagen läßt.

Dagegen ist eine andere Form der Anwendung von Wahrscheinlichkeitstheorie in der Psychologie sehr verbreitet: Die *Inferenzstatistik*, bei der es um die Frage geht, welche Schlußfolgerungen aus den Daten einer Stichprobe gezogen werden können. Ein einfaches, aber typisches Beispiel wurde bereits in der Einleitung von Statistik I behandelt. In einem Experiment aus der Therapieforschung erhielten die Personen einer Experimentalgruppe eine positive Vorinformation über den Therapeuten, während die Personen der Kontrollgruppe keine Vorinformation erhielten. Es zeichnete sich eine Tendenz ab, daß Personen der Experimentalgruppe in der fünften Therapiesitzung höhere Werte auf einer Skala der Selbstexploration erzielten als die Personen der Kontrollgruppe; aber bevor solche Ergebnisse über die Stichprobe hinaus verallgemeinert werden, hat man sich zu fragen, mit welcher Wahrscheinlichkeit eine solche (oder noch ausgeprägtere) Tendenz rein zufällig entstehen kann, z.B. dadurch, daß bei der Aufteilung der Gruppen zufällig mehr Personen mit hoher "mitgebrachter" Fähigkeit und Bereitschaft zur Selbstexploration in die Experimentalgruppe gekommen sind als in die Kontrollgruppe. Wenn diese Wahrscheinlichkeit beträchtlich ist, dann wäre eine über die Stichprobe hinausgehende Verallgemeinerung voreilig.

Man kann geteilter Meinung sein, inwieweit die Rolle, die diese Denkweise in der Psychologie spielt, dem Gegenstand adäquat ist. Noch mehr gilt das für die Begründungen. Auch in renommierten wissenschaftlichen Zeitschriften findet man manchmal Artikel, in denen entsprechende Berechnungen zwar korrekt durchgeführt sind, bei denen aber die Interpretation der Ergebnisse Zweifel aufkommen läßt, ob den Verfassern bewußt ist, warum solche Berechnungen erforderlich sind, und dann wird die Einhaltung der "in der Wissenschaft üblichen Regeln" tatsächlich zu einem unverstandenen Ritual.

Diese Gefahr ist noch größer geworden, seit es leicht verfügbare Computerprogramme gibt, bei denen man mit ein paar Mausklicks Berechnungen anfordern kann, deren Sinn man überhaupt nicht verstanden hat. In dieser Situation ist es weniger wichtig geworden, möglichst viele Rechentechniken kennenzulernen, die einem der Computer abnehmen kann. Umso wichtiger ist es dagegen, die wahrscheinlichkeitstheoretischen und logischen Grundlagen der schlußfolgernden Statistik zu kennen. Das ist nicht nur nötig, um selbst Daten sinnvoll auswerten und interpretieren zu können (z.B. im Empirischen Praktikum oder in der Diplomarbeit). Auch eine kritische Lektüre publizierter Untersuchungen ist nur möglich, wenn man mit diesen Grundlagen vertraut ist; denn nur dann kann man beurteilen, wo möglicherweise ein Feuerwerk von oberflächlich eindrucksvollen Berechnungen nichts als Blendwerk ist.

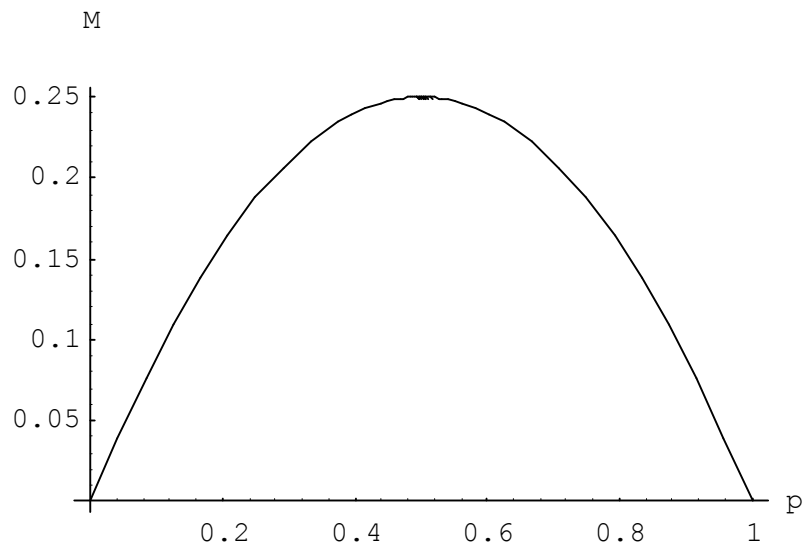
Von einigen Kritikern wird behauptet, daß die Anwendung der Wahrscheinlichkeitstheorie in der Psychologie typisch für eine "Sicht von außen" ist, die mit einer "subjektwissenschaftlichen" Herangehensweise unvereinbar sein soll. Deshalb soll noch eine weitere Form der

Anwendung von Wahrscheinlichkeitstheorie in der Psychologie an einem Beispiel demonstriert werden, bei dem deutlich die Perspektive des Subjekts eingenommen wird. Es handelt sich dabei um eine geringfügig vereinfachte Variante der Theorie der Leistungsmotivation von Atkinson (1964). Es geht um die Frage, wie motivierend eine bestimmte Aufgabe für eine Person ist. Ein typisches, intensiver untersuchtes Beispiel für die Aufgabe ist es, einen Ring über einen Stab zu werfen, wobei die Schwierigkeit natürlich mit der Entfernung des Stabs zunimmt. Der psychologische Gehalt der Theorie läßt sich etwa folgendermaßen formulieren: Steht der Stab sehr nahe, dann ist ein Erfolg zwar sehr wahrscheinlich, aber er ist auch wenig wert. Steht der Stab dagegen weit weg, dann wäre ein Erfolg zwar viel wert, aber er ist auch sehr unwahrscheinlich. In beiden Fällen ist die Aufgabe also wenig motivierend. Motivierend ist die Aufgabe dagegen bei einer mittleren Erfolgswahrscheinlichkeit.

Die Subjekt-Perspektive besteht dabei zunächst darin, daß die Erfolgswahrscheinlichkeit als der Grad der Gewißheit gesehen wird, mit dem die Person einen Erfolg erwartet. (In Abschnitt @B.I.1.a wird für so verstandene Wahrscheinlichkeiten die Bezeichnung "subjektive Wahrscheinlichkeit" eingeführt.) Bezeichnet man diese Wahrscheinlichkeit als p , dann stehen $p=0$ und $p=1$ für die Extremfälle "ich bin sicher, daß ich es nicht schaffen werden" und "ich bin sicher, daß ich es schaffen werde". Aber auch der "Wert" eines Erfolgs wird in dieser Theorie aus der Perspektive des Subjekts formuliert: Es wird angenommen, daß ein Erfolg - falls er eintritt - für die Person umso mehr wert ist, je geringer sie die Erfolgswahrscheinlichkeit einschätzt. Etwas vereinfachend könnte man die Differenz $W = 1 - p$ als Maß für den subjektiven Wert eines Erfolgs ansehen. Das ist im wesentlichen auch die Annahme Atkinson's. Nimmt man weiter an, daß das Produkt aus Erfolgswahrscheinlichkeit und Wert des Erfolgs die Motivationsstärke M ergibt, dann folgt daraus die Beziehung

$$M = p \cdot W = p \cdot (1 - p),$$

die in der folgenden Abbildung graphisch dargestellt wird:



Zusammenhang von subjektiver Erfolgswahrscheinlichkeit p und Motivationsstärke M nach der Gleichung

$$M = p \cdot (1 - p)$$

(vereinfachte Version des Modells von Atkinson, 1964).

Was aus dieser Abbildung vor allem deutlich wird, ist die Übereinstimmung mit dem Ergebnis der vorangehenden Überlegung, daß die Motivationsstärke bei einer mittleren subjektiven Erfolgswahrscheinlichkeit am größten ist.

<<

Wie erwähnt, handelt es sich hier um eine Vereinfachung des Modells von Atkinson (1964), für den noch weitere Variablen eine Rolle spielen.

Ohne Atkinson in diesem Zusammenhang ausdrücklich zu nennen, weist Holzkamp (1993, S. 85) darauf hin, daß früher Vertreter solcher Ansätze versucht haben, das "Erwartungskonzept" nicht subjektiv, sondern "explanativ" zu sehen. Historisch ist das zutreffend: Solche Versuche stammen aus einer Zeit, als die Beschäftigung mit "Subjektivem" als unwissenschaftlich galt. Umsomehr überrascht es, daß gerade Holzkamp, der in anderen Zusammenhängen die historische Bedingtheit psychologischer Theorien hervorhebt, die Distanzierung vom Vorwurf der Subjektivität als notwendige "Konsequenz dieser Mathematisierung des Erwartungskonzepts" interpretiert. Warum Holzkamp das alte Vorurteil von der Unvereinbarkeit von Mathematik und Hinwendung zum Subjekt aufrechterhält, kann hier natürlich nicht ausführlich diskutiert werden. Jedenfalls gibt es inzwischen eine Vielzahl von Ansätzen, "subjektive Wahrscheinlichkeiten" und "erwarteten subjektiven Nutzen" in einer psychologischen Entscheidungstheorie zu verknüpfen (vgl. z.B. Stephan, 1989 für eine Systematisierung solcher Ansätze).

>>

I. Die Begriffe "Wahrscheinlichkeit" und "Zufall"

1) *Wahrscheinlichkeit*

a) Wahrscheinlichkeitsbegriffe

Wenn wir eine gleichmäßige¹ Münze werfen, so beträgt die Wahrscheinlichkeit, daß der "Adler" oben liegen bleibt, 50%.

Was bedeutet die numerische Angabe 50%?

Es gibt auf diese Frage verschiedene richtige Antworten, die auf verschiedene Definitionen des Wahrscheinlichkeitsbegriffs hinauslaufen.

Wenn wir mehrere Versuche durchführen, so erwarten wir, daß die prozentuale Häufigkeit des Ereignisses "Adler" 50% ist. Die in Prozent angegebene Wahrscheinlichkeit wäre also die bei einer längeren Versuchsreihe zu erwartende prozentuale Häufigkeit des Ereignisses, dessen Wahrscheinlichkeit angegeben wird. In der Wahrscheinlichkeitstheorie gibt man Wahrscheinlichkeiten jedoch meistens nicht in Prozent, sondern als zu erwartende relative Häufigkeiten an. Man würde also beispielsweise die Wahrscheinlichkeit des Ereignisses "Adler" beim Münzwurf als 0.50 angeben.

Man könnte die Angabe der Wahrscheinlichkeit des Ereignisses "Adler" als 50% (oder 0.50) aber auch anders interpretieren: Da die Münze gleichmäßig ist, sind "Adler" und "Zahl" gleich wahrscheinlich; d.h. die Gewißheit, mit der wir "Adler" und "Zahl" erwarten, ist gleich groß. Da sich beide Wahrscheinlichkeiten offenbar² zu 100% (bzw. zu 1) ergänzen müssen, ist jede einzelne Wahrscheinlichkeit 50% (bzw. 0.5). Die Wahrscheinlichkeit eines Ereignisses wäre dann eine numerische Angabe über das Ausmaß der Gewißheit, mit der man das Eintreffen des Ereignisses erwartet.

Damit ergibt sich ein wesentlicher Unterschied zwischen den beiden Wahrscheinlichkeitsbegriffen: Die in einer längeren Versuchsreihe zu erwartende prozentuale oder relative Häufigkeit wird meist als eine Eigenschaft des Betrachtungs-Gegenstands (hier: der Münze) gesehen; dagegen ist das Ausmaß der Gewißheit, mit der man das Eintreffen eines Ereignisses erwartet, eine Aussage über das Subjekt, das den Vorgang betrachtet. Daher spricht man im ersten Fall von objektiven, im zweiten Fall von subjektiven Wahrscheinlichkeiten. Drei weitere Bezeichnungen für die objektive Wahrscheinlichkeit sind ebenfalls gebräuchlich. Die Bezeichnung "frequentistischer" Wahrscheinlichkeitsbegriff bedarf keiner weiteren Erläuterung. Die relative Häufigkeit, mit der z.B. eine Münze mit dem Adler oder ein Würfel mit der sechs nach oben liegen bleibt, kann als die "Disposition" der Münze bzw. des Würfels zur

¹d.h. eine Münze, deren Konstruktion zu keiner Bevorzugung von "Adler" oder "Zahl" führt.

²Wir werden dieses "offenbar" in einem späteren Abschnitt noch expliziter diskutieren.

Realisierung des entsprechenden Ereignisses aufgefaßt werden. Daher spricht man auch von einem "dispositionellen" Wahrscheinlichkeitsbegriff. Anders formuliert, geht es darum, wie stark die Münze bzw. der Würfel dazu "neigt", mit dem Adler oder der sechs nach oben liegen zu bleiben; darum spricht man auch von einer propensity-Interpretation der Wahrscheinlichkeit. (Die wörtliche Übersetzung von "propensity" wäre "Geneigtheit"; sie ist in diesem Zusammenhang aber ungebräuchlich. Allenfalls spricht man von der "Propensität".)

Neben dem objektiven und dem subjektiven gibt es noch einen abstrakten Wahrscheinlichkeitsbegriff, mit dem die mathematische Wahrscheinlichkeitstheorie heute arbeitet. Nach diesem Begriff ist eine Wahrscheinlichkeit eine Zahl zwischen 0 und 1, die einem Ereignis zugeordnet ist und die bestimmte Axiome erfüllt, auf die wir noch eingehen werden. Die Angabe der Wahrscheinlichkeit des Ereignisses "Adler" als 0.50 bedeutet für den Mathematiker also lediglich, daß dem Ereignis "Adler" die Zahl 0.50 zugeordnet ist.

Man könnte diese Zahl auch ein Maß nennen, das einem Ereignis zugeordnet ist und bestimmte Axiome erfüllt; denn ein Maß ist ja eine Zahl, die einem Ereignis (oder einem Objekt) zugeordnet ist. Entsprechend ist die mathematisch axiomatische Wahrscheinlichkeitstheorie ein spezielles Anwendungsgebiet der mathematischen Maßtheorie.

b) Gegenseitige Abwägung der drei Wahrscheinlichkeitsbegriffe

Der abstrakte mathematische Wahrscheinlichkeitsbegriff hat den Vorteil, wesentlich genereller anwendbar zu sein, als die beiden anderen Wahrscheinlichkeitsbegriffe. Hinsichtlich der "objektiven Wahrscheinlichkeit" ergeben sich gerade in der Psychologie logische Schwierigkeiten. Wenn wir beispielsweise sagen, die Wahrscheinlichkeit, daß eine Vp eine Aufgabe aus einem Intelligenztest in 30 Sekunden löst, sei 0.70, dann kann dies kaum bedeuten, daß die Vp die Aufgabe bei 1000-maliger Vorlage etwa 700 mal lösen und 300 mal nicht lösen wird. Bereits durch die erste Bearbeitung der Aufgabe ändert sich die Wahrscheinlichkeit der Lösung in späteren Aufgaben. Die Wahrscheinlichkeit von 0.70 gilt also nur für die erste Bearbeitung. Man könnte zwar als Ausweg sagen, die Wahrscheinlichkeitsangabe 0.70 bedeute, daß bei Vorlage der Aufgabe an 1000 Vpn, die die gleiche Intelligenz haben wie unsere Vp, etwa 700 die Aufgabe in 30 Sekunden lösen würden; aber dann müßten wir erst einmal definieren, was gleiche Intelligenz ist. Es wäre problematisch, zu definieren, daß gleiche Intelligenz zweier Personen dann vorliegt, wenn die Wahrscheinlichkeit der Lösung einer Aufgabe gleich ist; denn dann würden wir uns leicht im Kreise drehen: Die Definition der Wahrscheinlichkeit würde den Begriff gleicher Intelligenz voraussetzen, und die Definition gleicher Intelligenz den der Wahrscheinlichkeit.

Allgemein: Der Begriff der "objektiven" Wahrscheinlichkeit" also der Wahrscheinlichkeit als zu erwartender relativer Häufigkeit eines Ereignisses bei mehrfacher Wiederholung eines Versuchs, stößt in der Psychologie auf Schwierigkeiten, da die von diesem Begriff vorausgesetzte Wiederholbarkeit von Versuchen nicht immer gegeben ist.

Aber auch der Begriff der "subjektiven Wahrscheinlichkeit" hat seine Problematik in der Psychologie. Bei einer gleichmäßigen Münze oder einem gleichmäßigen (also ungefälschtem) Würfel ist es leicht, den Grad der subjektiven Gewißheit anzugeben, mit der wir das Ereignis "Adler" oder "6 Augen" erwarten. Ein Mensch ist aber kein so einfach durchschaubares Objekt wie eine gleichmäßige Münze oder ein gleichmäßiger Würfel, und daher wäre die numerische Angabe der subjektiven Gewißheit, daß ein Mensch in einer Situation eine gewisse Reaktion zeigen wird, keineswegs leicht - wenn überhaupt möglich.

Ein Nachteil des mathematischen Wahrscheinlichkeitsbegriffs ist seine Abstraktheit. Was nützt es uns, wenn wir dem Ereignis "Patient wird durch die Therapie geheilt" die Zahl 0.70 zuordnen können und wissen, daß bestimmte Axiome erfüllt sind, aber wegen der Abstraktheit des Begriffs nicht sagen können, was diese Wahrscheinlichkeit dann noch bedeutet?

In der Abstraktheit des mathematischen Wahrscheinlichkeitsbegriffs ist aber auch ein entscheidender Vorteil begründet. Es läßt sich nämlich leicht aufzeigen, daß die beiden "konkreten" Wahrscheinlichkeitsbegriffe (also der objektive und der subjektive) als Konkretisierungen des mathematisch abstrakten Wahrscheinlichkeitsbegriffs aufgebaut werden können. Beispielsweise ist jede relative Häufigkeit auch eine Zahl zwischen 0 und 1, und wir werden noch sehen, daß auch die übrigen Axiome des mathematischen Wahrscheinlichkeitsbegriffs erfüllt sind. Folglich lassen sich alle aus dem mathematischen Wahrscheinlichkeitsbegriff und seinen Axiomen abgeleiteten Lehrsätze auch auf objektive Wahrscheinlichkeiten übertragen. Ähnlich sind auch subjektive Wahrscheinlichkeiten meist so definiert, daß die Axiome der mathematischen Wahrscheinlichkeitstheorie erfüllt sind.

Für den Mathematiker ergibt sich daher kaum das Dilemma der Wahl zwischen verschiedenen Wahrscheinlichkeitsbegriffen. Er verwendet den abstrakten Begriff und überläßt die Frage nach möglichen Konkretisierungen des Begriffs und der daraus abgeleiteten Lehrsätze dem Fachwissenschaftler, der die Wahrscheinlichkeitstheorie anwendet. Er sagt ihm also: Du kannst dir unter einer Wahrscheinlichkeit vorstellen, was du willst. Solange deine Wahrscheinlichkeiten Zahlen zwischen 0 und 1 sind und meine übrigen Axiome erfüllt sind, gelten für deine Wahrscheinlichkeiten alle Lehrsätze, die ich aus meinem Begriff ableite.

Wir wollen in unserem Kurs eine ähnliche, allerdings etwas modifizierte Position einnehmen. Wir wollen solche Ergebnisse der Wahrscheinlichkeitstheorie behandeln, die für jeden Wahrscheinlichkeitsbegriff gültig sind, der die mathematischen Axiome erfüllt. Zur Vereinfachung des Verständnisses wollen wir die Lehrsätze jedoch an konkreten Beispielen herleiten. Daß die am konkreten Beispiel hergeleiteten Lehrsätze und Begriffe auch sinnvoll sind, wenn wir sie auf andere Beispiele anwenden, können wir dem Mathematiker glauben.³ Die Ergebnisse einer anderen Wissenschaft glauben kann man wohl bei kaum einer anderen Wissenschaft so gut wie bei der Mathematik.

³Natürlich können wir es auch in Lehrbüchern der mathematischen Statistik überprüfen, in denen die Lehrsätze unabhängig von allen Konkretisierungen hergeleitet werden.

Probleme ergeben sich dagegen in der Frage der Anwendbarkeit auf die Psychologie. Der Mathematiker kann nur beweisen, daß unter bestimmten Voraussetzungen bestimmte Schlußfolgerungen gelten. Die Anwendbarkeit der Ergebnisse des Mathematikers hängt davon ab, ob die von ihm gemachten Voraussetzungen in dem Bereich, in dem man die Mathematik anwendet, als erfüllt betrachtet werden können. Die Diskussion dieser Frage im Hinblick auf die Psychologie wollen wir auf einen späteren Zeitpunkt verschieben.

2) Der Begriff des Zufalls

Zufall ist ein Sammelbegriff für alle die Einflüsse, die auf ein Objekt oder ein Ereignis einwirken, ohne daß wir ihre Wirkungsweise im einzelnen verfolgen. Ob und inwieweit ein bestimmtes Ereignis auf Zufall zurückgeht, hängt also nicht nur von den objektiven Gegebenheiten ab, sondern auch von der Sichtweise, die wir an das Objekt herantragen.

3) Modelle zur Veranschaulichung von Wahrscheinlichkeit und Zufall

Wir haben uns vorgenommen, die Begriffe und Lehrsätze der Wahrscheinlichkeitstheorie nicht (oder jedenfalls nicht nur) in abstrakter Form, sondern in Konkretisierungen zu erarbeiten. Es wäre natürlich günstig, wenn wir hierfür Beispiele aus unserem Fachgebiet wählen würden. Das hätte jedoch den Nachteil, daß alle unsere Beispiele mit der bisher unbeantworteten Frage nach der Anwendbarkeit der Wahrscheinlichkeitstheorie in der Psychologie behaftet wären. Da das ständige Mitschleppen dieser Frage unsere Arbeit erheblich verkomplizieren würde, wollen wir die allgemeine Wahrscheinlichkeitstheorie anhand von einfach zu durchschauenden Modellen wie dem Münzwurf oder dem Würfel erörtern. Wenn wir - wie bereits angekündigt - in einem späteren Abschnitt die Frage der Anwendbarkeit der Wahrscheinlichkeitstheorie in der Psychologie diskutieren, dann wird diese Frage auch lauten müssen: Gibt es zumindest Teilaspekte des menschlichen Erlebens und Verhaltens, die sich mit ähnlichen Methoden mathematisch analysieren lassen wie Münzen und Würfel?

Neben diesen beiden Modellen für Zufallsprozesse wollen wir zwei andere verwenden, die ebenfalls in der Statistik eingeführt sind: Das Urnen-Modell und das Venn-Diagramm.

a) Das Urnen-Modell

Beim Urnen-Modell wird davon ausgegangen, daß sich in einer Urne verschiedene Kugeln befinden, die sich in verschiedenen Eigenschaften (z.B. Farbe, Material, aufgedruckte Symbole usw.) unterscheiden. Die Kugeln werden gut gemischt, und dann wird (bzw. werden) mit geschlossenen Augen eine Kugel (oder mehrere) herausgezogen. Soweit nicht ausdrücklich etwas anderes gesagt wird, wird beim Urnenmodell von der Annahme ausgegangen, daß alle

Kugeln in ihren physikalischen Eigenschaften soweit aufeinander abgestimmt sind, daß jede einzelne Kugel mit gleicher Wahrscheinlichkeit gezogen werden kann. Es wird also beispielsweise vorausgesetzt, daß nicht einige Kugeln leichter sind als andere - denn dann würden sich ja beim Schütteln die schwereren Kugeln unten in der Urne sammeln, und eine leichte Kugel würde mit größerer Wahrscheinlichkeit gezogen werden als eine schwere.

Die Wahrscheinlichkeit, eine Kugel mit bestimmten Eigenschaften zu ziehen, hängt natürlich von der Zusammensetzung des Inhalts der Urne ab. Enthält die Urne beispielsweise 70 rote und 130 schwarze Kugeln, dann ist die Wahrscheinlichkeit, eine rote Kugel zu ziehen, gleich $70/200$ oder 0.35 .

Können Sie das begründen?

Da jede Kugel die gleiche Wahrscheinlichkeit hat und diese Wahrscheinlichkeit über alle 200 Kugeln summiert offenbar⁴ 1 ergeben müssen, wird jede einzelne Kugel mit einer Wahrscheinlichkeit von $1/200$ gezogen. Die Wahrscheinlichkeit, irgendeine der 70 roten Kugeln zu ziehen, erhalten wir offenbar, indem wir die Wahrscheinlichkeiten aller roten Kugeln addieren.

Verallgemeinert: Befinden sich in einer Urne n Kugeln und ist f_A die Zahl der Kugeln, die zu einer Klasse A gehören, dann gilt für jede einzelne Kugel die Ziehungs-Wahrscheinlichkeit $1/n$, und die Wahrscheinlichkeit, eine Kugel der Klasse A zu ziehen ist dann

$$\frac{f_A}{n}$$

Um dies als Gleichung zu formulieren, verwenden wir den in der Wahrscheinlichkeitstheorie üblichen Buchstaben p als Symbol für "Wahrscheinlichkeit"⁵. Wir können das Ereignis, dessen Wahrscheinlichkeit wir angeben, durch einen Index kennzeichnen. Z.B. könnten wir für "Wahrscheinlichkeit, eine rote Kugel zu ziehen" oder "Wahrscheinlichkeit, eine Kugel der Klasse A zu ziehen" auch p_r bzw. p_A schreiben. Dann ergibt sich in symbolischer Schreibweise für unsere bisherigen Überlegungen:

$$p_r = \frac{f_r}{n} = \frac{70}{200}$$

und allgemein

$$p_A = \frac{f_A}{n}$$

⁴An die Stelle des hier und im folgenden Satz gebrauchten Wortes "offenbar" kann man in der mathematischen Statistik auch einen Beweis setzen, den wir später kennenlernen werden.

⁵Für englisch "probability" = Wahrscheinlichkeit. In deutschsprachigen Büchern wird stattdessen manchmal der Buchstabe W verwendet.

Wir können aber auch eine andere Schreibweise verwenden, bei der man das Ereignis, dessen Wahrscheinlichkeit angegeben wird, in Klammern hinter den Buchstaben p setzt. Häufig verwendet man für das Ereignis ein Symbol, das natürlich vorher definiert sein muß. In unserem Fall könnten wir definieren: Das Ereignis A liegt dann vor, wenn eine Kugel der Klasse A gezogen wird. Nach dieser Definition können wir schreiben:

$$p(A) = \frac{f_A}{n}$$

Werden aus einer Urne mehrere Kugeln gezogen, so muß man zwei "Spielregeln" unterscheiden: Das "Ziehen mit Zurücklegen" und das "Ziehen ohne Zurücklegen". Beim "Ziehen mit Zurücklegen" wird immer eine einzelne Kugel gezogen, ihre Eigenschaften werden festgestellt. Dann wird die Kugel in die Urne zurückgelegt, die Urne wird geschüttelt und eine weitere Kugel wird gezogen. Beim "Ziehen ohne Zurücklegen" wird eine Kugel nach der anderen (oder auch mehrere gleichzeitig) gezogen, wobei die schon gezogenen Kugeln draußen bleiben. Damit ändert sich natürlich die Zusammensetzung des Inhalts der Urne von Zug zu Zug. Beim "Ziehen mit Zurücklegen" ist dagegen die Zusammensetzung des Urneninhalts bei jedem Zug gleich. Dieser Unterschied hat teilweise erhebliche Auswirkungen auf die Wahrscheinlichkeiten. Ziehen wir beispielsweise aus einer Urne, in der sich eine rote und eine schwarze Kugel befinden, zweimal eine Kugel "mit Zurücklegen", dann ist die Wahrscheinlichkeit zweimal rot zu ziehen, 0.25. Ziehen wir dagegen zweimal "ohne Zurücklegen", dann ist die Wahrscheinlichkeit, zweimal rot zu ziehen, 0 - denn wenn nur eine rote Kugel in der Urne ist, ist es unmöglich, ohne Zurücklegen zweimal eine rote Kugel zu ziehen.

Wir haben bisher nur eine Anwendungsform des Urnenmodells kennengelernt: Die Angabe von Wahrscheinlichkeiten bestimmter Zugfolgen bei bekannter Zusammensetzung des Inhalts der Urne. Außerdem verwendet man das Urnenmodell aber auch in der Form, daß aus den Eigenschaften der gezogenen Kugeln Aussagen über die unbekannte Zusammensetzung des Urneninhalts abgeleitet werden. Wenn wir beispielsweise aus einer Urne mit 100 Kugeln acht mal ohne Zurücklegen ziehen und dabei sieben rote und eine schwarze Kugel ziehen, dann können wir mit Sicherheit nur sagen, daß in der Urne mindestens sieben rote und eine schwarze Kugel waren. Außerdem können wir auch sagen, daß wohl mehr rote als schwarze Kugeln in der Urne waren; denn sonst wäre es äußerst unwahrscheinlich, mit 8-maligem Ziehen sieben rote zu bekommen. Solche Überlegungen kommen vor allem in der Stichprobentheorie und bei der Schätzung von Populationskennziffern zur Anwendung. Es ist ja eine ähnliche Situation gegeben, wenn wir fragen, welche Rückschlüsse auf die Population aller Ehefrauen gezogen werden können, nachdem von einer Stichprobe von 400 Ehefrauen nur 4.3% angaben, daß die Hausarbeit zwischen ihnen und ihrem Ehemann etwa gleichmäßig verteilt ist.

b) Das Venn-Diagramm

In unserer wahrscheinlichkeitstheoretischen Anwendung ist ein Venn-Diagramm eine graphische Darstellung von Wahrscheinlichkeiten, bei der eine Gesamtfläche so in Unterflächen aufgeteilt ist, daß jeder Unterfläche ein bestimmtes Ereignis entspricht und daß die Größe der verschiedenen Unterflächen proportional der Wahrscheinlichkeit des von der Unterfläche dargestellten Ereignisses ist. In der folgenden Abbildung finden Sie verschiedene solche Venn-Diagramme:

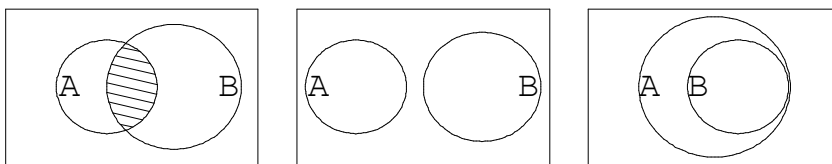


Abb. 1: Venn-Diagramme

In der linken Abbildung sind die Wahrscheinlichkeiten zweier Ereignisse A und B dargestellt, die einander nicht ausschließen. Es könnte sich z.B. um die Wahrscheinlichkeit der Lösung zweier Aufgaben eines Tests durch eine bestimmte Vp handeln. Wir können verschiedene Unterflächen unterscheiden, die alle ein bestimmtes Ereignis darstellen. Der linke Kreis entspricht dem Ereignis "Vp löst Aufgabe A richtig", der rechte Kreis dem Ereignis "Vp löst Aufgabe B richtig".

Welchem Ereignis entspricht dann die schraffierte Fläche, die zu den beiden Kreisen gehört?

Diese Fläche entspricht dem Ereignis "Vp löst beide Aufgaben richtig".

Welchem Ereignis entspricht der nicht schraffierte Teil des rechten Kreises?

Er entspricht dem Ereignis "Vp löst B richtig, dagegen Aufgabe A nicht richtig".

Welchem Ereignis entspricht die gesamte von den beiden Kreisen erfaßte Fläche?

Dem Ereignis "Vp löst entweder A oder B (oder beide) richtig".

Ursprünglich entwickelt wurde das Venn-Diagramm nicht für die Darstellung von Wahrscheinlichkeiten, sondern im Rahmen der Mengenlehre. Auch logische Beziehungen lassen sich mit dem Venn-Diagramm darstellen. So stellt das mittlere Venn-Diagramm in Abb. 1 den Fall dar, daß sich die Ereignisse A und B gegenseitig ausschließen (es gibt keine Unterfläche, die beiden Kreisen angehört, wodurch der Fall eines gemeinsamen Auftretens beider Kreise als unmöglich dargestellt wird).

Welche logische Beziehung zwischen den Ereignissen A und B wird im dritten Venn-Diagramm von Abb.1 dargestellt?

Da die gesamte Fläche des Kreises B auch zu A gehört, ergibt sich die logische Beziehung: Immer wenn B vorliegt, liegt auch A vor. Die Umkehrung dieses Satzes gilt bei diesem Venn-

Diagramm aber nicht, denn es gibt durchaus Teile des A-Kreises, die nicht zum B-Kreis gehören.

Neben logischen und mengentheoretischen Beziehungen lassen sich aber auch Wahrscheinlichkeiten mit Hilfe der Venn-Diagramme darstellen. Ein ideales Venn-Diagramm ist dabei so angelegt, daß jede Teilfläche proportional der Wahrscheinlichkeit des von dieser Teilfläche dargestellten Ereignisses ist.

Welche Aussage über das Verhältnis der Schwierigkeiten der Aufgaben A und B könnte man dann aus dem linken Venn-Diagramm in Abb. 1 entnehmen?

Die Teilfläche A ist kleiner als die Teilfläche B, folglich ist die Wahrscheinlichkeit des Ereignisses "Aufgabe A richtig gelöst" kleiner als die Wahrscheinlichkeit des Ereignisses "Aufgabe B richtig gelöst". Also fällt die Aufgabe B zumindest der Vp, für die unser Diagramm gilt, leichter als Aufgabe A.

Sind in der beschriebenen Weise alle Teilflächen proportional den Wahrscheinlichkeiten der zugehörigen Ereignisse, dann ist - wie sich mathematisch leicht beweisen läßt - die Wahrscheinlichkeit jedes Ereignisses gleich dem relativen Anteil der Teilfläche, die dem Ereignis entspricht, an der gesamten Fläche des Diagramms. In mathematischen Symbolen: Bezeichnen wir die Größe der Fläche, die dem Ereignis A zugeordnet ist, als f_A und die Gesamtfläche als n , so gilt wieder:

$$p(A) = \frac{f_A}{n}$$

Man kann sich fragen, welche Funktion ein Venn-Diagramm hat, wenn alle Flächen proportional den Wahrscheinlichkeiten sein müssen. Können wir es dann überhaupt in einer Situation verwenden, in der wir etwas über Wahrscheinlichkeiten erfahren wollen und sie noch nicht kennen? Wir können dann zwar noch kein ideales Venn-Diagramm anfertigen, aber immerhin eine Skizze. Eine solche Skizze hat eine ähnliche Funktion wie die skizzenmäßigen Dreiecke in der Geometrie, die man sich anlegt, um ein erst noch zu konstruierendes Dreieck zu planen. Genauso wie eine solche Skizze mit dem zu konstruierenden Dreieck noch nicht identisch ist und doch Rückschlüsse zuläßt, die auch für das richtige Dreieck gelten, können wir auch aus unserem skizzenhaften Venn-Diagramm schon einiges entnehmen. Aus dem rechten Diagramm in Abb. 1 können wir beispielsweise entnehmen, daß in dem dort skizzierten Fall die Wahrscheinlichkeit des Ereignisses B kleiner sein muß als die Wahrscheinlichkeit des Ereignisses A.

Wir haben bisher zweimal die Wahrscheinlichkeit des Ereignisses A mit der gleichen Formel angegeben:

$$p(A) = \frac{f_A}{n}$$

Im einen Fall gaben f_A und n Kugelzahlen an, im anderen Fall Flächen; n bezog sich auf die Gesamtheit, f_A auf den dem Ereignis entsprechenden Teil der Kugeln bzw. der Gesamtfläche. Wir können aber noch eine weitere Interpretation der Formel geben: Bei der Angabe der Wahrscheinlichkeit als zu erwartende relative Häufigkeit könnte man n als die Gesamtzahl der Versuche (z.B. 1000 Münzwürfe) interpretieren, und f_A als die zu erwartende absolute Häufigkeit des Ereignisses A (z.B. 500 "Adler"). Dann ist f_A / n die zu erwartende relative Häufigkeit des Ereignisses A (z.B. 0.50), also die objektive Wahrscheinlichkeit des Ereignisses A.

c) Zur Bedeutung mechanischer Wahrscheinlichkeitsmodelle in der Psychologie

In der Einleitung von Abschnitt @B.I.3 wurde die Verwendung mechanischer Wahrscheinlichkeitsmodelle vor allem damit begründet, daß fachnähere Beispiele mit der unbeantworteten Frage nach der Anwendbarkeit der Wahrscheinlichkeitstheorie in der Psychologie belastet wären. Daneben haben solche Modelle aber noch eine andere Funktion, die in diesem Abschnitt an einem Beispiel demonstriert werden soll.

Wir werden in Abschnitt @B.II.1.i die folgende Frage behandeln: Wenn ein Multiple-Choice-Test aus fünf Aufgaben besteht, bei denen jeweils sechs Alternativen zur Wahl gestellt werden, und eine Vp einfach würfelt, welche sie ankreuzen soll, wie groß ist dann die Wahrscheinlichkeit, daß sie zwei richtige Lösungen erzielt?

Auf den ersten Blick mag diese Fragestellung alle Bedenken wachrufen, die es gegen die Anwendung der Wahrscheinlichkeitstheorie in der Psychologie gibt. (Tatsächlich wurde die Wahrscheinlichkeitstheorie ursprünglich für die Analyse der Gewinnchancen bei Glücksspielen entwickelt!) Bedarf es überhaupt einer Diskussion, daß solche Fragestellungen recht wenig mit den Problemen zu tun haben, um deretwillen man Psychologie studiert?

Bevor man aufgrund derartiger Bedenken die Anwendung der Wahrscheinlichkeitstheorie in der Psychologie als "nicht Gegenstands-adäquat" ablehnt, lohnt sich wohl ein genaueres Hinsehen, welche Rolle eine solche Frage in einem größeren Kontext spielen kann. Wenn etwa ein "Hellseher" behauptet, trotz "gelegentlicher Fehler" angeben zu können, welche von 6 Spielkarten in einem benachbarten Raum gerade umgedeckt auf dem Tisch liegt, dann könnte man diesen Anspruch überprüfen, indem man diesen Versuch mehrmals durchführt. Natürlich wäre es auch ohne alle hellseherischen Fähigkeiten zu erwarten, daß dieser Proband *etwa* bei jedem sechsten Versuch rein zufällig die richtige Karte rät. Wenn er nun aber bei fünf Versuchen zwei Treffer erzielt, dann ist das etwas mehr, als "etwa ein Treffer bei jedem sechsten Versuch". Drei Fehler in fünf Versuchen ist zwar mehr, als die Einschränkung "trotz

gelegentlicher Fehler" zugibt; aber genügt es nicht für einen Nachweis besonderer "parapsychologischer" Fähigkeiten, wenn er mehr Treffer erzielt, als per Zufall zu erwarten wäre - also mehr als "etwa einen Treffer bei jedem sechsten Versuch"? Sollte man dann nicht am sinnvollsten annehmen, daß irgendwelche "parapsychischen" Signale ankommen und daß lediglich die Umsetzung dieser Signale in konkrete Antworten "verrauscht" ist?

An dieser Stelle lohnt es sich wohl, das Ergebnis der Berechnungen in Abschnitt @B.II.1.i vorwegzunehmen: Die Wahrscheinlichkeit, in dem o.g. Test mit Würfeln zwei richtige Lösungen zu erzielen, beträgt immerhin 0.1608. Nun kann man fünf Versuche, anzugeben, welche von sechs Karten im Nachbarraum umgedreht auf dem Tisch liegt, als einen Multiple-Choice-Test mit fünf Aufgaben und sechs Alternativen pro Aufgabe betrachten. Wenn man aber weiß, daß in einer solchen Situation selbst beim Würfeln die Wahrscheinlichkeit von zwei Treffern immerhin 0.1608 beträgt, dann wird man diese zwei Treffer noch nicht als ausreichenden Beleg für hellseherische Fähigkeiten ansehen.

An diesem Beispiel sieht man, daß es an der Logik eines solchen Gedankengangs völlig vorbeigehen würde, wenn man ihn einfach mit dem Argument vom Tisch fegen würde, daß man doch gesehen hat, daß der Proband nicht gewürfelt hat. Die Bedeutung, die das Würfel-Modell in diesem Fall hat, kann man darin sehen, daß die zur Diskussion stehende hellseherische Fähigkeit von diesem Modell abgegrenzt wird: Eine solche Fähigkeit wird man erst erst bei einer Trefferzahl annehmen, die beim Würfeln äußerst unwahrscheinlich wäre.

<<

Wie wir später sehen werden, würde eine noch präzisere Formulierung lauten: ... bei einer Trefferzahl, bei der die Wahrscheinlichkeit, diese oder eine noch höhere mit Würfeln zu erzielen, äußerst gering wäre.

>>

Das läßt sich folgendermaßen verallgemeinern: Die Bedeutung von Wahrscheinlichkeitsmodellen in der Psychologie besteht oft nicht darin, das von einer Theorie vermutete Erleben oder Verhalten zu modellieren. Stattdessen spezifizieren psychologische Theorien oft, in welcher Weise Meßwerte oder statistische Kennziffern von den Werten abweichen, die unter einem bestimmten Wahrscheinlichkeitsmodell zu erwarten wären.

d) Die Angabe von Wahrscheinlichkeiten durch "Odds"

<<

Bisher haben wir Wahrscheinlichkeiten immer in der Form angegeben, daß wir ein betrachtetes Ereignis irgendwie in Verhältnis zu allen möglichen Ereignissen gesetzt haben. Es gibt aber noch eine andere Form einer solchen Angabe. Bei der Urne mit 70 roten und 130 schwarzen Kugeln sagt man manchmal auch, die Wahrscheinlichkeit, eine rote Kugel zu ziehen, betrage "70 zu 130". Dabei werden also die 70 roten Kugeln nicht im Verhältnis zu allen Kugeln (einschließlich der roten) betrachtet, sondern im Verhältnis zu allen nicht-roten. Diese Art der Angabe von Wahrscheinlichkeiten kommt vor allem bei Wetten zur Anwendung. Wenn man z.B. sagt, die Wetten auf einen Regierungswechsel nach einer bevorstehenden Bundestagswahl stünden 3 zu 1, dann heißt das: Wer auf "Wechsel" wettet, muß das Dreifache dessen setzen, der auf "kein Wechsel" wettet, und darin drückt sich aus, daß das Ereignis "Wechsel" für drei

mal so wahrscheinlich gehalten wird wie das Ereignis "kein Wechsel". Weil das Verhältnis der Wetteinsätze angibt, wie "schief" die Wette ist, spricht man im Englischen auch von "odds".⁶⁾ Die Umrechnung solcher odds in die bisher mit p bezeichneten "üblichen" Wahrscheinlichkeiten läßt sich mit den folgenden Formeln angeben, bei denen $o(A)$ für "odds von A " steht, also für das Verhältnis des Einsatzes auf Ereignis A zum Einsatz auf das Ereignis $\neg A$:

$$p(A) = \frac{o(A)}{1 + o(A)}$$

und

$$o(A) = \frac{p(A)}{1 - p(A)}$$

Mit der zweiten Formel könnte man grundsätzlich alle Wahrscheinlichkeiten in odds umrechnen. Die meisten Formeln der Wahrscheinlichkeitstheorie sind einfacher, wenn man die bisher betrachteten p -Wahrscheinlichkeiten verwendet, und deshalb haben diese sich durchgesetzt. In einigen wenigen Bereichen sind aber odds einfacher, und deshalb ist es nicht nur für das Wetten nützlich, auch odds zu kennen.

>>

4) Bedingte Wahrscheinlichkeiten

a) Der Begriff der bedingten Wahrscheinlichkeit

Wir nehmen an, wir hätten in einer Urne 200 Kugeln, von denen 80 schwarz und 120 rot sind. Die Kugeln unterscheiden sich außerdem in ihrem Material: Die einen sind aus Holz, die anderen aus Plastik. Holz- und Plastikugeln sind jedoch so ähnlich gefärbt und poliert, daß man sie mit bloßem Auge nicht unterscheiden kann. Die Zusammensetzung des Inhalts der Urne wird von der folgenden Tabelle angegeben.

	Holz	Plastik	gesamt
schwarz	50	30	80
rot	40	80	120
gesamt	90	110	200

Tabelle 1: Häufigkeiten verschiedener Kugeln in einer Urne

Wie groß ist die Wahrscheinlichkeit eine Holzkugel zu ziehen?

⁶⁾Daß ein englisches Wort verwendet wird, beruht natürlich darauf, daß das Wetten in Wettbüros im angloamerikanischen Bereich viel verbreiteter ist als bei uns. - Bei solchen Wettbüros kommt natürlich zum eigentlichen Einsatz noch ein (oft versteckter) Gebührenanteil hinzu, von dem der Inhaber des Büros lebt. Die obigen Berechnungen gelten nach Abzug dieses Gebührenanteils.

Wir definieren: Das Ereignis h besteht im Ziehen einer hölzernen Kugel. Dann können wir für die gesuchte Wahrscheinlichkeit schreiben:

$$p(h) = \frac{f_h}{n} = \frac{90}{200} = 0.45$$

Wir nehmen nun an, wir hätten eine Kugel gezogen und festgestellt, daß es eine rote ist. Da Kugeln aus Holz und aus Plastik sich im Aussehen nicht unterscheiden, erhebt sich die Frage:

Wie groß ist die Wahrscheinlichkeit, daß unsere Kugel, deren rote Farbe wir schon festgestellt haben, aus Holz ist?

Wir können diese Wahrscheinlichkeit ähnlich herleiten wie in Abschnitt 3a. Wir wissen, daß es sich um eine der 120 roten Kugeln handelt; jede rote Kugel hat die gleiche Wahrscheinlichkeit, die jetzt $1/120$ beträgt, denn offenbar müssen sich diese Wahrscheinlichkeiten über alle 120 roten Kugeln addiert zu 1 ergänzen. Ziehen wir die Wahrscheinlichkeiten aller Holzkugeln, die noch in Frage kommen (also aller 40 roten Holzkugeln), zusammen, dann ergibt sich $40/120 = 0.33\dots$ als "bedingte Wahrscheinlichkeit, daß es sich um eine Holzkugel handelt".

In der mathematischen Symbolik verwendet man für die Angabe einer bedingten Wahrscheinlichkeit folgendes Symbol: $p(A/B)$ ist die Wahrscheinlichkeit, daß Ereignis A eintritt, gegeben daß Ereignis B eintritt. Hinter dem Schrägstrich (für den man häufig auch einen senkrechten Strich setzt) ist also die Bedingung angegeben, vor dem Strich das Ereignis, dessen Wahrscheinlichkeit angegeben wird.

In Worten spricht man auch einfach von der "bedingten Wahrscheinlichkeit von A gegeben B". In unserem Beispiel könnten wir schreiben:

$$p(h/r) = \frac{40}{120} = 0.33\dots$$

Das Symbol r steht dabei für das Ereignis "rote Kugel". Verallgemeinern wir den Gedankengang - zunächst jedoch nur für das Urnenmodell. Wenn wir schon wissen, daß eine Kugel der Klasse B gezogen wird, dann ist die Wahrscheinlichkeit jeder einzelnen Kugel der Klasse B gleich $1/f_B$. Die Zahl der Kugeln der Klasse A, die dann noch in Frage kommen, ist $f_{A \wedge B}$ (\wedge ist das logische Zeichen für "und".) Das Symbol " $f_{A \wedge B}$ " steht für: Zahl der Kugeln, die sowohl zur Klasse A (z.B. "Holzkugel") als auch zur Klasse B (z.B. "rote Kugel") gehören. Die bedingte Wahrscheinlichkeit, eine Kugel der Klasse A zu ziehen, gegeben es wird eine Kugel der Klasse B gezogen, ist dann:

$$p(A/B) = \frac{f_{A \wedge B}}{f_B}$$

Wie groß ist die Wahrscheinlichkeit, aus der oben beschriebenen Urne eine Plastikkugel

zu ziehen - gegeben die Kugel ist rot?

Wir definieren: Das Symbol pl steht für "Plastikkugel". Dann können wir schreiben:

$$p(pl/r) = \frac{f_{pl \wedge r}}{f_r} = \frac{80}{120} = 0.67$$

Wir können nun diese Symbole auch auf andere Veranschaulichungen des Wahrscheinlichkeitsbegriffs anwenden. Beim Venn-Diagramm wäre $f_{A \wedge B}$ diejenige Teilfläche, die sowohl zur Fläche A als auch zur Fläche B gehört (also z.B. der schraffierte Teil des linken Venn-Diagramms in der Abbildung aus Abschnitt @ I.3.b). Auch bei erwarteten Häufigkeiten können wir entsprechend operieren. Wenn wir etwa mit einem ungefälschten Würfel würfeln, so können wir definieren: Ereignis A besteht darin, daß wir eine gerade Zahl würfeln; Ereignis B besteht darin, daß wir eine Zahl würfeln, die größer als 2 ist.

Wie groß sind dann f_A und f_B (also die erwarteten absoluten Häufigkeiten der Ereignisse A und B) bei 600 Würfen?

Da unter den sechs Zahlen des Würfels drei gerade und vier größer als zwei sind, ergibt sich offenbar:

$$f_A = 300 \quad f_B = 400$$

Was bedeutet hier wohl $f_{A \wedge B}$ und wie groß ist diese Zahl?

$f_{A \wedge B}$ ist hier die erwartete absolute Häufigkeit von Würfeln, bei denen sowohl Ereignis A als auch Ereignis B auftritt. Da von den sechs Zahlen des Würfels zwei sowohl gerade als auch größer als 2 sind, ist $f_{A \wedge B}$ offenbar 200.

Die bedingte Wahrscheinlichkeit des Ereignisses A "gerade Zahl" gegeben "Zahl größer als 2" können wir angeben als:

$$p(A/B) = \frac{f_{A \wedge B}}{f_B} = \frac{200}{400} = 0.50$$

In Worten: wir erwarten, daß von den ca. 400 Würfeln, in denen Ereignis B auftritt, bei etwa 200 (also bei einem relativen Anteil von 0.50) auch das Ereignis A eintritt.

Unsere Formel für die bedingte Wahrscheinlichkeit können wir noch vereinfachen. Wir kürzen den Bruch hinter dem Gleichheitszeichen durch n und erhalten:

$$p(A/B) = \frac{f_{A \wedge B} / n}{f_B / n}$$

Der Nenner des neuen Bruchs ist nun die Wahrscheinlichkeit des Ereignisses B, und der Zähler ist die Wahrscheinlichkeit, daß die Ereignisse A und B beide eintreffen. In Symbolen können wir also auch schreiben:

$$p(A/B) = \frac{p(A \wedge B)}{p(B)}$$

Dies ist die mathematische Definition der bedingten Wahrscheinlichkeit. Sie gilt unabhängig von allen Konkretisierungen des Wahrscheinlichkeitsbegriffs. Wie jede Definition kann man auch diese nicht beweisen. Unsere Herleitung des Begriffs hatte also nicht die Funktion, die Definition zu beweisen, sondern zu zeigen, welchen Sinn sie hat.

<<

Aus dieser Definition folgt dann auch: Ist die Wahrscheinlichkeit $p(B)$ null, dann ist die bedingte Wahrscheinlichkeit $p(A/B)$ nicht definiert; denn sie würde ja eine Division durch 0 erfordern. Das ist auch sinnvoll; denn meist gibt es kaum Anhaltspunkte für die Angabe, wie groß die Wahrscheinlichkeit des Ereignisses A ist, wenn das "fast unmögliche" Ereignis B doch einmal eintritt.⁷

>>

Zu ergänzen ist noch, daß man die Wahrscheinlichkeit $p(A)$ zur Unterscheidung von bedingten Wahrscheinlichkeiten auch die "unbedingte" Wahrscheinlichkeit oder - in Analogie zum Begriff der Randverteilung - die *Randwahrscheinlichkeit* des Ereignisses A nennt. Die Wahrscheinlichkeit, daß mehrere Ereignisse (z.B. A und B) beide auftreten, nennt man eine *Verbundwahrscheinlichkeit*.

Mi diesen Begriffen läßt sich die obige formelmäßige Definition der bedingten Wahrscheinlichkeit folgendermaßen in Worte fassen:

Bedingte Wahrscheinlichkeit gleich Verbundwahrscheinlichkeit dividiert durch Randwahrscheinlichkeit der Bedingung

b) Unabhängige und abhängige Ereignisse

Definition: Die Ereignisse A und B heißen unabhängig, wenn die Wahrscheinlichkeit des Ereignisses A unabhängig von Bedingungen über das Ereignis B ist, wenn also gilt:

$$p(A/B) = p(A/\neg B)$$

Das in dieser Formel verwendete Zeichen " \neg " ist das logische Zeichen für "nicht".

Es läßt sich mathematisch beweisen, daß dann auch die folgenden Gleichungen gelten:

⁷In Abschnitt @ B.IV wird genauer gezeigt, was damit gemeint ist, daß ein Ereignis mit Wahrscheinlichkeit 0 nicht immer unmöglich, sondern möglicherweise nur "fast unmöglich" ist.

$$p(A/B) = p(A)$$

und

$$p(B/A) = p(B/\neg A) = p(B).$$

Außerdem läßt sich beweisen, daß immer dann, wenn eine dieser Gleichungen zutrifft, auch alle übrigen zutreffen.⁸ Wir können also jede der Gleichungen als Definitionsgrundlage für den Begriff der Unabhängigkeit nehmen. Wählen wir etwa die zweite Gleichung, dann würde die Definition der Unabhängigkeit lauten:

Zwei Ereignisse heißen unabhängig voneinander, wenn die bedingte Wahrscheinlichkeit des einen, gegeben das andere, gleich der Wahrscheinlichkeit des erstgenannten ist.

Trifft andererseits eine dieser Gleichungen nicht zu, dann treffen auch die anderen nicht zu, und damit sind die Ereignisse voneinander abhängig.

Beispiel: In einer Urne befinden sich 36 schwarze und 44 rote Holzkugeln sowie 54 schwarze und 66 rote Plastikkugeln. Wie sich leicht nachprüfen läßt, gilt dann

$$p(h / r) = p(h / \neg r) = p(h) = 0.4.$$

Es ist zu betonen, daß der Begriff der Abhängigkeit auch hier - genau so wie bei Korrelation, Regression und Determination - keine Kausalaussage beinhaltet. Wenn wir sagen, A sei von B abhängig, dann bedeutet das nur, daß die Wahrscheinlichkeit sich durch die Angabe einer Bedingung über B ändert. Auf welcher Kausalbeziehung das beruht, spielt keine Rolle. Um dies zum Ausdruck zu bringen, spricht man auch von statistisch (oder stochastisch) abhängigen bzw. unabhängigen Ereignissen.

<<

In der mathematischen Wahrscheinlichkeitstheorie wird die Unabhängigkeit von Ereignissen meist etwas anders definiert. Dies wird in Abschnitt @B.II.1.a näher erläutert. Dasselbe gilt auch für die folgende Unterscheidung von paarweiser und vollständiger Unabhängigkeit.

>>

<<

c) Paarweise und vollständige Unabhängigkeit von Ereignissen

Haben wir die stochastische Abhängigkeit oder Unabhängigkeit von mehr als zwei Ereignissen zu

⁸Eine Ausnahme tritt natürlich ein, wenn eines der Ereignisse die Wahrscheinlichkeit 0 hat. Gilt dies z.B. für das Ereignis B, dann ist die bedingte Wahrscheinlichkeit $p(A/B)$ nicht definiert, weil ja im Nenner der Definitionsgleichung eine 0 stünde.

prüfen, dann sind zwei Fragen zu unterscheiden. Man kann einmal fragen: Ist jedes einzelne dieser Ereignisse im Sinne der bisherigen Definition unabhängig von jedem einzelnen anderen? D.h. besteht innerhalb eines jeden Paares von zwei Ereignissen, die beide zu den zu überprüfenden Ereignissen gehören, das bisher diskutierte Unabhängigkeitsverhältnis? Wenn diese Unabhängigkeit für jedes dieser Paare gilt, dann nennt man die Ereignisse auch paarweise unabhängig.

Beispiel: Mit einem ungefälschten Würfel wird viermal gewürfelt. Das Ereignis " s_1 " bedeutet: Beim ersten Wurf wird eine Sechs gewürfelt. Entsprechend bedeutet " s_2 " das Würfeln einer Sechs beim zweiten Wurf usw.

Sind die Ereignisse s_1, s_2, s_3 und s_4 voneinander paarweise unabhängig?

Innerhalb des Paares " $s_1 s_2$ " besteht offenbar stochastische Unabhängigkeit, denn die Angabe einer Bedingung über das Ereignis s_1 ändert nichts an der Wahrscheinlichkeit des Ereignisses s_2 und umgekehrt:

$$p(s_2) = p(s_2/s_1) = p(s_2/\neg s_1)$$

$$p(s_1) = p(s_1/s_2) = p(s_1/\neg s_2)$$

Entsprechendes gilt für jedes andere Paar von Ereignissen, das aus den Ereignissen s_1, s_2, s_3 und s_4 gebildet wird, und daher sind die Ereignisse paarweise unabhängig.

Von der Frage der paarweisen Unabhängigkeit ist die Frage nach der vollständigen Unabhängigkeit zu unterscheiden. Dazu eine

Definition: Mehrere Ereignisse sind vollständig unabhängig, wenn sich durch eine Bedingung über das Eintreffen einer beliebigen Untergruppe dieser Ereignisse nichts ändert an der bedingten Wahrscheinlichkeit eines weiteren der Ereignisse, das nicht zur Untergruppe gehört.

Liegt bei den von uns definierten Ereignissen s_1, s_2, s_3 und s_4 auch vollständige Unabhängigkeit vor?

Zur Veranschaulichung greifen wir eine Untergruppe dieser Ereignisse heraus - die Untergruppe " $s_2 \wedge s_3$ ". Wir fragen uns: Ändert eine Angabe über das Eintreffen der Ereignisse dieser Untergruppe etwas an der bedingten Wahrscheinlichkeit eines anderen der von uns diskutierten Ereignisse, das nicht zur Untergruppe gehört? Ist also

$$p(s_1/s_2 \wedge s_3) = p(s_1) \text{ und}$$

$$p(s_4/s_2 \wedge s_3) = p(s_4)?$$

Wir können diese Frage nicht einfach bejahen, weil wir paarweise Unabhängigkeit der Ereignisse festgestellt haben. Wir werden nämlich noch sehen, daß es durchaus Fälle gibt, in denen mehrere Ereignisse zwar paarweise, aber nicht vollständig unabhängig sind. Wir können unsere Frage aber aufgrund unserer Sachkenntnis über den Vorgang des Würfeln beantworten: Da bei jedem Wurf der Zufall wieder neu ins Spiel kommt und das Auftreten einer Sechs in einem oder mehreren Würfeln in keiner Weise die Wahrscheinlichkeit einer Sechs in einem weiteren Wurf verändert, solange "fair" gewürfelt wird, können wir das Zutreffen unserer beiden obigen Gleichungen aufgrund dieses Sachwissens bestätigen. Das gleiche gilt auch für jede andere bedingte Wahrscheinlichkeit eines der Ereignisse s_1, s_2, s_3 und s_4 , bei der die Bedingung aus einer Angabe über das Auftreten oder Nicht-Auftreten der übrigen Ereignisse besteht. Daher sind unsere vier Ereignisse nicht nur paarweise, sondern auch vollständig unabhängig voneinander.

Die Bedeutung der Unterscheidung von paarweiser und vollständiger Unabhängigkeit liegt in der Tatsache, daß es Situationen gibt, in denen zwar paarweise, aber keine vollständige Unabhängigkeit vorliegt. Beispiel: In einer Urne befinden sich 2000 Kugeln, die sich in Material (Holz - Plastik), Farbe (schwarz - rot) und in einem aufgedruckten Symbol (Dreieck - Viereck) unterscheiden. Die Zusammensetzung der Urne ist folgendermaßen:

1200 Kugeln mit aufgedrucktem Dreieck:

	Holz	Plastik	zusammen
schwarz	224	616	840
rot	16	344	360
gesamt	240	960	1200

800 Kugeln mit aufgedrucktem Viereck:

	Holz	Plastik	zusammen
schwarz	56	504	560
rot	104	136	240
gesamt	160	640	800

Fassen wir die Kugeln mit aufgedrucktem Dreieck und aufgedrucktem Viereck zusammen, so ergibt sich:

	Holz	Plastik	zusammen
schwarz	280	1120	1400
rot	120	480	600
gesamt	400	1600	2000

Wir stellen zunächst fest, ob die Ereignisse "Holzkugel ziehen" (h), "schwarze Kugel ziehen" (s) und "Kugel mit aufgedrucktem Dreieck ziehen" (d) paarweise unabhängig sind.

Versuchen Sie, diese Überprüfung selbst durchzuführen!

Es ist zu überprüfen, ob die Angabe einer Bedingung über eines der Ereignisse etwas an der Wahrscheinlichkeit eines anderen ändert. Die Randwahrscheinlichkeit des Ereignisses h können wir aufgrund der 3. Tabelle bestimmen: Von insgesamt 2000 Kugeln sind 400 aus Holz. Es ist also:

$$p(h) = 400/2000 = 0.2$$

Wie groß ist die bedingte Wahrscheinlichkeit, eine Holzkugel zu ziehen, gegeben, es wird eine schwarze Kugel gezogen?

Wir können wieder die 3. Tabelle heranziehen. Von insgesamt 1400 schwarzen Kugel sind 280 aus Holz. Es ergibt sich also:

$$p(h/s) = 280/1400 = 0.2$$

Wie groß ist die bedingte Wahrscheinlichkeit, eine Holzkugel zu ziehen, gegeben, es wird eine Kugel mit aufgedrucktem Dreieck gezogen?

Aus der ersten Tabelle entnehmen wir, daß von den 1200 Kugeln mit aufgedrucktem Dreieck 240 aus Holz sind. Daraus folgt:

$$p(h/d) = 240/1200 = 0.2$$

Zusammenfassend können wir schreiben:

$$p(h) = p(h/s) = p(h/d) = 0.2$$

Wir könnten die hier für das Ereignis h angestellten Überlegungen auch für die Ereignisse s und d anstellen und kämen zu den Ergebnissen:⁹

⁹Wollen Sie übungshalber eine dieser drei Gleichungen überprüfen?

$$p(s) = p(s/h) = p(s/d) = 0.7$$

$$p(d) = p(d/h) = p(d/s) = 0.6$$

In Worten besagen die drei letzten Gleichungen: Bei keinem der Ereignisse h, s und d ändert sich die Wahrscheinlichkeit durch Hinzufügen einer Bedingung über das Auftreten eines der anderen Ereignisse. Demnach sind die Ereignisse paarweise unabhängig.

Trotzdem sind die drei Ereignisse nicht vollständig unabhängig. Wenn wir etwa eine Kugel ziehen und feststellen, daß sie schwarz ist und ein aufgedrucktes Dreieck hat, dann können wir die bedingte Wahrscheinlichkeit, daß diese Kugel aus Holz ist, folgendermaßen bestimmen: Nach unserer ersten Tabelle sind von den 840 Kugeln, die die Bedingung "schwarz mit aufgedrucktem Dreieck" erfüllen, 224 aus Holz. Damit ist:

$$p(h/s \wedge d) = 224/840 = 0.3666\dots$$

In Worten: Es gibt eine Untergruppe von Ereignissen (nämlich die Untergruppe "s \wedge d"), deren Auftreten die bedingte Wahrscheinlichkeit des übrigen Ereignisses h verändert; denn die bedingte Wahrscheinlichkeit $p(h/s \wedge d)$ ist nicht mehr gleich der Randwahrscheinlichkeit $p(h)$. Folglich sind die Ereignisse h, s und d nicht völlig unabhängig; denn wenn sie es wären, dürfte auch die Angabe einer Bedingung über das Auftreten einer Untergruppe der Ereignisse nichts an der Wahrscheinlichkeit eines weiteren ändern.

Als Gesamtergebnis dieses Urnenbeispiels können wir festhalten: Es gibt Fälle, in denen mehrere Ereignisse zwar paarweise, aber nicht völlig unabhängig sind. Der umgekehrte Fall ist dagegen nicht möglich. Es gilt vielmehr der (im Grunde triviale)

Lehrsatz: Sind mehrere Ereignisse vollständig unabhängig, dann sind sie auch paarweise unabhängig.

Die Unterscheidung von paarweiser und vollständiger Unabhängigkeit ist bei einigen Modellen in der Psychodiagnostik von Bedeutung. Bei früheren Ansätzen versuchte man, möglichst viele Symptome ausfindig zu machen, die mit einem bestimmten Persönlichkeitsbild (z.B. einer Geisteskrankheit) paarweise Abhängigkeit aufzeigten. Grob gesagt, würde man nach einem solchen Ansatz die Wahrscheinlichkeit des Vorliegens einer Geisteskrankheit um so größer ansetzen, je mehr Symptome vorliegen, bei denen ein stochastischer Zusammenhang mit der Geisteskrankheit besteht. Die Unterscheidung von paarweiser und vollständiger Abhängigkeit von Ereignissen legt aber die Vermutung nahe, daß ein solcher Ansatz die in den Symptomen steckende Information nicht völlig ausschöpft: Es wäre denkbar, daß eine Geisteskrankheit (G) von jedem der Symptome A und B paarweise unabhängig ist, daß aber ihr gemeinsames Auftreten doch zu einer erhöhten Wahrscheinlichkeit des Vorliegens der Geisteskrankheit führt. In Formelsprache ausgedrückt:

Es ist zwar:

$$p(G/A) = p(G/B) = p(G), \text{ aber} \\ p(G/A \wedge B) > p(G)$$

Aufgrund des älteren Ansatzes würde man beide Symptome als unbrauchbar für die Diagnose dieser Geisteskrankheit ansehen, denn das Vorliegen eines einzelnen ändert nichts an der Wahrscheinlichkeit der Geisteskrankheit (erste Gleichung). Trotzdem kann das gemeinsame Auftreten beider Symptome zu einer erhöhten Wahrscheinlichkeit der Geisteskrankheit führen, und diese Information geht verloren, wenn man aus der paarweisen Unabhängigkeit voreilig auf völlige Unabhängigkeit schließt. Der hier beschriebene Ansatz ist in der Psychodiagnostik unter dem Namen "Konfigurationsfrequenzanalyse" bekannt. Es werden nicht nur einzelne Symptome isoliert voneinander betrachtet, sondern ganze Symptomkonfigurationen analysiert.

II. Kombinatorik

Unter dem Begriff der Kombinatorik faßt man verschiedene mathematische Überlegungen zusammen, die es gestatten, aus einigen bekannten Wahrscheinlichkeiten andere zu berechnen. Es handelt sich dabei einmal um Lehrsätze über die Wahrscheinlichkeiten von logisch kombinierten Ereignissen; daneben geht es um Berechnungen der Anzahl möglicher Ereignisse, aus denen man Rückschlüsse auf die Wahrscheinlichkeit dieser Ereignisse ziehen kann.

1) Wahrscheinlichkeitssätze

a) Der Multiplikationssatz

Der Multiplikationssatz der Wahrscheinlichkeitsrechnung befaßt sich mit der Wahrscheinlichkeit, daß mehrere Ereignisse auftreten. Beim Urnenmodell könnte man beispielsweise fragen: Wie groß ist die Wahrscheinlichkeit, aus einer Urne mit drei roten und zwei schwarzen Kugeln bei zweimaligem Ziehen ohne Zurücklegen zuerst eine schwarze und dann eine rote Kugel zu ziehen? Wir gehen aus von der Definition der bedingten Wahrscheinlichkeit des Ereignisses B, gegeben Ereignis A:

$$p(B/A) = \frac{p(A \wedge B)}{p(A)}$$

Multiplizieren wir beide Seiten mit $p(A)$, so erhalten wir:

$$p(A \wedge B) = p(A) \cdot p(B/A)$$

Können Sie die Wahrscheinlichkeit " $p(A \wedge B)$ " auch aufgrund der bedingten Wahrscheinlichkeit $p(A/B)$ herleiten?

Aus der Definitionsgleichung

$$p(A/B) = \frac{p(A \wedge B)}{p(B)}$$

ergibt sich nach Multiplikation mit $p(B)$:

$$p(A \wedge B) = p(B) \cdot p(A/B)$$

In Worten: Die Wahrscheinlichkeit, daß zwei Ereignisse beide auftreten, ist gleich dem Produkt der Randwahrscheinlichkeiten des einen multipliziert mit der bedingten Wahrscheinlichkeit des anderen - gegeben das erste liegt vor.

Um den Satz auf unser Beispiel anwenden zu können, definieren wir: Das Ereignis A besteht darin, daß die erste Kugel schwarz ist. Das Ereignis B besteht darin, daß die zweite Kugel rot ist.

Können Sie jetzt die Wahrscheinlichkeit angeben, daß die erste Kugel schwarz und die zweite Kugel rot ist, wenn man ohne Zurücklegen zieht?

Wir haben den Multiplikationssatz in zwei Fassungen kennengelernt. Im einen Fall benötigen wir $p(A)$ und $p(B/A)$, im anderen Fall $p(B)$ und $p(A/B)$. Wenn man eine der beiden anwenden will und nicht weiß, welche von beiden besser paßt, dann fragt man sich, welche Wahrscheinlichkeiten man angeben kann. Da wir eine Randwahrscheinlichkeit und eine bedingte Wahrscheinlichkeit brauchen und bedingte Wahrscheinlichkeiten meist schwerer anzugeben sind als Randwahrscheinlichkeiten, ist es eine gute Taktik, sich zuerst zu fragen, welche bedingte Wahrscheinlichkeit man leichter angeben kann: $p(A/B)$ oder $p(B/A)$.

In unserem Fall ist es leichter, $p(B/A)$ anzugeben, also die Wahrscheinlichkeit, beim zweiten Zug eine rote Kugel zu erhalten, wenn die erste schwarz war. Wenn nämlich die erste Kugel schwarz war, befinden sich beim Ziehen ohne Zurücklegen in der Urne noch drei rote und eine schwarze Kugel. Die Wahrscheinlichkeit, beim zweiten Zug eine rote Kugel zu ziehen, wenn die erste schwarz war, können wir also angeben als

$$p(B/A) = 0.75$$

Es läuft also auf eine Anwendung unserer ersten Gleichung hinaus. Wir brauchen außerdem noch $p(A)$ - die Wahrscheinlichkeit, im ersten Zug eine schwarze Kugel zu ziehen. Diese ist offenbar $2/5 = 0.40$. Die Wahrscheinlichkeit, daß die Ereignisse A und B eintreffen, können wir dann angeben als:

$$p(A \wedge B) = p(A) \cdot p(B/A) = 0.40 \cdot 0.75 = 0.30$$

Ein Spezialfall des Multiplikationssatzes ist der Multiplikationssatz für unabhängige Ereignisse. Sind die Ereignisse A und B statistisch unabhängig, dann können wir für die bedingten Wahrscheinlichkeiten auch die Randwahrscheinlichkeiten einsetzen. Dann ergibt sich der folgende "einfache Multiplikationssatz":

$$p(A \wedge B) = p(A) \cdot p(B)$$

Dieser gilt aber nur bei stochastischer Unabhängigkeit der Ereignisse A und B. Das läßt sich auf mehr als zwei Ereignisse verallgemeinern, wenn diese vollständig unabhängig sind. Dann ist

$$p(A \wedge B \wedge C \wedge \dots) = p(A) \cdot p(B) \cdot p(C) \cdot \dots$$

Wir können also *in Worten* formulieren:

Die Wahrscheinlichkeit, daß zwei (oder mehr) vollständig unabhängige Ereignisse beide (bzw. alle) eintreffen, ist gleich dem Produkt der Randwahrscheinlichkeiten der Ereignisse.

<<

Wir haben den einfachen Multiplikationssatz aus einem Begriff der Unabhängigkeit abgeleitet, der an bedingten Wahrscheinlichkeiten orientiert ist. Das ist recht anschaulich, hat aber einen "Schönheitsfehler": Ist die Wahrscheinlichkeit von B null, dann ist die bedingte

Wahrscheinlichkeit $p(A/B)$ nicht definiert, und damit ist nach unserer bisherigen Definition auch nicht zu klären, ob die Ereignisse A und B unabhängig sind. In der mathematischen Wahrscheinlichkeitstheorie wird daher das Verhältnis von Definition und Konsequenz umgekehrt:

- Die Definition der Unabhängigkeit lautet: Endlich viele Ereignisse sind unabhängig, wenn die Verbundwahrscheinlichkeit gleich dem Produkt der Randwahrscheinlichkeiten ist.
- Daraus folgt dann: Ist $0 < p(B) < 1$, dann sind die Ereignisse A und B genau dann unabhängig, wenn die Gleichung $p(A/B) = p(A/\bar{B}) = p(A)$ gilt.

Daraus kann man dann schließen: Wenn das Auftreten oder Nicht-Auftreten von B keinen Einfluß auf die bedingte Wahrscheinlichkeit von A hat, dann gilt der einfache Multiplikationssatz.

>>

b) Der einfache Additionssatz

Wenn zwei Ereignisse A und B sich gegenseitig ausschließen, so ist die Wahrscheinlichkeit, daß entweder das Ereignis A oder das Ereignis B auftritt, gleich der Summe der Randwahrscheinlichkeiten der Ereignisse. Als Gleichung:

$$p(A \vee B) = p(A) + p(B)$$

Das in dieser Gleichung verwendete Zeichen " \vee " ist das logische Zeichen für "oder".

Das läßt sich auf mehr als zwei einander ausschließende Ereignisse ausdehnen: Die Wahrscheinlichkeit, daß eines von ihnen auftritt, ist die Summe der Einzelwahrscheinlichkeiten.

$$p(A \vee B \vee C \vee \dots) = p(A) + p(B) + p(C) + \dots$$

c) Die Axiome der mathematischen Wahrscheinlichkeitstheorie

<<

Dieser Abschnitt ist ganz "für Spezialisten".

Wir haben den Additionssatz für einander ausschließende Ereignisse bereits an mehreren früheren Stellen stillschweigend vorausgesetzt.

Können Sie angeben, an welchen Stellen?

Als wir die Wahrscheinlichkeit ableiteten, aus einer Urne mit 70 roten und 130 schwarzen Kugeln bei einmaligem Ziehen eine rote Kugel zu erhalten (Abschnitt I.3.a), haben wir gesagt: Die Wahrscheinlichkeit, irgendeine der 70 roten Kugeln zu ziehen, erhalten wir "offenbar", indem wir die Wahrscheinlichkeiten aller 70 roten Kugeln zusammenziehen. Das erschien uns plausibel, aber wir haben es nicht begründet.

In der mathematischen Statistik wird der einfache Additionssatz auch nicht bewiesen, sondern als Axiom vorausgesetzt. Es ist nicht nur das Recht, sondern die Aufgabe der Mathematik, bestimmte nicht weiter begründete Sätze (eben die Axiome) vorauszusetzen und zu überprüfen, welche Schlußfolgerungen sich daraus ziehen lassen. Im allgemeinen sind solche Axiome so gehalten, daß sie in sich plausibel sind und daher auch bei den meisten praktischen Anwendungen zutreffen. Das entbindet nicht von der Verpflichtung, ihre Gültigkeit in dem speziellen

Anwendungsgebiet zu überprüfen.

Eine solche Überprüfung ist möglich, wenn wir den Begriff der Wahrscheinlichkeit konkretisieren als erwartete relative Häufigkeit. In diesem Fall läßt sich die Gültigkeit des Additionssatzes beweisen, wie wir im vorigen Abschnitt II.1.b gesehen haben.

Bei anderen Anwendungen des Wahrscheinlichkeitsbegriffs wird man sich von der Gültigkeit des Axioms auf andere Weise vergewissern. Bei der Angabe von subjektiven Wahrscheinlichkeiten kann man es sich selbst zur Regel machen, sie so anzugeben, daß der einfache Additionssatz gilt. Wenn man beispielsweise die subjektive Gewißheit des Ereignisses "Patient ist nach drei Monaten Psychotherapie geheilt" mit 0.50 und die subjektive Wahrscheinlichkeit des Ereignisses "Patient ist nach drei Monaten Psychotherapie nicht geheilt, aber im Zustand gebessert" mit 0.30 angibt, wird man auch sagen: Die subjektive Gewißheit, daß der Patient nach drei Monaten Psychotherapie geheilt oder zumindest im Zustand gebessert ist, beträgt $0.50 + 0.30 = 0.80$. Solange man sich in dieser Art die Regel setzt, alle Wahrscheinlichkeiten so anzugeben, daß der Additionssatz gilt, dann sind die aus dem als Axiom verstandenen Additionssatz abgeleiteten Lehrsätze der Wahrscheinlichkeitstheorie ebenfalls gültig.

Ganz entsprechendes gilt für ein letztes Axiom der mathematischen Wahrscheinlichkeitstheorie, mit dem vorausgesetzt wird, daß die Wahrscheinlichkeit eines sicheren Ereignisses 1 ist. Auch dieses Axiom läßt sich nicht allgemeingültig beweisen, wohl aber für die Konkretisierung der Wahrscheinlichkeit als erwartete relative Häufigkeit. Bei subjektiven Wahrscheinlichkeiten kann man es sich wieder zur Regel machen, sie so anzugeben, daß die Wahrscheinlichkeit eines sicheren Ereignisses 1 ist.

Abschließend können wir die drei Axiome der mathematischen Wahrscheinlichkeitstheorie zusammenstellen

1. Jede Wahrscheinlichkeit ist eine Zahl, die einem Ereignis zugeordnet ist und die nicht kleiner als 0 und nicht größer als 1 ist.
2. Die Wahrscheinlichkeit, daß eines von abzählbar unendlich vielen einander ausschließenden Ereignissen auftritt, ist die Summe der Einzelwahrscheinlichkeiten dieser Ereignisse.
3. Die Wahrscheinlichkeit eines sicheren Ereignisses ist 1.

>>

d) Der erweiterte Additionssatz

Schließen die Ereignisse A und B einander nicht aus, so läßt sich mit Hilfe des einfachen Additionssatzes beweisen, daß

$$p(A \vee B) = p(A) + p(B) - p(A \wedge B)$$

Es ist zu betonen, daß das Wort "oder" im allgemeinen Sprachgebrauch in verschiedener Bedeutung verwendet wird. Wenn man sagt, es gebe für einen Studenten an einem bestimmten Abend nur die Wahl, entweder zu arbeiten oder zu feiern, dann wird das Wort "oder" verwendet, um die Möglichkeit einer gleichzeitigen Realisierung beider Ereignisse auszuschließen. In diesem Fall spricht man vom "exklusivem oder". Wenn wir dagegen von einem Menschen sagen, daß er jeden Samstag tagsüber sein Auto putzt oder abends vorm Fernsehschirm sitzt, dann ist die Möglichkeit, daß er beides tut, in der Regel nicht ausgeschlossen, sondern inbegriffen. Bei einer solchen Verwendung des Wortes "oder", bei der die Möglichkeit der Realisierung beider mit

"oder" verbundenen Ereignisse mit inbegriffen ist, spricht man vom "inkluisiven oder". Der erweiterte Additionssatz gilt für den Fall des "inkluisiven oder".

Für welche der folgenden Wahrscheinlichkeiten kann man den erweiterten Additionssatz verwenden:

a) Wahrscheinlichkeit, daß eine schwarze Kugel (= Ereignis A) oder eine hölzerne Kugel (= Ereignis B) aus der Urne gezogen wird, wobei die Fälle, in denen eine schwarze Holzkugel gezogen wird, mitgemeint sind.

b) Wahrscheinlichkeit, entweder eine schwarze Kugel (= Ereignis A) oder eine hölzerne (= Ereignis B), aber nicht eine schwarze Holzkugel zu ziehen.

Im ersten Fall wird das "oder" durch den nachfolgenden Satz als "inkluisives oder" spezifiziert. Der erweiterte Additionssatz ist also anwendbar. Im zweiten Fall wird das "oder" durch den Zusatz als "exklusives oder" spezifiziert, und daher können wir den erweiterten Additionssatz nicht anwenden.

e) Negation

Die Wahrscheinlichkeit, daß das Ereignis A nicht auftritt, läßt sich angeben als

$$p(\neg A) = 1 - p(A)$$

f) Das Bayes-Theorem

Bei zwei Ereignissen A und B müssen die beiden bedingten Wahrscheinlichkeiten $p(A|B)$ und $p(B|A)$ gut unterschieden werden. Manchmal gibt es Situationen, in denen man die eine kennt, aber die andere braucht. Man möchte sozusagen die bedingte Wahrscheinlichkeit "umdrehen". Dazu dient das Bayes-Theorem. Die einfachste Formel dafür ist

$$p(B|A) = \frac{p(B) \cdot p(A|B)}{p(A)}$$

Weiterhin gilt auch:

$$p(B|A) = \frac{p(B) \cdot p(A|B)}{p(B) \cdot p(A|B) + p(\neg B) \cdot p(A|\neg B)}$$

<< Nach dem erweiterten Multiplikationssatz läßt sich die Wahrscheinlichkeit des Ereignisses $A \wedge B$ darstellen als $p(A \wedge B) = p(A) \cdot p(B|A)$ oder als $p(A \wedge B) = p(B) \cdot p(A|B)$. Da beides für dieselbe Wahrscheinlichkeit steht, muß auch gelten $p(A) \cdot p(B|A) = p(B) \cdot p(A|B)$. Dividieren wir beide Seiten dieser Gleichung durch $p(A)$, dann erhalten wir die erste Gleichung.

Die zweite Gleichung entsteht aus der ersten, indem wir das Ereignis A im Nenner "zerlegen" in die Ereignisse $B \wedge A$ und $\neg B \wedge A$, wie in der ersten Zeile der folgenden Gleichung:

$$\begin{aligned} p(A) &= p((B \wedge A) \vee (\neg B \wedge A)) \\ &= p(B \wedge A) + p(\neg B \wedge A) \\ &= p(B) \cdot p(A|B) + p(\neg B) \cdot p(A|\neg B). \end{aligned}$$

Da die beiden Ereignisse $B \wedge A$ und $\neg B \wedge A$ sich gegenseitig ausschließen, können wir die Oder-Verbindung nach dem einfachen Additionssatz auflösen (zweite Zeile), und dann können wir auf die beiden Und-Verbindungen den erweiterten Multiplikationssatz anwenden (dritte Zeile).

>>

Anwendungsbeispiel: In der Diagnostik stellt sich häufig die Frage, wie groß die Wahrscheinlichkeit eines "latenten Zustands" ist, wenn ein bestimmtes "manifestes Symptom" vorliegt. Z.B. liegt bei vielen "hypermotorischen" Kindern ("Zappelphilipp") eine "minimale cerebrale Dysfunktion (MCD) vor. Dann stellt sich die Frage: Wie groß ist die Wahrscheinlichkeit, daß bei einem hypermotorischen Kind, das in unsere Beratung kommt, eine solche MCD vorliegt. Formalisiert:

Ereignis A = Das zu untersuchende Kind ist hypermotorisch.

Ereignis B = Das Kind hat MCD.

Die bedingte Wahrscheinlichkeit $p(B/A)$ läßt sich nach dem Bayes-Theorem ermitteln, wenn wir die auf der rechten Seite einer der obigen Gleichungen auftretenden Wahrscheinlichkeiten kennen (z.B. aus bisherigen Untersuchungen). Wenn wir z.B. aus vorherigen Untersuchungen wissen, daß 50% der MCD-Kinder und 10% der nicht-MCD-Kinder, die in unsere Beratungsstelle kommen, hypermotorisch sind, und daß 30% der Kinder, die in unsere Beratungsstelle kommen, MCD-Kinder sind, dann können wir aus diesen Angaben berechnen, wie groß die Wahrscheinlichkeit ist, daß ein Kind ein MCD-Kind ist, wenn es hypermotorisch ist. Wir können schreiben:

$$p(B) = 0.30; \text{ also: } p(\neg B) = 0.70$$

$$p(A/B) = 0.50$$

$$p(A/\neg B) = 0.10$$

Also gilt:

$$p(B/A) = \frac{0.30 \cdot 0.50}{0.30 \cdot 0.50 + 0.70 \cdot 0.10} = \frac{15}{22} \approx 0.68$$

Die Wahrscheinlichkeit, daß ein Kind, das hypermotorisch ist, ein MCD-Kind ist, beträgt also 0.68.

<< Das Hauptproblem beim Bayes-Theorem ist meistens die Bestimmung der Wahrscheinlichkeit $p(A)$. Diese Wahrscheinlichkeit durch Zerlegung des Ereignisses A zu bestimmen, ist möglich, führt aber zu einer Gleichung, die nur schwer "im Kopf" anzuwenden ist. Bei einer anderen Formel, der sog. "Quotienten-Version" des Bayes-Theorems, ist das etwas einfacher, und diese Formel ist auch aus anderen Gründen beachtenswert. Wenn wir noch ein drittes Ereignis C in Betrachtung ziehen, dann können wir zunächst in der ersten Formel für das Bayes-Theorem das dortige Ereignis B durch C ersetzen, und dann erhalten wir

$$p(C/A) = \frac{p(C) \cdot p(A/C)}{p(A)}$$

Der "Trick" der Quotientenversion des Bayes-Theorems besteht nun darin, die erste Formel durch die neue zu dividieren, so daß die unbequeme Wahrscheinlichkeit $p(A)$ sich wegekürzt. Dann erhalten wir

$$\frac{p(B/A)}{p(C/A)} = \frac{p(B)}{p(C)} \cdot \frac{p(A/B)}{p(A/C)}$$

Das gilt allgemein für beliebige Ereignisse A , B und C . Am wichtigsten ist der Fall, in dem C die Negation von B ist. Dann können wir in der letzten Gleichung C durch $\neg B$ ersetzen und erhalten

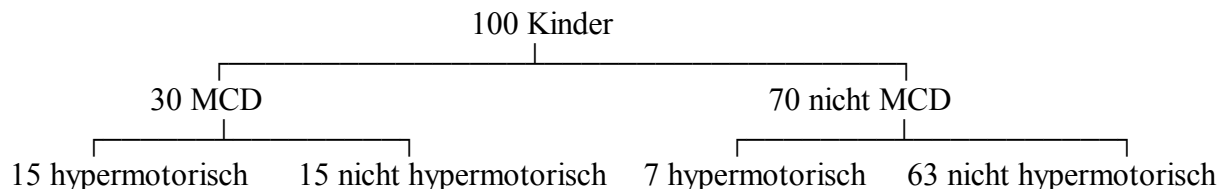
$$\frac{p(B/A)}{p(\neg B/A)} = \frac{p(B)}{p(\neg B)} \cdot \frac{p(A/B)}{p(A/\neg B)}$$

Diese Gleichung wird nun überschaubar und leichter zu merken, wenn man den drei Brüchen Namen gibt. Der einfachste ist derjenige unmittelbar nach dem Gleichheitszeichen. Unter Rückgriff auf den in Abschnitt @B.I.3.d eingeführten Begriff "odds" können wir sagen, daß es sich um die odds von Ereignis B handelt. Auch bei dem Bruch vor dem Gleichheitszeichen handelt es sich um odds. Wir könnten von "bedingten odds" sprechen, aber eine andere Bezeichnung hat sich eher durchgesetzt: Da es sich um die odds "nach Kenntnisnahme des Ereignisses A " handelt, spricht man von "posteriori odds", und entsprechend wird bei dem ersten Bruch hinter dem Gleichheitszeichen von "priori odds" gesprochen, da es sich um die odds vor der Kenntnisnahme von A handelt. Der letzte Bruch hat ebenfalls einen Namen, der vor allem dann verwandt wird, wenn es sich - wie bei dem Beispiel zu Hypermotorik und MCD - bei A um "manifeste Daten" handelt und bei B und $\neg B$ (bzw. C) um "latente Zustände" oder um Hypothesen handelt. Dann nennt man das Verhältnis der Wahrscheinlichkeiten der Daten unter den beiden Hypothesen die "likelihood ratio". Mit diesen Bezeichnungen kann man den Aussagegehalt der Quotienten-Version des Bayes-Theorems in folgendem Satz zusammenfassen:

Die posteriori odds erhält man, indem man die priori odds mit der likelihood ratio multipliziert.

Übungsaufgabe: Bestimmen Sie (aus den zuvor angegebenen Wahrscheinlichkeiten) die priori odds für MCD und die posteriori odds nach Kenntnisnahme von "Hypermotorik". Überzeugen Sie sich, daß nach Umrechnung der posteriori odds in eine Wahrscheinlichkeit (nach der entsprechenden Formel aus Abschnitt @B.I.3.d) dieselbe bedingte Wahrscheinlichkeit herauskommt wie zuvor.

Es gibt noch eine andere Form, die Anwendung des Bayes-Theorems anschaulicher zu machen: Man ersetzt die Wahrscheinlichkeiten durch absolute Häufigkeiten und stellt diese in einem "Baum-Diagramm" der folgenden Art dar:



In Worten: Wenn wir insgesamt 100 Kinder unserer Beratungsstelle betrachten und annehmen, daß sie sich genau im Verhältnis der zugrundegelegten Wahrscheinlichkeiten auf die vier Gruppen verteilen, dann sind darunter 30 MCD-Kinder und 70 Nicht-MCD-Kinder. Die 30 MCD-Kinder teilen sich auf in 15 hypermotorische und 15 nicht hypermotorische, und die 70 Nicht-MCD-Kinder in 7 hypermotorische und 63 nicht hypermotorische. (Überzeugen Sie sich, daß dies den zunächst angenommenen Wahrscheinlichkeiten entspricht!) Wenn nun ein Kind hypermotorisch ist, dann können wir feststellen: Insgesamt haben wir $15 + 7 = 22$ hypermotorische Kinder in der

letzten Zeile. Unter diesen 22 hypermotorischen Kindern sind 15 MCD-Kinder und 7 Nicht-MCD-Kinder. Also beträgt die bedingte Wahrscheinlichkeit von MCD, gegeben Hypermotorik, wieder 15 / 22.

>>

g) Die Anwendung der Wahrscheinlichkeitssätze auf bedingte Wahrscheinlichkeiten

<<

Alle von uns behandelten Wahrscheinlichkeitssätze lassen sich auch auf bedingte Wahrscheinlichkeiten anwenden. Man muß nur die gleiche Bedingung bei allen Wahrscheinlichkeiten anfügen - sowohl auf der linken als auch auf der rechten Seite der Gleichungen, mit denen die Wahrscheinlichkeitssätze formuliert werden.

Beispiel: Den Additionssatz für einander nicht ausschließende Ereignisse kann man auch verwenden, um anzugeben, wie groß die Wahrscheinlichkeit ist, daß entweder Ereignis A oder Ereignis B auftritt, gegeben Ereignis C tritt auf:

$$p(A \vee B / C) = p(A / C) + p(B / C) - p(A \wedge B / C)$$

Wie würde man entsprechend den Negationssatz anwenden, um die bedingte Wahrscheinlichkeit anzugeben, daß Ereignis A nicht auftritt, gegeben Ereignis C tritt auf?

$$p(\neg A / C) = 1 - p(A / C)$$

Bei den Wahrscheinlichkeitssätzen, die schon bedingte Wahrscheinlichkeiten enthalten, würde man eine zusätzliche Bedingung mit "und" an die bereits vorhandene Bedingung anschließen.

Beispiel: Aus dem Multiplikationssatz für abhängige Ereignisse

$$p(A \wedge B) = p(A) \cdot p(B / A)$$

wird bei Hinzufügung der Bedingung C eine Gleichung für die Wahrscheinlichkeit, daß A und B auftreten, gegeben C tritt auf:

$$p(A \wedge B / C) = p(A / C) \cdot p(B / A \wedge C)$$

Was wird aus dem Bayes-Theorem, wenn wir überall die Bedingung C anfügen?

$$p(B / A \wedge C) = \frac{p(B / C) \cdot p(A / B \wedge C)}{p(B / C) \cdot p(A / B \wedge C) + p(\neg B / C) \cdot p(A / \neg B \wedge C)}$$

Anwendungsbeispiel hierfür: Wir greifen auf das in Abschnitt @ f behandelte psychodiagnostische Problem zurück:

Ereignis A = Das zu untersuchende Kind ist hypermotorisch.

Ereignis B = Das Kind hat MCD.

Ereignis C = Es handelt sich um einen Jungen.

Mit Hilfe unseres modifizierten Bayes-Theorems könnten wir nun die Wahrscheinlichkeit angeben, daß ein zu untersuchendes Kind ein MCD-Kind ist, wenn es hypermotorisch und ein Junge ist.

>>

h) Erweiterung von Multiplikationssatz und Additionssatz bei mehr als zwei Ereignissen

<<

Anwendungsbeispiel: In Erziehungsberatungsstellen taucht nicht selten das Problem des Bettnässens auf. Dabei gibt es verschiedene Behandlungsmethoden: In einer rein somatischen Behandlung würde man die Intaktheit der hier relevanten Organe überprüfen und eventuelle Schäden mit medizinischen Methoden zu heilen versuchen. Außerdem würde man in der somatischen Behandlung die Eß-, Trink- und Schlafgewohnheiten des Kindes und seiner Familie

in einem Beratungsgespräch klären, um auf eventuelles Fehlverhalten hinzuweisen. Andererseits kann man in einer psychoanalytischen Behandlung versuchen, durch das Durcharbeiten von unbewußten Konflikten, deren Ausdruck das Bettnässen ist, auch zu einer Heilung des Symptoms zu kommen. Schließlich könnte man ein von der Lernpsychologie vorgeschlagenes Verfahren verwenden, bei dem jede Befeuchtung des Leintuchs einen elektrischen Stromkreis schließt, der dann eine als Wecker wirkende Klingel auslöst. Viele Kinder lernen dann nach einigen Nächten, schon auf Grund der Erwartung des lauten Klingeltons aufzuwachen, noch bevor ein Unglück passiert ist. In einer bestimmten Erziehungsberatungsstelle wurden alle drei Verfahren angewandt: Zunächst eine somatische Behandlung, dann bei Mißerfolg die lernpsychologische "Verhaltenstherapie", und wenn auch diese nicht hilft, wird eine psychoanalytische Behandlung durchgeführt. Es ist zu fragen, wie groß die Wahrscheinlichkeit ist, daß keine dieser Behandlungsmethoden Erfolg hat. Wir definieren: Ereignis A besteht darin, daß die somatische Behandlung erfolglos ist; Ereignis B besteht darin, daß die lernpsychologische Methode keinen Erfolg hat, und Ereignis C besteht darin, daß die psychoanalytische Methode keinen Erfolg hat. Gesucht ist also die Wahrscheinlichkeit, daß die drei Ereignisse A, B und C alle drei auftreten. Diese Wahrscheinlichkeit können wir durch mehrfache Anwendung des Multiplikationssatzes angeben. Wir definieren zunächst: das Ereignis D besteht darin, daß A und B beide eintreffen. Zunächst ist dann

$$p(A \wedge B \wedge C) = p(D \wedge C)$$

denn die Ereignisse, deren Wahrscheinlichkeiten links und rechts vom Gleichheitszeichen angegeben werden, sind offenbar identisch.

Wie könnte man die auf der rechten Seite stehende Wahrscheinlichkeit auflösen in lauter Wahrscheinlichkeiten, in denen nur die Ereignisse A, B und C auftreten?

Nach dem Multiplikationssatz gilt zunächst:

$$p(D \wedge C) = p(D) \cdot p(C/D)$$

Durch nochmalige Anwendung des Multiplikationssatzes erhalten wir:

$$p(D) = p(A \wedge B) = p(A) \cdot p(B/A)$$

Setzen wir diese Wahrscheinlichkeit in die vorhergehende Gleichung ein und bedenken außerdem, daß wir anstelle von $p(C/D)$ auch schreiben können $p(C/A \wedge B)$

Die linke Seite dieser Gleichung ist aber identisch mit der rechten Seite unserer Ausgangsgleichung, und damit erhalten wir:

$$p(A \wedge B \wedge C) = p(A) \cdot p(B/A) \cdot p(C/A \wedge B)$$

Dieses Ergebnis hatte besondere Bedeutung für das Anwendungsbeispiel, von dem wir ausgingen. Die Gesamtwahrscheinlichkeit eines Mißerfolgs ergibt sich nicht einfach aus den Erfolgs- bzw. Mißerfolgswahrscheinlichkeiten der drei Einzelverfahren - also aus den Randwahrscheinlichkeiten $p(A)$, $p(B)$ und $p(C)$: Es sind neben einer Randwahrscheinlichkeit zwei bedingte Wahrscheinlichkeiten im Spiel.

Ist Ihrer Vermutung nach $p(B/A)$ größer oder kleiner als die Randwahrscheinlichkeit $p(B)$?

Die Randwahrscheinlichkeit $p(B)$ wäre die Wahrscheinlichkeit eines Mißerfolgs der lernpsychologischen Methode, wenn diese ohne Vorinformation eines Mißerfolgs der somatischen Methode unter sonst gleichen Bedingungen durchgeführt würde. Die "Verhaltenstherapie" dürfte bei Kindern mit somatischen Schäden recht wenig erfolgreich sein. Wenn man also die Kinder, bei denen in Wirklichkeit eine somatische Behandlung angebracht wäre, ebenfalls mit der Verhaltenstherapie behandeln würde, wäre die Mißerfolgswahrscheinlichkeit $p(B)$ größer, als bei Kindern, bei denen die somatische Behandlung keinen Erfolg brachte, also größer als $p(B/A)$. Entsprechend ist auch die Mißerfolgswahrscheinlichkeit der psychoanalytischen Methode größer, wenn man diese Methode

auch bei Kindern anwendet, bei denen das Bettnässen an somatischen Schäden oder an mangelndem Aufwecktraining liegt, als wenn man diese Kinder vorher ausgeschieden hat. Setzt man außerdem voraus, daß die Psychoanalyse auch in Fällen wirksam ist, in denen die VT nichts ausrichtet (wenn also "die Probleme tiefer liegen"), dann ist $p(C)$ größer als $p(C/A \wedge B)$.

Zusammengefaßt ist also die Mißerfolgswahrscheinlichkeit der kombinierten Behandlung kleiner als das Produkt der Mißerfolgswahrscheinlichkeiten der drei isoliert verwendeten Verfahren.

Zurück zur Wahrscheinlichkeitstheorie. Unsere Gleichung für $p(A \wedge B \wedge C)$ gilt natürlich ganz allgemein für das Eintreffen dreier Ereignisse A, B und C. Entsprechend würde man erhalten, wenn es um vier Ereignisse A, B, C und D geht:

$$p(A \wedge B \wedge C \wedge D) = p(A) \cdot p(B/A) \cdot p(C/A \wedge B) \cdot p(D/A \wedge B \wedge C)$$

Ganz allgemein erhält man die Wahrscheinlichkeit, daß mehrere Ereignisse alle eintreffen, indem man ein Produkt aus einer unbedingten und mehreren bedingten Wahrscheinlichkeiten bildet. Für jedes der Ereignisse wird dabei einmal die Wahrscheinlichkeit angegeben, und bei allen folgenden Wahrscheinlichkeiten taucht das Ereignis dann als Bedingung auf. Soweit nach dieser Regel mehrere Ereignisse als Bedingung auftauchen (also vom dritten Faktor des Produkts an), sind diese mit "und" verbunden. Die Reihenfolge, in der die Ereignisse in dieser Kette auftauchen, ist übrigens beliebig, solange für jedes Ereignis einmal die Wahrscheinlichkeit angegeben wird und es dann in den folgenden Wahrscheinlichkeiten als Bedingung auftritt. Wir können also auch schreiben:

$$p(A \wedge B \wedge C) = p(B) \cdot p(A/B) \cdot p(C/A \wedge B)$$

Sind die Ereignisse voneinander unabhängig, so kann man auf die Angabe der Bedingung verzichten und erhält das Produkt der Randwahrscheinlichkeiten.

Genügt hier paarweise Unabhängigkeit oder müssen die Ereignisse vollständig unabhängig sein?

Wenn wir nur die Randwahrscheinlichkeiten einsetzen und auf die Angabe der Bedingungen verzichten, dann setzen wir voraus, daß nicht nur eine Bedingung über ein einzelnes Ereignis, sondern auch eine Bedingung über das Eintreffen mehrerer Ereignisse an der Wahrscheinlichkeit eines weiteren Ereignisses nichts ändert, und diese Voraussetzung ist nicht erfüllt, wenn nur paarweise Unabhängigkeit besteht.

Wir können also festhalten:

Sind mehrere Ereignisse vollständig unabhängig, so ist die Wahrscheinlichkeit des Auftretens aller Ereignisse gleich dem Produkt der Randwahrscheinlichkeiten.

Ähnlich kann man auch den Additionssatz auf mehr als zwei Ereignisse ausdehnen. Bei drei Ereignissen könnte man folgendes Venn-Diagramm zeichnen:

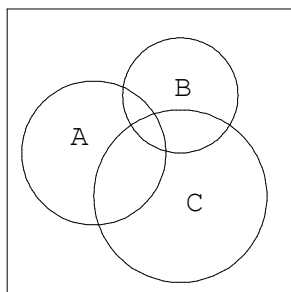


Abb. 2: Venn-Diagramm für drei Ereignisse

Versuchen Sie, die Wahrscheinlichkeit $p(A \vee B \vee C)$ anzugeben!

Als Ergebnis erhalten wir

$$\begin{aligned} p(A \vee B \vee C) &= p(A) + p(B) + p(C) \\ &\quad - p(A \wedge B) - p(A \wedge C) - p(B \wedge C) \\ &\quad + p(A \wedge B \wedge C) \end{aligned}$$

Wir wählen die Fläche des Gesamtdiagramms als Einheit der Flächenangabe, damit jede Teilfläche gleich der entsprechenden Wahrscheinlichkeit ist. Dann können wir die obige Gleichung folgendermaßen begründen: Wenn wir die Summe $p(A) + p(B) + p(C)$ bilden, dann sind alle Fälle (d.h. alle Teilflächen), bei denen zwei oder drei Ereignisse vorliegen, doppelt verrechnet, und die Fälle, in denen die drei Ereignisse A, B und C alle drei vorliegen, sind sowohl in $p(A)$ als auch in $p(B)$ als auch in $p(C)$ enthalten, also dreifach verrechnet. Subtrahieren wir jetzt in der zweiten Zeile die Wahrscheinlichkeiten, daß A und B, A und C bzw. B und C vorliegen, dann sind die bisher doppelt verrechneten Fälle berichtigt - genau wie beim Additionssatz für zwei Ereignisse.

Wie oft sind jetzt die Fälle verrechnet, in denen alle drei Ereignisse auftreten?

Sie sind sowohl in $p(A \wedge B)$ als auch in $p(A \wedge C)$ als auch in $p(B \wedge C)$ enthalten. In der ursprünglichen Summe waren sie dreifach verrechnet, und in der zweiten Zeile haben wir sie 3 mal subtrahiert. Also sind sie jetzt gar nicht mehr in dem Ausdruck enthalten, und daher müssen wir sie in einem weiteren Summanden wieder addieren.

Zu dem gleichen Ergebnis kämen wir auf einem ganz anderen Weg, der dem beim Multiplikationssatz angewandten entspricht.

Wir definieren:

Das Ereignis D besteht darin, daß entweder A oder B zutrifft. Dann erhalten wir:

$$p(A \vee B \vee C) = p(D \vee C) = p(D) + p(C) - p(D \wedge C)$$

Versuchen Sie, diesen Ausdruck selbst solange umzuformen, bis die rechte Seite der obigen Gleichung entsteht!

Die Wahrscheinlichkeit des Ereignisses D können wir nach dem Additionssatz für zwei Ereignisse bestimmen:

$$p(D) = p(A \vee B) = p(A) + p(B) - p(A \wedge B)$$

Etwas schwieriger ist die Bestimmung der Wahrscheinlichkeit $p(D \wedge C)$. Nach dem Multiplikationssatz können wir zunächst schreiben

$$p(D \wedge C) = p(C) \cdot p(D/C)$$

Da das Ereignis D aber identisch ist mit "A oder B" können wir auch sagen:

$$p(D/C) = p(A \vee B/C)$$

Jetzt erinnern wir uns daran, daß man den Additionssatz auch für bedingte Wahrscheinlichkeiten verwenden kann, und erhalten:

$$\begin{aligned} p(D/C) &= p(A \vee B/C) \\ &= p(A/C) + p(B/C) - p(A \wedge B/C) \end{aligned}$$

Setzen wir diesen Ausdruck anstelle von $p(D/C)$ weiter oben ein, so ergibt sich:

$$\begin{aligned} p(D \wedge C) &= p(C) \cdot [p(A/C) + p(B/C) - p(A \wedge B/C)] \\ &= p(C) \cdot p(A/C) + p(C) \cdot p(B/C) - p(C) \cdot p(A \wedge B/C) \\ &= p(A \wedge C) + p(B \wedge C) - p(A \wedge B \wedge C) \end{aligned}$$

Das letzte Gleichheitszeichen erhalten wir durch Anwendung des Multiplikationssatzes auf die Produkte der vorhergehenden Zeile. Setzen wir nun diese Wahrscheinlichkeit $p(D \wedge C)$ und die schon vorher abgeleitete Wahrscheinlichkeit $p(D)$ in die erste Gleichung ein, so erhalten wir

$$\begin{aligned} p(A \vee B \vee C) &= p(A) + p(B) - p(A \wedge B) + p(C) \\ &\quad - [p(A \wedge B) + p(A \wedge C) - p(A \wedge B \wedge C)] \end{aligned}$$

Aus dieser Gleichung ergibt sich die anhand des Venn-Diagramms hergeleitete Gleichung durch Auflösung der eckigen Klammer und Vertauschung der Reihenfolge der Summanden.

Auf ähnlichem Weg, also entweder mit dem Venn-Diagramm oder auf algebraischem Weg, könnten wir auch die Wahrscheinlichkeiten herleiten, daß irgendeines der vier Ereignisse W, X, Y und Z eintritt:

$$\begin{aligned} p(W \vee X \vee Y \vee Z) &= p(W) + p(X) + p(Y) + p(Z) \\ &\quad - p(W \wedge X) - p(W \wedge Y) - p(W \wedge Z) - p(X \wedge Y) - p(X \wedge Z) - p(Y \wedge Z) \\ &\quad + p(W \wedge X \wedge Y) + p(W \wedge X \wedge Z) + p(W \wedge Y \wedge Z) + p(X \wedge Y \wedge Z) \\ &\quad - p(W \wedge X \wedge Y \wedge Z) \end{aligned}$$

In Worten - und verallgemeinert auf beliebig viele Ereignisse: Die Wahrscheinlichkeit, daß eines von mehreren Ereignissen auftritt, erhalten wir, indem wir zunächst die Randwahrscheinlichkeiten der Ereignisse addieren. Dann subtrahieren wir die Wahrscheinlichkeiten aller Kombinationen von zwei Ereignissen, addieren die Wahrscheinlichkeiten aller Kombinationen von drei Ereignissen, subtrahieren die Wahrscheinlichkeiten aller Kombinationen von vier Ereignissen u.s.w D.h.: Die Wahrscheinlichkeiten aller Kombinationen einer ungeraden Zahl von Ereignissen müssen addiert und die Wahrscheinlichkeiten aller Kombinationen einer geraden Zahl von Ereignissen müssen subtrahiert werden.

Was wird aus diesem Additionssatz, wenn sich die Ereignisse alle gegenseitig ausschließen?
In diesem Falle haben alle Kombinationen aus mehreren Ereignissen eine Wahrscheinlichkeit von 0 und damit ergibt sich der

Additionssatz für mehrere einander ausschließende Ereignisse:

Die Wahrscheinlichkeit, daß eines von mehreren einander ausschließenden Ereignissen eintritt, ist die Summe der Randwahrscheinlichkeiten dieser Ereignisse.

Ein Anwendungsbeispiel für diesen Additionssatz werden wir im nächsten Abschnitt kennenlernen.

Wie wir in Abschnitt C gesehen haben, ist die Gültigkeit des Additionssatzes für einander ausschließende Ereignisse in der mathematischen Statistik ein Axiom. Genauer gesagt ist es ein Axiom, daß er für "abzählbar unendlich viele" Ereignisse gilt. Daraus läßt sich bereits beweisen, daß er für endlich viele Ereignisse gilt.

>>

i) Ein Beispiel für die kombinierte Anwendung mehrerer Wahrscheinlichkeitssätze

Wie bereits in Abschnitt @B.I.3.c angekündigt, können wir uns nun der folgenden Frage zuwenden: Wenn ein Multiple-Choice-Test aus fünf Aufgaben besteht, bei denen jeweils sechs Alternativen zur Wahl gestellt werden, und eine Vp einfach würfelt, welche sie ankreuzen soll, wie groß ist dann die Wahrscheinlichkeit, daß sie zwei richtige Lösungen erzielt?

Wollen Sie versuchen, diese Wahrscheinlichkeit selbst herzuleiten?

Wenn man ein solches Problem nicht auf Anhieb lösen kann, dann lohnt es sich oft, zu fragen, für welche einfachen Ereignisse, die mit dem fraglichen Ereignis etwas zu tun haben, man denn eine Wahrscheinlichkeit angeben kann. Hier könnte man zunächst fragen:

Wie groß ist die Wahrscheinlichkeit, daß die Vp die erste Aufgabe richtig löst?

Diese Wahrscheinlichkeit beträgt offenbar 1/6, denn unabhängig davon, welche Alternative die richtige ist, beträgt die Wahrscheinlichkeit, daß gerade die Zahl gewürfelt wird, die der richtigen

Alternative entspricht, $1/6$. Wir können als Formel schreiben:

$$p(\text{Aufgabe 1 richtig gelöst}) = 1/6$$

Zur Vereinfachung der Schreibweise wollen wir für das Ereignis "Vp löst Aufgabe 1 richtig" das Symbol R_1 verwenden. Entsprechend bezeichnet R_2 das Ereignis "Vp löst Aufgabe 2 richtig" usw.

Wie würde man die eben ermittelte Wahrscheinlichkeitsangabe mit diesen Symbolen schreiben?

$$p(R_1) = 1/6$$

Entsprechendes gilt natürlich für die 2. Aufgabe:

$$p(R_2) = 1/6$$

und allgemein für die i -te Aufgabe:

$$p(R_i) = 1/6 \quad (i = 1 \text{ bis } 5)$$

Für das Ereignis "Vp löst Aufgabe 1 falsch" wollen wir entsprechend das Symbol F_1 verwenden und entsprechend F_2, F_3 usw. für das Falschlösen der übrigen Aufgaben.

Können Sie die Wahrscheinlichkeit des Ereignisses F_1 mit Hilfe eines uns bekannten Wahrscheinlichkeitssatzes aus der Wahrscheinlichkeit des Ereignisses " R_1 " herleiten?

Das Ereignis ist gleichbedeutend mit "nicht R_1 ". Nach dem Negationssatz ergibt sich daher:

$$p(F_1) = p(\neg R_1) = 1 - p(R_1) = 1 - 1/6 = 5/6$$

Allgemein gilt für die i -te Aufgabe:

$$p(F_i) = 5/6 \quad (i = 1, \dots, 5)$$

Wir haben damit die Wahrscheinlichkeit jedes einzelnen Aufgabenausgangs bestimmt. Wir wenden uns nun der Wahrscheinlichkeit bestimmter Muster von richtigen und falschen Lösungen zu. Wir können uns zum Beispiel fragen: Wie groß ist die Wahrscheinlichkeit, daß die Vp die ersten beiden Aufgaben richtig löst und die letzten drei falsch?

Wie könnte man dieses Ereignis mit den von uns definierten Symbolen schreiben?

Das Ereignis "Vp löst die ersten beiden Aufgaben richtig und die letzten drei falsch" ist identisch mit:

$$R_1 \wedge R_2 \wedge F_3 \wedge F_4 \wedge F_5$$

Wir können uns nun leicht überzeugen, daß die Ereignisse R_1, R_2, F_3, F_4, F_5 vollständig unabhängig voneinander sind. Jedes Ereignis bezieht sich ja auf einen anderen Wurf, und verschiedene Würfe eines Würfels sind vollständig unabhängig voneinander (vgl. auch Seite @). Die Wahrscheinlichkeit des Ereignisses "Vp löst die ersten beiden Aufgaben richtig und die letzten drei falsch" können wir demnach aufgrund des Multiplikationssatzes für mehrere vollständig unabhängige Ereignisse bestimmen:

$$\begin{aligned} p(R_1 \wedge R_2 \wedge F_3 \wedge F_4 \wedge F_5) &= p(R_1) \cdot p(R_2) \cdot p(F_3) \cdot p(F_4) \cdot p(F_5) \\ &= 1/6 \cdot 1/6 \cdot 5/6 \cdot 5/6 \cdot 5/6 \\ &= (1/6)^2 \cdot (5/6)^3 \end{aligned}$$

Damit haben wir die Wahrscheinlichkeit eines ganz bestimmten Lösungsmusters mit zwei richtigen Lösungen festgestellt. Unsere Ausgangsfrage lautete aber: Wie groß ist die Wahrscheinlichkeit, bei diesem "Test" zwei richtige Lösungen zu erreichen? - und zwei richtige Lösungen kann man auch mit anderen Mustern von richtigen und falschen Lösungen erzielen - z.B. indem man die erste und dritte Aufgabe richtig löst und die drei restlichen falsch. Bestimmen wir auch die Wahrscheinlichkeit dieses Lösungsmusters:

$$\begin{aligned} p(R_1 \wedge F_2 \wedge R_3 \wedge F_4 \wedge F_5) &= p(R_1) \cdot p(F_2) \cdot p(R_3) \cdot p(F_4) \cdot p(F_5) \\ &= 1/6 \cdot 5/6 \cdot 1/6 \cdot 5/6 \cdot 5/6 \\ &= (1/6)^2 \cdot (5/6)^3 \end{aligned}$$

Wir erhalten also die gleiche Wahrscheinlichkeit für dieses Lösungsmuster wie für das erste - und dieselbe Wahrscheinlichkeit würden wir für jede andere Abfolge von zwei richtigen und drei falschen Lösungen erhalten: Die Wahrscheinlichkeit eines jeden solchen Lösungsmusters besteht - genau wie in unseren Beispielen - aus einem Produkt, in welchem den zwei richtigen Lösungen je einmal der Bruch 1/6 und den drei falschen je einmal der Bruch 5/6 entspricht.

Nach welchem Wahrscheinlichkeitssatz würden wir die Wahrscheinlichkeit bestimmen, daß irgendeines dieser Lösungsmuster mit zwei richtigen und drei falschen Lösungen auftritt?

Wir können jedes dieser Lösungsmuster als ein "Ereignis" auffassen. Gesucht ist dann die Wahrscheinlichkeit, daß irgendeines von mehreren Ereignissen eintritt - also das Ereignis " R_1, R_2, F_3, F_4, F_5 " oder das Ereignis " R_1, F_2, R_3, F_4, F_5 " oder.....oder.....oder.....

Das läuft offenbar auf die Anwendung des Additionssatzes hinaus. In einer solchen Situation fragt man natürlich:

Schließen sich die Ereignisse, deren Wahrscheinlichkeiten mit dem Additionssatz kombiniert werden sollen, gegenseitig aus?

Es ist nicht möglich, sowohl das Lösungsmuster " R_1, R_2, F_3, F_4, F_5 " als auch das Lösungsmuster " R_1, F_2, R_3, F_4, F_5 " zu erzielen. Diese beiden Ereignisse schließen sich also gegenseitig aus - und dasselbe ist offenbar für alle übrigen Paare von Lösungsmustern der Fall. Wir hätten also den einfachen Additionssatz anzuwenden. Wir wissen weiter, daß die Randwahrscheinlichkeiten der verschiedenen mit dem Additionssatz zu kombinierenden Ereignisse (d.h. der verschiedenen Lösungsmuster) alle gleich sind; und dann können wir anstelle einer Addition von lauter gleichen Wahrscheinlichkeiten auch schreiben:

$$(1/6)^2 \cdot (5/6)^3 + (1/6)^2 \cdot (5/6)^3 + \dots \text{ usw. für jedes Lösungsmuster einmal}$$

$$= (1/6)^2 \cdot (5/6)^3 \cdot (\text{Anzahl der möglichen Abfolgen von 2 richtigen und 3 falschen Lösungen}).$$

Wir brauchen also nur noch die Anzahl der möglichen Abfolgen von zwei richtigen und drei falschen Lösungen zu bestimmen. In Abschnitt @B.II.2.c werden wir eine Formel kennenlernen, mit der man diese Anzahl bestimmen kann. Im vorliegenden Beispiel ist es aber auch möglich, diese Abfolgen systematisch hinzuschreiben und auszuzählen. ("Systematisch" ist erforderlich, damit wir keine Möglichkeit übersehen.) Ein solches System wäre: Wir schreiben zunächst alle

"Muster" hin, bei denen die erste richtige Lösung in der ersten Aufgabe auftaucht, dann die Muster mit der ersten richtigen Lösung in der zweiten Aufgabe usw. Wir erhalten:

R_1, R_2, F_3, F_4, F_5 R_1, F_2, R_3, F_4, F_5 R_1, F_2, F_3, R_4, F_5 R_1, F_2, F_3, F_4, R_5
 F_1, R_2, R_3, F_4, F_5 F_1, R_2, F_3, R_4, F_5 F_1, R_2, F_3, F_4, R_5
 F_1, F_2, R_3, R_4, F_5 F_1, F_2, R_3, F_4, R_5
 F_1, F_2, F_3, R_4, R_5

Zählen wir ab, so erhalten wir insgesamt 10 Muster mit zwei richtigen Lösungen. Damit können wir die Wahrscheinlichkeit, daß die Zahl der richtigen Lösungen (X) gleich 2 ist, angeben als

$$p(X=2) = 10 \cdot (1/6)^2 \cdot (5/6)^3 = 0.1608.$$

Wir haben damit die gesuchte Wahrscheinlichkeit, in dem genannten Test durch Würfeln zwei richtige Lösungen zu erzielen.

<<

Wir werden dieses Ergebnis im folgenden weiter verwenden. Außerdem veranschaulicht die Aufgabe einige Prinzipien, die bei der Berechnung der Wahrscheinlichkeit komplexer Ereignisse zu beachten sind, insbesondere:

1. Bei solchen Aufgaben ist es oft nötig, mehrere Wahrscheinlichkeitssätze ineinander zu verschachteln.
2. Häufig sieht man es der Formulierung des Problems nicht unmittelbar an, welche Wahrscheinlichkeitssätze zum Einsatz kommen.

Kommen in der Formulierung des Problems die Worte "und", "oder" oder "nicht" vor, dann liegt die Vermutung nahe, daß man den Multiplikationssatz (bei "und"-Formulierung), den Additionssatz (bei "oder"-Formulierungen) oder den Negationssatz (bei "nicht"-Formulierungen) verwenden kann. Das ist aber nur eine Vermutung - denn man muß zur Verwendung dieser Sätze die Wahrscheinlichkeiten der durch "und", "oder" oder "nicht" verknüpften Ereignisse angeben können.

Ist der optimale Ansatz aus der Formulierung des Problems nicht unmittelbar ersichtlich, dann kommt es darauf an, das Problem so umzuformulieren, daß es in der beschriebenen Weise durchsichtig wird. Manchmal gelingt einem das auf Anhieb. Vielleicht haben Sie selbst zu Beginn das Problem sofort umformuliert in die Frage: Wie groß ist die Wahrscheinlichkeit, entweder die beiden ersten Aufgaben oder die erste und dritte oder die erste und vierte usw. richtig zu lösen? Wenn man mit dieser Umformulierung angefangen hat, dann geht man in der Regel auch weiter "analytisch" vor: Man löst dann das Ereignis "erste zwei Aufgaben richtig gelöst" auf in die von uns als R_1, R_2, F_3, F_4, F_5 bezeichneten Einzelereignisse, deren Wahrscheinlichkeit man leicht angeben und mit Hilfe von Multiplikations- und Additionssatz zu der Gesamtwahrscheinlichkeit kombinieren kann. Ob man in dieser Weise "auf Anhieb" den richtigen Ansatz zur Auflösung findet, ist z.T. eine Frage der Übung im Umgang mit Wahrscheinlichkeiten; aber auch dem versierten Wahrscheinlichkeitsrechner kommen Probleme in die Hand, die er nicht sofort analytisch lösen, also in eine "und", "oder", "nicht"-Formulierung von Ereignissen mit bekannter (oder leicht zu bestimmender) Wahrscheinlichkeit umwandeln kann. In einem solchen Fall würde man in der beschriebenen Weise vorgehen und fragen: Für welche einfachen Ereignisse, die mit dem zu tun haben, um das es geht, läßt sich eine Wahrscheinlichkeit angeben? (In unserem Fall waren das die Ereignisse R_1, F_2, R_3, F_4, F_5 , usw.). Man kann sich dann weiter fragen, ob und ggf. wie man aus den Ereignissen mit bekannter Wahrscheinlichkeit das Ereignis, dessen Wahrscheinlichkeit gefragt ist, kombinieren kann. Ist dies nicht möglich, dann kann man sich

doch fragen: Für welche etwas komplexeren Ereignisse, die dem Problem schon etwas näher kommen, können wir die Wahrscheinlichkeit angeben? (In unserem Fall konnten wir die Wahrscheinlichkeit eines solchen Musters von Richtig- und Falschlösungen mit Hilfe des Multiplikationssatzes bestimmen). Hat man auf diese Weise einen besseren Einblick in die Wahrscheinlichkeitsstruktur der Situation bekommen, dann fällt es meist leichter, nun das komplexe Ereignis, dessen Wahrscheinlichkeit bestimmt werden soll, so umzuformulieren, daß eine "und", "oder" oder "nicht"-Verbindung von einfachen oder zumindest weniger komplexen Ereignissen auftritt, deren Wahrscheinlichkeit man bestimmen kann.

Zusammengefaßt: Man wird bei der Lösung von Wahrscheinlichkeitsproblemen entweder vom Problem ausgehen und dieses direkt vereinfachen durch Umwandlung in eine Formulierung mit "und", "oder" oder "nicht"; wenn das nicht weiterführt, wird man herauszufinden versuchen, für welche einfacheren Ereignisse man Wahrscheinlichkeiten angeben kann, um sich so mit der Problemlage vertraut zu machen. In Frageform lauten die beiden Ansätze: In welche einfach strukturierten Einzelereignisse kann ich das Gesamtergebnis auflösen? Bzw.: Wie kann ich Ereignisse, deren Wahrscheinlichkeiten ich schon kenne, zu dem fraglichen Ergebnis zusammenfügen? In der Praxis verlaufen natürlich beide Denkprozesse parallel bzw. abwechselnd. Abschließend wollen wir uns vergegenwärtigen, daß dies natürlich kein automatisches Erfolgsrezept ist; aber man kann dem produktiven Einfall, der die Situation klarer macht, doch mit den erwähnten Methoden etwas nachhelfen. Zum Abschluß einer solchen Analyse tut man gut daran, sich einige kritische Fragen zu stellen; insbesondere die folgenden drei:

- Habe ich mit meiner Umformulierung alle Möglichkeiten der Realisierung des Ereignisses erfaßt, um das es geht?
- Sind irgendwelche Möglichkeiten doppelt oder mehrfach erfaßt?
- Sind in meiner Umformulierung irgendwelche Ereignisse enthalten, die in dem Ereignis, um das es geht, nicht enthalten sein sollen?

Die Bedeutung der zweiten Frage haben wir schon im Zusammenhang mit dem Additionssatz für einander nicht ausschließende Ereignisse kennengelernt: Die erste Frage dient vor allem der Vermeidung eines häufig gemachten Fehlers: Man bestimmt nur die Wahrscheinlichkeit einer einzigen Realisationsmöglichkeit des problematischen Ereignisses und gibt sich damit zufrieden. (In unserem Beispiel läge dieser Fehler vor, wenn wir nur die Wahrscheinlichkeit des Lösungsmusters R_1, R_2, F_3, F_4, F_5 bestimmt hätten. Dies ist zwar eine Möglichkeit zwei richtige Lösungen zu erzielen, aber die Wahrscheinlichkeit dieses Lösungsmusters ist noch nicht die gesuchte, weil es eben noch andere Möglichkeiten gibt, zwei richtige Lösungen zu erzielen. Als Faustregel könnte man sich merken: Wenn ich nur die Wahrscheinlichkeit eines einzigen Realisierungsbeispiels des kritischen Ereignisses angebe, bin ich noch nicht weit genug. Selbst wenn alle übrigen Realisierungsmöglichkeiten des kritischen Ereignisses eine Wahrscheinlichkeit von null haben, so daß das gewählte Beispiel alle Möglichkeiten erschöpft, müßte man dies noch gesondert nachweisen.

Auch die dritte Frage (habe ich irgendwelche Ereignisse erfaßt, die nicht erfaßt werden sollten) dient der Vermeidung eines häufig gemachten Fehlers. Bei unserem Beispiel könnte es vorkommen, daß man sagt: Die Wahrscheinlichkeit, die erste und die zweite Aufgabe richtig zu lösen, bestimme ich nach dem Multiplikationssatz als:

$$p(R_1 \wedge R_2) = p(R_1) \cdot p(R_2) = 1/6 \cdot 1/6$$

Diese Aussage ist zwar richtig, aber sie enthält auch Fälle, in denen nicht nur die ersten beiden, sondern auch noch weitere Aufgaben richtig gelöst werden. Wir haben diesen Fehler vermieden, indem wir noch zusätzlich gefordert haben, daß die drei übrigen Aufgaben falsch gelöst werden. An diesem Beispiel sieht man auch, daß es sich lohnen kann, die genannten kritischen Fragen

nicht nur am Ende der Analyse, sondern auch nach jedem wichtigen Zwischenschritt zu stellen. Unser Beispiel eines mit Würfeln gelösten Tests veranschaulichte neben der häufigen Notwendigkeit der Anwendung von mehreren Wahrscheinlichkeitssätzen und der Frage nach einer günstigen Strategie zur Umformung von Problemen, die nicht in "und"-, "oder"- oder "nicht"-Verbindungen von Ereignissen mit bekannter Wahrscheinlichkeit bestehen, noch ein drittes: Bei vielen Wahrscheinlichkeitsaussagen kommt es darauf an zu bestimmen, wie viele mögliche Ereignisse einer bestimmten Art es gibt. In unserem Beispiel war zu bestimmen, wie viele Abfolgen von zwei richtigen und drei falschen Lösungen es gab. Wir konnten diese Anzahl bestimmen, indem wir alle Möglichkeiten systematisch aufschrieben und dann abzählten. Dieses Verfahren ist aber sehr umständlich, wenn die Zahl der Möglichkeiten größer ist als hier, und solche Probleme gibt es in der Wahrscheinlichkeitsrechnung recht häufig. Daher wurden verschiedene Formeln entwickelt, mit denen man solche Anzahlen möglicher Ereignisse bestimmen kann. Diese Methoden sind Gegenstand des nächsten Abschnitts.

>>

2) Anzahlen

In der in Abschnitt @B.II.1.i behandelten Aufgabe war zu bestimmen, wie viele Abfolgen von zwei richtigen und drei falschen Lösungen es gab. Wir konnten diese Anzahl bestimmen, indem wir alle Möglichkeiten systematisch aufschrieben und dann abzählten. Dieses Verfahren ist aber sehr umständlich, wenn die Zahl der Möglichkeiten größer ist als hier, und solche Probleme gibt es in der Wahrscheinlichkeitsrechnung recht häufig. Es geht dabei jeweils um die Frage, wie viele Ereignisse einer bestimmten Art es gibt. Dafür wurden - je nach Art der Fragestellung - verschiedene Formeln entwickelt, die in den folgenden Unterabschnitten behandelt werden.

a) Permutationen

Es gibt viele Fragen nach Anzahlen, die sich reduzieren lassen auf die Frage: Wie viele Möglichkeiten gibt es, eine bestimmte Zahl (N) von Elementen, die alle voneinander verschieden sind, in eine Reihenfolge zu bringen.

Beispiel: Um zu überprüfen, ob das Urteil von Lehrern über die Intelligenz ihrer Schüler etwas mit der von Intelligenztests gemessenen Intelligenz zu tun hat, wird ein Lehrer gebeten, fünf Schüler in eine Rangfolge hinsichtlich der mutmaßlichen Intelligenz zu bringen. Die Schüler werden außerdem mit Intelligenztests getestet. Nehmen wir einmal an, die vom Lehrer angegebene Rangfolge stimme genau mit der Rangfolge überein, die sich aufgrund des Intelligenztests ergibt. Man wird dann zu fragen haben, ob das noch Zufall sein kann.

Bei zwei Schülern beträgt die Wahrscheinlichkeit, die richtige Reihenfolge durch Zufall anzugeben, offenbar 50%, denn für zwei Schüler A und B gibt es nur zwei mögliche Rangfolgen: A B oder B A. Bei drei Schülern gibt es schon sechs mögliche Rangfolgen: Wir können in den beiden möglichen Rangfolgen von A und B den Schüler C an drei verschiedenen Stellen einschieben; so daß er erster, zweiter oder dritter wird. Aus der Rangfolge A B für zwei Schüler

ergeben sich somit die folgenden drei Rangfolgen von drei Schülern:

C A B (= Einfügen von C als 1.)

A C B (= " " C " 2.)

und

A B C (= " " C " 3.)

Welche Rangfolgen ergeben sich entsprechend aus B A?

Es ergeben sich:

C B A

B C A

und

B A C

Damit sind offenbar¹⁰ alle Möglichkeiten einer Rangfolge von drei Elementen A, B und C erschöpft. Insgesamt gibt es also sechs mögliche Rangfolgen der Buchstaben A, B und C.

Wir können nun fortfahren und auf ähnliche Weise bestimmen, wie viele Rangfolgen von vier Elementen A, B, C und D es gibt. Wir könnten uns zunächst fragen:

Wie viele Rangfolgen kann man aus herleiten, indem man den Buchstaben D an unterschiedlichen Stellen der Folge A B C einfügt?

Offenbar können wir den Buchstaben D so einfügen, daß er an 1., 2., 3. oder 4. Stelle zu stehen kommt. So entstehen die Rangfolgen

D A B C (D an die 1. Stelle gesetzt)

A D B C (D " " 2. " ")

A B D C (D " " 3. " ")

und

A B C D (D " " 4. " ")

Wir könnten in ähnlicher Weise alle Rangfolgen aufzählen, die sich durch Einfügung von D in die Folge A C B erzielen lassen usw. Aber es geht uns ja nicht um das Aufzählen der möglichen Rangfolgen.

Können Sie auch ohne Aufzählen angeben, wie viele Reihenfolgen der Buchstaben A, B, C und D es gibt?

Aus jeder der sechs möglichen Reihenfolgen der Buchstaben A, B und C können wir durch Einfügen von D an vier verschiedenen Stellen neue Rangfolgen machen. Also gibt es $6 \cdot 4 = 24$ mögliche Reihenfolgen der Buchstaben A, B, C und D.

Jetzt können wir auch - wieder nach demselben System - angeben, wie viele mögliche

¹⁰Man kann es auch beweisen. Man braucht sich nur zu vergegenwärtigen, daß man jede Rangfolge von drei Elementen A, B und C darstellen kann durch Einfügen des Elements C in eine Rangfolge von A und B. Da wir alle diese Einfügungsmöglichkeiten erschöpft haben, haben wir auch alle möglichen Rangfolgen aufgezählt. Entsprechendes gilt für die späteren Aussagen über die Vollzähligkeit aller Rangfolgen von mehreren Elementen.

Rangfolgen von fünf Schülern A, B, C, D und E es gibt.

In jeder der 24 Reihenfolgen der Buchstaben A, B, C und D kann man den Buchstaben E an 5 verschiedenen Stellen einsetzen. So entstehen $24 \cdot 5 = 120$ mögliche Rangfolgen der Schüler. Es ist demnach ziemlich unwahrscheinlich, daß unser Lehrer nur durch Zufall die richtige Rangfolge getroffen hat: Bestünde kein Zusammenhang zwischen Lehrerurteil und Intelligenztest, dann wäre jede andere Rangfolge genauso wahrscheinlich wie die vom Intelligenztest angegebene; jede einzelne hätte also die Wahrscheinlichkeit $1/120$. In der Sprache der schlußfolgernden Statistik ausgedrückt: Unter der Hypothese, daß überhaupt kein Zusammenhang zwischen Lehrerurteil und Intelligenztest besteht, beträgt die Wahrscheinlichkeit, daß unsere Daten (nämlich die völlige Übereinstimmung von Lehrerurteil und Intelligenztests) nur einen Zufallstreffer darstellen, $1/120$, also weniger als 1%. Wir können daher die Hypothese eines fehlenden Zusammenhangs verwerfen. (Diese Technik des Verwerfens einer Hypothese aufgrund von Daten, die nur mit sehr geringer Wahrscheinlichkeit zufällig entstanden sein könnten, wenn die Hypothese gilt, werden wir noch häufig verwenden).

Kehren wir zu unserem allgemeineren Problem zurück. Wir können recht allgemein sagen: Kennen wir die Zahl der möglichen Reihenfolgen von $N-1$ Elementen, dann können wir die Zahl der möglichen Reihenfolgen von N Elementen durch Multiplikation mit N ermitteln. Da es für ein Element nur eine mögliche Reihenfolge gibt, gibt es für zwei Elemente $1 \cdot 2 = 2$, für 3 Elemente $1 \cdot 2 \cdot 3$ und allgemein für N Elemente $1 \cdot 2 \cdot 3 \cdot 4$ usw. bis N mögliche Reihenfolgen.

Für das dabei auftretende Produkt aller ganzen Zahlen von 1 bis N ist in der Mathematik das Symbol " $N!$ " eingeführt.

Definition: $N!$ (gesprochen: N Fakultät) ist das Produkt aller ganzen Zahlen von 1 bis N .

Es gilt dabei noch die

Zusatzdefinition: $0! = 1$

Den Sinn dieser Zusatzdefinition werden wir in den nächsten Abschnitten kennenlernen. Mit der Definition der Fakultät können wir den *Lehrsatz* aufstellen:¹¹

Die Zahl der Möglichkeiten, N verschiedene Elemente in eine Reihenfolge zu bringen, ist gleich $N!$

Häufig bezeichnet man auch jede solche Reihenfolge als eine Permutation der Elemente. Dann würde man sagen: Die Zahl der möglichen Permutationen von N Elementen ist $N!$.

In unserem Lehrsatz taucht die Voraussetzung auf, daß es sich um N *verschiedene* Elemente handeln muß. Es ist anzumerken, daß damit nicht objektive Verschiedenheit, sondern

¹¹Aus dem folgenden Lehrsatz kann man auch eine Erklärung für die Verwendung des Wortes Fakultät herleiten: Eine der wichtigsten Anwendungsmöglichkeiten des Produkts aller ganzen Zahlen von 1 bis N ist eben die dort formulierte "Zahl der Möglichkeiten" (lat. facultas = Möglichkeit).

Verschiedenheit unter dem Gesichtspunkt des Anwendens dieses Lehrsatzes gemeint ist. Ein Beispiel kann uns das veranschaulichen:

In einer Urne befinden sich drei Kugeln: eine rote und zwei schwarze. Es wird drei mal ohne Zurücklegen gezogen, so daß also alle Kugeln gezogen werden. Gefragt ist nach der Zahl der möglichen Reihenfolgen von Kugeln.

Um die zwei schwarzen Kugeln unterscheiden zu können, wollen wir sie "Fritz" und "Franz" nennen. Die rote Kugel soll auch einen Namen haben: Sie heißt "Minna".

Jetzt kommt es darauf an, ob man beispielsweise die Reihenfolgen "Fritz, Minna, Franz" und "Franz, Minna, Fritz" als zwei verschiedene Reihenfolgen mitgezählt haben möchte oder nicht. Man könnte ja auch sagen: In beiden Fällen ist die Abfolge der Farben dieselbe (nämlich schwarz, rot, schwarz), und nur diese Abfolge interessiert uns. Wenn einen in dieser Weise nur die Abfolge der Farben interessiert, dann betrachtet man die Kugeln "Fritz" und "Franz" als gleiche Kugeln. Dann können wir nicht einfach den Lehrsatz für die Zahl der möglichen Reihenfolgen von N verschiedenen Elementen anwenden. In einem anderen Zusammenhang könnte man aber an der Zahl der möglichen Reihenfolgen nicht der Kugelfarben, sondern der einzelnen Kugel interessiert sein. Dann würde man die Reihenfolgen "Fritz, Minna, Franz" und "Franz, Minna, Fritz" als zwei verschiedene Möglichkeiten einer Reihenfolge verrechnen. Bei dieser Problemstellung werden also die beiden schwarzen Kugeln als verschiedene Kugeln betrachtet, und damit könnte man unseren Lehrsatz anwenden.

Ganz entsprechendes gilt auch für die in den folgenden Abschnitten zu behandelnden Lehrsätze: Wenn dort von "verschiedenen Elementen" die Rede ist, dann geht es nicht um objektive Verschiedenheit, sondern um Verschiedenheit unter den von der Fragestellung her nahegelegten Gesichtspunkten des Anwendens.

Es ist weiterhin zu ergänzen, daß unser Lehrsatz über mögliche Permutationen auch angewandt werden kann, um Probleme zu lösen, bei denen es nicht unmittelbar aus der Formulierung hervorgeht, daß die Zahl möglicher Reihenfolgen gefragt ist.

Dies gilt insbesondere für sogenannte Zuordnungsprobleme, bei denen wir zwei getrennte Gruppen von gleich vielen Elementen haben; jedes Element der einen Gruppe ist einem Element der 2. Gruppe so zuzuordnen, daß sich ebensoviele Paare ergeben, wie vorher Elemente pro Gruppe vorhanden waren.

Beispiel: In einem Roman mit dem Problem "Wer kriegt wen?" könnte sich der Autor fragen, wie viele Möglichkeiten er denn für das Happy-End hat. Er könnte dabei so verfahren, daß er sich die Männer in einer willkürlichen festen Reihenfolge aufschreibt - z.B. Kasimir, Melchior, Balthasar und David. Jeder möglichen Reihenfolge der Frauen entspricht dann ein mögliches Happy-End. Der Reihenfolge "Katja, Katrin, Karin, Käthe" entspräche also das Happy-End "Kasimir bekommt Katja, Melchior bekommt Katrin, usw.". Das ganze reduziert sich also auf die Frage, in wie viele

mögliche Reihenfolgen sich die vier Frauen bringen lassen¹², und da alle vier als verschieden betrachtet werden, können wir die Zahl dieser möglichen Reihenfolgen als $4! = 24$ angeben. (Es bedarf wohl keiner Erwähnung, daß es solche Probleme auch in ernsteren Situationen gibt, z.B. wenn eine Vp in einem Wortschatztest vier Begriffe und vier Definitionen einander zuordnen soll und nach der Zahl der möglichen Zuordnungen gefragt wird).

<< Schließlich ist zu ergänzen, daß das Problem der Berechnung von Fakultäten größerer Zahlen (d.h. großer Rechenaufwand!) sich erübrigt, da es für mittelgroße Zahlen (zumindest bis 300) Tabellen der zugehörigen Fakultäten gibt. Bei noch größeren (aber auch schon ab $N=7$) kann man die folgende Approximationsformel benutzen:

$$N! \approx \sqrt{2\pi} \cdot N^{(N + 0.5)} \cdot e^{\left(\frac{1}{12N} - N - \frac{1}{360N^2}\right)}$$

Dabei ist π die Kreisconstante 3.14... und $e = 2.718...$ (Basis der natürlichen Logarithmen). Die Herleitung dieser Approximationsformel verlangt einiges an höherer Mathematik und soll uns hier nicht weiter interessieren.

>>

b) Teilsequenzen

Im vorigen Abschnitt haben wir alle N Elemente einer Menge in eine Reihenfolge gebracht. Manchmal will man aber auch wissen, wie viele Möglichkeiten es gibt, eine Sequenz von weniger als N Elementen zu bilden.

Beispiel: Eine Gruppe besteht aus 10 Mitgliedern. Sie wählt sich einen Vorstand, der aus 3 Mitgliedern besteht: Einem 1. Vorsitzenden, einem 2. Vorsitzenden und einem 3. Vorsitzenden. Es ist zu fragen, wie viele mögliche Vorstände es gibt. Dabei ist auch die Rollenverteilung zu beachten: Wenn beispielsweise Fritz 1. Vorsitzender, Franz 2. Vorsitzender und Peter 3. Vorsitzender ist, dann ist dies ein anderer Vorstand, als wenn dieselben Personen mit anderer Postenverteilung den Vorstand bilden.

Wir können dann einfach fragen: Wie viele Möglichkeiten gibt es, aus 10 verschiedenen Elementen eine Teilsequenz von drei Elementen zu bilden, in der jedes Element höchstens einmal vorkommen darf?

Geben wir den 10 Elementen (Gruppenmitgliedern) einmal die Buchstaben A bis J als Namen. Wir können dann zunächst angeben, daß es $10!$ mögliche Permutationen aller 10 Elemente gibt. Aber es geht uns ja jetzt um Sequenzen von drei Elementen. Wir können aber feststellen, daß jede 3-er Sequenz an der Spitze verschiedener Permutationen auftaucht: Die Sequenz A B C finden

¹²Die Männer braucht man nicht in verschiedene Reihenfolgen zu bringen, denn dabei ergeben sich lauter Kombinationen, die schon da waren. Beispiel: Mit den Reihenfolgen "David, Melchior, Kasimir, Balthasar" und "Käthe, Katrin, Katja, Karin" käme man zum selben Happy-End wie eben. - Genau so gut könnte man natürlich die Frauen in eine feste Reihenfolge bringen und die Männer in allen möglichen Reihenfolgen dahinterschreiben. Das Ergebnis bliebe dasselbe.

wir z.B. an der Spitze der Permutationen A B C D E F G H I J oder A B C E D F G H I J und vieler anderer.

An der Spitze von wieviel Permutationen taucht die Sequenz A B C auf?

Wir könnten uns auch fragen: Wie viele Möglichkeiten gibt es, die Sequenz A B C mit den Buchstaben D bis J fortzusetzen? Da es von den $10 - 3 = 7$ Buchstaben von D bis J $7!$ mögliche Reihenfolgen gibt, ist dies auch die gesuchte Zahl der möglichen Fortsetzungen der Sequenz A B C. Das gleiche gilt für die übrigen möglichen Dreiersequenzen: Jede taucht an der Spitze von $7!$ der $10!$ Permutationen der Buchstaben A bis J auf.

Anders formuliert: Wenn wir die $10!$ Permutationen der Buchstaben A bis J alle hinschreiben würden und dann überall die letzten 7 Buchstaben durchstreichen würden, so blieben lauter Dreiersequenzen stehen, und jede Dreiersequenz stünde dann $7!$ mal da.

Also gilt:

$10! = \text{Zahl der Dreiersequenzen} \cdot 7!$

oder:

$$\text{Zahl der Dreiersequenzen} = \frac{10!}{7!} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9 \cdot 10}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7} = 8 \cdot 9 \cdot 10 = 720$$

Es gibt also 720 mögliche Dreiersequenzen von 10 Elementen, wenn jedes Element höchstens einmal in der Sequenz auftauchen darf - also 720 mögliche Vorstände unserer Gruppe.

Allgemein: Die Zahl der Möglichkeiten, aus N voneinander verschiedenen Elementen eine Teilsequenz von r voneinander verschiedenen Elementen zu ziehen, ist

$$\frac{N!}{(N-r)!}$$

Manchmal werden auch diese Teilsequenzen Permutationen genannt. Man spricht dann von der Zahl der Permutationen von r aus N verschiedenen Elementen.

Der Beweis des allgemeinen Satzes verläuft genau so wie in unserem Spezialfall: Wenn wir eine Liste aller $N!$ Permutationen anfertigen würden und immer die letzten $N - r$ Elemente wegstreichen würden, dann bliebe eine Liste von $N!$ Teilsequenzen stehen, in der aber jede Teilsequenz $(N - r)!$ mal vorkäme.

Die Zahl der möglichen Reihenfolgen im Sinne von Abschnitt a kann man im übrigen als Spezialfall unseres Lehrsatzes über Sequenzen ansehen: Es ist von den N Elementen eine Sequenz von $r = N$ Elementen zu bilden. Die Zahl der möglichen Sequenzen wäre nach unserem Lehrsatz:

$$\frac{N!}{(N-r)!} = \frac{N!}{(N-N)!} = \frac{N!}{0!}$$

Da definitionsgemäß $0! = 1$ ist, können wir dann auch schreiben:

Zahl der möglichen Sequenzen = $N!/1 = N!$

Wir sehen hieran, daß die Definition $0! = 1$ sinnvoll ist: Sie paßt zu einem aus den Fakultäten abgeleiteten Lehrsatz; und das gleiche gilt auch von den folgenden Lehrsätzen.

c) Untergruppen (Kombinationen ohne Reihenfolge)

Wir hätten im vorigen Abschnitt auch fragen können: Wie groß ist die Zahl der möglichen Vorstände ohne Berücksichtigung der Reihenfolge. D.h. es kommt uns jetzt nicht mehr darauf an, wer im Vorstand welche Position hat, sondern nur darauf, wer rein kommt und wer nicht. Zwei Vorstände, die sich nicht in der Zusammensetzung, sondern nur in der Verteilung der Positionen unterscheiden, wären also jetzt nicht mehr als zwei verschiedene Vorstände zu zählen.

Das ist ein Spezialfall des allgemeineren Problems: Wie viele Möglichkeiten gibt es, aus N verschiedenen Elementen eine Untergruppe von r verschiedenen Elementen herauszuziehen? Wir können dieses Problem ziemlich einfach auf das Problem der Sequenzen reduzieren. Jede Vorstandsgruppe taucht mit verschiedenen Positionsverteilungen auf.

Wie groß ist die Zahl der möglichen Positionsverteilungen jeder Vorstandsgruppe?

Da es für jede Vorstandsgruppe von drei Mitgliedern $3!$ mögliche Reihenfolgen gibt, ist dies auch die Zahl der möglichen Positionsverteilungen (= Zuordnungen zu Positionen) innerhalb der Vorstandsgruppe. Wir erhalten also:

$$\text{Zahl der möglichen Dreiergruppen} = \frac{\text{Zahl d. möglichen Dreiersequenzen}}{3!} = \frac{720}{6} = 120$$

Das allgemeine Problem der Untergruppen von r aus N Elementen können wir ähnlich lösen: Wenn wir uns eine Liste aller Teilsequenzen von r Elementen aus N anlegen würden, dann würde jede Untergruppe von r Elementen in Form mehrerer Sequenzen auf dieser Liste stehen, und zwar jede dieser Untergruppen $r!$ mal. Die Zahl der Untergruppen von r Elementen aus N verschiedenen ergibt sich demnach folgendermaßen:

$$\text{Zahl der Untergruppen} = \frac{\text{Zahl der Teil-Sequenzen}}{r!} = \frac{N!}{(N-r)! \cdot r!}$$

Für den dabei auftretenden Bruch $\frac{N!}{(N-r)! \cdot r!}$ verwendet man auch das Symbol $\binom{N}{r}$,

gesprochen "N über r".¹³

Man nennt dies einen "Binomialkoeffizienten", weil solche Ausdrücke auch bei der Ausrechnung von Binomen vorkommen.¹⁴

Übungsaufgabe: Wie groß ist die Zahl der Möglichkeiten, aus den Zahlen von 1 bis 49 eine Untergruppe von 6 Zahlen herauszuziehen (also die Zahl der möglichen Ergebnisse im Zahlenlotto "6 aus 49" - ohne Zusatzzahl, Superzahl usw.)?

Aus einer Obergruppe von $N = 49$ Elementen ist eine Untergruppe von 6 Elementen herauszuziehen. Die Zahl der Möglichkeiten dieser Untergruppe ist also

$$\binom{N}{r} = \binom{49}{6} = \frac{49!}{(49-6)! \cdot 6!} = \frac{49!}{43! \cdot 6!}$$

Wir können diesen Bruch nun erheblich vereinfachen, so daß wir keine Fakultätentabellen und keine Approximationsformeln zur Berechnung der Fakultäten brauchen, sondern zur Not "mit Hand" ausrechnen können. In dem Produkt der Zahlen von 1 bis 49 (also in $N!$) ist das Produkt der Zahlen von 1 bis 43 (also $(N - r)!$) ganz enthalten. Würden wir also die Fakultäten im Zähler und im Nenner in Produkte auflösen, dann könnten wir das ganze Produkt der Zahlen von 1 bis 43 wegkürzen. Nach der Vertauschung der Reihenfolge der Faktoren im Zähler kämen wir dann zu dem Bruch

$$\binom{49}{6} = \frac{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6}$$

Diesen Vorgang kann man bei allen Berechnungen von Binomialkoeffizienten durchführen, und daraus ergibt sich als praktische Rechenformel:

$$\binom{N}{r} = \frac{N \cdot (N-1) \cdot (N-2) \cdot \text{u.s.w. insgesamt } r \text{ Zahlen}}{r!}$$

Als Zusammenfassung unserer Überlegungen können wir festhalten:

Die Zahl der möglichen Untergruppen von r verschiedenen Elementen aus N verschiedenen Elementen ist $\binom{N}{r}$. Dabei ist

¹³Zu beachten: Im Englischen steht "N over r" nicht als Verbalisierung dieses Symbols, sondern für den Bruch n/r .

¹⁴Ein Binom ist ein Ausdruck der Form $(a + b)^n$, dessen Spezialfall für $n=2$ man in der Schule kennenlernt: $(a + b)^2 = a^2 + 2ab + b^2$.

$$\binom{N}{r} = \frac{N!}{(N-r)! \cdot r!}$$

Auch hier gibt es in verschiedenen Büchern verschiedene Begriffe: Was wir als Zahl der "Untergruppen" bezeichneten, wird in anderen Büchern als Zahl der "Kombinationen von r aus N Elementen" aufgeführt. Andere Bücher sprechen hier von "Kombinationen ohne Berücksichtigung der Reihenfolge", während unsere Teilsequenzen als "Kombinationen mit Berücksichtigung der Reihenfolge" bezeichnet werden. Wir wollen uns im folgenden an die Begriffe "Sequenzen" und "Untergruppen" halten, da diese eindeutig und außerdem vielleicht etwas einfacher und anschaulicher sind als die übrigen.

Wir haben bisher immer die Voraussetzung gemacht, daß jedes Element höchstens einmal in der Teilsequenz oder Untergruppe auftauchen darf. Man spricht hier auch von "Kombinationen ohne Wiederholung". Dagegen lägen "Kombinationen mit Wiederholung" vor, wenn eine Person mehrere Vorstandsämter haben darf oder wenn beim Lotto 6 aus 49 die Möglichkeit bestünde, daß eine Zahl mehrmals gezogen wird. Auch für diese Kombinationen mit Wiederholung gibt es Spezialformeln, die hier aber nicht behandelt werden sollen.

Wir haben schon in einem früheren Abschnitt einmal eine Anzahl von möglichen Untergruppen, wie sie Gegenstand dieses Abschnitts waren, berechnet. Können Sie angeben an welcher Stelle?

Als wir bestimmten, wie viele Möglichkeiten es gibt, von fünf Aufgaben zwei richtig und drei falsch zu lösen, ging es im Grunde darum, wie viele Untergruppen von zwei richtig gelösten Aufgaben man aus fünf Aufgaben bilden kann. Nach unserem jetzigen Stand würden wir angeben: Die Zahl der möglichen Untergruppen und damit die Zahl der Abfolge von zwei richtigen und drei falschen Lösungen ist

$$\binom{5}{2} = \frac{5 \cdot 4}{1 \cdot 2} = 10$$

Wie nicht anders zu erwarten war, kommen wir zum gleichen Ergebnis wie beim Aufzählen. Wir können jetzt auch allgemein angeben, wieviel "Lösungsmuster" mit x richtigen Lösungen es gibt: Wir müssen einfach fragen, wie viele Untergruppen von x Aufgaben man aus fünf Aufgaben bilden kann. In jedem Fall ist die Zahl der Lösungsmuster mit x richtigen Lösungen $\binom{5}{x}$.

Interessant ist das Verhalten der Formel in den Grenzfällen x=0 und x=5. Bei der Herleitung der Formel hatten wir ja nur Fälle im Auge, bei denen die Untergruppen mehr als 0 und weniger als N Elemente umfaßten. Bei x=0 ergibt sich aber hier:

$$\binom{5}{0} = \frac{5!}{(5-0)! \cdot 0!} = 1$$

Tatsächlich gibt es nur ein Lösungsmuster mit 0 richtigen Lösungen, und zwar F_1, F_2, F_3, F_4, F_5 . Ähnlich gilt für x=5:

$$\binom{5}{5} = \frac{5!}{(5-5)! \cdot 5!} = 1$$

Auch für fünf richtige Lösungen gibt es nur ein Lösungsmuster, nämlich R_1, R_2, R_3, R_4, R_5 . Unsere Definition $0!=1$ hat sich also wieder bewährt beim Einsatz in Grenzfällen einer Formel, die wir zunächst gar nicht für diese Grenzfälle entwickelt haben.

<< Wie bereits angemerkt wurde, beruht die Bezeichnung "Binomialkoeffizient" auf der Verwendung in einer allgemeinen Formel für sog. "Binome", also Ausdrücke der Form $(a + b)^n$. Diese allgemeine Formel lautet

$$(a + b)^n = \sum_{r=0}^n \binom{n}{r} \cdot a^r \cdot b^{n-r}$$

Übungsaufgabe: Überprüfen Sie, daß sich für $n = 2$ die aus der Schulmathematik bekannte Formel $(a + b)^2 = a^2 + 2ab + b^2$ ergibt.

>>

III. Diskrete Wahrscheinlichkeitsverteilungen

1) Eindimensionale diskrete Wahrscheinlichkeitsverteilungen

a) Der Begriff der Wahrscheinlichkeitsverteilung - hergeleitet an einem Beispiel

Wir haben schon mehrfach eine Vp betrachtet, die einen Test mit Würfeln bearbeitet. Der Test bestand aus fünf multiple-choice-Aufgaben mit sechs Alternativen. Wir hatten bereits angegeben, daß die Wahrscheinlichkeit, zwei richtige Lösungen zu erzielen, gleich 0.1608 ist. Bezeichnet man die Zahl der richtigen Lösungen mit X , dann könnten wir schreiben:

$$p(X=2) = 0.1608$$

In Worten: Die Wahrscheinlichkeit des Ereignisses " $X = 2$ " ist 0.1608.

Mit demselben Ansatz, mit dem wir dieses Ergebnis erzielten, können wir auch eine allgemeine Formel für die Wahrscheinlichkeit irgendeiner anderen Zahl (allgemein: a) richtiger Lösungen ableiten - also für die Wahrscheinlichkeit, daß die "zufällige Größe" X einen bestimmten Wert a annimmt:

$$p(X=a) = \binom{5}{a} \cdot \left(\frac{1}{6}\right)^a \cdot \left(\frac{5}{6}\right)^{5-a}$$

Wollen Sie versuchen, diese Formel selbst herzuleiten?

Genauso, wie wir gesehen haben, daß die Wahrscheinlichkeit jedes einzelnen Lösungsmusters mit zwei richtigen Lösungen gleich $\left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right)^3$ ist, würde man auch zeigen, daß jedes

Lösungsmuster mit a richtigen (und damit $5 - a$ falschen Lösungen) eine Wahrscheinlichkeit hat, die sich nach dem Multiplikationssatz für unabhängige Ereignisse berechnet als ein Produkt, in dem a mal der Faktor $1/6$ und $5 - a$ mal der Faktor $5/6$ vorkommt, also als

$$\left(\frac{1}{6}\right)^a \cdot \left(\frac{5}{6}\right)^{5-a}$$

Da es - wie wir im letzten Abschnitt sahen - $\binom{5}{a}$ solche Lösungsmuster mit a richtigen Lösungen gibt, und da sich diese gegenseitig ausschließen, berechnet sich die Wahrscheinlichkeit, daß irgendeines dieser Lösungsmuster auftritt, nach dem Additionssatz als Summe der Einzelwahrscheinlichkeiten der Lösungsmuster. Da bei dieser Summe aber $\binom{5}{a}$ mal dieselbe Einzelwahrscheinlichkeit (nämlich die oben genannte) auftreten würde, ergibt sich:

$p(X=a) = p(\text{irgendeines der Lösungsmuster mit } a \text{ richtigen Lösungen}).$

$$= \binom{5}{a} \cdot \left(\frac{1}{6}\right)^a \cdot \left(\frac{5}{6}\right)^{5-a}$$

Die nach dieser Gleichung berechneten Wahrscheinlichkeiten sind in der folgenden Tabelle 2 zusammengefaßt:

x_k	p_k
0	0.4019
1	0.4019
2	0.1608
3	0.0322
4	0.0032
5	0.0001

Tabelle 2: Wahrscheinlichkeitsverteilung der Zahl der richtigen Lösungen im "Würfeltst".

In der ersten Spalte ist jeweils der x -Wert jeder Klasse (x_k) angegeben, in der 2. Spalte die Wahrscheinlichkeit des x -Wertes der jeweiligen Klasse (p_k).

Eine solche Zuordnung von Wahrscheinlichkeiten zu den verschiedenen Klassen einer Skala nennt man eine Wahrscheinlichkeitsverteilung, genauso wie man eine Zuordnung von Häufigkeiten zu den verschiedenen Klassen einer Skala eine Häufigkeitsverteilung nennt. Wahrscheinlichkeitsverteilungen haben mit Häufigkeitsverteilungen vieles gemeinsam, wie wir noch sehen werden. Zunächst sollte aber noch einmal der Unterschied zwischen Häufigkeits- und

Wahrscheinlichkeitsverteilungen betont werden.

Eine Häufigkeitsverteilung, wie wir sie in der deskriptiven Statistik behandelt haben, gibt an, mit welcher Häufigkeit die Werte einer bestimmten Klasse beobachtet wurden. Sie beruht also auf empirischen Beobachtungen und wird daher auch eine *empirische Verteilung* genannt. Dagegen gibt eine Wahrscheinlichkeitsverteilung an, mit welcher Wahrscheinlichkeit die Werte einer bestimmten Klasse zu erwarten sind. Diese Wahrscheinlichkeiten werden theoretisch aus einem Modell hergeleitet - in unserem Fall aus dem Modell des Würfels mit einem völlig gleichmäßigen Würfel. Aus diesem Grund nennt man Wahrscheinlichkeitsverteilungen auch *theoretische Verteilungen*.

Wie wir später (in Abschnitt @) sehen werden, gibt es spezielle Probleme, wenn die Skala der Größe, für die wir Wahrscheinlichkeiten angeben, kontinuierlich ist. Daher beschränken wir uns zunächst auf den Fall einer diskreten Skala, der ja in unserem Beispiel vorliegt. In diesem Fall spricht man dann von einer diskreten Wahrscheinlichkeitsverteilung. Dafür gilt die

Definition: Eine diskrete Wahrscheinlichkeitsverteilung besteht in einer Zuordnung von Wahrscheinlichkeiten zu den Klassen einer diskreten Skala.

Die Größe (bzw. Variable), für deren Werte Wahrscheinlichkeiten angegeben ("zugeordnet") werden, nennt man eine *zufällige Größe* oder *Zufallsvariable*, weil man annimmt, daß ihr Wert vom Zufall bestimmt wird.

Welche Größe wird in unserer obigen Wahrscheinlichkeitsverteilung als Zufallsvariable betrachtet?

Die Wahrscheinlichkeitsverteilung gibt Wahrscheinlichkeiten für verschiedene Werte der Größe "Zahl der richtigen Lösungen" an. Diese Größe ist also die von der Wahrscheinlichkeitsverteilung behandelte zufällige Größe bzw. Zufallsvariable.

Die kurze Schreibweise $X = a$ für das Ereignis "die Zahl der Treffer (also der Wert der Zufallsvariablen X) beträgt a " ist recht praktisch, vor allem wenn es darum geht, für die Wahrscheinlichkeit eines solchen Ereignisses eine Formel anzugeben. Entsprechend kann man natürlich $X \leq a$ schreiben für das Ereignis "die Zahl der Treffer beträgt höchstens a " usw.. Zu beachten ist, daß dabei die Buchstaben X und a eine recht unterschiedliche Bedeutung haben: X steht für die Worte "der Wert der Zufallsvariablen X ", und dafür kann man dann bei konkreten Anwendungen eine Formulierung wie "die Zahl der Treffer" setzen. Dagegen steht a immer für eine Zahl (mathematisch ausgedrückt: für einen möglichen Wert der Zufallsvariablen). Um diese Rollen auseinanderzuhalten, hat sich eine Schreibkonvention entwickelt (die allerdings nicht immer ganz konsequent eingehalten wird):

Schreibkonvention: Für Zufallsvariablen werden (meistens) Großbuchstaben verwendet;

dagegen bezeichnen Kleinbuchstaben (meistens) mögliche Werte der Zufallsvariablen.

Damit kann man also auch $X = x$ schreiben für das Ereignis "der Wert der Zufallsvariablen X beträgt x ", wobei dann x irgendeine Zahl sein kann. Hat man sich an diese Konvention einmal gewöhnt, dann ist es sogar recht hilfreich, für Zufallsvariablen und ihre möglichen Werte denselben Buchstaben zu verwenden und die beiden unterschiedlichen "Rollen" nur durch die Unterscheidung von Groß- und Kleinbuchstaben zum Ausdruck zu bringen. Deshalb wird diese Konvention auch im folgenden zugrundegelegt.

Da die Einhaltung einer solchen Regel etwas "gewöhnungsbedürftig" ist, wird es nicht als "klausurrelevantes Lernziel" erwartet, daß jeder perfekt damit umgehen kann. Hauptsache ist, daß man beim Lesen mit einem Ausdruck wie $p(X = x)$ etwas anfangen kann.

b) Relative Häufigkeiten als Grundlage der Beziehungen zwischen empirischen Verteilungen und Wahrscheinlichkeitsverteilungen

Das verbindungsstiftende Element zwischen Häufigkeits- und Wahrscheinlichkeitsverteilungen ist die Interpretation der Wahrscheinlichkeit als zu erwartende relative Häufigkeit.

In unserer Wahrscheinlichkeitsverteilung wird die Wahrscheinlichkeit des Ereignisses " $X=2$ " als 0.1608 angegeben. Wie könnte man diese Wahrscheinlichkeitsangabe im Sinne einer zu erwartenden relativen Häufigkeit interpretieren?

Um die Wahrscheinlichkeit eines Ereignisses als zu erwartende relative Häufigkeit zu interpretieren, muß man annehmen, daß der "Versuch", bei dem das genannte Ereignis mit der angegebenen Wahrscheinlichkeit auftritt, sehr oft (zumindest aber mehrfach) durchgeführt wird. In unserem Fall müßte man also fragen: Bei welchem Versuch ist die Wahrscheinlichkeit des Ereignisses " $X = 2$ " gleich 0.1608?

Dieser Versuch besteht in einer Bearbeitung des aus fünf Aufgaben bestehenden Tests mit Würfeltechnik. Wenn wir also diesen Versuch mehrfach durchführen würden (d.h., wenn die V_p den Test mehrmals mit Würfeltechnik bearbeiten würde oder wenn mehrere V_{pn} dies täten), dann hätten wir das Ereignis " $X=2$ " mit einer relativen Häufigkeit von 0.1608 zu erwarten.

In gleicher Weise könnten wir auch die anderen Wahrscheinlichkeitsangaben aus unserer Wahrscheinlichkeitsverteilung interpretieren, und dann könnte man auch die gesamte Wahrscheinlichkeitsverteilung als erwartete Verteilung relativer Häufigkeiten interpretieren.

Wir können nun viele Berechnungen, die wir bei Häufigkeitsverteilungen kennengelernt haben, auch mit Wahrscheinlichkeitsverteilungen durchführen. Wir müssen uns dabei auf diejenigen Formeln stützen, die auf den relativen Häufigkeiten aufbauen. Im Laufe dieser Berechnungen entsteht die folgende Tabelle, deren einzelnen Spalten wir in den folgenden Abschnitten behandeln werden.

x_k	p_k	$p_{cum\ k}$	$p_k \cdot x_k$	$p_k \cdot x_k^2$	$p_k \cdot (x_k - \mu_x)^2$
0	0.4019	0.4019	0.0000	0.0000	0.2791
1	0.4019	0.8038	0.4019	0.4019	0.0112
2	0.1608	0.9645	0.3215	0.6430	0.2188
3	0.0322	0.9967	0.0965	0.2894	0.1509
4	0.0032	0.9999	0.0129	0.0514	0.0322
5	0.0001	1.0000	0.0006	0.0032	0.0022
Summe	1.0000		0.8333...	1.3888...	0.6944...

Tabelle 3 Berechnungstabelle für Kennziffern einer Wahrscheinlichkeitsverteilung

In der letzten Zeile der Tabelle 3 sind diejenigen Summen angegeben, die sich ergeben würden, wenn keine Rundungsfehler aufträten. Die in den beiden letzten Spalten angegebenen Brüche würde man erhalten, wenn man die bei der Berechnung der Wahrscheinlichkeiten auftretenden Brüche nicht in Dezimalbrüche umformen, sondern nach den normalen Regeln der Bruchrechnung weiterrechnen würde.

Wenn wir hier die Wahrscheinlichkeitsverteilungen in eine enge Beziehung zu den Häufigkeitsverteilungen setzen, dann bedeutet dies nicht, daß unsere Überlegungen nur da anwendbar sind, wo den Wahrscheinlichkeitsverteilungen der Begriff der Wahrscheinlichkeit als zu erwartender relativer Häufigkeit zugrundeliegt. Die von uns herzuleitenden Begriffe und Formeln enthalten vor allem Definitionen, die unabhängig von dem zugrundeliegenden Wahrscheinlichkeitsbegriff gültig sind. Unsere Bezugnahme auf relative Häufigkeit hat die Absicht, den Sinn dieser Definitionen aufzuzeigen.

Können Sie sich eine Wahrscheinlichkeitsverteilung vorstellen, die auf "subjektiven Wahrscheinlichkeiten" beruht?

Wenn wir beispielsweise einen Patienten bei der Aufnahme in eine Klinik angeben lassen, wie wahrscheinlich ihm verschiedene Dauern eines Klinikaufenthaltes sind¹⁵, dann könnte man daraus möglicherweise eine subjektive Wahrscheinlichkeitsverteilung der zufälligen Größe "Dauer des Klinikaufenthalts" entnehmen. Für diese Wahrscheinlichkeitsverteilung wären die in den folgenden Abschnitten zu besprechenden Begriffe und Kennziffern genauso definiert wie in dem hier zu besprechenden Beispiel, in dem Wahrscheinlichkeiten der Anschaulichkeit halber als erwartete relative Häufigkeiten interpretierbar sind.

¹⁵Solche Angaben könnte man beispielsweise erheben, um eventuelle Fehlerwartungen über kurzfristige Erfolge rechtzeitig zu erfahren und damit Enttäuschungen verhindern zu können.

c) Kumulative Wahrscheinlichkeitsverteilungen und Zentile

Bei den empirischen Verteilungen war jeder Klasse eine kumulative relative Häufigkeit zugeordnet. Sie war definiert als die relative Häufigkeit, mit der ein Meßwert auftrat, der in die betreffende Klasse oder darunter fiel und konnte berechnet werden als Summe der relativen Häufigkeiten aller Klassen bis einschließlich der Klasse, deren kumulative Häufigkeit anzugeben war.

Entsprechend gilt bei den Wahrscheinlichkeitsverteilungen die

Definition: Die kumulative Wahrscheinlichkeit $p_{\text{cum } k}$ ist die Wahrscheinlichkeit, mit der die in der Wahrscheinlichkeitsverteilung erfaßte zufällige Größe X in die k -te Klasse oder darunter fällt. Als Formel:

$$P_{\text{cum } k} = P(X \leq x_k)$$

Um die Parallelen zwischen relativen Häufigkeiten und Wahrscheinlichkeiten herauszustellen, verwenden wir bewußt das gleiche Symbol $p_{\text{cum } k}$ für die kumulative Wahrscheinlichkeit der k -ten Klasse wie für die entsprechende kumulative relative Häufigkeit. Ähnlich wie in der beschreibenden Statistik können wir die kumulative Wahrscheinlichkeit der k -ten Klasse auch als Summe der Wahrscheinlichkeiten aller Klassen bis zur k -ten berechnen. Als kumulative Wahrscheinlichkeit für die 3. Klasse ergäbe sich z.B.:¹⁶

$$p_{\text{cum } 3} = p_1 + p_2 + p_3 = 0.4019 + 0.4019 + 0.1608 \approx 0.9645$$

Nach welchem Wahrscheinlichkeitssatz ist diese additive Berechnung der kumulativen Wahrscheinlichkeit (also z.B. der Wahrscheinlichkeit, daß X in die 3. Klasse oder darunter fällt) möglich?

Da die Ereignisse " $X=0$ ", " $X=1$ " und " $X=2$ " sich gegenseitig ausschließen, können wir nach dem Additionssatz für einander ausschließende Ereignisse schreiben:

$$p(X=0 \vee X=1 \vee X=2) = p(X=0) + p(X=1) + p(X=2) = p_1 + p_2 + p_3$$

Diese Überlegung gilt natürlich nicht nur für die dritte, sondern für jede weitere Klasse.

Kann man auch die Behauptung, daß die Summe der relativen Häufigkeiten aller Klassen der Skala gleich 1 sein muß, auf Wahrscheinlichkeitsverteilungen übertragen?

Eine Skala muß laut Definition "erschöpfend" sein (also alle Möglichkeiten umfassen); d.h.: Unsere zufällige Größe fällt mit Sicherheit in eine der Klassen. Da aber die Wahrscheinlichkeit

¹⁶Bei genauem Nachrechnen ergibt sich $0.4019 + 0.4019 + 0.1608 = 0.9646$. Die Abweichung dieser Summe von dem in der Tabelle angegebenen Wert der kumulativen Wahrscheinlichkeit von 0.9645 beruht auf Rundungsfehlern; der Wert 0.9645 ist genauer. (Wer will, kann es nachprüfen: Für p_k keine Dezimalbrüche einsetzen, sondern die Brüche, die sich aus der Formel für $p(X=a)$ ergeben, aufsummieren und erst dann die Summe in einen Dezimalbruch umwandeln. Das ergibt dann 0.9645.)

eines sicheren Ereignisses gleich 1 ist, gilt dies auch für die Wahrscheinlichkeit, daß die zufällige Größe in irgendeine der Klassen unserer Skala fällt - in die erste oder zweite oder...oder..., und diese Wahrscheinlichkeit ist - ganz analog zu unseren Überlegungen zu den kumulativen Wahrscheinlichkeiten - nach dem Additionssatz für einander ausschließende Ereignisse gleich der Summe der Wahrscheinlichkeiten aller einzelnen Klassen. *Zusammenfassend* ergibt sich:

Die Summe aller Klassenwahrscheinlichkeiten muß bei einer diskreten Wahrscheinlichkeitsverteilung 1 sein. In mathematischen Symbolen geschrieben: $\sum_k p_k = 1$.

Diese Regel gilt auch dann noch, wenn unsere Skala unendlich viele Klassen hat. Ist die Zahl der Klassen dagegen endlich, so können wir auch sagen: Die kumulative Wahrscheinlichkeit der letzten Klasse muß gleich 1 sein - sonst haben wir einen Fehler gemacht.¹⁷

Aufgrund der kumulativen Wahrscheinlichkeiten können wir nun - genau wie bei Häufigkeitsverteilungen - Zentile angeben. Wir könnten hierzu die kumulativen Wahrscheinlichkeiten alle mit 100 multiplizieren, so daß sie prozentuale Wahrscheinlichkeiten angeben (die kumulative Wahrscheinlichkeit der 2. Klasse in unserem Beispiel könnten wir ja auch als 80.38% angeben).

Können Sie sich vorstellen, wie dann die Definition des 80. Zentils (also des C_{80}) in einer Wahrscheinlichkeitsverteilung lautet?

Diese Definition würde lauten: Der 80. Zentil (C_{80}) ist der Skalenwert, dessen prozentuale kumulative Wahrscheinlichkeit 80% ist. In unserem Fall wäre dies etwa der Skalenwert 1.

Wäre es nicht sinnvoller, den 80. Zentil fast ganz am "oberen Intervallende", also kurz unterhalb von 1.500 anzusetzen?

Die Interpolation von Zentilen ist nur bei künstlich diskreten, nicht aber bei natürlich diskreten Skalen sinnvoll. Die Klassen unserer Wahrscheinlichkeitsverteilung kann man kaum als Zusammenfassung von größeren Intervallen betrachten: Es gibt kaum kontinuierliche Übergänge zwischen "bei zwei Aufgaben richtig würfeln" und "bei drei Aufgaben richtig würfeln". Eine Interpolation von Zentilen ist daher nicht sinnvoll. Genauso wie bei den empirischen Verteilungen legt man bei Wahrscheinlichkeitsverteilungen über echt diskrete Skalen die Zentile auf den ersten Skalenwert, dessen kumulative prozentuale Wahrscheinlichkeit nicht mehr kleiner als p ist.

Wo würde man also in unserer Beispiel-Verteilung den C_{50} ansetzen?

Der erste Skalenwert, dessen kumulative prozentuale Wahrscheinlichkeit nicht mehr kleiner als 50 ist, ist ebenfalls bei $X=1$. Es ist also hier $C_{50} = 1$.

Ist dagegen unsere Skala künstlich diskret, dann kann man die Zentile nach den gleichen

¹⁷In der 3. Spalte unserer Tabelle 3 ist die kumulative Wahrscheinlichkeit der letzten Klasse mit 1.001 angegeben. Das liegt an den beim Rechnen mit Dezimalzahlen meist unvermeidlichen Rundungsfehlern. Würde man die Klassenwahrscheinlichkeiten nach der Formel für $p(X = a)$ als Brüche berechnen und diese nicht in Dezimalzahlen umwandeln, dann ergäbe sich genau 1. (Wer will, kann's nachprüfen.)

Prinzipien und Formeln wie bei den empirischen Verteilungen interpolieren.

Es braucht wohl kaum erwähnt zu werden, daß Wahrscheinlichkeitsverteilungen sich ebenso wie Häufigkeitsverteilungen graphisch darstellen lassen. Die folgende Abbildung gibt als ein Beispiel die kumulative Wahrscheinlichkeitsverteilung unseres Beispiels als Ogive an.¹⁸

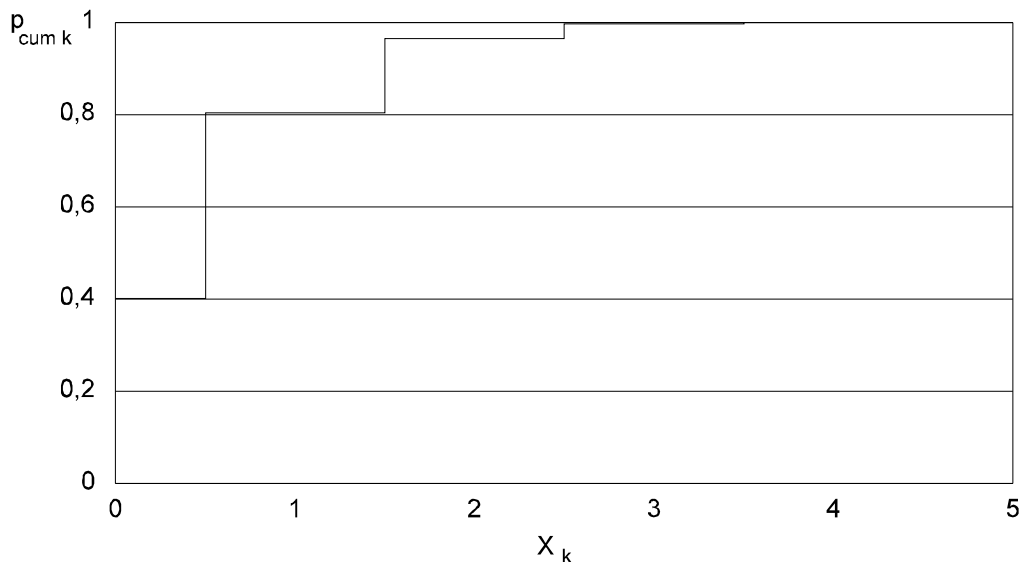


Abbildung 3: Kumulative Wahrscheinlichkeitsverteilung

d) Kennziffern der Zentralen Tendenz in diskreten Wahrscheinlichkeitsverteilungen

Auch zu den Kennziffern der Häufigkeitsverteilungen gibt es Parallelen in den Wahrscheinlichkeitsverteilungen. Bei diskreten Verteilungen erhält man sie jeweils, indem man in den aus der deskriptiven Statistik stammenden Formeln, die auf den empirischen relativen Häufigkeiten der Klassen aufbauen, das Symbol p_k uminterpretiert als Wahrscheinlichkeit. Zur Unterscheidung von den Kennziffern der empirischen Verteilungen wählt man für die Kennziffern der Wahrscheinlichkeitsverteilungen meist griechische Buchstaben als Symbole.

Genauso wie bei den Häufigkeitsverteilungen gibt es zunächst Kennziffern der zentralen Tendenz, die alle in irgendeiner Form die Mitte der Verteilung angeben. Am einfachsten übertragbar sind die Begriffe des Modalwerts und des Medians: Der Median ist identisch mit dem C_{50} und der Modalwert ist der Skalenwert mit der größten Wahrscheinlichkeit.

¹⁸Allerdings wäre es sinnvoller, wenn die "Sprungstellen" nicht bei 0.5, 1.5 usw. lägen, sondern bei den ganzen Zahlen. Beispiel: Die Wahrscheinlichkeit des Ereignisses $X \leq 1.9$ ist noch gleich der Wahrscheinlichkeit des Ereignisses $X \leq 1$, und erst bei dem Ereignis $X \leq 2$ erfolgt ein Sprung auf die nächsthöhere kumulative Wahrscheinlichkeit. Aus technischen Gründen ist eine Korrektur der Abbildung aber schwierig und deshalb für eine spätere Auflage vorgemerkt.

Welche Formel für das arithmetische Mittel von X könnte man am ehesten auf Wahrscheinlichkeitsverteilungen übertragen?

Es ist eine Formel zu suchen, in der das arithmetische Mittel aufgrund der p_k berechnet wird. Diese Formel lautet:

$$aM(X) = \sum_k p_k \cdot x_k$$

Diese Definition könnten wir auf die Wahrscheinlichkeitsverteilungen übertragen. Üblicherweise reserviert man jedoch die Begriffe des arithmetischen Mittels und des Durchschnitts für empirische Verteilungen und spricht bei Wahrscheinlichkeitsverteilungen vom Erwartungswert. Als Symbole für den Erwartungswert der zufälligen Größe X verwendet man $E(X)$ oder - mit einem griechischen Buchstaben, da es sich um eine Kennziffer von Wahrscheinlichkeitsverteilungen handelt - das Symbol μ_x .

<<

Bei konsequenter Anwendung der Regel, für Zufallsvariablen grundsätzlich Großbuchstaben zu verwenden (vgl. S. @) würde man eher μ_X schreiben. Es hat sich aber μ_x "eingebürgert", und deshalb wird diese Schreibweise beibehalten. Entsprechendes gilt auch für die im folgenden behandelten Kennziffern.

>>

Die Definition des Erwartungswerts entspricht ganz der rechten Seite unserer obigen Formel für das arithmetische Mittel. Zusammenfassend können wir also festhalten:

In Analogie zum arithmetischen Mittel von empirischen Verteilungen ist der Erwartungswert von diskreten Wahrscheinlichkeitsverteilungen definiert als:

$$\mu_x := E(X) := \sum_k p_k \cdot x_k$$

In Tabelle 3 auf Seite @ ist dieser Erwartungswert schon berechnet. Suchen Sie diese Berechnung und geben Sie den Erwartungswert an.

In der 4. Spalte der Tabelle 3 ist für jede Klasse das Produkt $p_k \cdot x_k$ berechnet. Der Erwartungswert ergibt sich durch Addition dieser Produkte, also als:

$$E(X) = 5/6 = 0.8333...$$

Häufig haben wir in der beschreibenden Statistik das arithmetische Mittel einer von X abgeleiteten Größe $g(X)$ berechnet. Eine entsprechende Formel, die mit relativen Häufigkeiten arbeitet, lautet:

$$aM(g(X)) = \sum_k p_k \cdot g(x_k)$$

Ganz entsprechend gilt:

Regel: Ist X eine Zufallsvariable mit diskreter Wahrscheinlichkeitsverteilung, dann erhält man den Erwartungswert einer von X abgeleiteten Größe $g(X)$, indem man für jede Klasse die Klassenwahrscheinlichkeit p_k mit dem entsprechend umgerechneten x -Wert der Klasse (also mit $g(x_k)$) multipliziert und diese Produkte über alle Klassen addiert.

Als Formel:

$$E(g(X)) = \sum_k p_k \cdot g(x_k)$$

Wie würde man z.B. bei der bisher (in den Tabellen @2 und 3) behandelten Zufallsgröße X den Erwartungswert von X^2 bestimmen?

Dieser Erwartungswert ist in unserer Tabelle 3 bereits ausgerechnet: Man multipliziert jede Klassenwahrscheinlichkeit mit dem quadrierten x -Wert der Klasse. Diese Produkte stehen in der vorletzten Spalte von Tabelle 3. Die Summe dieser Produkte ist dann der Erwartungswert von X^2 :

$$E(X^2) = \sum_k p_k \cdot x_k^2 = 0.4019 \cdot 0^2 + 0.4019 \cdot 1^2 + 0.1608 \cdot 2^2 \text{ usw.} = 1.3888\dots$$

<< Mit der Definition des Erwartungswerts einer von X abgeleiteten Größe $g(X)$ kann man auch das harmonische Mittel einer Wahrscheinlichkeitsverteilungen angeben:

$$hM(X) = \frac{1}{E\left(\frac{1}{X}\right)} = \frac{1}{\sum_k p_k \cdot \frac{1}{x_k}}$$

Wenn wir dies für unsere Beispielverteilung berechnen wollten, würden wir allerdings in der 1. Klasse auf eine Schwierigkeit stoßen: Wir würden $1/0$ für $1/x_k$ einsetzen müssen, und dieser Quotient existiert nicht. Entsprechend existiert auch der Erwartungswert von $1/X$ nicht.

In der mathematischen Statistik ist dies auch bei unendlich vielen Klassen der Skala manchmal der Fall: Es ist hier denkbar, daß die Summe $\sum_k p_k \cdot x_k$ (oder $\sum_k p_k \cdot g(x_k)$) unendlich groß ist. Dann sagt man ebenfalls, der Erwartungswert von X (bzw. der Erwartungswert der von X abgeleiteten Größe $g(X)$) existiere nicht. Schließlich ist anzumerken, daß in Fällen, in denen X (oder $g(X)$) negative Werte annehmen kann, der Erwartungswert von X bzw. $g(X)$ definitionsgemäß auch dann nicht existiert, wenn die Summe $\sum_k p_k \cdot |x_k|$ (bzw. $\sum_k p_k \cdot |g(x_k)|$) nicht endlich ist. Unsere Definitionen des Erwartungswerts von X und von $g(X)$ gelten also nur, wenn keine der genannten Endlichkeitsbedingungen verletzt ist. Ansonsten sagt man: Die zufällige Größe X (bzw. die von der zufälligen Größe X abgeleitete Funktion $g(X)$) hat keinen Erwartungswert.

Können Sie schließlich eine Formel für das geometrische Mittel angeben, die sich auf Wahrscheinlichkeitsverteilungen übertragen läßt?

Es gibt zwei solche Ansätze: Zunächst kann man die Definitionsformel des geometrischen Mittels schreiben als

$$gM(X) = \sqrt[N]{\prod_k x_k^{f_k}} = \prod_k x_k^{\frac{f_k}{N}} = \prod_k x_k^{p_k}$$

Der letzte Ausdruck läßt sich dann auf Wahrscheinlichkeitsverteilungen übertragen. Ein anderer Ansatz geht von der Beziehung des geometrischen Mittels zum arithmetischen Mittel des Logarithmus von X aus. Unsere diesbezüglichen Überlegungen auf S.@ erlauben uns zu schreiben:

$$gM(X) = 10^{aM(\log(X))}$$

und darauf aufbauend können wir eine Definition des geometrischen Mittels einer Wahrscheinlichkeitsverteilung aufstellen:

$$gM(X) = 10^{E(\log(X))}$$

Übungsaufgabe: Überprüfen Sie, daß die beiden Definitionen des geometrischen Mittels identisch sind.

Es genügt zu zeigen, daß der Logarithmus des geometrischen Mittels nach beiden Definitionen identisch ist.

>>

e) Kennziffern der Variabilität in diskreten Wahrscheinlichkeitsverteilungen

Die Kennziffern der Variabilität in empirischen Verteilungen geben in konzentrierter Form an, wie große Abweichungen der Meßwerte von den Kennziffern der zentralen Tendenz man empirisch vorgefunden hat. Entsprechend geben sie bei Wahrscheinlichkeitsverteilungen an, welche derartigen Abweichungen wahrscheinlich sind.

Man könnte alle von uns behandelten Kennziffern der Variabilität übersetzen für Wahrscheinlichkeitsverteilungen. Wir wollen dies nur für die Varianz, die Standardabweichung und für die mittlere Quartilabweichung (*QD*) tun. Bei der letzteren ist die Übertragung am einfachsten: Die Quartile Q_1 und Q_3 sind wie bei Häufigkeitsverteilungen definiert als C_{75} und C_{25} , und damit ist die Definition direkt übertragbar:

$$QD = \frac{|C_{75} - C_{50}| + |C_{50} - C_{25}|}{2} = \frac{C_{75} - C_{25}}{2}$$

Die Varianz ist bei Häufigkeitsverteilungen definiert als das arithmetische Mittel der quadrierten Abweichungen aller Beobachtungen vom Stichprobendurchschnitt. Als Formel:

$$s_x^2 = aM[(X - \bar{x})^2] = \sum_k p_k \cdot (x_k - \bar{x})^2$$

Wie würde man diese Definition auf Wahrscheinlichkeitsverteilungen übertragen?

Die *Definition* lautet hier: Die Varianz einer Wahrscheinlichkeitsverteilung ist der Erwartungswert der quadrierten Abweichung der zufälligen Größe vom Erwartungswert der Verteilung.

Als Symbol für die Varianz in Wahrscheinlichkeitsverteilungen ist σ^2 bzw. σ_x^2 üblich.¹⁹

Dann ergibt sich die Formel:

$$\sigma_x^2 = E[(X - \mu_x)^2] = \sum_k p_k \cdot (x_k - \mu_x)^2$$

¹⁹Der dabei verwendete griechische Buchstabe σ heißt "Sigma".

Wie würde man nach dieser Formel die Varianz unserer Wahrscheinlichkeitsverteilung aus Tabelle 3 berechnen?

Da $\mu_x = 0.8333$, hätten wir zu rechnen (vgl. auch letzte Spalte von Tabelle 3 @)

$$\sigma_x^2 = 0.4019 \cdot (0 - 0.8333)^2 + 0.4019 \cdot (1 - 0.8333)^2 + 0.1608 \cdot (2 - 0.8333)^2 + \text{usw.}$$

Wir hätten allerdings auch auf eine andere Varianzformel zurückgreifen können (vgl. S.@):

$$\sigma_x^2 = \alpha M(X^2) - [\alpha M(X)]^2 = \sum_k p_k \cdot x_k^2 - \left(\sum_k p_k \cdot x_k \right)^2$$

Übertragen auf Wahrscheinlichkeitsverteilungen:

$$\sigma_x^2 = E(X^2) - [E(X)]^2 = \sum_k p_k \cdot x_k^2 - \left(\sum_k p_k \cdot x_k \right)^2$$

Wie würde die Varianz in unserer Tabelle 3 nach dieser Formel zu berechnen sein?

Den Erwartungswert von X^2 haben wir schon in einem früheren Abschnitt aufgrund der vorletzten Spalte unserer Tabelle als 1.3888... bestimmt. Da der Erwartungswert von X gleich 0.8333... ist, ergibt sich:

$$\sigma_x^2 = 1.3888... - (0.8333...)^2 = 0.69444...$$

Beide Definitionen der Varianz führen natürlich (wie bei empirischen Verteilungen) zu identischen Ergebnissen - was sich auch mathematisch beweisen läßt.

<<

Bisher haben wir die Gleichung @ $\sigma_x^2 = E(X^2) - [E(X)]^2$ nur als Formel zur Berechnung der Varianz kennengelernt. Dazu muß man dann $E(X^2)$ und $E(X)$ kennen bzw. ermitteln und beide Erwartungswerte in die Formel eintragen. In manchen Situationen ist es aber anders: Man kennt den Erwartungswert und die Varianz einer Zufallsgröße X und braucht den Erwartungswert von X^2 . Dann kann man einfach den Ausdruck $[E(X)]^2$ auf beiden Seiten der obigen Gleichung addieren und erhält die Gleichung

$$E(X^2) = \sigma_x^2 + [E(X)]^2.$$

Ein Beispiel für die Anwendung dieser Gleichung werden wir in Abschnitt @D.II.1 kennenlernen.

>>

Die *Standardabweichung* einer Wahrscheinlichkeitsverteilung ist wie bei empirischen Verteilungen definiert als Quadratwurzel der Varianz. So ergibt sich:

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{0.69444...} = 0.8333...$$

Auch der Vorgang der Standardisierung (Normierung, Umrechnung in z-Werte) läßt sich auf Wahrscheinlichkeitsverteilungen übertragen. Wir könnten uns beispielsweise fragen, wieviel Standardabweichungen ein X -Wert von 0 unter (bzw. über) dem Erwartungswert μ_x liegt. In der beschreibenden Statistik wurden solche Fragen durch Umrechnung in z-Werte nach der Formel

$$z_X = \frac{X - \bar{x}}{s_x}$$

beantwortet. Diese Formel können wir fast wörtlich für Wahrscheinlichkeitsverteilungen übernehmen und brauchen nur die Symbole \bar{x} und s_x durch die entsprechenden Bezeichnungen μ_x und σ_x für Erwartungswert und Standardabweichung einer Wahrscheinlichkeitsverteilung zu ersetzen. Dann ergibt sich:

$$z_X = \frac{X - \mu_x}{\sigma_x} = \frac{0 - 0.8333\dots}{0.8333\dots} = -1$$

Ebenso könnten wir feststellen, daß ein X -Wert von 4 genau 3.8 Standardabweichungen über dem Erwartungswert μ_x liegt:

$$z_X = \frac{X - \mu_x}{\sigma_x} = \frac{4 - 0.8333\dots}{0.8333\dots} = 3.8000$$

f) Schiefe, Exzeß und Momente

Wenn wir uns die in Tab. 3 dargestellte Wahrscheinlichkeitsverteilung in Form eines Histogramms darstellen würden, dann würde sich eine deutliche Schiefe herausstellen, die wir auch schon den Zahlen der Tabelle ansehen: Im unteren Skalenbereich sind die Wahrscheinlichkeiten geballt, im oberen ergibt sich ein langer Schwanz.

Wäre dies eine positive oder eine negative Schiefe?

Das Vorzeichen einer Schiefe richtet sich danach, wohin der "lange Schwanz" zeigt. Wenn er nach rechts ("nach Plus") zeigt, handelt es sich um eine positive Schiefe.

<<

Die Definition der Kennziffer für die Schiefe entspricht ganz der bei den empirischen Verteilungen behandelten:

$$\text{Schiefe} = aM(z_X^3)$$

Bei Wahrscheinlichkeitsverteilungen gilt entsprechend:

$$\text{Schiefe} = E(z_X^3)$$

Wie könnte man diesen Erwartungswert mit Hilfe der Wahrscheinlichkeiten der Klassen

(also mit Hilfe der p_k) in eine Summe auflösen?

z_X^3 ist eine von X abgeleitete Größe. Wie man den Erwartungswert einer solchen von X abgeleiteten Größe bestimmt, haben wir in Abschnitt II.1.d behandelt: Man multipliziert für jede Klasse die Klassenwahrscheinlichkeit p_k mit dem entsprechend umgerechneten x -Wert der Klasse und addiert diese Produkte über alle Klassen. Damit ergibt sich hier:

$$\text{Schiefe} = E(z_X^3) = E\left[\left(\frac{X - \mu_x}{\sigma_x}\right)^3\right] = \sum_k p_k \cdot \left(\frac{x_k - \mu_x}{\sigma_x}\right)^3$$

Wollen Sie zur Veranschaulichung die ersten drei Glieder dieser Summe für unsere in Tabelle 3 angegebene Wahrscheinlichkeitsverteilung aufschreiben?

Wenn wir die in Tabelle 3 für die verschiedenen Klassen angegebenen x_k - und p_k -Werte sowie die für diese Verteilung berechneten Kennziffern μ_x und σ_x in die obige Formel einsetzen, so ergibt sich:

$$\begin{aligned} E(z_X^3) &= \sum_k p_k \cdot \left(\frac{x_k - \mu_x}{\sigma_x}\right)^3 \\ &= 0.4019 \cdot \left(\frac{0 - 0.8333\dots}{0.8333\dots}\right)^3 + 0.4019 \cdot \left(\frac{1 - 0.8333\dots}{0.8333\dots}\right)^3 + \\ &\quad 0.1608 \cdot \left(\frac{2 - 0.8333\dots}{0.8333\dots}\right)^3 + \text{u.s.w.} \end{aligned}$$

Ganz ähnlich läßt sich die Formel für den Exzeß übertragen: Aus der Formel

$$\text{Exzeß} = aM(z_X^4) - C,$$

(in der C eine Konstante ist, die von verschiedenen Autoren verschieden angesetzt wird), entsteht jetzt die Formel:

$$\text{Exzeß} = E(z_X^4) - C = \sum_k p_k \cdot \left(\frac{x_k - \mu_x}{\sigma_x}\right)^4 - C$$

Schließlich läßt sich auch der Begriff der "Momente" einer Verteilung auf Wahrscheinlichkeitsverteilungen übertragen: Das r -te Moment²⁰ um den Punkt a war bei Häufigkeitsverteilungen definiert als $aM[(X - a)^r]$

²⁰Wir verwenden hier den Buchstaben r statt des auf S.@ verwendeten k , um Verwechslungen mit dem Klassenindex k in den Symbolen p_k und x_k zu vermeiden.

Wie würde die entsprechende Definition bei Wahrscheinlichkeitsverteilungen aussehen?

Wir können aus dem arithmetischen Mittel einen Erwartungswert machen und diesen dann in eine Summe auflösen. So ergibt sich die Definition des r-ten Moments um den Punkt a als

$$E[(X-a)^r] = \sum_k p_k \cdot (x_k - a)^r$$

Es versteht sich fast von selbst, daß die Begriffe der zentralen Momente (also der Momente um den Stichprobendurchschnitt \bar{x}) und der absoluten Momente (= Momente um den Punkt 0) sowie unsere Überlegungen über die Beziehungen der Momente zu Varianz, Schiefe und Exzeß (Vgl. S.@) sich ohne weiteres auf Wahrscheinlichkeitsverteilungen übertragen lassen.

Anzumerken ist noch, daß es spezielle mathematische Methoden gibt, mit denen man aus der formelmäßigen Angabe einer Wahrscheinlichkeitsverteilung (vgl. dazu Abschnitt h @) die Momente einer Verteilung ableiten kann. Diese kann man unter dem Stichwort "momenterzeugende Funktionen" (engl.: "moment generating functions") in Lehrbüchern der mathematischen Statistik nachschlagen.

>>

g) Die Übertragbarkeit der Lehrsätze über Kennziffern von Verteilungen

Wir haben in der beschreibenden Statistik verschiedene Lehrsätze über unsere Kennziffern kennengelernt, die sich alle ähnlich wie die Definitionen sinngemäß auf Wahrscheinlichkeitsverteilungen übertragen lassen.

Dies gilt insbesondere für die Lehrsätze über die Änderung der Kennziffern bei linearen Skalentransformationen. Formelmäßig zusammengefaßt lauteten diese:

Ist $X' = b \cdot X + a$ (wobei b und a Konstanten sind), so ist

$$\begin{aligned}\bar{x}' &= b \cdot \bar{x} + a \\ s_{x'}^2 &= b^2 \cdot s_x^2 \\ s_{x'} &= |b| \cdot s_x\end{aligned}$$

Schiefe und standardisierte Meßwerte ändern ihren Betrag nicht und ihr Vorzeichen nur bei negativem b; der Exzeß ändert sich überhaupt nicht.

Diese Lehrsätze lassen sich auf Wahrscheinlichkeitsverteilungen übertragen:

Unterwerfen wir die zufällige Größe X der linearen Skalentransformation

$X' = b \cdot X + a$ (wobei b und a Konstanten sind), so ist

$$\begin{aligned}\mu_{x'} &= b \cdot \mu_x + a \\ \sigma_{x'}^2 &= b^2 \cdot \sigma_x^2 \\ \sigma_{x'} &= |b| \cdot \sigma_x\end{aligned}$$

Schiefe, standardisierte Werte und Exzeß ändern sich nicht mit der Ausnahme einer Vorzeichenumkehrung von Schiefe und standardisierten Werten bei negativem b.

An Stelle der Gleichung für den Erwartungswert können wir auch schreiben:

$$E(b \cdot X + a) = b \cdot E(X) + a$$

Selbstverständlich gelten diese Gleichungen auch für $b = 1$ oder für $a = 0$; so ergeben sich die Spezialfälle:

$$E(X + a) = E(X) + a$$

$$\text{und } E(b \cdot X) = b \cdot E(X)$$

Wir können diese Formeln auf das Beispiel unseres "Würfeltests" anwenden: Bezeichnen wir die Zahl der durch Würfeln falsch gelösten Aufgaben als X' , so gilt offenbar:

$$X' = 5 - X$$

oder - was dasselbe ist -

$$X' = (-1) \cdot X + 5$$

Wie können wir demnach den Erwartungswert, die Varianz und die Standardabweichung der Zahl der Falschlösungen bestimmen?

Unsere letzte Gleichung für X' ist eine Lineartransformation mit $b = -1$ und $a = 5$. Demnach gilt:

$$\mu_{X'} = (-1) \cdot \mu_X + 5 = (-1) \cdot 0.8333... + 5 = 4.1666...$$

$$\sigma_{X'}^2 = (-1)^2 \cdot \sigma_X^2 = 1 \cdot 0.69444... = 0.69444...$$

$$\sigma_{X'} = |-1| \cdot \sigma_X = 1 \cdot 0.8333... = 0.8333...$$

Für einen X -Wert von 4 stellten wir ein z_X von 3.8 fest. Welcher standardisierte Wert entspricht demnach der zugehörigen Zahl von $X' = 1$ Falschlösung?

Wir könnten natürlich von den Kennziffern der Verteilung von X' ausgehend schreiben

$$z_{X'} = \frac{X' - \mu_{X'}}{\sigma_{X'}} = \frac{1 - 4.1666...}{0.8333...} = -3.8$$

Zu dem gleichen Ergebnis kämen wir auch durch Anwendung des einschlägigen Lehrsatzes: Bei Lineartransformation mit negativem b ändert sich nur das Vorzeichen des z -Wertes, so daß wir von einem z_X von 3.8 unmittelbar auf ein $z_{X'}$ von -3.8 kommen.

Bisher haben wir immer gesagt, b und a müßten Konstanten sein. In der beschreibenden Statistik bedeutete dies, daß wir für alle Vpn die gleichen Werte b und a zu verwenden hatten. Bei Wahrscheinlichkeitsverteilungen genügt es entsprechend, wenn die Größen b und a feste Größen sind, d.h. Größen, deren Wert feststeht unabhängig von dem Ausgang des "Zufallsexperiments", dessen mögliche Ausgänge von unserer Wahrscheinlichkeitsverteilung beschrieben werden. Es ist dabei nicht erforderlich, daß der Wert der Größe bekannt ist. Wir können beispielsweise für jeden Test, der aus n Aufgaben besteht, die nur richtig oder falsch gelöst werden können, für die Beziehung der Zahl der richtigen Lösungen (X) und die Zahl der Falschlösungen (X') schreiben:

$$X' = (-1) \cdot X + n$$

Damit gilt die Gleichung

$$\mu_{X'} = (-1) \cdot \mu_X + n$$

auch dann, wenn n eine unbekannte Größe ist, z.B. wenn es uns darum geht, das Verhalten von V_{pn} in einem noch zu entwickelnden Test wahrscheinlichkeitstheoretisch zu beschreiben, bevor die Zahl der Aufgaben feststeht. Hauptsache ist, daß n zum Zeitpunkt der Testbearbeitung feststeht und unabhängig ist von Zufälligkeiten, von denen X abhängt.

<< Noch eine weitere Eigenschaft des Durchschnitts \bar{x} läßt sich sinngemäß auf den Erwartungswert μ_x übertragen: Die Least-Squares Eigenschaft. In @ Statistik I stellte sich die Frage, bei welcher Zahl a die Summe $\sum_i (X_i - a)^2$ der Abweichungsquadrate am kleinsten ist, und es ergab sich, daß dies genau dann der Fall ist, wenn a der Durchschnitt \bar{x} ist. Um dieses Ergebnis sinngemäß auf diskrete Wahrscheinlichkeitsverteilungen zu übertragen, müssen wir es wieder - wie bei vielen bisherigen und folgenden Überlegungen - so umformulieren, daß aus der Summe $\sum_i \dots$ eine Summe $\sum_k p_k \cdot \dots$ wird, in der p_k eine relative Häufigkeit ist, die dann als Wahrscheinlichkeit uminterpretiert wird. Nun kann man leicht die folgende Gleichung herleiten, die bei allen Stichproben für jede beliebige Zahl a gilt (Übungsaufgabe):

$$\sum_i (X_i - a)^2 = \sum_k f_k \cdot (x_k - a)^2 = N \cdot \sum_k p_k \cdot (x_k - a)^2.$$

Wenn aber die Summe $\sum_i (X_i - a)^2$ für jede Zahl a genau N mal so groß ist wie die Summe $\sum_k p_k \cdot (x_k - a)^2$, dann werden beide Summen durch dieselbe Zahl a minimiert. M.a.W.: Die Forderung, die Summe $\sum_i (X_i - a)^2$ durch die Wahl einer geeigneten Zahl a möglichst klein zu machen, ist äquivalent mit der Forderung, daß die Summe $\sum_k p_k \cdot (x_k - a)^2$ möglichst klein sein soll. Im Kontext der beschreibenden Statistik wäre die Forderung kaum zu motivieren, daß die Summe $\sum_k p_k \cdot (x_k - a)^2$ durch die Wahl einer geeigneten Zahl a möglichst klein werden soll; aber diese Formulierung der Least-Squares-Forderung läßt sich auf diskrete Wahrscheinlichkeitsverteilungen übertragen, indem man p_k als Wahrscheinlichkeit uminterpretiert.

Dann ergibt sich die folgende

Regel: In diskreten Wahrscheinlichkeitsverteilungen ist die Zahl a , bei der die Summe $\sum_k p_k \cdot (x_k - a)^2$ am kleinsten wird, der Erwartungswert μ_x .

>>

h) Kleinere Besonderheiten von Wahrscheinlichkeitsverteilungen

Wir haben die Wahrscheinlichkeit eines x -Werts der k -ten Klasse bisher in Analogie zu den Häufigkeitsverteilungen als p_k bezeichnet. Daneben ist aber auch die Bezeichnung $f(x)$ für die Wahrscheinlichkeit eines möglichen Wertes x der Zufallsvariablen X üblich.

Zur Erinnerung: Nach einer in Abschnitt @ eingeführten Konvention steht ein großes X für die Zufallsgröße, und ein kleines x für einen möglichen Wert dieser Größe, also eine Zahl. Daher können wir auch schreiben:

$$f(x) = p(X = x).$$

Wichtig: Wenn wir p_k schreiben, haben wir als k die "fortlaufende Nummer" der Klasse einzusetzen. Bei der hier behandelten anderen Schreibweise ist dagegen nicht die Klassennummer, sondern der zugehörige Wert der x -Skala einzusetzen.

Bei unserer Verteilung aus Tabelle 3 könnte man beispielsweise schreiben: $p_3 = 0.1608$.

Wie könnte man dies in der neuen Schreibweise zum Ausdruck bringen?

Der zugehörige x -Wert ist 2; wir hätten also zu schreiben:

$$f(2) = 0.1608$$

Entsprechend ist $F(x)$ die kumulative Wahrscheinlichkeit eines x -Werts.

Wie könnte man also die aus Tabelle 3 zu entnehmende Aussage $p_{cum\ 2} = 0.8038$ auch schreiben?

Da es sich um die kumulative Häufigkeit des x -Werts 1 handelt, wäre zu schreiben:

$$F(1) = 0.8038$$

<<

Mit der Schreibweise $f(x)$ läßt sich auch eine Formel für den Erwartungswert angeben, bei der keine Klassenindices mehr auftauchen:

$$E(X) = \sum_x x \cdot f(x).$$

Dabei steht \sum_x für "Summe über alle Zahlen x , die als Werte der Zufallsvariablen infragekommen. Bei unserer Beispielsverteilung sind das die ganzen Zahlen von 0 bis 5.

Entsprechend kann man den Erwartungswert einer von X abgeleiteten Größe $g(X)$ angeben als

$$E[g(X)] = \sum_x g(x) \cdot f(x).$$

>>

Bisher haben wir Wahrscheinlichkeitsverteilungen genau so wie die empirischen Verteilungen in Form von Tabellen angegeben. Häufig ist es jedoch möglich, statt dessen eine Formel anzugeben, mit der man die Wahrscheinlichkeiten der verschiedenen x -Werte berechnen kann.

Zu Beginn von Abschnitt a haben wir beispielsweise eine solche Formel für unsere Wahrscheinlichkeitsverteilungen aufgestellt. Wollen Sie versuchen, diese Formel umzuwandeln in eine Formel für $f(x)$?

Bei der dortigen Schreibweise haben wir für den Wert, dessen Wahrscheinlichkeit anzugeben war den Buchstaben a eingesetzt. Bei unserer neuen Schreibweise steht dagegen der Buchstabe x für diesen Wert, dessen Wahrscheinlichkeit angegeben wird. Demnach können wir schreiben:

$$f(x) = \binom{5}{x} \cdot \left(\frac{1}{6}\right)^x \cdot \left(\frac{5}{6}\right)^{5-x}$$

Der Sinn dieser Gleichung wird wohl klar, wenn wir für x einen bestimmten Wert, z.B. 2 einsetzen. Dann gibt die Gleichung die schon früher hergeleitete Formel für die Wahrscheinlichkeit dieses x -Wertes an:

$$f(2) = \binom{5}{2} \cdot \left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right)^3$$

Die Beschreibung von Wahrscheinlichkeitsverteilungen in Gleichungsform hat den Vorteil, daß die gesamte Information kurz und übersichtlich dargeboten wird, und daß daher mathematische Ableitungen leichter möglich sind. Vor allem ermöglicht es diese Darstellungsweise, eine Formel anzugeben, in der man eine ganze "Familie" von Wahrscheinlichkeitsverteilungen zusammenfassen kann. Die Besprechung einer solchen Familie ist Gegenstand des nächsten Abschnitts. Dabei wird sich dann auch die genaue Definition des Begriffs einer Familie ergeben.

2) Die Binomialverteilung

a) Die Voraussetzungen und die Formel der Binomialverteilung

Unter dem Begriff "Binomialverteilung" wird eine Familie von Verteilungen zusammengefaßt, die alle eines gemeinsam haben: Es wird eine Reihe von "Versuchen" durchgeführt, bei denen ein bestimmtes "kritisches Ereignis" auftreten kann. Gezählt wird die Zahl der Versuche, bei denen ein kritisches Ereignis auftritt. Diese Zahl der Versuche, in denen das "kritische Ereignis" auftritt, wird als zufällige Größe X aufgefaßt und hat unter den im folgenden beschriebenen Bedingungen eine Binomialverteilung.

Die verschiedenen Anwendungsfälle der Binomialverteilung unterscheiden sich zunächst durch die verschiedene Art der "Versuche" und der "kritischen Ereignisse". Es kann sich um das "Ziehen einer schwarzen Kugel" (= kritisches Ereignis) beim Kugelziehen aus einer Urne (= Versuch) oder um das Auftreten einer 6 (= kritisches Ereignis) beim Würfeln (= Versuch) oder um viele andere Versuche und kritische Ereignisse handeln. Diese inhaltlichen Unterschiede machen an der Wahrscheinlichkeitsverteilung der Zahl der kritischen Ereignisses nichts aus. Diese hängt von zwei Größen ab (die allerdings von den "inhaltlichen Unterschieden" nicht ganz unabhängig sind): Von der Zahl der durchgeführten Versuche, für die wir das Symbol n verwenden und von der sogenannten "Grundwahrscheinlichkeit p ". Letzteres ist die Wahrscheinlichkeit, daß das Ereignis in einem bestimmten Versuch auftritt. Wir haben einen Spezialfall der Binomialverteilung schon kennengelernt, als es um die Lösung unseres 5-Aufgaben-Tests mit Würfeltechnik ging.

Sehen Sie eine Möglichkeit, eine einmalige Durchführung dieses Tests in "Versuche" und "kritische Ereignisse" aufzuteilen?

Wir haben an anderer Stelle den "Würfeltest" schon in anderer Form in "Versuche" aufgeteilt; In Abschnitt 1b @ ging es an einer Stelle um eine mehrfache Durchführung dieses Tests, Hier geht es aber darum, eine einmalige Durchführung des Tests in Versuche und kritische Ereignisse aufzuteilen.

Wir könnten beispielsweise sagen: Das Bearbeiten einer Aufgabe des Tests stellt einen Versuch dar, und das kritische Ereignis besteht darin, daß die Aufgabe richtig gelöst wird.

Wie groß sind hier die Größen n und p ?

n ist die Zahl der durchgeführten Versuche (also 5), und p ist die Wahrscheinlichkeit, daß bei einem bestimmten Versuch (z.B. beim 1.) ein kritisches Ereignis (also eine richtige Lösung) auftritt - in unserem Fall also $1/6$.

Damit die Formel für die Binomialverteilung angewandt werden kann, müssen die folgenden Voraussetzungen erfüllt sein:

- 1. Die Wahrscheinlichkeit des Auftretens eines kritischen Ereignisses muß bei allen n Versuchen gleich sein. (Diese Wahrscheinlichkeit ist dann die "Grundwahrscheinlichkeit*

p'').

2. Die verschiedenen Versuche müssen (zumindest hinsichtlich des kritischen Ereignisses) völlig unabhängig voneinander sein.

Sind die Voraussetzungen erfüllt, dann kann man die *Wahrscheinlichkeitsverteilung* der zufälligen Größe X (d.h. der Zahl der Versuche, bei denen das kritische Ereignis auftritt) angeben als

$$f(x) = \binom{n}{x} \cdot p^x \cdot (1 - p)^{n-x}$$

Häufig bezeichnet man die Wahrscheinlichkeit, daß das kritische Ereignis in einem Versuch *nicht* auftritt, als q . Nach dem Negationssatz gilt dann offenbar:

$$q = 1 - p$$

Mit dieser Definition kann man die Formel für die Wahrscheinlichkeitsverteilung noch etwas vereinfachen:

$$f(x) = \binom{n}{x} \cdot p^x \cdot q^{n-x}$$

Wollen Sie sich vor dem Beweis dieser Gleichung davon überzeugen, daß sie zumindest für unseren Spezialfall einer Binomialverteilung ("Würfetest") zutrifft?

Wenn wir die hier gültigen Werte $n=5$, $p=1/6$ und $q=5/6$ in die Gleichung einsetzen, erhalten wir die schon im Abschnitt 1b @ aufgestellte Formel.

<<

Der Beweis der Richtigkeit dieser Formel ist eine Verallgemeinerung unseres Beweises über die Wahrscheinlichkeit, im Würfetest zwei richtige Lösungen zu erzielen.

Wollen Sie den Beweis selbst versuchen?

Genauso wie wir beim Würfetest zunächst das Ereignis "zwei richtige Lösungen" aufteilen in verschiedene Abfolgen von richtigen und falschen Lösungen, so teilen wir jetzt das Ereignis " x Versuche mit kritischem Ereignis" auf in verschiedene Abfolgen von Versuchen mit und ohne kritische Ereignisse.

Insgesamt gibt es offenbar $\binom{n}{x}$ Möglichkeiten, aus den n Versuchen eine Untergruppe von x Versuchen herauszuziehen, so daß bei diesen das kritische Ereignis auftritt. Anders formuliert: Es gibt insgesamt $\binom{n}{x}$ Abfolgen, in denen das kritische Ereignis bei x Versuchen auftritt und bei den übrigen $n - x$ Versuchen nicht auftritt. Die Wahrscheinlichkeit jeder einzelnen solchen Abfolge ergibt sich - ganz analog unseren Überlegungen beim Würfetest - nach dem Multiplikationssatz als ein Produkt, in dem x -mal der Faktor p und $(n - x)$ -mal der Faktor q (oder $(1 - p)$) vorkommt. Also ist die Wahrscheinlichkeit jeder Abfolge mit x kritischen Ereignissen $p^x \cdot q^{n-x}$. Die Wahrscheinlichkeit, daß irgendeine dieser Abfolgen mit x kritischen Ereignissen auftritt, kann man nach dem Additionssatz für einander ausschließende Ereignisse²¹ durch Zusammenfassung

²¹Es ist auch hier selbstverständlich, daß die verschiedenen möglichen Abfolgen einander ausschließende Ereignisse darstellen.

der $\binom{n}{x}$ einander gleichen Einzelwahrscheinlichkeiten bestimmen als

$$f(x) = \binom{n}{x} \cdot p^x \cdot q^{n-x}$$

und das war zu beweisen.

Die Bezeichnung "Binomialverteilung" kann man zunächst damit begründen, daß in dieser Verteilung Binomialkoeffizienten eine wichtige Rolle spielen. Wenn man aber die (am Ende von Abschnitt @II.2.c angegebene) allgemeine Formel für Binome auf den Ausdruck $(p + q)^n$ anwendet, dann erkennt man leicht eine noch weitergehende Begründung für die Bezeichnung "Binomialverteilung": Nach dieser Formel ist

$$(p + q)^n = \sum_{x=0}^n \binom{n}{x} \cdot p^x \cdot q^{n-x}$$

Daß wir den "Laufindex" der Summe - abweichend von der allgemeinen Formel - als x und nicht als r bezeichnet haben, stellt natürlich keinen sachlichen Unterschied dar; es erleichtert aber die Feststellung: Die Summanden in der Summe auf der rechten Seite sind die Wahrscheinlichkeiten der einzelnen x -Werte.

>>

Anhand der allgemeinen Gleichung für die Binomialverteilung läßt sich nun auch der Begriff einer "Familie" von Verteilungen präziser fassen:

Eine Familie von Verteilungen ist eine Menge von Verteilungen, bei denen dieselbe Formel für $f(x)$ gilt, wobei in dieser Formel aber zusätzlich zu x noch weitere Symbole auftauchen, die für unterschiedliche Zahlen stehen können.

Die Zahlen, für die diese Symbole stehen, werden auch die "Parameter" der zu der Familie gehörenden Verteilungen genannt. Deswegen spricht man auch von einer "parametrischen Familie von Verteilungen".

<<

Wir wollen nun noch etwas im einzelnen auf die Voraussetzungen der Binomialverteilung eingehen, und zwar zunächst auf den Begriff des *Versuchs*. Veranschaulichen wir uns das Problem nochmals an unserem "Würfetest". Wir können einerseits sagen, daß der ganze Test eine Versuchsreihe von fünf Versuchen (= Aufgaben) darstellt. Wenn nun die Vp den Test mehrmals mit Würfeln bearbeitet, dann können wir zunächst die genannte Betrachtungsweise beibehalten und für jede einzelne Durchführung mit Hilfe der Binomialverteilung die Wahrscheinlichkeit des Ereignisses "zwei richtige Lösungen" als 0.1608 angeben. Wenn dagegen gefragt wird, wie groß ist die Wahrscheinlichkeit, bei 12-maliger Durchführung des Würfetests vier mal zwei richtige zu erzielen, dann könnte man auch diese Frage durch eine Änderung der Betrachtungsweise - insbesondere hinsichtlich der Definition des "Versuchs" und des "kritischen Ereignisses" - mit Hilfe der Binomialverteilung beantworten.

Auf welche Änderung der Betrachtungsweise wird hier wohl angespielt?

Wir könnten wieder wie in Abschnitt @ III.2.a jede der 12 Durchführungen des Tests als einen Versuch auffassen, bei dem das "kritische Ereignis 2 richtige Lösungen" mit einer für alle

Versuche gleichen Wahrscheinlichkeit von 0.1608 auftritt. Da die verschiedenen Versuche (Testergebnisse) offenbar wieder völlig unabhängig voneinander sind, können wir die Binomialverteilung anwenden.

Wie groß wäre also die Wahrscheinlichkeit, bei 12-maliger Durchführung des Tests viermal zwei richtige Lösungen zu erzielen?

Mit einem $n=12$ und einer Grundwahrscheinlichkeit von 0.1608 ist die Wahrscheinlichkeit, viermal das kritische Ereignis "zwei richtige Lösungen" zu erzielen, anzugeben als

$$f(4) = \binom{12}{4} \cdot 0.1608^4 \cdot (1 - 0.1608)^{12-4}$$

Was wir an diesem Beispiel vor allem sehen können, ist folgendes: Es ist nicht objektiv vorgegeben, was "Versuch", "Versuchsreihe" und "kritisches Ereignis" ist. Es richtet sich nach der Betrachtungsweise, die von der zu beantwortenden Frage nahegelegt wird: Das Ereignis, das gemäß einer zu suchenden Wahrscheinlichkeit mit einer bestimmten Häufigkeit auftreten soll, ist das "kritisches Ereignis" und die Prozedur, bei der das kritische Ereignis mit einer immer gleichen Wahrscheinlichkeit auftritt, ist ein Versuch.

Es ist dabei nicht erforderlich, daß die verschiedenen Versuche und die dabei auftretenden kritischen Ereignisse alle identisch sind. In unserem Würfeltest bestand z.B. bei einmaliger Durchführung das kritische Ereignis darin, die Aufgabe richtig zu lösen. Das wird aber in verschiedenen Aufgaben in der Regel bei einer verschiedenen Augenzahl des Würfels der Fall sein. Die fünf Versuche (Aufgaben) sind also nicht identisch, und bei jedem ist das kritische Ereignis in gewisser Hinsicht ein anderes. Es ist Sache des Anwenders der Binomialverteilung, zu entscheiden, welche verschiedenen möglichen Versuchsdurchgänge er zu der Äquivalenzklasse "kritisches Ereignis" zählt. Es muß nur gewährleistet sein, daß die beiden Voraussetzungen (gleiche Wahrscheinlichkeit des kritischen Ereignisses in allen Versuchen und Unabhängigkeit der Versuche) erfüllbar sind. Hinsichtlich der Unabhängigkeitsvoraussetzung ist zu ergänzen, daß die Versuche nicht in jeder Hinsicht, sondern nur in Bezug auf das kritische Ereignis unabhängig sein müssen. Machen wir uns das wieder an einem Beispiel klar: Wir wollen annehmen, daß eine Vp unseren 5-Aufgaben-Test nicht mehr mit einem Würfel, sondern mit üblichem Raten löst. Der VI (= Versuchsleiter) hat aber vorher für jede Aufgabe mit Würfeln entschieden, auf welche der Alternative A bis F die richtige Lösung kommt. Es ist nun bekannt, daß Vpn beim Raten in multiple-choice-Tests Wiederholungen der gleichen Alternativen vermeiden. Die Wahrscheinlichkeit, daß die Vp in der 2. Aufgabe Alternative A wählt, ist also geringer, wenn sie in der 1. Aufgabe schon A gewählt hat, als wenn sie dort eine andere angegeben hat. Die verschiedenen Aufgaben (Versuche) sind also hinsichtlich des Verhaltens der Vp nicht unabhängig.

Trotzdem gilt aber die stochastische Unabhängigkeit in unserem Fall noch hinsichtlich des Ereignisses "richtige Lösung". Man kann sich das plausibel machen, indem man dieses Ereignis umformuliert in: Der VI würfelt bei der betreffenden Aufgabe die Nummer derjenigen Alternative, die die Vp später raten wird. Die Unabhängigkeit kommt dann durch das Würfeln des VI zustande: Die Wahrscheinlichkeit, daß der VI für die 2. Aufgabe die von der Vp später angegebene Alternativnummer als Position für die richtige Alternative erwürfelt, ist offenbar unabhängig davon, ob ein entsprechendes Ereignis beim vorhergehenden Wurf eintrat oder nicht. Genau so kann man auch die völlige Unabhängigkeit aller "Versuche" hinsichtlich des Ereignisses "richtige Lösung" beweisen.

Dieses Beispiel ist aus zwei Gründen interessant: Einmal im Zusammenhang der Binomialverteilung, deren 2. Voraussetzung (vollständige Unabhängigkeit der Versuche) spezifiziert wird als Unabhängigkeit hinsichtlich des kritischen Ereignisses. Außerdem

demonstriert dieses Beispiel eine wesentliche Möglichkeit, Menschen wahrscheinlichkeitstheoretisch zu analysieren: Eine ratende Vp verhält sich nicht wie ein Würfel - das Raten bei der 1. und bei der 2. Aufgabe ist nicht stochastisch unabhängig -, und trotzdem kann man die Anwendbarkeit von wahrscheinlichkeitstheoretischen Modellannahmen (wie z.B. der stochastischen Unabhängigkeit von Richtig- und Falschlösungen in verschiedenen Aufgaben) durch geeignete versuchsplanerische Maßnahmen herstellen (nämlich durch die zufällige Bestimmung der Position der richtigen Lösung). Über solche Techniken erfährt man näheres in einigen Lehrbüchern der Experimentellen Psychologie unter dem Stichwort "Randomisierung".

Eine weitere Bemerkung betrifft die Definition der zufälligen Größe X in der Binomialverteilung. Bisher haben wir entweder von der "Zahl der kritischen Ereignisse" oder von der "Zahl der Versuche, in denen das kritische Ereignis auftritt" gesprochen. Beide Formulierungen sind natürlich äquivalent, wenn in keinem Versuch das kritische Ereignis mehrmals auftreten kann. Andernfalls ist X die Zahl der Versuche, in denen das kritische Ereignis auftritt.

Beispiel: Ein "Versuch" besteht darin, per Zufall eine Person aus der Population zu ziehen und sie einen schwierigen Satz lesen zu lassen. Das kritische Ereignis besteht darin, daß sich die Vp verliert. Dieses kritische Ereignis kann natürlich in einem Versuch mehrmals auftreten. Man wiederholt den Versuch mehrmals. Hat man durch geeignete Maßnahmen sichergestellt, daß die Versuche voneinander unabhängig sind und daß die Wahrscheinlichkeit der kritischen Ereignisse bei allen Versuchen gleich ist,²² dann ist die Zahl der Versuche, in denen das Ereignis "Verlesen" auftritt, binomialverteilt. D.h.: Die Wahrscheinlichkeit, daß sich x Vpn verlesen, läßt sich mit Hilfe der Binomialverteilung angeben. Für die Wahrscheinlichkeit, daß x kritische Ereignisse (Lesefehler) auftreten, gilt dies jedoch nur unter ganz speziellen Voraussetzungen.

All diese Probleme bekommen einen anderen Stellenwert, wenn man die Rolle von Definition und daraus abgeleiteter Konsequenz vertauscht, und das ist in der mathematischen Behandlung der Binomialverteilung (und anderer Verteilungsfamilien) üblich. Unsere obige Darstellungsweise könnte man folgendermaßen zusammenfassen:

- Die Definition lautet: Eine Zufallsvariable X ist binomialverteilt, wenn es n Versuche mit folgenden Eigenschaften gibt:
 - Die Wahrscheinlichkeit, daß in einem Versuch ein "kritisches Ereignis auftritt", ist bei allen Versuchen gleich.
 - Die Versuche sind hinsichtlich des kritischen Ereignisses vollständig unabhängig voneinander.
 - Die Zufallsvariable X ist die Zahl der Versuche, in denen ein kritisches Ereignis auftritt.
- Daraus haben wir dann die Formel der Binomialverteilung abgeleitet.

Für die Mathematik gilt:

- Definition: Eine Zufallsgröße X ist binomialverteilt, wenn es eine natürliche Zahl n und eine Wahrscheinlichkeit p derart gibt, daß bei jeder ganzen Zahl von 0 bis n die Formel

$$p(X = x) = f(x) = \binom{n}{x} \cdot p^x \cdot (1 - p)^{n-x}$$

für die Wahrscheinlichkeit des Ereignisses $X = x$ gilt.

- Daraus wird dann als Konsequenz der "Lehrsatz" hergeleitet: Ist n eine Zahl von Versuchen

²²Solche Maßnahmen werden wir im Abschnitt "Stichprobentheorie" behandeln.

... (usw., wie in unserer Definition vorausgesetzt), dann ist die Wahrscheinlichkeitsverteilung der Zufallsgröße X eine Binomialverteilung.

Diese Herangehensweise hat einen erheblichen Vorteil: Alle mathematischen Schlußfolgerungen über die Binomialverteilung lassen sich aus der Wahrscheinlichkeits-Formel herleiten. Damit stehen sie aber auf einem solideren Fundament, als wenn wir die unterschiedlich ausdeutbaren Voraussetzungen mit n Versuchen usw. zugrundelegen.

>>

b) Kennziffern der Binomialverteilung

Es lassen sich mathematisch ziemlich einfach die folgenden allgemeinen Formeln für den Erwartungswert und die Varianz der Binomialverteilung herleiten:²³

$$\mu_x = E(X) = n \cdot p$$

$$\sigma_x^2 = n \cdot p \cdot (1 - p) = n \cdot p \cdot q$$

Wollen Sie als Anwendungsbeispiel für diese Formeln unsere Berechnungen von Erwartungswert und Varianz der zufälligen Größe "richtige Lösungen" im Würfeltest mit fünf Aufgaben überprüfen?

Die zufällige Größe X (also die Zahl der richtigen Lösungen) hat, wie wir im Abschnitt II.2.a gesehen haben, eine Binomialverteilung mit $n = 5$ und $p = 1/6$, also $q = 5/6$. Damit ergibt sich:

$$\mu_x = E(X) = 5 \cdot 1/6 = 0.8333\dots$$

$$\sigma_x^2 = 5 \cdot 1/6 \cdot 5/6 = 25/36 = 0.69444\dots$$

<< An diesem Beispiel können wir einen weiteren Vorteil der formelmäßigen (und nicht tabellarischen) Angabe von Wahrscheinlichkeitsverteilungen demonstrieren: Häufig ist es möglich, aus der Formel einer Wahrscheinlichkeitsverteilung Formeln für ihre Kennziffern herzuleiten. Für den Erwartungswert der Binomialverteilung gilt nach der allgemeinen Formel für Erwartungswerte zunächst:

$$\mu_x = \sum_{x=0}^n x \cdot \binom{n}{x} \cdot p^x \cdot q^{n-x}$$

Wir nehmen nun einige Umformungen vor, die nur dadurch motiviert sind, daß sie zu einem überschaubaren Ergebnis führen. Zunächst ist ja der erste Summand (also derjenige für $x = 0$) auf jeden Fall 0, da mit x multipliziert wird. Daher können wir auf diesen Summanden auch verzichten und schreiben

$$\mu_x = \sum_{x=1}^n x \cdot \binom{n}{x} \cdot p^x \cdot q^{n-x}$$

²³Die Herleitungen finden sich "für Spezialisten" am Ende dieses Abschnitts.

Wenn wir den Binomialkoeffizienten $\binom{n}{x}$ ausschreiben, erhalten wir

$$\mu_x = \sum_{x=1}^n x \cdot \frac{n!}{(n-x)! \cdot x!} \cdot p^x \cdot q^{n-x}$$

Nun können wir anstelle von $x!$ im Nenner auch $(x-1)! \cdot x$ schreiben und das x vor dem Bruch wegkürzen. Dann ergibt sich

$$\mu_x = \sum_{x=1}^n \frac{n!}{(n-x)! \cdot (x-1)!} \cdot p^x \cdot q^{n-x}$$

Eine weitere Umformung der Summe ergibt sich aus folgender Überlegung: Statt den Index x von 1 bis n laufen zu lassen, können wir auch einen Index r von 0 bis $n-1$ laufen lassen und dafür überall $r+1$ anstelle von x schreiben. Insbesondere ist p^x durch p^{r+1} zu ersetzen, und dafür können wir auch $p \cdot p^r$ schreiben. Ergebnis:

$$\mu_x = \sum_{r=0}^{n-1} \frac{n!}{(n-r-1)! \cdot r!} \cdot p \cdot p^r \cdot q^{n-r-1}$$

Außerdem definieren wir die Größe m als $m := n - 1$. Dann können wir überall $m+1$ anstelle von n schreiben, und aus $n!$ wird $n \cdot m!$. Damit erhalten wir:

$$\mu_x = \sum_{r=0}^m n \cdot \frac{m!}{(m-r)! \cdot r!} \cdot p \cdot p^r \cdot q^{m-r}$$

Nun können wir den Bruch als Binomialkoeffizienten $\binom{m}{r}$ schreiben. Außerdem enthält jeder Summand die Faktoren n und p , die wir vor die Summe ziehen ("ausklammern") können. Dann erhalten wir:

$$\mu_x = n \cdot p \cdot \sum_{r=0}^m \binom{m}{r} \cdot p^r \cdot q^{m-r}$$

Damit hat die Summe eine "handliche" Form bekommen: Nach der allgemeinen Binom-Formel ist sie gleich $(p + q)^m$, und da die Gleichung $p + q = 1$ unmittelbar aus der Definition von q folgt, ist die Summe gleich 1. Dann bleibt die einfache Gleichung $\mu_x = n \cdot p$ übrig.

Den einen oder anderen "Super-Spezialisten" mag es reizen, jetzt auch noch die Formel für die Varianz herzuleiten. Dazu aber nur ein Tip: Man leitet auf ähnliche Weise den Erwartungswert $E[X \cdot (X-1)]$ her und verwendet die Gleichung $E[X \cdot (X-1)] = E(X^2) - E(X)$.

Nur erwähnt sei, daß es für die Herleitung von Formeln für die Momente beliebiger Verteilungen eine Art Standardmethode gibt. Diese Methode der "momenterzeugenden Funktionen", die allerdings einige Ansprüche an die mathematische Versiertheit ihrer Benutzer stellt, ist z.B. bei Kreyszic (1965) dargestellt.

>>

3) Einige weitere eindimensionale diskrete Wahrscheinlichkeitsverteilungen

Dieser Abschnitt, der ganz "für Spezialisten" ist, wird überarbeitet und nachgeliefert.

4) Mehrdimensionale diskrete Wahrscheinlichkeitsverteilungen

a) Der Begriff der mehrdimensionalen diskreten Wahrscheinlichkeitsverteilung, dargestellt an einem Beispiel

Genauso wie es enge Beziehungen zwischen eindimensionalen Häufigkeits- und Wahrscheinlichkeitsverteilungen gibt, gibt es mehrdimensionale Wahrscheinlichkeitsverteilungen als Parallele zu den mehrdimensionalen Häufigkeitsverteilungen. Eine zweidimensionale Häufigkeitsverteilung bestand in einer Zuordnung von empirischen absoluten (oder relativen) *Häufigkeiten* zu Meßwertpaaren bzw. zu den 2-dimensionalen Klassen einer 2-dimensionalen Skala. Auch hier wollen wir uns zunächst nur mit diskreten Verteilungen befassen. Dazu gilt die folgende

Definition: Eine diskrete 2-dimensionale Wahrscheinlichkeitsverteilung besteht in einer Zuordnung von Wahrscheinlichkeiten zu Meßwertpaaren bzw. zu den 2-dimensionalen Klassen einer diskreten 2-dimensionalen Skala.

Als Beispiel für eine 2-dimensionale Wahrscheinlichkeitsverteilung wollen wir wieder eine Vp betrachten, die einen Test mit Würfeltechnik löst. Der Test besteht diesmal aus drei Aufgaben, bei denen jeweils fünf Antwortmöglichkeiten zur Wahl gestellt werden. Die Vp würfelt für jede Aufgabe einmal und kreuzt die der gewürfelten Augenzahl entsprechende Antwortalternative als Lösung an. Würfelt sie eine 6, so "überschlägt" sie die Aufgabe. Sie geht zur nächsten Aufgabe über, ohne eine Lösung anzukreuzen.

Wir können nun zwei Zufallsvariablen definieren:

X = Zahl der richtigen Lösungen; Y = Zahl der falschen Lösungen.

Wir können nun versuchen, für jedes denkbare Wertepaar der Zufallsvariablen X und Y die Wahrscheinlichkeit zu bestimmen. Bei dieser Berechnung würde man im Prinzip ganz ähnlich vorgehen wie bei dem früheren Würfeltest, und daher brauchen wir diese Berechnung hier nicht von neuem vorzuführen. (Wollen Sie trotzdem als Übung in der Kombinatorik die folgende Gleichung überprüfen: $p(X=1 \wedge Y=2) = 24/216 = 0.2222\dots?$).

Die Ergebnisse dieser Berechnung sind in der folgenden Tabelle zusammengetragen:

$Y \setminus X \rightarrow$	0	1	2	3	Randverteilung
3	0.2963	0	0	0	0.2963
2	0.2222	0.2222	0	0	0.4444
1	0.0556	0.1111	0.0556	0	0.2222
0	0.0046	0.0139	0.0139	0.00463	0.0370
Rand- verteilung	0.5787	0.3472	0.0694	0.00463	1.0000

Tabelle 6: Zweidimensionale Wahrscheinlichkeitsverteilung der richtigen Lösungen (X) und der falschen Lösungen (Y) in einem "Würfeltest".

Warum ist im oberen rechten Teil überall die Wahrscheinlichkeit 0 angegeben?

In diesem Teil der Tabelle stehen lauter unmögliche Ereignisse. Es ist beispielsweise unmöglich, $X=1$ richtige Lösungen und $Y=3$ falsche Lösungen zu erzielen, da der Test nur drei Aufgaben hat.

Genau so wie aus einer 2-dimensionalen Häufigkeitsverteilung können wir aus einer 2-dimensionalen Wahrscheinlichkeitsverteilung die Randverteilung von X und Y herausziehen, indem wir die Eintragungen in der Tabelle entweder senkrecht oder waagerecht addieren. Diese Randverteilungen stellen die "Randwahrscheinlichkeiten" bestimmter x - bzw. y -Werte dar.

Beispiel: Die Wahrscheinlichkeit des Ereignisses $X = 1$ erhalten wir durch Addition der Wahrscheinlichkeiten aller Werte-Paare mit $X=1$:

$$p(X=1) = p(X=1 \wedge Y=0) + p(X=1 \wedge Y=1) + \text{usw.} = 0.0139 + 0.1111 + 0.2222 + 0 = 0.3472$$

Mit welchem Wahrscheinlichkeitsansatz können wir diese additive Berechnung der Randverteilung begründen?

Die Ereignisse " $X=1 \wedge Y=0$ " oder " $X=1 \wedge Y=1$ " usw., deren Wahrscheinlichkeiten in der Tabelle stehen, schließen sich gegenseitig aus. Wir verwenden also den Additionssatz für einander ausschließende Ereignisse, aus dem sich die obige Additionsregel für die Randverteilung ohne

weiteres ergibt.²⁴

Genau so wie bei den eindimensionalen gibt es auch bei den 2-dimensionalen Wahrscheinlichkeitsverteilungen viele weitere, fast wörtliche Parallelen zu den entsprechenden Häufigkeitsverteilungen, die sich alle erklären lassen, indem man die Wahrscheinlichkeiten als erwartete relative Häufigkeiten interpretiert. Auf diese Parallelen wollen wir im übernächsten Abschnitt eingehen. Daneben gibt es ähnliche Besonderheiten der 2-dimensionalen Wahrscheinlichkeitsverteilungen wie bei den eindimensionalen:

1. Für die Wahrscheinlichkeit des Ereignisses $X = x \wedge Y = y$ schreibt man auch $f(x,y)$.
Was würde beispielsweise das Symbol $f(2, 1)$ bedeuten und wie groß wäre diese Wahrscheinlichkeit?

Definitionsgemäß ist $f(2, 1)$ die Wahrscheinlichkeit des Ereignisses " $X=2 \wedge Y=1$ ". Aus unserer Tabelle können wir entnehmen $f(2,1) = 0.0556$.

2. Neben der tabellarischen Darstellung gibt es auch bei 2-dimensionalen Wahrscheinlichkeitsverteilungen oft die Möglichkeit einer formelmäßigen Angabe der Wahrscheinlichkeiten aller möglichen Meßwertpaare.

<< Für unsere Verteilung könnte man z.B. die folgende Formel angeben:

$$f(x,y) = \frac{3!}{x! \cdot y! \cdot (3 - x - y)!} \cdot \left(\frac{1}{6}\right)^x \cdot \left(\frac{4}{6}\right)^y \cdot \left(\frac{1}{6}\right)^{3-x-y}$$

Eine Herleitung dieser Formel würde ganz ähnlich verlaufen wie bei dem früher behandelten Würfeltest. Natürlich gilt diese Formel nur für $x + y \leq 3$. Bei $x + y > 3$ haben wir dagegen ein unmögliches Ereignis und damit eine Wahrscheinlichkeit von 0.

- >>
3. Genau so wie es bei den eindimensionalen Verteilungen Familien von Verteilungen gibt, die sich nur dadurch unterscheiden, daß in die allgemeine Formel verschiedene spezielle Zahlen anstelle von Buchstaben eingesetzt werden,²⁵ gibt es auch bei den 2-dimensionalen

²⁴ Andere in Tabelle @6 angegebene Randwahrscheinlichkeiten sind nicht genau gleich der Summe der Eintragungen in der entsprechenden Spalte oder Zeile. Das liegt an Rundungsfehlern. Für genauere Berechnungen kann man Brüche mit Nenner 216 bilden. Für die Wahrscheinlichkeiten im Inneren der Tabelle lautet dann der Zähler

64	0	0	0
48	48	0	0
12	24	12	0
1	3	3	1

Beispiel: Die in der Tabelle angegebene Wahrscheinlichkeit für $X=0$ und $Y=3$ steht für $64/216$.

Bei Berechnungen in den folgenden Abschnitten sind diese exakten Wahrscheinlichkeiten zugrundegelegt. Daher können sich beim Nachrechnen mit dem Taschenrechner kleine Abweichungen ergeben, wenn man mit den Zahlen aus Tabelle @6 rechnet.

²⁵Beispiel: Familie der Binomialverteilungen, die sich nur dadurch voneinander unterscheiden, daß man in die allgemeine Formel der Binomialverteilung verschiedene Werte für n und p einsetzt.

Verteilungen solche Gruppen von Verteilungen. Unsere Verteilung ist z.B. ein Spezialfall der sogenannten Multinomialverteilung (einer Verallgemeinerung der Binomialverteilung, die bei Hays, 19@, näher beschrieben ist).

b) Bedingte Verteilung

Neben der Randverteilung können auch andere eindimensionale Verteilungen aus der "gemeinsamen Verteilung" von X und Y herausgezogen werden: Die bedingten Verteilungen. Wir könnten uns z.B. fragen, welche "bedingte" Wahrscheinlichkeitsverteilung die zufällige Größe Y hat, wenn wir wissen, daß $X=1$ ist. Wie groß ist z.B. die bedingte Wahrscheinlichkeit $p(Y=0 / X=1)$? Nach der Formel:

$$p(A/B) = \frac{p(A \wedge B)}{p(B)}$$

können wir schreiben:

$$p(Y=0/X=1) = \frac{p(Y=0 \wedge X=1)}{p(X=1)} = \frac{0.0139}{0.3472} = 0.040$$

Entsprechend können wir die bedingte Wahrscheinlichkeit aller möglichen y -Werte unter der Bedingung $X=1$ angeben. Sie sind in den beiden ersten Spalten von Tabelle 7 zusammengestellt. Dabei steht $f(y / X=1)$ für die bedingte Wahrscheinlichkeit $p(Y=y / X=1)$.

y	$f(y / X=1)$	$y \cdot f(y / X=1)$
0	0.040	0.000
1	0.320	0.320
2	0.640	1.280
3	0.000	0.000
Σ	1.000	1.600

Tabelle 7 Bedingte Wahrscheinlichkeitsverteilung von Y für $X=1$.

Den Erwartungswert dieser bedingten Verteilung können wir berechnen, indem wir die in der zweiten Spalte der Tabelle aufgeführten bedingten Wahrscheinlichkeiten mit dem entsprechenden y -Wert (erste Spalte) multiplizieren und diese Produkte (dritte Spalte) aufsummieren. Das Ergebnis - das in unserem Fall 1.600 beträgt - nenn man auch den "bedingten Erwartungswert von Y , gegeben $X=0$ ".

Für die bedingten Wahrscheinlichkeiten der y -Werte und den entsprechenden bedingten Erwartungswert sind die Symbole $f(y/X=0)$ und $E(Y/X=0)$ üblich. Allgemein schreibt man auch $f(y/X)$ und $E(Y/X)$; diese bedingten Wahrscheinlichkeiten und Erwartungswerte sind natürlich

meistens eine Funktion des gegebenen Wertes von X . Für unseren Fall gilt z.B. die Formel²⁶

$$E(Y/X) = 4/5 \cdot (3-X)$$

Wesentlich ist, daß ein bedingter Erwartungswert der Form $E(Y/X=0)$ eine feste Zahl ist (hier: 1.600), während $E(Y/X)$ eine Funktion von X ist.

<<

Die obige Herleitung einer bedingten Wahrscheinlichkeitsverteilung läßt sich in der allgemeinen Formel

$$f(y/X=x) = \frac{f(x,y)}{f(x)}$$

zusammenfassen. Sie hat allerdings einen "Schönheitsfehler": Der Buchstabe f taucht mit drei verschiedenen Bedeutungen auf; aber das Gemeinte sollte klar sein.

>>

Genau so kann man natürlich auch bedingte Verteilungen der Y -Werte für einen gegebenen X -Wert, und die zugehörigen bedingten Erwartungswerte angeben.

In den Häufigkeitsverteilungen gab es eine gewisse Parallele zu den bedingten Verteilungen und bedingten Erwartungswerten. Worin bestand diese Parallele?

Im Zusammenhang mit dem Eta-Koeffizienten (vgl. Skript Statistik I, Abschnitt @AII2e(ii)) wurde in einer zweidimensionalen Häufigkeitsverteilung (Tabelle aus Abschnitt @AIII1) die Verteilung von Y -Werten derjenigen V_{pn} betrachtet, die einen bestimmten X -Wert erzielten. Diese Verteilung ist eine Art bedingte Verteilung von Y , gegeben ein bestimmter X -Wert. Die Durchschnitte dieser Teilverteilungen (die in der Tabelle mit Kreuzchen markiert sind), entsprechen den bedingten Erwartungswerten von Y für die X -Werte der jeweiligen Spalte.

c) Weitere Parallelen von 2-dimensionalen Wahrscheinlichkeits- und Häufigkeitsverteilungen

Genau so wie bei 2-dimensionalen Häufigkeitsverteilungen kann man auch bei 2-dimensionalen Wahrscheinlichkeitsverteilungen neben den Kennziffern der Randverteilungen weitere Kennziffern und Formeln angeben, die diejenige Information auswerten, die nur in der gemeinsamen Verteilung und nicht in den Randverteilungen steckt. Eine Produkt-Moment-Korrelation, eine Kovarianz, Regressionsgeraden und im Zusammenhang damit Schätzfehler-varianzen, Standardschätzfehler sowie einen Schätzungseffekt. Was dazu wichtig ist, ist vor allem das "Grundrezept", das wir schon von den eindimensionalen Wahrscheinlichkeitsverteilungen her

²⁶Statt eines Beweises: Kennt man die Zahl der richtigen Lösungen (X), dann erwartet man, daß von übrigen $(3 - X)$ Aufgaben etwa $4/5$ falsch gelöst und $1/5$ überschlagen werden. (Die Wahrscheinlichkeit einer Falschlösung ist 4 mal so groß wie die des Überschlagens!)

kennen: Man nehme eine Formel aus der deskriptiven Statistik, in der ein arithmetisches Mittel zugrundegelegt wird, und ersetze "arithmetisches Mittel von ..." durch "Erwartungswert von ...". Natürlich kann man jeden solchen Erwartungswert auch als eine Summe schreiben, und für diejenigen, denen solche Summenformeln das Verständnis erleichtern, sind diese im folgenden auch ausgeschrieben.

<<

Für dieses Ausschreiben von Erwartungswerten in Summenformeln müssen wir eine kleine Ergänzung unserer Schreibweise einführen, die an einem Beispiel veranschaulicht werden soll. In einer Formel für die Kovarianz tauchte in der beschreibenden Statistik das arithmetische Mittel des Produkts $X \cdot Y$ auf, und dafür konnten wir dort einfach die Summe $\sum_i X_i \cdot Y_i$ durch N dividieren. Nun gibt es ja bei Wahrscheinlichkeitsverteilungen keine "Individuen", über die man summieren könnte, und deshalb konnten wir auch bei den eindimensionalen Verteilungen nur diejenigen Formeln aus der deskriptiven Statistik sinngemäß übertragen, die auf einer Summation über Klassen beruhen und bei denen der Ausdruck, dessen arithmetisches Mittel zu bilden ist, mit der Klassenwahrscheinlichkeit multipliziert wird. Da wir solche Formeln für zweidimensionale Häufigkeitsverteilungen nicht eingeführt haben, soll dies hier kurz nachgeholt werden.

Dazu vereinbaren wir, daß die x -Klassen - wie bisher - mit dem Index k durchnummeriert werden und daß in gleicher Weise für die y -Klassen der Index m verwendet wird. x_k ist also der x -Wert der k -ten x -Klasse, und y_m der y -Wert der m -ten y -Klasse. Nun besteht der eigentliche Informationsgehalt zweidimensionaler Häufigkeitsverteilungen ja in der Häufigkeit, mit der bestimmte x -Werte zusammen mit bestimmten y -Werten auftreten. Damit diese Häufigkeiten in Formeln verwendet werden können, soll f_{km} die (absolute) Häufigkeit bezeichnen, mit der ein x -Wert der k -ten x -Klasse gepaart mit einem y -Wert der m -ten y -Klasse auftritt, und p_{km} ist die entsprechende relative Häufigkeit. Tritt z.B. in einer Stichprobe von $N=200$ Personen bei 18 Personen ein x -Wert der zweiten x -Klasse gepaart mit einem y -Wert der dritten y -Klasse auf, dann ist $f_{23} = 18$ und $p_{23} = 18 / 200 = 0.09$. Man spricht dies "f zwei drei" bzw. "p zwei drei" (und nicht "f dreiundzwanzig" oder "p dreiundzwanzig").

Mit dieser Vereinbarung können wir nun zum arithmetischen Mittel des Produkts $X \cdot Y$ zurückkehren und die übliche Überlegung anstellen: Statt für jede einzelne Person das Produkt $X_i \cdot Y_i$ zu bilden und diese Produkte aufzuaddieren, können wir auch für die aus der k -ten x -Klasse und der m -ten y -Klasse gebildete "zweidimensionale Klasse" das Produkt $x_k \cdot y_m$ mit f_{km} multiplizieren und diese Produkte über alle solchen zweidimensionalen Klassen aufsummieren. Als Formel geschrieben:

$$\sum_i X_i \cdot Y_i = \sum_{km} f_{km} \cdot x_k \cdot y_m.$$

Das Zeichen \sum_{km} bedeutet dabei: Summe über alle aus einem k und einem m gebildeten zweidimensionalen Klassen. Hätten wir beispielsweise 3 x -Klassen und 4 y -Klassen, dann wäre diese Summe auszuschreiben als

$$\begin{aligned} \sum_{km} f_{km} \cdot x_k \cdot y_m &= f_{11} \cdot x_1 \cdot y_1 + f_{12} \cdot x_1 \cdot y_2 + f_{13} \cdot x_1 \cdot y_3 + f_{14} \cdot x_1 \cdot y_4 \\ &+ f_{21} \cdot x_2 \cdot y_1 + f_{22} \cdot x_2 \cdot y_2 + f_{23} \cdot x_2 \cdot y_3 + f_{24} \cdot x_2 \cdot y_4 \\ &+ f_{31} \cdot x_3 \cdot y_1 + f_{32} \cdot x_3 \cdot y_2 + f_{33} \cdot x_3 \cdot y_3 + f_{34} \cdot x_3 \cdot y_4. \end{aligned}$$

Da diese Summe mit der Summe der Produkte $X_i \cdot Y_i$ übereinstimmt, können wir für das arithmetische Mittel des Produkts $X \cdot Y$ auch schreiben:

$$aM(X \cdot Y) = \frac{\sum_i X_i \cdot Y_i}{N} = \frac{\sum_{km} f_{km} \cdot x_k \cdot y_m}{N} = \sum_{km} p_{km} \cdot x_k \cdot y_m$$

Die Begründung für das letzte Gleichheitszeichen ist dabei die gleiche, die wir schon aus anderen

solchen Zusammenhängen kennen: Statt die gesamte Summe durch N zu dividieren, können wir das auch mit jedem einzelnen Summanden tun, und das geschieht faktisch, wenn wir f_{km} durch p_{km} ersetzen.

Diese Überlegung läßt sich natürlich in gleicher Weise anstellen, wenn wir nicht das arithmetische Mittel des Produkts $X \cdot Y$ bilden wollen, sondern irgendeiner anderen aus X und Y abgeleiteten Größe $g(X, Y)$, wobei $g(X, Y)$ z.B. das Produkt der z -Werte oder der Abweichungswerte sein kann. (Im letzteren Fall wäre also $g(X, Y) = (X - \bar{x}) \cdot (Y - \bar{y})$). Dann können wir denselben Gedankengang anstellen und kommen zu dem Ergebnis

$$aM [g(X, Y)] = \frac{\sum_i g(X_i, Y_i)}{N} = \frac{\sum_{km} f_{km} \cdot g(x_k, y_m)}{N} = \sum_{km} p_{km} \cdot g(x_k, y_m)$$

Nun haben wir eine Formel, die sich auf zweidimensionale Wahrscheinlichkeitsverteilungen übertragen läßt.

>>

Für den Erwartungswert einer von X und Y abgeleiteten Größe $g(X, Y)$ (z.B. des Produkts $X \cdot Y$) gilt die allgemeine Gleichung

$$E [g(X, Y)] = \sum_{km} p_{km} \cdot g(x_k, y_m)$$

Dabei bezeichnet p_{km} die Wahrscheinlichkeit des gemeinsamen Auftretens eines x -Werts der k -ten x -Klasse und eines y -Werts der m -ten y -Klasse, und die Summe \sum_{km} bedeutet, daß der rechts daneben stehende Ausdruck $p_{km} \cdot g(x_k, y_m)$ für alle aus einer k -ten x -Klasse und einer m -ten y -Klasse gebildeten zweidimensionalen Klasse zu bilden und aufzusummieren ist.

Beispiel: In der zweidimensionalen Wahrscheinlichkeitsverteilung aus Tabelle @6 soll der (in einer späteren Kovarianz-Formel verwendete) Erwartungswert des Produkts $X \cdot Y$ gebildet werden. Wir haben also $g(X, Y) = X \cdot Y$. Daher haben wir für jedes Feld im Inneren der Tabelle 6 das Produkt $p_{km} \cdot x_k \cdot y_m$ zu bilden und diese Produkte aufzusummieren. Die Produkte sind in der folgenden Tabelle berechnet (exakte Werte der Produkte: s. Fußnote zu Tabelle 6!):

$Y \setminus X \rightarrow$	0	1	2	3
3	$0.2963 \cdot 0 \cdot 3 = 0$	$0.0000 \cdot 1 \cdot 3 = 0$	$0.0000 \cdot 2 \cdot 3 = 0$	$0.0000 \cdot 3 \cdot 3 = 0$
2	$0.2222 \cdot 0 \cdot 2 = 0$	$0.2222 \cdot 1 \cdot 2 = 0.4444$	$0.0000 \cdot 2 \cdot 2 = 0$	$0.0000 \cdot 3 \cdot 2 = 0$
1	$0.0556 \cdot 0 \cdot 1 = 0$	$0.1111 \cdot 1 \cdot 1 = 0.1111$	$0.0556 \cdot 2 \cdot 1 = 0.1111$	$0.0000 \cdot 3 \cdot 1 = 0$
0	$0.0046 \cdot 0 \cdot 0 = 0$	$0.0139 \cdot 1 \cdot 0 = 0$	$0.0139 \cdot 2 \cdot 0 = 0$	$0.0046 \cdot 3 \cdot 0 = 0$

Zu Tabelle @6: Berechnung der Produkte $p_{km} \cdot x_k \cdot y_m$.

Die Summe dieser Produkte ist 0.666..., und das ist dann der Erwartungswert $E(X \cdot Y)$.

<<

Mit der Schreibweise $f(x, y)$ für die Wahrscheinlichkeit des Ereignisses $X = x \wedge Y = y$ kann man die Formel für den Erwartungswert einer von X und Y abgeleiteten Zufallsgröße $g(X, Y)$ auch als

$$E [g(X, Y)] = \sum_x \sum_y f(x, y) \cdot g(x, y)$$

angeben. Diese Darstellung wird später (in Abschnitt @B.IV.4.c) hilfreich sein, um den Erwartungswert abgeleiteter Zufallsgrößen in weiteren zweidimensionalen Wahrscheinlichkeitsverteilungen zu bestimmen.

>>

Nach diesen Vorbemerkungen sollen nun die Kennziffern und ihre Formeln kurz vorgestellt werden. (Vorschlag: Wollen Sie sich zum besseren Verständnis zu jeder Formel die entsprechende Formel aus der beschreibenden Statistik suchen?)

In unserer 2-dimensionalen Wahrscheinlichkeitsverteilung aus Abschnitt @II.4.a gibt es eine gegenläufige Beziehung von X und Y : Bei großen X -Werten sind vor allem kleine Y -Werte wahrscheinlich und umgekehrt. Man sieht es in der Tabelle, und es ist auch plausibel: Wer mehr Aufgaben richtig löst, macht tendenziell weniger Fehler. Diese gegenläufige Beziehung drückt sich in einer negativen Produkt-Moment-Korrelation aus. Als Symbol für diese Korrelation wird wieder ein griechischer Buchstabe herangezogen, da es sich um eine Kennziffer einer Wahrscheinlichkeitsverteilung handelt: ρ_{xy} (sprich: rho x y) ist die Produkt-Moment-Korrelation der zufälligen Größen X und Y .²⁷

Die Formel der Produkt-Moment-Korrelation in Wahrscheinlichkeitsverteilungen baut genau so wie bei Häufigkeitsverteilungen auf dem Produkt von z_X und z_Y auf:

$$\rho_{xy} = E(z_X \cdot z_Y) = \sum_{km} p_{km} \cdot z_{xk} \cdot z_{ym}$$

Dabei ist z_{xk} der zum x -Wert der k -ten x -Klasse gehörende z -Wert und z_{ym} entsprechend der zum y -Wert der m -ten y -Klasse gehörende z -Wert.

Beispiel: Der x -Wert der 1. x -Klasse in unserer 2-dimensionalen Verteilung aus Abschnitt @II.4.a ist 0. Den Erwartungswert und die Standardabweichung von X kann man aus der Randverteilung von X berechnen²⁸ als $\mu_x = 0.50$ und $\sigma_x = 0.65$.

Wie groß ist z.B. z_{x1} , also der zu dem x -Wert der 1. x -Klasse gehörende z -Wert?

$$z_{x_1} = \frac{x_1 - \mu_x}{\sigma_x} = \frac{0 - 0.50}{0.65} = - 0.77$$

Für Y ergibt sich entsprechend $\mu_y = 2$ und $\sigma_y = 0.67$, und für den Wert der 1. y -Klasse ergibt sich

²⁷Um Mißverständnisse beim Lesen anderer Bücher zu vermeiden, ist darauf hinzuweisen, daß das Symbol ρ_{xy} manchmal für die Spearman'sche Rangkorrelation und nicht für die Produkt-Moment-Korrelation in 2-dimensionalen Wahrscheinlichkeitsverteilungen steht.

²⁸Wollen Sie beweisen, daß die Randverteilung von X eine Binomialverteilung mit $n=3$ und $p=1/6$ ist?

Aufgrund dieser Tatsache lassen sich die oben genannten Kennziffern auch nach den Formeln der Binomialverteilung berechnen.

Übungsaufgabe: Machen Sie ähnliche Aussagen über die Randverteilung der Y -Werte.

$$z_{y_1} = \frac{y_1 - \mu_y}{\sigma_y} = \frac{0 - 2.00}{0.67} = - 3.00$$

In dieser Weise könnte man für jede x-Klasse und jede y-Klasse einen z-Wert bestimmen und dann die Korrelation ρ_{xy} nach der obigen Formel bestimmen. Dazu könnte man eine ähnliche Tabelle bilden wie für die Bestimmung des Erwartungswerts $E(X \cdot Y)$: In jedes Feld würde man das Produkt $p_{km} \cdot z_{xk} \cdot z_{yk}$ eintragen, und die Summe dieser Produkte wäre die Korrelation ρ_{xy} .

Die Parallelen zur beschreibenden Statistik gehen aber weiter. Genau so wie dort, ist auch bei zweidimensionalen Wahrscheinlichkeitsverteilungen eine andere Formel für die Produkt-Moment-Korrelation ρ_{xy} günstiger für Berechnungen. Diese Formel baut auf der Kovarianz σ_{xy} auf, die analog zur Kovarianz s_{xy} der beschreibenden Statistik definiert ist als:

$$\begin{aligned}\sigma_{xy} &= E [(X - \mu_x) \cdot (Y - \mu_y)] \\ &= \sum_{km} p_{km} \cdot (x_k - \mu_x) \cdot (y_m - \mu_y) \\ &= E (X \cdot Y) - E(X) \cdot E(Y) \\ &= \sum_{km} p_{km} \cdot x_k \cdot y_m - (\sum_k p_k \cdot x_k) \cdot (\sum_m p_m \cdot y_m).\end{aligned}$$

Zur Berechnung können wir die vorletzte Zeile dieser Gleichung zugrundelegen, da der Erwartungswert $E(X \cdot Y) = 0.666...$ schon an früherer Stelle berechnet wurde (S. ?). Kombinieren wir diesen Wert mit den Erwartungswerten der Randverteilungen $E(X) = 0.5$ und $E(Y) = 2.0$, dann erhalten wir

$$\sigma_{xy} = 0.666... - 0.5 \cdot 2.0 = - 0.333....$$

Es läßt sich dann - genau wie in der beschreibenden Statistik - beweisen, daß die obige Definition der Produkt-Moment-Korrelation ρ_{xy} identisch ist mit:

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$$

Für das Beispiel ergibt sich also

$$\rho_{xy} = \frac{-0.333...}{0.65 \cdot 0.82} = -0.63$$

Das negative Vorzeichen dieser Korrelation drückt die gegenläufige Beziehung der Zufallsgrößen X und Y aus, auf die bereits als Anlaß für die Definition und Berechnung der Produkt-Moment-Korrelation und der Kovarianz in zweidimensionalen Wahrscheinlichkeitsverteilungen hingewiesen wurde.

Auch die Schätzungen durch lineare Regressionsgleichungen lassen sich auf zweidimensionale Wahrscheinlichkeitsverteilungen übertragen: Kennen wir den Wert, den die zufällige Größe X annimmt, dann können wir die zufällige Größe Y schätzen als:

$$\hat{Y} = b_{yx} \cdot X + a_{yx}$$

mit

$$b_{yx} = \rho_{xy} \cdot \frac{\sigma_y}{\sigma_x}$$

und

$$a_{yx} = \mu_y - b_{yx} \cdot m_{yx}$$

Entsprechend lautet die Gleichung für die Schätzung von X aus Y :

$$\hat{X} = b_{xy} \cdot Y + a_{xy}$$

mit

$$b_{xy} = \rho_{yx} \cdot \frac{\sigma_x}{\sigma_y}$$

und

$$a_{xy} = \mu_x - b_{xy} \cdot m_{xy}$$

Welchen Fehler wir bei diesen Schätzungen in Kauf nehmen, können wir wieder angeben mit der Schätzfehlervarianz

$$\sigma_{y.x}^2 = E[(Y - \hat{Y})^2] = \sigma_y^2 \cdot (1 - \rho_{xy}^2)$$

bzw.

$$\sigma_{x.y}^2 = E[(X - \hat{X})^2] = \sigma_x^2 \cdot (1 - \rho_{xy}^2)$$

oder mit dem Standardschätzfehler

$$\sigma_{y.x} = \sqrt{\sigma_{y.x}^2} = \sigma_y \cdot \sqrt{1 - \rho_{xy}^2}$$

bzw.

$$\sigma_{x.y} = \sqrt{\sigma_{x.y}^2} = \sigma_x \cdot \sqrt{1 - \rho_{xy}^2}$$

Der Schätzungseffekt würde wieder angeben, um wieviel % dieser Standardschätzfehler geringer

ist, als bei der "Blindschätzung":

$$\hat{Y} = \mu_y \text{ bzw. } \hat{X} = \mu_x$$

Schätzungseffekt:

$$\left(1 - \sqrt{1 - \rho_{xy}^2}\right) \cdot 100\%$$

Diese linearen Regressionsgleichungen haben ähnliche Eigenschaften wie bei der beschreibenden Statistik: Der Erwartungswert des Schätzfehlers ist 0, und jede andere lineare Schätzgleichung hätte eine größere Schätzfehlervarianz als unsere lineare Regressionsgleichung.

<<

Die letztgenannte Eigenschaft entspricht der Least-squares-Eigenschaft der linearen Regressionsgleichungen.

Schließlich läßt sich auch die Unterscheidung von linearen und nichtlinearen stochastischen Abhängigkeiten auf 2-dimensionale Wahrscheinlichkeitsverteilungen übertragen. Das Pendant zu den "Mittelwerten der Teilverteilungen", die wir bei der beschreibenden Statistik verwendeten, sind die bedingten Erwartungswerte (vgl. Abschnitt III.4.b). Entsprechend lautet die *Definition* der linearen Abhängigkeit:

Definition: Die Abhängigkeit der zufälligen Größe X von der zufälligen Größe Y ist linear, wenn der bedingte Erwartungswert von X, gegeben ein Y-Wert, eine lineare Funktion des gegebenen Y-Wertes ist.

Für die umgekehrte Abhängigkeit gilt Entsprechendes. Genau wie bei der beschreibenden Statistik muß auch hier getrennt festgestellt werden, ob X von Y und ob Y von X linear abhängig ist.

>>

Im Zusammenhang mit den gemeinsamen Häufigkeitsverteilungen haben wir verschiedene Lehrsätze über das Verhalten der Kennziffern bei Lineartransformationen und bei Addition oder Linearkombination von mehreren Größen kennengelernt (Zur Addition vgl. Abschnitt @A.II.5 von Statistik I.) Die Lehrsätze und Formeln lassen sich (bis auf einen Austausch von griechischen und lateinischen Buchstaben) so genau auf die mehrdimensionalen Wahrscheinlichkeitsverteilungen übertragen, daß wir sie hier nicht nochmals im Detail aufzuzählen brauchen.

d) Unabhängige und unkorrelierte zufällige Größen

Die Definition von unabhängigen zufälligen Größen ist in gewisser Hinsicht ein Spezialfall der Definition von unabhängigen Ereignissen.

Wie lautete diese?

Von mehreren möglichen Definitionen wollen wir uns hier auf die folgende beziehen: Zwei Ereignisse sind stochastisch unabhängig, wenn die Angabe einer Bedingung über das Eintreffen (oder Nichteintreffen) des einen Ereignisses nichts an der Wahrscheinlichkeit des anderen

Ereignisses ändert, wenn also die entsprechenden bedingten Wahrscheinlichkeiten gleich den zugehörigen Randwahrscheinlichkeiten sind.

Ganz entsprechend gilt für den Begriff der unabhängigen zufälligen Größen die folgende

Definition: Zwei Zufallsvariablen sind stochastisch unabhängig, wenn die Angabe von Bedingungen über den Wert der einen zufälligen Größe nichts an der bedingten Wahrscheinlichkeitsverteilung der anderen ändert, wenn also die entsprechenden bedingten Verteilungen gleich den zugehörigen Randverteilungen sind.

Beispiel. Wir führen den in früheren Abschnitten behandelten Würfeltest, bei dem es in fünf Aufgaben nur richtige oder falsche Lösungen gab, 2-mal durch und bezeichnen als X und Y die Zahl der richtigen Lösungen im 1. und 2. Durchgang. Offenbar sind diese beiden zufälligen Größen stochastisch unabhängig; denn die Angabe eines X -Wertes ändert nichts an der bedingten Verteilung der Y -Werte und umgekehrt: Alle bedingten Verteilungen von Y , gegeben ein Wert von X , sind gleich der Randverteilung von Y und umgekehrt.

Ein weiteres Beispiel für zwei unabhängige Zufallsgrößen ist in Tabelle @6a dargestellt. Es wird dreimal gewürfelt und ausgezählt, bei wie vielen Würfeln eine gerade Augenzahl gewürfelt wird (X) und bei wie vielen Würfeln eine Augenzahl, die größer als 2 ist (Y).

$Y \setminus X \rightarrow$	0	1	2	3	Randverteilung
3	8/216	24/216	24/216	8/216	8/27
2	12/216	36/216	36/216	12/216	12/27
1	6/216	18/216	18/216	6/216	6/27
0	1/216	3/216	3/216	1/216	1/27
Randverteilung	1/8	3/8	3/8	1/8	

Tabelle 6a: Zweidimensionale Wahrscheinlichkeitsverteilung unabhängiger Zufallsvariablen: Gerade Augenzahlen (X) und Augenzahlen, die größer als 2 sind (Y), bei dreimaligem Würfeln.

<<

Die Berechnung dieser Wahrscheinlichkeiten kann man als "Fleißarbeit" betreiben:

- Man stellt leicht fest, daß es $6^3 = 216$ mögliche Abfolgen von Augenzahlen gibt. Diese schreibt man auf.
- Für jede dieser Abfolgen bestimmt man die Werte der Zufallsvariablen X und Y .
- Für jede mögliche Kombination von X und Y zählt man aus, bei wievielen Abfolgen von Augenzahlen sie vorliegt. Die Kombination $X = 1 \wedge Y = 0$ liegt z.B. bei den drei Abfolgen 112, 121 und 211 vor.
- Da jede der 216 möglichen Abfolgen von Augenzahlen gleich wahrscheinlich ist, erhält man z.B. die Wahrscheinlichkeit von $X = 1 \wedge Y = 0$, indem man die Anzahl der entsprechenden Abfolgen durch 216 dividiert, also als $3/216$. Entsprechend natürlich bei allen anderen Werte-Kombinationen der Zufallsvariablen X und Y .

Natürlich gibt es auch kombinatorische Ansätze, die einem diese Fleißarbeit ersparen würden. Darauf braucht aber hier nicht näher eingegangen zu werden.

>>

Im Zusammenhang des Begriffs unabhängiger Zufallsgrößen kommt es nicht darauf an, wie man zu den in der Tabelle angegebenen Wahrscheinlichkeiten kommt. Es soll vor allem demonstriert werden, was es konkret bedeutet, daß alle bedingten Verteilungen mit den entsprechenden Randverteilungen übereinstimmen.

Wollen Sie in ähnlicher Weise, wie die ersten beiden Spalten von Tabelle @7 aus Tabelle @6 abgeleitet wurden, auch für dieses Beispiel die bedingte Wahrscheinlichkeitsverteilung von X für $Y = 0$ berechnen?

Wenn man z.B. für die bedingte Wahrscheinlichkeit $p(X = 0/Y = 0)$ die Zahlen der hiesigen Tabelle in die zuvor zugrundegelegte Formel einsetzt, dann ergibt sich:

$$p(X=0/Y=0) = \frac{p(X=0 \wedge Y=0)}{p(Y=0)} = \frac{1/216}{1/27} = 1/8$$

Führt man entsprechende Berechnungen auch für die übrigen möglichen Werte von X durch, dann ergibt sich eine bedingte Verteilung, die mit der Randverteilung von X genau übereinstimmt. Dasselbe gilt auch, wenn wir anstelle von $Y = 0$ eine andere Bedingung über Y zugrundelegen (z.B. $Y = 2$ oder auch $Y \geq 2$).

<<

Für das "bedingende Ereignis" $Y \geq 2$ kann man natürlich $Y = 2 \vee Y = 3$ schreiben. Damit ergibt sich

$$p(Y \geq 2) = p(Y = 2 \vee Y = 3) = 12/27 + 8/27,$$

wobei das letzte Gleichheitszeichen auf dem einfachen Additionssatz beruht. Entsprechend kann man für das Ereignis $X = 0 \wedge Y \geq 2$ auch $(X = 0 \wedge Y = 2) \vee (X = 0 \wedge Y = 3)$ schreiben und erhält die Gleichung

$$p(X = 0 \wedge Y \geq 2) = p((X = 0 \wedge Y = 2) \vee (X = 0 \wedge Y = 3)) = 12/216 + 8/216.$$

Damit ergibt sich

$$p(X = 0 / Y \geq 2) = p(X = 0 \wedge Y \geq 2) / p(Y \geq 2) = 1/8 = p(X = 0),$$

und aufgrund entsprechender Berechnungen ist die gesamte bedingte Verteilung von X unter der Bedingung $Y \geq 2$ identisch mit der Randverteilung von X .

>>

Ebenso kann man für beliebige Bedingungen über X die bedingte Wahrscheinlichkeitsverteilung von Y bestimmen und erhält als bedingte Verteilung immer die (in der letzten Spalte von Tabelle @6a angegebene) Randverteilung von Y .

Spätestens bei derartigen Berechnungen stellt man eine Eigenschaft der zweidimensionalen Wahrscheinlichkeitsverteilung aus Tabelle @6a fest, auf der die Übereinstimmung von bedingten Verteilungen und Randverteilungen beruht: Jede im Inneren der Tabelle eingetragene Wahrscheinlichkeit ist das Produkt der zugehörigen Randwahrscheinlichkeiten. Und das ist nicht nur in diesem Beispiel so, sondern gilt entsprechend bei allen unabhängigen Zufallsgrößen. Der Bezug zum einfachen Multiplikationssatz ist auch intuitiv einleuchtend: Sind X und Y unabhängige Zufallsvariablen und bezieht ein Ereignis A sich auf X und Ereignis B auf Y , dann sind die Ereignisse A und B unabhängig, und daher ist die Verbundwahrscheinlichkeit das

Produkt der Randwahrscheinlichkeiten.

<<

Für die mathematische Wahrscheinlichkeitstheorie ist dies sogar Grundlage der Definition unabhängiger Zufallsvariablen. Nach dieser Definition sind zwei Zufallsvariablen X und Y unabhängig, wenn für jedes auf die Zufallsvariable X bezogene Ereignis A und jedes auf die Zufallsvariable Y bezogene Ereignis B die Gleichung $p(A \wedge B) = p(A) \cdot p(B)$ gilt.

Diese Definition ist zwar weniger anschaulich als die auf den bedingten Verteilungen aufbauende; sie hat aber einen ähnlichen formalen Vorteil wie die (in Abschnitt @B.II.1.a erwähnte) Definition unabhängiger Ereignisse: Wenn die Randwahrscheinlichkeit eines auf die Zufallsvariable Y bezogenen Ereignisses B null ist, dann können wir keine bedingte Verteilung von X gegeben B bilden; denn dazu müßten wir durch $p(B)$ dividieren.

>>

Genau so wie bei mehreren Ereignissen muß man auch bei mehr als zwei zufälligen Größen unterscheiden zwischen paarweiser und vollständiger Unabhängigkeit: Paarweise Unabhängigkeit liegt vor, wenn die oben genannte Unabhängigkeitsbedingung für jedes Paar von Zufallsvariablen gilt; vollständige Unabhängigkeit liegt aber erst dann vor, wenn die bedingte Verteilung jeder Zufallsvariablen auch bei Angabe von Bedingungen über die Werte aller übrigen Zufallsvariablen nicht verändert wird.

Ein anderer von der Unabhängigkeit sorgfältig zu unterscheidende Begriff ist die Unkorreliertheit:

Definition: Zwei zufällige Größen sind unkorreliert, wenn ihre Kovarianz (und damit auch ihre Produkt-Moment-Korrelation) 0 ist.

Es gilt hierzu der folgende, in sich plausible

Lehrsatz: Sind zwei zufällige Größen unabhängig, dann sind sie auch unkorreliert.

Dagegen gilt die Umkehrung dieses Lehrsatzes nicht. Wir haben schon in der beschreibenden Statistik Häufigkeitsverteilungen kennengelernt, bei denen trotz stochastischer Abhängigkeit eine Produkt-Moment-Korrelation von $r = 0$ vorlag. Ähnliche Fälle, in denen zwei zufällige Größen zwar unkorreliert, aber nicht unabhängig sind, gibt es auch bei 2-dimensionalen Wahrscheinlichkeitsverteilungen.

IV Kontinuierliche Wahrscheinlichkeitsverteilungen

1) Grundlegende Begriffe und Beziehungen

Bei den bisher behandelten diskreten Wahrscheinlichkeitsverteilungen handelte es sich (definitionsgemäß) um die Zuordnung von Wahrscheinlichkeiten zu den Klassen einer diskreten Skala. Die Beschränkung auf diesen Typ von Wahrscheinlichkeitsverteilungen ermöglichte es, anschauliche Begriffe aus der beschreibenden Statistik ganz unmittelbar zu übertragen: Es genügte im Prinzip, das Symbol p_k , das in der beschreibenden Statistik für die relative Besetzungshäufigkeit der k -ten Klasse stand, umzuinterpretieren als die Wahrscheinlichkeit, daß der Wert einer Zufallsgröße in diese Klasse fällt.

Auch wenn eine Zufallsgröße alle Werte einer kontinuierlichen Skala annehmen kann, kann vieles aus der beschreibenden Statistik sinngemäß übertragen werden. Während aber bei den diskreten Wahrscheinlichkeitsverteilungen die Übertragung am direktesten aufgrund der Formeln erfolgte, geht bei den kontinuierlichen Wahrscheinlichkeitsverteilungen ein recht anschaulicher Weg über die graphischen Darstellungen. Dazu stellte sich in Statistik I heraus, daß Ogiven oft recht informativ sind, wenn man sie nur zu lesen versteht. Das gilt erst recht bei kontinuierlichen Wahrscheinlichkeitsverteilungen; deshalb werden diese im folgenden über die Ogive eingeführt.

Die wohl am häufigsten benutzten kontinuierlichen Verteilungen gehören entweder selbst zur Familie der "Normalverteilungen", oder sie sind davon abgeleitet. Bei diesen Verteilungen ist aber der Einstieg über die Ogive etwas erschwert. Deshalb wird im folgenden ein Verteilungstyp zugrundegelegt, der in der Psychologie vor allem bei der Modellbildung über kognitive Prozesse eine Rolle spielt.

<<

Ein Standardwerk zu diesem Ansatz ist ein Buch von Luce (1986). Dort wird dargestellt, wie man aus Modellannahmen über den Verlauf kognitiver Prozesse Schlußfolgerungen über die Wahrscheinlichkeitsverteilung von Antwortzeiten ableiten kann. Das Buch arbeitet mit Kenntnissen über kontinuierliche Wahrscheinlichkeitsverteilungen, die über den Rahmen unserer Einführung hinausgehen. Diese Voraussetzungen werden in einem Einleitungskapitel dargestellt, das mit etwas Hintergrundwissen über kontinuierliche Wahrscheinlichkeitsverteilungen sowie über Differential- und Integralrechnung lesbar ist.

In einer systematischen mathematischen Abhandlung müßte natürlich am Anfang eine genaue Definition einer kontinuierlichen Wahrscheinlichkeitsverteilung stehen. Diese läßt sich auch leicht formulieren: Die Wahrscheinlichkeitsverteilung einer Zufallsvariablen X ist kontinuierlich, wenn die (im nächsten Unterabschnitt näher eingeführte) kumulative Verteilungsfunktion

$$F(x) = p(X \leq x)$$

eine differenzierbare Funktion von x ist.²⁹ Aus dieser Definition lassen sich die im folgenden

²⁹In der üblichen mathematischen Definition wird sogar auch noch zugelassen, daß $F(x)$ "an höchstens abzählbar unendlich vielen Stellen" nicht differenzierbar, aber stetig ist. Das erweitert den Anwendungsbereich, ist aber eine mathematische Feinheit, auf die es in einer ersten Einführung nicht ankommt.

dargestellten Eigenschaften systematisch ableiten. Für einen ersten Zugang ist es aber wohl sinnvoller, diese Eigenschaften an einem Beispiel aufzuzeigen, statt sie systematisch herzuleiten.

>>

a) Kumulative Wahrscheinlichkeiten und Ogive

Kumulative Wahrscheinlichkeiten sind bei kontinuierlichen Wahrscheinlichkeitsverteilungen genau so zu interpretieren wie bei diskreten. Wenn man z.B. die von einer Person für die Lösung einer Aufgabe benötigte Zeit (die "Antwortzeit") als Zufallsvariable X bezeichnet³⁰, dann ist die kumulative Wahrscheinlichkeit des Meßwerts 5 Sekunden die Wahrscheinlichkeit, daß die betrachtete Person die Aufgabe in einer Zeit von bis zu 5 Sekunden löst. Wie bei diskreten Wahrscheinlichkeitsverteilungen wird auch bei kontinuierlichen die kumulative Wahrscheinlichkeit einer Zahl x als $F(x)$ bezeichnet. Es gilt also

$$F(x) = p(X \leq x).$$

In Abb. @2a sind nun diese kumulativen Wahrscheinlichkeiten für eine kontinuierliche Wahrscheinlichkeitsverteilung dargestellt, die in diesem Abschnitt und den nächsten als Demonstrationsbeispiel "unter die Lupe genommen" werden soll.³¹ Eine solche graphische

³⁰Statt "Antwortzeit" (engl.: response time) wird oft auch von "Reaktionszeit" gesprochen. Dieser Begriff läßt aber leicht die Vermutung aufkommen, daß es sich um die Zeit für einfachste Reaktionen handelt. Der Begriff "Antwortzeit" bringt dagegen ausdrücklich zum Ausdruck, daß auch komplexere Prozesse (z.B. Problemlösen) gemeint sein können.

Üblicherweise wird Zeit als Zufallsvariable mit T bezeichnet, und eine bestimmte Zeit (z.B. 4 sec) mit t . Da es aber im hiesigen Zusammenhang um eine allgemeine Einführung in kontinuierliche Wahrscheinlichkeitsverteilungen geht, werden die auch in anderen Bereichen verwendbaren Bezeichnungen X und x benutzt.

³¹Die Frage, woher man denn wissen kann, daß die Wahrscheinlichkeitsverteilung der Antwortzeiten gerade die in Abb. @2a und der nachfolgenden Formel dargestellte ist, ist berechtigt. Tatsächlich weiß man das nie genau. Im größeren Zusammenhang von Untersuchungen gibt es immer eine Vielzahl von Wahrscheinlichkeitsverteilungen, die mit den tatsächlich festgestellten Antwortzeiten (den "Daten") vereinbar sind. Um beispielsweise festzustellen, ob die in Abb. @2a dargestellte Wahrscheinlichkeitsverteilung mit den erhaltenen Daten vereinbar ist, würde sich die Frage stellen, ob die Wahrscheinlichkeit dieser Daten bei dieser Verteilung groß genug ist oder ob die Daten nicht unter einer anderen Annahme viel wahrscheinlicher wären. In einem solchen Zusammenhang müßten dann alle Schlußfolgerungen, die aus einer solchen Annahme gezogen werden, streng genommen mit einer Formulierung wie "wenn die Wahrscheinlichkeitsverteilung der Antwortzeiten die in Abb. @2a dargestellte ist, dann gilt folgendes" beginnen. Um diese umständliche Formulierung zu vermeiden, soll ein für allemal festgehalten werden, daß bei allen im folgenden angestellten Überlegungen und Schlußfolgerungen diese Formulierung stillschweigend mitzudenken ist.

Darstellung kumulativer Wahrscheinlichkeiten nennt man - genau wie die entsprechenden Darstellungen bei Häufigkeitsverteilungen oder diskreten Wahrscheinlichkeitsverteilungen - eine *Ogive*, und die dargestellte Funktion $F(x)$ nennt man auch die *kumulative Verteilungsfunktion* der Zufallsvariablen X .

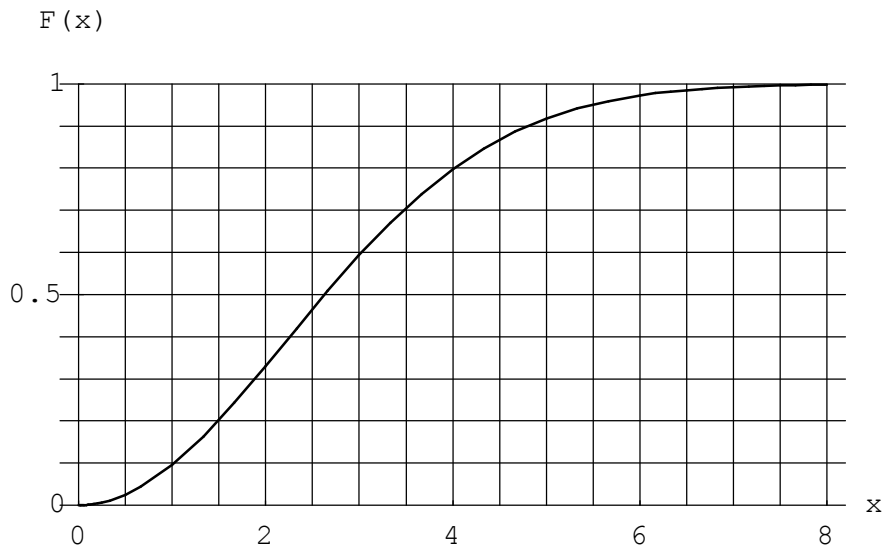


Abb. 2a: Ogive zu einer Wahrscheinlichkeitsverteilung von Antwortzeiten (in Sekunden).

Ablesebeispiel: Die Wahrscheinlichkeit, daß die Antwortzeit X höchstens 4 Sekunden beträgt, ist etwa 0.8. Diese Bestimmung kumulativer Wahrscheinlichkeiten aus der Ogive ist natürlich recht ungenau. Deshalb sind in der folgenden Tabelle @6b die kumulativen Wahrscheinlichkeiten $F(x)$ zu einigen Antwortzeiten x zusammengefaßt.

x	F(x)	x	F(x)	x	F(x)	x	F(x)
0.1	0.001000	2.1	0.356607	4.1	0.813812	6.1	0.975790
0.2	0.003992	2.2	0.383687	4.2	0.828642	6.2	0.978592
0.3	0.008960	2.3	0.410806	4.3	0.842606	6.3	0.981108
0.4	0.015873	2.4	0.437858	4.4	0.855720	6.4	0.983361
0.5	0.024690	2.5	0.464739	4.5	0.868006	6.5	0.985375
0.6	0.035360	2.6	0.491352	4.6	0.879487	6.6	0.987170
0.7	0.047819	2.7	0.517609	4.7	0.890190	6.7	0.988768
0.8	0.061995	2.8	0.543424	4.8	0.900141	6.8	0.990187
0.9	0.077806	2.9	0.568721	4.9	0.909373	6.9	0.991443
1.0	0.095163	3.0	0.593430	5.0	0.917915	7.0	0.992553
1.1	0.113966	3.1	0.617490	5.1	0.925801	7.1	0.993533
1.2	0.134112	3.2	0.640845	5.2	0.933063	7.2	0.994394
1.3	0.155491	3.3	0.663447	5.3	0.939735	7.3	0.995151
1.4	0.177988	3.4	0.685257	5.4	0.945850	7.4	0.995814
1.5	0.201484	3.5	0.706242	5.5	0.951442	7.5	0.996393
1.6	0.225858	3.6	0.726376	5.6	0.956544	7.6	0.996899
1.7	0.250988	3.7	0.745639	5.7	0.961187	7.7	0.997339
1.8	0.276750	3.8	0.764018	5.8	0.965403	7.8	0.997721
1.9	0.303021	3.9	0.781507	5.9	0.969223	7.9	0.998052
2.0	0.329680	4.0	0.798103	6.0	0.972676	8.0	0.998338

Tabelle 6b: Kumulative Wahrscheinlichkeiten $F(x)$ einiger Antwortzeiten (x Sekunden).

Die angegebenen Zeiten x sind dabei als exakte (also nicht gerundete) Dezimalzahlen für die Antwortzeit in Sekunden zu interpretieren. Die erste Eintragung gibt z.B. die kumulative Wahrscheinlichkeit einer Antwortzeit von 0.10000.. Sekunden als 0.001000 an. Dagegen sind die angegebenen kumulativen Wahrscheinlichkeiten $F(x)$ nach den üblichen Regeln auf- oder abgerundet.

<<

Die kumulative Wahrscheinlichkeit von 0.998338 für eine Antwortzeit von 8.0 Sekunden zeigt, daß die in Abb. 2a wiedergegebene Ogive streng genommen unvollständig ist. Nach der zugrundeliegenden Modellgleichung (die gleich dargestellt wird) hat jede noch so große Antwortzeit eine kumulative Wahrscheinlichkeit, die kleiner als 1 ist. Deshalb würde eine vollständige Ogive bis in's Unendliche gehen; aber daß man von Kurven nur den interessanten Teil graphisch darstellt, ist ja gängige Praxis. Daß alle kumulativen Wahrscheinlichkeiten kleiner als 1 sind, bedeutet natürlich auch, daß die Wahrscheinlichkeit einer Antwortzeit von über 200 Jahren noch minimal größer als 0 ist. Das entspricht natürlich nicht der Realität; aber das kann man schmunzelnd übergehen, da man in der Praxis ohnehin froh sein muß, wenn die von einem

Modell angenommenen Wahrscheinlichkeiten wenigstens näherungsweise stimmen.³²

Eine Formel für die kumulativen Wahrscheinlichkeiten, die in unserem Demonstrationsbeispiel zugrundegelegt ist, lautet

$$F(x) = 1 - \frac{1}{e^{0.1 \cdot x^2}}$$

Dabei ist e die (vor allem als Basis der "natürlichen Logarithmen" bekannte) "Euler'sche Zahl" $e \approx 2.718282$.

Wie schon angedeutet wurde, wird eine solche Formel nicht "bewiesen". Sie ergibt sich aus Modellannahmen, die in einem Datensatz überprüft werden.

>>

Zurück zur Ogive. Man kann sie natürlich auch "rückwärts ablesen", also bestimmen, für welche Antwortzeit die kumulative Wahrscheinlichkeit einen bestimmten Wert erreicht. Die waagerechte Linie in Höhe einer Wahrscheinlichkeit von 0.5 schneidet die Ogive an einer Stelle, die für eine Antwortzeit von etwa 2.6 Sekunden steht. Diese Antwortzeit mit einer kumulativen Wahrscheinlichkeit von 0.5 ist dann der C_{50} , also der Median. Man kann diese Zahl natürlich auch aus der Tabelle bestimmen, indem man sie rückwärts liest. Die kumulative Wahrscheinlichkeit bei $x = 2.6$ ist noch etwas kleiner als 0.5, aber bei $x = 2.7$ ist sie schon größer. Also muß der x-Wert mit einer kumulativen Wahrscheinlichkeit von 0.5 zwischen 2.6 und 2.7 liegen. In gleicher Weise kann man natürlich auch andere Centile bestimmen.

<<

Man könnte auch die aus Statistik I bekannte Formel für die Interpolation von Zentilen verwenden, um zu etwas genaueren Zahlen zu kommen. Für den C_{50} ist z.B. der Intervallanfang 2.6, die Intervallbreite 0.1 Sekunden, und die kumulative Wahrscheinlichkeit am Intervallanfang entnehmen wir der Tabelle @6b als 0.491352. Die Wahrscheinlichkeit eines Werts im Intervall ergibt sich als Differenz der kumulativen Wahrscheinlichkeiten am Intervallende und am Intervallanfang, also als $0.517609 - 0.491352 = 0.026257$.

Noch genauer wäre es allerdings, den x-Wert mit kumulativer Wahrscheinlichkeit p (z.B. 0.5) zu bestimmen, indem man die Gleichung für F(x) nach x auflöst. Es ist also der x-Wert gesucht, für den die Gleichung

$$p = 1 - \frac{1}{e^{0.1 \cdot x^2}}$$

gilt. Daraus ergibt sich zunächst

$$e^{0.1 \cdot x^2} = \frac{1}{1 - p}$$

dann

³²Das gilt - wohlbemerkt - nur, wenn es lediglich um die Wahrscheinlichkeit geht. Auf Folgeprobleme, die sich ergeben können und die keinesfalls mit ähnlicher Leichtigkeit übergangen werden können, wird an späterer Stelle eingegangen (s. Abschnitt @B.IV.2.c: Existenz des Erwartungswerts.)

$$0.1 \cdot x^2 = \ln\left(\frac{1}{1-p}\right)$$

(wobei ln für "natürlicher Logarithmus" steht), und schließlich

$$x = \sqrt{10 \cdot \ln\left(\frac{1}{1-p}\right)}$$

Setzen wir in diese Formel z.B. $p = 0.5$ ein, dann erhalten wir für die Antwortzeit mit kumulativer Wahrscheinlichkeit 0.5 - also den Median - einen Wert von 2.6328.

>>

Aus der Interpretation von Ogiven bei Häufigkeitsverteilungen und diskreten Wahrscheinlichkeitsverteilungen kennen wir die Regel, daß die Sprunghöhe in einem Punkt der Besetzungshäufigkeit der entsprechenden Klasse bzw. der Wahrscheinlichkeit des entsprechenden x-Werts entspricht. Da die Ogive in Abbildung @2a keine Sprünge enthält, sondern überall glatt ansteigt, stellt sich die Frage, ob das etwa bedeutet, daß jeder einzelne x-Wert die Wahrscheinlichkeit 0 hat. Das ist in der Tat so, auch wenn das auf den ersten Blick überraschen mag. Man kann einen Vergleich ziehen, von dem sich später sogar herausstellen wird, daß er ziemlich genau stimmt. Geschwindigkeit ist ja zurückgelegte Strecke pro Zeit. Aber auch bei einem Körper, der sich sehr schnell bewegt, ist die in einem bestimmten Zeitpunkt zurückgelegte Strecke 0. Das schließt aber nicht aus, daß dieser Körper in dem betrachteten Zeitpunkt eine hohe Geschwindigkeit hat. Was dieser "Geschwindigkeit in einem Zeitpunkt" bei den kontinuierlichen Wahrscheinlichkeitsverteilungen entspricht, wird in Abschnitt @B.IV.1.b gezeigt. Im Augenblick ist die folgende Verallgemeinerung festzuhalten:

Bei einer kontinuierlichen Wahrscheinlichkeitsverteilung ist die Wahrscheinlichkeit, daß eine Zufallsgröße X einen ganz bestimmten Wert x hat, für jedes x null.

Das kann sogar als Definition verstanden werden: Eine Wahrscheinlichkeitsverteilung, bei der diese Aussage zutrifft, ist eine kontinuierliche Wahrscheinlichkeitsverteilung, und eine Zufallsvariable, bei der das gilt, ist eine kontinuierlich Zufallsvariable. Entsprechend wird eine Zufallsvariable mit diskreter Wahrscheinlichkeitsverteilung auch als diskrete Zufallsvariable bezeichnet.

<<

Ein solcher Vergleich veranschaulicht einen Sachverhalt, aber er beweist natürlich nichts. Daß in unserer Beispielsverteilung die Wahrscheinlichkeit eines Werts der Zufallsvariablen X in jedem Zeitpunkt x tatsächlich 0 ist, kann man folgendermaßen beweisen. Wenn x irgendeine bestimmte Antwortzeit ist und x' eine zweite, kleinere Antwortzeit, dann läßt sich das Ereignis $X \leq x$ zerlegen in die Ereignisse $X \leq x'$, $x' < X < x$ und $X = x$. Da diese Ereignisse sich gegenseitig ausschließen, können wir nach dem einfachen Additionssatz schreiben:

$$\begin{aligned} p(X \leq x) &= p(X \leq x' \vee x' < X < x \vee X = x) \\ &= p(X \leq x') + p(x' < X < x) + p(X = x). \end{aligned}$$

Schreibt man nun $F(x)$ bzw. $F(x')$ für die Wahrscheinlichkeiten $p(X \leq x)$ und $p(X \leq x')$, dann ergibt sich aus der obigen Gleichung mit einfachen Umformungen

$$F(x') = F(x) - p(X = x) - p(x' < X < x).$$

Da aber die letzte Wahrscheinlichkeit $p(x' < X < x)$ in dieser Gleichung keine negative Zahl sein kann, gilt auch

$$F(x') \leq F(x) - p(X = x).$$

Da dies für jede Antwortzeit x' gilt, die kleiner als x ist, ergibt sich: Wäre die Wahrscheinlichkeit $p(X = x)$ nicht 0, dann müßte die Ogive im Punkt X einen Sprung mindestens um den Betrag $p(X = x)$ machen, damit für jedes x' , das kleiner als x ist, die obige Beziehung gelten kann. Daß die Ogive keine Sprünge macht, sieht man "schon mit bloßem Auge"; aber das schließt natürlich nicht aus, daß es Sprünge gibt, die so klein sind, daß man sie nicht sieht. Verlässlicher ist die folgende Überlegung: Wenn die Wahrscheinlichkeit $p(X = x)$ nicht 0 wäre, dürfte es zum Beispiel kein x' mit der kumulativen Wahrscheinlichkeit $F(x') = F(x) - 0.5 \cdot p(X = x)$ geben; denn ein solches x' müßte kleiner als x sein (da seine kumulative Wahrscheinlichkeit kleiner wäre als die von x), und damit wäre die obige, für alle $x' < x$ gültige Ungleichung verletzt. Tatsächlich läßt sich aber mit der entsprechenden Formel ein x' -Wert mit kumulativer Wahrscheinlichkeit $F(x') = F(x) - 0.5 \cdot p(X = x)$ bestimmen:

$$x' = \sqrt{10 \cdot \ln \left(\frac{1}{1 - (F(x) - 0.5 \cdot p(X = x))} \right)}$$

Zusammengefaßt: Die Annahme einer von 0 verschiedenen Wahrscheinlichkeit $p(X = x)$ führt zu Konsequenzen, die mit unseren bisherigen Ergebnissen unvereinbar sind. Also kann die Wahrscheinlichkeit $p(X = x)$ nur 0 sein.

>>

Noch eine andere Eigenschaft der Ogiven von Häufigkeitsverteilungen und diskreten Wahrscheinlichkeitsverteilungen, die eng mit den Sprunghöhen zusammenhängt, läßt sich auf die Ogiven von kontinuierlichen Wahrscheinlichkeitsverteilungen übertragen: Sind x' und x'' zwei Zahlen, von denen x' die kleinere ist, dann entspricht die Zunahme der kumulativen Wahrscheinlichkeit von x' zu x'' (also die Höhendifferenz der Ogive) der Häufigkeit bzw. Wahrscheinlichkeit von Werten, die größer als x' und kleiner oder gleich x'' sind. Für die Antwortzeiten $x' = 4$ und $x'' = 6$ können wir z.B. der Ogive entnehmen, daß die kumulative Wahrscheinlichkeit von etwa 0.8 auf einen Wert knapp unterhalb von 1 ansteigt, und der Tabelle @6b können wir als genauere kumulative Wahrscheinlichkeiten die Werte $F(4,0) = 0.798103$ und $F(6,0) = 0.972676$ entnehmen. Die Wahrscheinlichkeit einer Antwortzeit, die größer als 4.0 und kleiner oder gleich 6.0 ist, beträgt also $0.972676 - 0.798103 = 0.174573$.

Ein kleiner Unterschied besteht, der aber eine Erleichterung darstellt, sozusagen eine

"Entwarnung": Bei Häufigkeitsverteilungen und diskreten Wahrscheinlichkeitsverteilungen mußte immer genau darauf geachtet werden, daß Fälle an der Untergrenze des Intervalls ausgeschlossen, solche an der Obergrenze dagegen einbezogen werden. (Es ging um die Häufigkeit von Werten, die größer als die Untergrenze x' und kleiner oder gleich der Obergrenze x'' des betrachteten Intervalls sind. Denn in die Höhendifferenz der Ogive an beiden Punkten gehen die Fälle an der Untergrenze nicht ein, wohl aber die an der Obergrenze.) Bei kontinuierlichen Wahrscheinlichkeitsverteilungen brauchen wir dagegen nicht so exakt anzugeben, ob die Untergrenze (4.0) und die Obergrenze (6.0) des Intervalls mitgemeint ist; denn da die Wahrscheinlichkeit jedes einzelnen Punkts null ist, ändert sich an den Wahrscheinlichkeiten eines Werts in einem Intervall nichts, wenn wir die Untergrenze zu dem Intervall hinzunehmen oder die Obergrenze herausnehmen.

Die Wahrscheinlichkeit, daß eine Zufallsgröße X in ein bestimmtes Intervall fällt, wird auch als die entsprechende "Intervallwahrscheinlichkeit" bezeichnet. Mit den Symbolen x_{IA} und x_{IE} für Intervallanfang (Untergrenze) und Intervallende (Obergrenze) können wir als allgemeines Ergebnis festhalten:

Differenzenregel für Intervallwahrscheinlichkeiten: Bei einer Zufallsvariablen X mit einer kontinuierlichen Wahrscheinlichkeitsverteilung ist die Wahrscheinlichkeit, daß X in das Intervall mit Untergrenze x_{IA} und Obergrenze x_{IE} fällt, gleich der Differenz $F(x_{IE}) - F(x_{IA})$ der kumulativen Wahrscheinlichkeiten.

<<

Für einen Beweis dieser Differenzenregel können wir uns auf die Wahrscheinlichkeit $p(x_{IA} < X \leq x_{IE})$ beschränken. Dazu stellen wir zunächst fest, daß sich das Ereignis $X \leq x_{IE}$ auch als $(X \leq x_{IA} \text{ oder } x_{IA} < X \leq x_{IE})$ formulieren läßt. Da die beiden Teilereignisse sich gegenseitig ausschließen, gilt nach dem einfachen Additionssatz:

$$\begin{aligned} p(X \leq x_{IE}) &= p(X \leq x_{IA} \vee x_{IA} < X \leq x_{IE}) \\ &= p(X \leq x_{IA}) + p(x_{IA} < X \leq x_{IE}). \end{aligned}$$

Daraus ergibt sich unmittelbar:

$$p(x_{IA} < X \leq x_{IE}) = p(X \leq x_{IE}) - p(X \leq x_{IA}) = F(x_{IE}) - F(x_{IA}).$$

>>

b) Die Dichtekurve

In Abbildungen im Skript Statistik I war manchmal anstelle eines Histogramms (das ja eine Linie mit Zacken ist) eine glatte Linie gezeichnet.³³ Das wurde dort so erklärt, daß die Ecken des

³³Solche Abbildungen befinden sich @in den Kapiteln zur Standardabweichung und zum Exzeß.

Histogramms einer Verteilung mit vielen Klassen geglättet sind, und bei Häufigkeitsverteilungen kann eine solche glatte Linie auch nur durch eine derartige "Überarbeitung" entstehen.

Bei kontinuierlichen Wahrscheinlichkeitsverteilungen wäre es wenig sinnvoll, ein Histogramm zu zeichnen. (Es wurde bereits gezeigt, daß die Wahrscheinlichkeit jedes einzelnen x -Werts null ist, und deshalb bestünde eine Kurve, die für jeden x -Wert seine Wahrscheinlichkeit angibt, aus einer Linie, die mit der x -Achse zusammenfällt. Da dies bei allen kontinuierlichen Wahrscheinlichkeitsverteilungen so ist, wäre diese Darstellung völlig uninformativ.) Stattdessen übernimmt bei kontinuierlichen Wahrscheinlichkeitsverteilungen eine andere Kurve eine ähnliche Rolle, wie wir sie bisher für das Histogramm kannten. Es handelt sich um die *Dichtekurve*. Was sie bedeutet, soll in diesem Abschnitt dargestellt werden.

Eine erste Annäherung an diese Bedeutung besteht darin, daß man aus der kontinuierlichen Skala der Antwortzeiten eine künstlich diskrete macht. Dabei wird aber etwas anders gerundet, als wir es aus Statistik I gewohnt sind. Man kann sich z.B. vorstellen, daß in dem Augenblick, in dem die Antwortzeit beginnt, eine Digitaluhr gestartet und im Augenblick der Antwort angehalten und abgelesen wird. Wenn diese Uhr die Zeit z.B. in Sekunden angibt, dann springt sie nach genau 2.00.. sec auf 2, und nach 3.00.. sec auf 3. Eine von dieser Uhr abgelesene Zeit von 2 sec entspricht also einer kontinuierlichen Antwortzeit im Intervall von 2 sec (einschließlich) bis 3 sec (ausschließlich). In der folgenden Abb. @2b sind entsprechende Histogramme für Digitaluhren unterschiedlicher Genauigkeit dargestellt. Aus Gründen, die bald erläutert werden, sind auf der Ordinaten-Achse ("y-Achse") aber nicht die Wahrscheinlichkeiten der verschiedenen Intervalle angegeben; vielmehr sind diese durch die Intervallbreite dividiert. (Beispiel: In der oberen linken Kurve wird mit einer Intervallbreite von 2 Sekunden gearbeitet. Für das Intervall von 4 bis 6 Sekunden haben wir in Abschnitt @B.IV.1.a eine Intervallwahrscheinlichkeit von 0.174573 festgestellt. Dividieren wir diese Wahrscheinlichkeit durch die Intervallbreite 2, dann erhalten wir den auf der Ordinaten-Achse für dieses Intervall angegebenen Wert. Übungsaufgabe: Berechnen Sie weitere Intervallwahrscheinlichkeiten nach der Differenzenregel für Intervallwahrscheinlichkeiten und vergleichen Sie das Ergebnis mit dem entsprechenden Histogramm.)

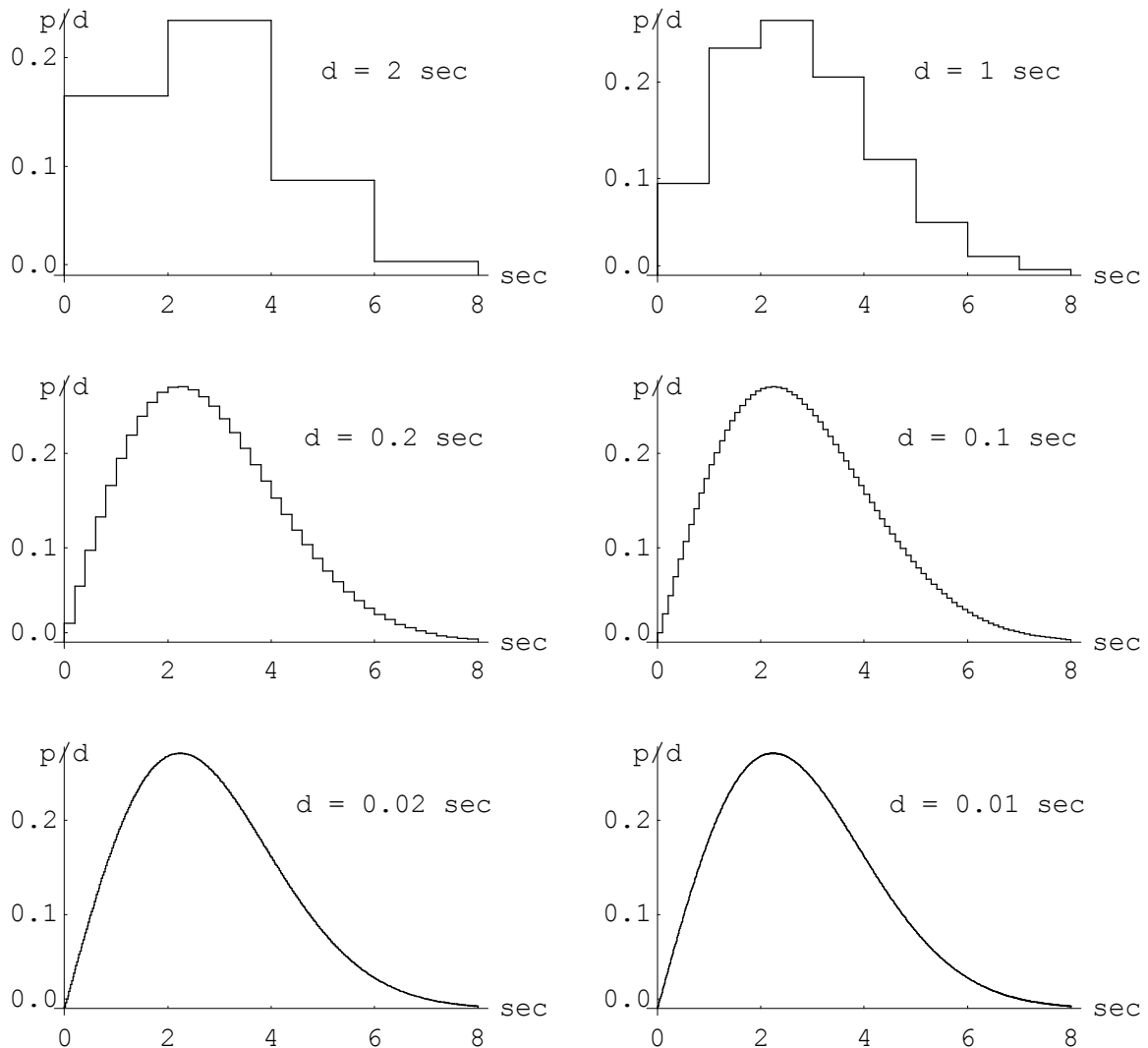


Abb. 2b: Annäherung an die Dichtekurve durch Histogramme für unterschiedliche Intervallbreiten. Dargestellt ist der Quotient aus Intervallwahrscheinlichkeit p und Intervallbreite d .

Was man an den verschiedenen Histogrammen sieht: Je kleiner die Intervalle werden, um so mehr nähern sich die Kurven einer glatten Linie. Diese glatte Kurve, die den Grenzfall für "infinitesimal kleine Intervallbreite d " darstellt, ist die Dichtekurve.

Um dieses Ergebnis zu erzielen, ist auf der Ordinatenachse der Quotient aus Intervallwahrscheinlichkeit und Intervallbreite angegeben. Hätten wir nämlich einfach die Intervallwahrscheinlichkeiten eingetragen, dann würde zwar die Form der Kurve ebenfalls gegen die glatte Kurve unten rechts streben, aber die Zahlen an der Ordinaten-Achse würden immer kleiner werden. Mit den Angaben auf der Ordinatenachse in Abb. @2b kann man dagegen sagen: Für jede Antwortzeit x (also jeden Wert auf der Abszissenachse ("x-Achse")) nähert sich der

entsprechende Wert auf der Ordinatenachse einer bestimmten Zahl, und diese Zahl ist die "Wahrscheinlichkeitsdichte im Punkt x ".

Die Bezeichnung "Dichte" kann man sich als eine Analogie zu dem Begriff der Kohlendioxyd-Dichte in folgendem Satz erklären: "Die Kohlendioxyd-Dichte nimmt mit zunehmendem Abstand von einer Emissionsquelle ab." Man kann auch an die Dichte als "Masse pro Volumen" denken. Manche Mathematiker sprechen sogar direkt von Wahrscheinlichkeitsmasse und würden z.B. sagen, daß sich in unserer Beispiel-Verteilung in dem Intervall von 4 sec bis 6 sec eine Wahrscheinlichkeitsmasse von 0.174573 befindet.

Das übliche Symbol für die Wahrscheinlichkeitsdichte im Punkt x ist $f(x)$. Bei diskreten Wahrscheinlichkeitsverteilungen steht dieses Symbol ja für die Wahrscheinlichkeit, daß eine Zufallsgröße genau den Wert x annimmt. Diese Interpretation wäre aber bei kontinuierlichen Wahrscheinlichkeitsverteilungen völlig fehl am Platz. In Abschnitt @B.IV.1.a wurde ja bereits festgestellt, daß diese Wahrscheinlichkeit eines bestimmten Skalenpunkts bei kontinuierlichen Wahrscheinlichkeitsverteilungen immer 0 ist.

Die Dichtekurve für unsere Beispielverteilung - also die glatte Kurve, der sich die Histogrammdarstellungen in Abb. @2b mit abnehmender Intervallbreite nähern - ist in Abb. @2c dargestellt. Die schraffierte Linie unter der Kurve über dem Intervall von 4 bis 6 leitet über zu einer Eigenschaft von Histogrammen, die in Statistik I (in Abschnitt @) behandelt wurde: Die Fläche unter dem Histogramm in einem bestimmten Skalenabschnitt ist proportional zur Zahl der Meßwerte (z.B. V_{pn}) in diesem Skalenabschnitt, und der Anteil dieser Fläche an der Gesamtfläche unter dem Histogramm ist gleich der relativen Häufigkeit von Meßwerten in diesem Skalenabschnitt. Wählen wir die Einheit, in der die Fläche angegeben wird, so, daß die Gesamtfläche unter dem Histogramm 1 ist, dann können wir auch einfach sagen: Die Fläche unter dem Histogramm in einem bestimmten Skalenabschnitt ist gleich der relativen Häufigkeit von Meßwerten in diesem Skalenabschnitt.

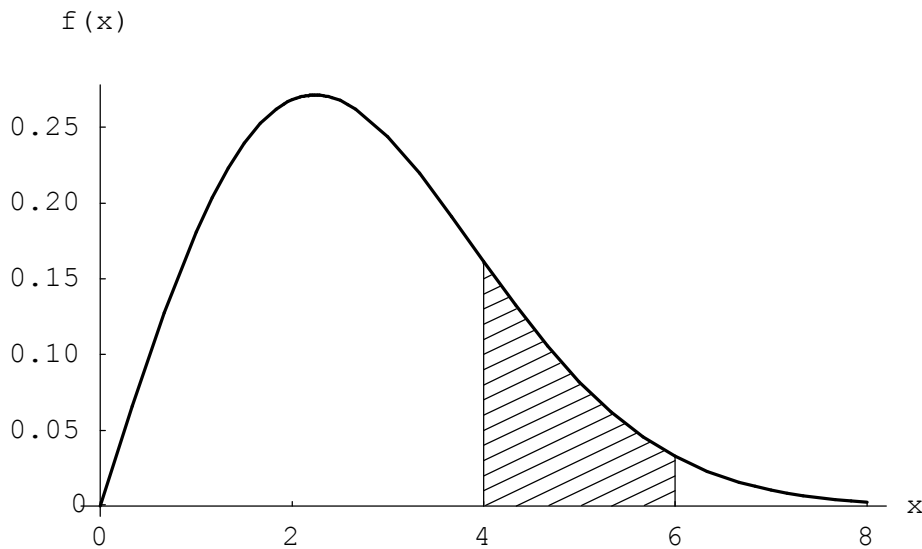


Abb. 2c: Dichtekurve für die Wahrscheinlichkeitsverteilung von Antwortzeiten. Die schraffierte Fläche entspricht der Wahrscheinlichkeit einer Antwortzeit im Intervall von 4 sec bis 6 sec.

Diese Eigenschaft von Histogrammen kann sinngemäß auf die Dichtekurve übertragen werden³⁴, wobei es kaum überrascht, daß "relative Häufigkeit" durch "Wahrscheinlichkeit" zu ersetzen ist. In Abb. @2c ist z.B. der Anteil der schraffierten Fläche an der Gesamtfläche unter der Kurve gleich der Differenz $F(6) - F(4) = 0.174573$, die wir als Wahrscheinlichkeit einer Antwortzeit X im Intervall von 4 sec bis 6 sec kennengelernt haben. Allgemein gilt die folgende

Flächenregel für Dichtekurven: Mit "Flächen unter der Dichtekurve" sind immer Flächen zwischen der Kurve und der Abszissenachse ("x-Achse") gemeint. Die Größe von Teilen der Fläche unter einer Dichtekurve wird als Anteil der Teilfläche an der Gesamtfläche unter dieser Kurve angegeben. Dann ist die Fläche unter der Dichtekurve in einem Skalenabschnitt gleich der Wahrscheinlichkeit eines Werts der betrachteten Zufallsvariablen in diesem Skalenabschnitt.

Diese Flächenregel ist das Wichtigste, was man über Dichtekurven wissen muß; denn beim späteren Umgang mit kontinuierlichen Wahrscheinlichkeitsverteilungen wird meistens nicht mit der Ogive gearbeitet, sondern mit der Dichtekurve, und dies vor allem in der Weise, daß Flächen unter der Kurve nach der obigen Regel als Wahrscheinlichkeiten interpretiert werden.

In unserer Beispielverteilung ist auch die kumulative Wahrscheinlichkeit leicht mit Hilfe der Flächenregel zu bestimmen. Das Ereignis "Antwortzeit bis 6 Sekunden" ist ja gleichbedeutend mit "Antwortzeit im Intervall von 0 bis 6 Sekunden". Daher ist die entsprechende kumulative Wahr-

³⁴Bisher ist das nur eine Analogie! Ein Beweis folgt später.

scheinlichkeit $F(6)$ gleich der Fläche unter der Dichtekurve von 0 bis 6 Sekunden. Man kann das auch folgendermaßen formulieren:

Spezialfall der Flächenregel: Die kumulative Wahrscheinlichkeit $F(x)$ eines Wertes x ist gleich der gesamten Fläche unter der Dichtekurve bis x .

In dieser Form gilt die Regel auch für kontinuierliche Wahrscheinlichkeitsverteilungen, die sich in einem Punkt von der Dichtekurve in Abb. @2c unterscheiden. Um diesen Unterschied zu beschreiben, ist zunächst auf eine bisher nicht weiter beachtete Eigenschaft der Dichtekurve aus Abb. @2c hinzuweisen. Im Skalenpunkt 8 hat die Dichtekurve schon fast die Abszissenachse ("x-Achse") erreicht, sie fällt aber nicht mit der Abszissenachse zusammen. Das bleibt auch so "bis in's Unendliche": Die Dichtekurve nähert sich der Achse immer mehr, erreicht sie aber bei keinem endlichen Skalenwert. Bekanntlich nennt man eine solche Linie, der eine Kurve beliebig nahe kommt, ohne sie je zu erreichen, eine Asymptote.

Am unteren Ende der Dichtekurve in Abb. @2c liegt dagegen eine andere Situation vor: Hier erreicht die Dichtekurve die Abszissenachse im Skalenpunkt 0, und unterhalb dieses Punktes liegt sie natürlich nirgendwo oberhalb dieser Achse. (Das würde ja nach der Flächenregel bedeuten, daß es eine von 0 verschiedene Wahrscheinlichkeit für negative Reaktionszeiten gäbe!) Daher kann man bei dieser Dichtekurve die "Fläche unter der Kurve bis zum Skalenpunkt x " gleichsetzen mit der Fläche vom Skalenpunkt 0 bis zum Skalenpunkt x .

Es gibt aber auch kontinuierliche Wahrscheinlichkeitsverteilungen, bei denen die Abszissenachse nicht nur "nach oben" (also in Richtung $+\infty$) Asymptote der Dichtekurve ist, sondern auch "nach unten" (in Richtung $-\infty$). Dazu gehören z.B. die Normalverteilungen, die in Abschnitt @B.IV.3 näher behandelt werden. Eine besonders wichtige Normalverteilung ist z.B. die sog. Standardnormalverteilung.³⁵ In Abb. @2d ist ihre Dichtekurve dargestellt, und die Fläche unter der Dichtekurve bis zu einem Skalenwert von 1 ist schraffiert. Das ist aber so zu verstehen, daß auch die gesamte Fläche zwischen Dichtekurve und Abszissenachse in Richtung $-\infty$ als schraffiert zu denken ist. Nur wenn man diesen "Schwanz" der Verteilung in die "Fläche unter der Dichtekurve bis zum Skalenpunkt x " einbezieht, gilt auch für eine solche Verteilung die oben als "Spezialfall der Flächenregel" formulierte Beziehung zur kumulativen Wahrscheinlichkeit.³⁶

³⁵Im Augenblick genügt es, zu wissen, daß die Standardnormalverteilung eine kontinuierliche Wahrscheinlichkeitsverteilung mit der in Abb. @2d dargestellten Dichtekurve ist. Genauer: Symmetrisch um den Nullpunkt, Wendepunkte bei -1 und $+1$.

³⁶Die Tatsache, daß die Abszissenachse Asymptote der Dichtekurve ist, bedeutet keineswegs, daß die Fläche zwischen beiden Linien unendlich ist; denn der Abstand wird ja auch immer kleiner. Es gilt die allgemeine Regel: Ob die Fläche zwischen einer asymptotischen Kurve und ihrer Asymptote endlich oder unendlich ist, hängt davon ab, wie schnell sich die Kurve der Asymptote nähert.

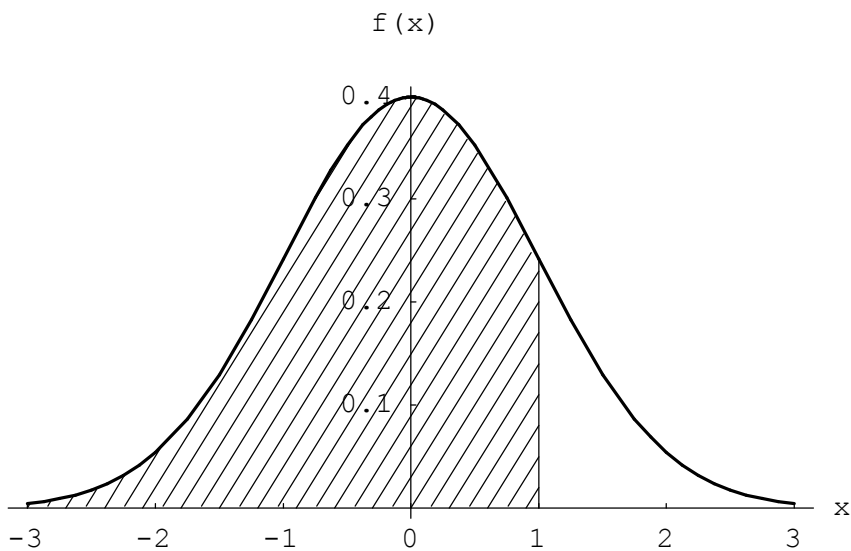


Abb. 2d: Die kumulative Wahrscheinlichkeit $F(1)$ in einer Standardnormalverteilung als Fläche unter der Dichtekurve bis zum Skalenpunkt 1.

<<

Aus der Flächenregel lassen sich noch weitere Feststellungen über kontinuierliche Wahrscheinlichkeitsverteilungen ableiten. Zunächst spiegelt sich darin eine Tatsache wider, die sich schon aus der Behandlung der Ogive in Abschnitt @B.IV.1.a ergab: Für jeden einzelnen x -Wert ist die Wahrscheinlichkeit, daß die Zufallsgröße X gerade diesen Wert annimmt, gleich 0. Aus der Flächenregel ergibt sich das unmittelbar, da die "Fläche" über einem Skalenpunkt lediglich aus einer Linie besteht, die natürlich den Flächeninhalt 0 hat.

Als weitere Konsequenz aus der Flächenregel ergibt sich eine recht präzise Definition des Begriffs einer kontinuierlichen Wahrscheinlichkeitsverteilung:

Definition: Eine Zufallsgröße hat eine kontinuierliche Wahrscheinlichkeitsverteilung, wenn es eine nach der obigen Regel interpretierbare Dichtekurve für diese Verteilung gibt.

>>

c) Näheres zur Bedeutung der Wahrscheinlichkeitsdichte

Im @vorigen Abschnitt wurde die Dichtekurve anschaulich als Grenzfall von Histogrammen hergeleitet. Das ist ein anschaulicher Zugang, der vor allem einen ziemlich direkten Weg zur wichtigsten Anwendung der Dichtekurve eröffnet: Zur Bestimmung von Intervallwahrscheinlichkeiten und kumulativen Wahrscheinlichkeiten über die Flächenregel. Was dabei etwas zu kurz kam, ist die genaue Bedeutung der Zahlen an der Ordinatenachse ("y-Achse"). Man kann z.B. fragen, was die Wahrscheinlichkeitsdichte von ca. 0.16 bedeutet, die man aus der Dichtekurve in Abb. @2c für den Skalenpunkt 4 ablesen kann. Solche Fragen zu beantworten, ist Aufgabe dieses

Abschnitts.

Einen Zugang zu einer Antwort können wir uns verschaffen, indem wir noch einmal zur Ogive zurückkehren und die am Ende von Abschnitt @B.IV.1.a aufgestellte Differenzenregel verwenden, um die Gestalt der Ogive in Abb. @2a näher zu interpretieren. Was bedeutet die Form eines schräggestellten S, also die Tatsache, daß die Ogive zunächst als linksdrehende Kurve beginnt und dann nach einem Wendepunkt³⁷ rechtsdrehend wird?

Wir haben bereits gesehen, daß der Höhenunterschied der Ogive (also die Zunahme der kumulativen Wahrscheinlichkeit) zwischen den Skalenpunkten 4 und 6 als die Wahrscheinlichkeit interpretiert werden kann, daß die Zufallsvariable X - die Antwortzeit unserer Vp - in dieses Intervall fällt. Wenn man in ähnlicher Weise auch die Intervalle zwischen den anderen an der x-Achse eingetragenen Zahlen betrachtet, dann stellt man einen Zusammenhang mit dem Anstieg der Kurve fest, der an ähnliche Eigenschaften früher behandelter Ogiven erinnert: Je steiler die Kurve in den verschiedenen 2-cm-Intervallen ist, um so größer ist die Wahrscheinlichkeitszunahme. Das mag zunächst trivial erscheinen, aber darin steckt der einfachste Zugang zum Konzept der Wahrscheinlichkeitsdichte.

Was die Drehrichtung der Ogive bedeutet, kann man leicht sehen, wenn man sich vorstellt, daß die Untergrenze des bisher betrachteten Intervalls von 4 bis 6 fest bei $x=4$ bleibt, während die Obergrenze kleiner wird. Natürlich wird dann der Höhenunterschied der Ogive zwischen Intervallanfang und Intervallende - also die Wahrscheinlichkeit einer Antwortzeit in dem betrachteten Intervall - immer kleiner, und dasselbe gilt selbstverständlich, wenn wir den Intervallanfang auf 0 legen. Aber die folgende Abb.@2e zeigt einen Unterschied: Zeichnet man eine gerade Verbindungsstrecke zwischen den Ogivenpunkten für die Endpunkte des betrachteten Intervalls, dann wird diese Verbindungslinie in dem oberen Bereich, wo die Ogive rechtsdrehend ist, mit kleiner werdender Intervall-Obergrenze immer steiler, während sie im linksdrehenden unteren Bereich immer flacher wird.

³⁷Wer etwas versiert in Differentialrechnung ist, kann aus der unmittelbar nach Tabelle @6b angegebenen Formel für die kumulative Wahrscheinlichkeit ableiten, daß der Wendepunkt bei $x = \sqrt{5}$ liegt, also etwa bei $x = 2.236$.

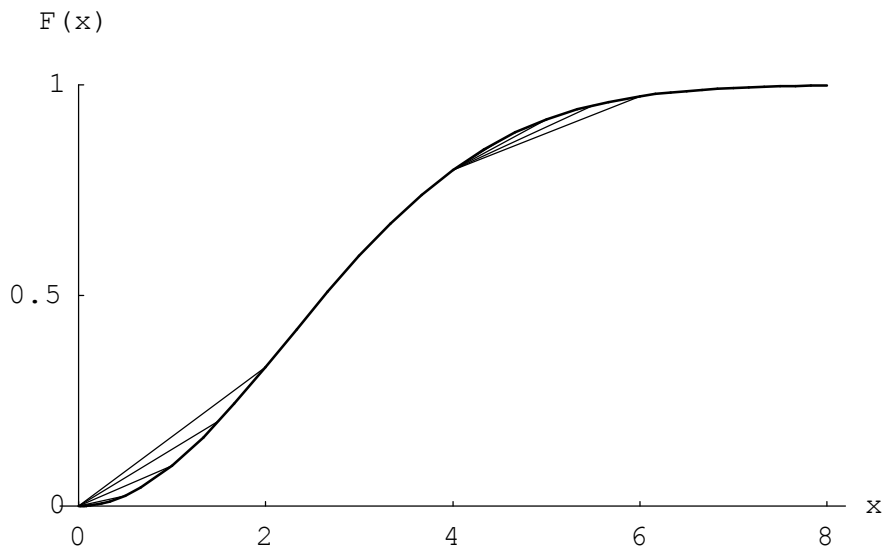


Abb. 2e: Unterschiedlicher Anstieg von Verbindungslinien zwischen Ogiven-Punkten

Zunächst ist das kaum mehr als eine Umschreibung dessen, was man unter einer rechtsdrehenden bzw. linksdrehenden Kurve versteht. Wenn man aber den Anstieg dieser Verbindungslinien für immer kleiner werdende Intervalle betrachtet, landet man schließlich beim Konzept der Wahrscheinlichkeitsdichte.

Es wurde bereits angedeutet, daß dieses Konzept einige Ähnlichkeit zum Begriff der Geschwindigkeit in einem Zeitpunkt hat. Auch diese läßt sich bekanntlich über den Anstieg immer kürzer werdender Verbindungslinien in einer vergleichbaren Kurve darstellen, und sich diesen bekannten Gedanken zu vergegenwärtigen, mag auch das Verständnis der Wahrscheinlichkeitsdichte erleichtern. Dazu ist in Abb. @2f dargestellt, welche Strecke $s(t)$, in Metern angegeben) ein Körper im "freien Fall" nach t Sekunden zurückgelegt hat.³⁸ Die Kurve wird mit zunehmender Zeit t immer steiler, da die Geschwindigkeit im freien Fall ja ständig zunimmt.

Der Anstieg der Verbindungslinien zwischen den Kurvenpunkten für $t=1$ und verschiedenen Punkten für höhere Zeiten entspricht der mittleren Geschwindigkeit in dem entspre-

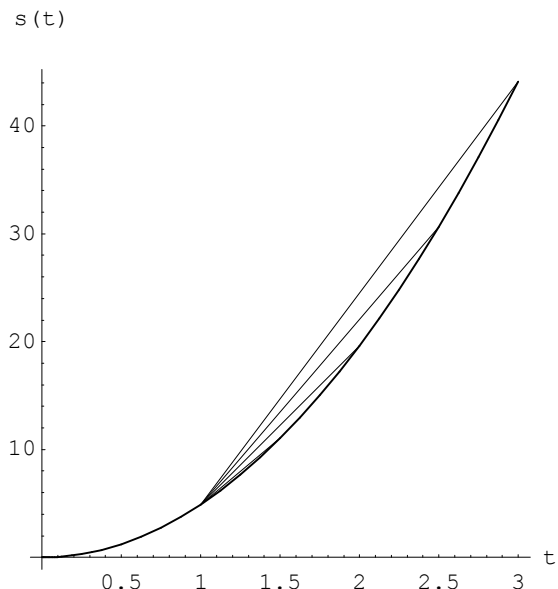


Abb. 2f: Zurückgelegte Strecke $s(t)$ (in Metern) nach t Sekunden im freien Fall.

³⁸Die Gleichung für eine solche Kurve lautet bekanntlich $s(t) = \frac{1}{2} g \cdot t^2$, wobei g die "Erdbeschleunigung" von 9.80665 m/sec^2 ist.

chenden Zeitintervall. In dem Zeitintervall von 1 sec bis 3 sec nach Beginn des Falls werden z.B. ca. 39.2 m zurückgelegt - das ist der Höhenunterschied der beiden Kurvenpunkte - und die mittlere Geschwindigkeit von 19.6 m/sec in diesem Zeitintervall ist genau der Anstieg der geraden Verbindungslinie zwischen den entsprechenden Kurvenpunkten. Läßt man nun die Obergrenze des Intervalls immer näher an den Wert 1 sec heranrücken, dann wird diese Verbindungslinie immer flacher. Die Geschwindigkeit in dem Zeitpunkt $t=1$ zu Beginn dieser Zeitintervalle ist der Anstieg der Kurve in diesem Punkt. Man erhält ihn bekanntlich, indem man den Wert bestimmt, gegen den die mittlere Geschwindigkeit (also der Quotient aus der in dem Zeitintervall zurückgelegten Strecke und der Breite des Zeitintervalls) strebt, wenn die Breite des Zeitintervalls gegen 0 strebt. Diese Geschwindigkeit nach 1 sec beträgt ca. 9.8 m/sec.

Zwischen dieser Kurve und der Ogive in Abb. 2a und 2e besteht ein Zusammenhang. Der Zunahme der zurückgelegten Strecke beim freien Fall entspricht die Zunahme der kumulativen Wahrscheinlichkeit bei der Ogive. In beiden Fällen handelt es sich um den Höhenunterschied der Kurve am Anfang und am Ende eines betrachteten Intervalls. Nun ergibt sich ja aus der Differenzenregel für Intervallwahrscheinlichkeiten aus Abschnitt @B.IV.1.a, daß bei der Ogive diese Zunahme (also die Differenz $F(x_{IE}) - F(x_{IA})$) die Wahrscheinlichkeit ist, daß die Zufallsgröße X in das Intervall von x_{IA} bis x_{IE} fällt. Was dabei herauskommt, wenn man den Quotienten aus dieser Intervallwahrscheinlichkeit und der Intervallbreite für immer kleiner werdende Intervalle bestimmt, wird in der folgenden Tabelle @6c demonstriert.

x_{IA}	x_{IE}	d	p	p/d
4	6.0000	2.0000	0.1745727955	0.0873
4	5.0000	1.0000	0.1198115194	0.1198
4	4.2000	0.2000	0.0305384591	0.1527
4	4.1000	0.1000	0.0157088228	0.1571
4	4.0200	0.0200	0.0032125863	0.1606
4	4.0100	0.0100	0.0016107315	0.1611
4	4.0020	0.0020	0.0003228568	0.1614
4	4.0010	0.0010	0.0001614728	0.1615
4	4.0002	0.0002	0.0000323019	0.1615
4	4.0001	0.0001	0.0000161481	0.1615

Tabelle @6c: Der Quotient aus Intervallwahrscheinlichkeit p und Intervallbreite d strebt gegen die Dichte.

Die Intervallwahrscheinlichkeit für das Intervall von 4 sec bis 6 sec Antwortzeit wurde auf S. @95 als $F(6) - F(4) = 0.174573$ angegeben. Mehr Nachkommastellen gibt die dort zugrundegelegte Tabelle @6b nicht her. Für die hiesigen Berechnungen ist aber eine höhere Genauigkeit erforderlich, damit auch in den letzten Zeilen der Quotient p/d (in der letzten Spalte) noch sinnvoll interpretierbar ist. Deshalb wurden die Intervallwahrscheinlichkeiten in der

vorletzten Spalte auf einem PC mit 10 Stellen hinter dem Komma berechnet, und zwar - genau wie auf S. @95 demonstriert - als Differenz der kumulativen Wahrscheinlichkeiten am Intervallende und am Intervallanfang. (Übungsaufgabe: Welcher Wert für die Intervallwahrscheinlichkeit des Intervalls von 4 bis 4.1 würde sich aus Tabelle @6b ergeben?)

Das wesentliche Ergebnis dieser Berechnungen ist die Tatsache, daß der in der letzten Spalte angegebene Quotient aus Intervallwahrscheinlichkeit und Intervallbreite bei einem Wert von 0.1615 "stehenbleibt". (Wenn man die Werte mit noch mehr Nachkommastellen berechnen würde, könnte man zwar feststellen, daß der Quotient noch geringfügig weiter zunimmt, aber erst hinter der vierten Stelle nach dem Komma.³⁹)

Diese Zahl 0.1615 ist die "Wahrscheinlichkeitsdichte" im Skalenpunkt 4 (oder genauer 4 sec). Sie ist noch etwas genauer als die Zahl 0.16, die wir zu Beginn @dieses Abschnitts aus der Dichtekurve abgelesen haben.

Bevor dieser Begriff der Wahrscheinlichkeitsdichte definiert wird, soll die Überlegung noch einmal in Bezug zur Ogive (@Abb. 2e) gesetzt werden. Es wurde bereits darauf hingewiesen, daß der Höhenunterschied der vom Kurvenpunkt für $x = 4$ ausgehenden geraden Linien die Wahrscheinlichkeit einer Antwortzeit in dem entsprechenden Intervall der x -Skala ist. Das sind also die in der vorletzten Spalte von @Tabelle 6c als p angegebenen Werte.⁴⁰ Die in der letzten Spalte der Tabelle angegebenen Quotienten p / d entsprechen dann dem Anstieg der geraden Verbindungslinien in der Abbildung. Läßt man nun - wie in der Tabelle - die Intervallbreite d immer kleiner werden, dann wird aus dem Anstieg der geraden Verbindungslinien der Anstieg der Ogive im entsprechenden Skalenpunkt.

Aus dieser Betrachtungsweise und aus der exemplarischen Berechnung in Tabelle @6c ergeben sich als Verallgemeinerung zwei Definitionen, die aber mathematisch auf dasselbe hinauslaufen:

Zwei Definitionen der Wahrscheinlichkeitsdichte:

- *Die Wahrscheinlichkeitsdichte im Skalenpunkt x ist der Anstieg der Ogive in diesem Punkt.*
- *Die Wahrscheinlichkeitsdichte im Skalenpunkt x ist der Grenzwert, gegen den der Quotient aus Intervallwahrscheinlichkeit und Intervallbreite strebt, wenn man Intervalle betrachtet, die im Punkt x beginnen und deren Intervallbreite gegen null geht.*

Etwas anschaulicher könnte man auch sagen: Die Wahrscheinlichkeitsdichte im Skalenpunkt x gibt an, wie dicht die Wahrscheinlichkeit in der unmittelbaren Umgebung von x gepackt ist.

³⁹Eine genauere Angabe für den Grenzwert ist 0.1615172144.

⁴⁰Allerdings sind in @Abbildung 2e größere Intervallbreiten d verwendet als in Tabelle 6c, da man bei den minimalen Intervallbreiten aus der Tabelle die Verbindungslinien noch schwerer erkennen könnte, als es ohnehin schon in der Abbildung der Fall ist.

Ähnliche Berechnungen wie in Tabelle @6c kann man natürlich auch für andere Punkte der Skala durchführen und auch bei Zufallsgrößen mit völlig anderen kontinuierlichen Wahrscheinlichkeitsverteilungen. Der Gedankengang, der allen derartigen Berechnungen zugrundeliegt, läßt sich folgendermaßen zusammenfassen: Um die Wahrscheinlichkeitsdichte im Skalenpunkt x zu berechnen, bestimmt man für immer kleiner werdende Intervallbreiten d die Intervallwahrscheinlichkeit des Intervalls von x bis $x+d$. Diese Intervallwahrscheinlichkeit dividiert man durch die Intervallbreite d und erhält einen Wert, den man als die durchschnittliche Wahrscheinlichkeitsdichte in dem jeweils betrachteten Intervall ansehen kann (ähnlich wie beim freien Fall, wo der Quotient aus zurückgelegter Strecke und Dauer eines Zeitintervalls die durchschnittliche Geschwindigkeit in dem Zeitintervall ist). Läßt man nun die Intervallbreite immer kleiner werden, so daß das Intervall "zu einem Skalenpunkt zusammenschrumpft", dann strebt der Quotient aus Intervallwahrscheinlichkeit und Intervallbreite gegen eine Zahl, welche die Wahrscheinlichkeitsdichte im Skalenpunkt x ist. Mit dem bereits im Zusammenhang der Dichtekurve eingeführten Symbol $f(x)$ für diese Wahrscheinlichkeitsdichte läßt sich der gesamte Gedankengang in der Definitions-Gleichung

$$f(x) := \lim_{d \downarrow 0} \frac{\text{Intervallwahrscheinlichkeit des Intervalls von } x \text{ bis } x+d}{d}$$

zusammenfassen.⁴¹

Die Schreibweise $f(x)$ deutet an, daß die Wahrscheinlichkeitsdichte als Funktion des Skalenwerts x betrachtet werden kann: Zu jedem⁴² Skalenpunkt x gibt es eine diesem Skalenpunkt zugeordnete Wahrscheinlichkeitsdichte $f(x)$. Daher spricht man auch von der *Dichtefunktion* $f(x)$, die durch die @obige Gleichung definiert wird.

<<

Diese Gleichung ermöglicht es, die Beziehung zwischen kumulativer Verteilungsfunktion und Dichtefunktion mit Hilfe der Differential- und Integralrechnung darzustellen. Nach der Differenzenregel für Intervallwahrscheinlichkeiten kann man für die auf der rechten Seite der Gleichung im Zähler stehende Wahrscheinlichkeit ja auch $F(x+d) - F(x)$ schreiben, und damit ergibt sich die folgende Fassung der Definitionsgleichung:

$$f(x) := \lim_{d \downarrow 0} \frac{F(x+d) - F(x)}{d}$$

⁴¹Die Angabe $d \downarrow 0$ (mit einem nach unten zeigenden Pfeil) unter dem Symbol \lim bedeutet, daß d "von oben" gegen 0 strebt. Für $d \leq 0$ macht der dahinterstehende Bruch ja keinen Sinn.

⁴²Es wurde bereits @an früherer Stelle erwähnt, daß die in der Mathematik übliche Definition einer kontinuierlichen Wahrscheinlichkeitsverteilung auch Situationen zuläßt, in denen die kumulative Verteilungsfunktion $F(x)$ an einigen Stellen nicht differenzierbar ist. Wegen des Bezugs der Dichte zur ersten Ableitung von $F(x)$ (der gleich gezeigt wird) bedeutet dies dann auch, daß es nicht zu jedem Skalenpunkt x eine Wahrscheinlichkeitsdichte gibt. Mathematisch exakt ausgedrückt: Die Dichtefunktion f ist eine Abbildung derjenigen Skalenpunkte, an denen $F(x)$ differenzierbar ist, in die Menge der reellen Zahlen.

Das bedeutet aber, daß die Dichtefunktion auch die erste Ableitung der kumulativen Verteilungsfunktion $F(x)$ ist. Für unsere Beispielverteilung ergibt sich daher die Dichtefunktion⁴³

$$f(x) = \frac{0.2 \cdot x}{e^{0.1 \cdot x^2}}$$

Allgemein ergibt sich aus der Eigenschaft der Dichtefunktion als erste Ableitung von $F(x)$ zunächst, daß die beiden Definitionen der Dichte von S. @106 auf das gleiche Ergebnis hinauslaufen; denn der Anstieg der Ogive ist ja die erste Ableitung der kumulativen Verteilungsfunktion.

Außerdem kann man die Flächenregel für Intervallwahrscheinlichkeiten auch in der Sprache der Integralrechnung schreiben; denn die Fläche unter einer Kurve zwischen zwei Punkten a und b ergibt sich ja aus dem "Integral von a bis b " der entsprechenden Funktion. Differenzenregel und Flächenregel kann man in der folgenden Gleichung zusammenfassen, die für beliebige Intervallgrenzen a und b mit der Eigenschaft $a \leq b$ gilt⁴⁴:

$$p(a < X \leq b) = F(b) - F(a) = \int_a^b f(x) dx$$

Damit ist übrigens die Flächenregel für Intervallwahrscheinlichkeiten (die bisher nur als Analogie zu Histogrammen eingeführt wurde) sehr einfach bewiesen: Das erste Gleichheitszeichen beruht ja auf der Differenzenregel, die sich allein aus dem einfachen Additionssatz ergibt (wie in Abschnitt @B.IV.1.a gezeigt wurde). Das zweite Gleichheitszeichen muß nach dem "Hauptsatz der Differential- und Integralrechnung" immer dann gelten, wenn die Funktion $f(x)$ erste Ableitung der Funktion $F(x)$ ist.

Entsprechend gilt auch für die kumulative Wahrscheinlichkeit als Spezialfall der Flächenregel⁴⁵:

$$p(X \leq b) = F(b) = \int_{-\infty}^b f(x) dx$$

Diese Gleichung ist eng verwandt mit der folgenden bei diskreten Verteilungen gültigen Gleichung (bei der die Angabe $x \leq b$ unter dem Summenzeichen bedeutet, daß über alle

⁴³Zur Bildung der ersten Ableitung der auf S. @ angegebenen kumulativen Verteilungsfunktion $F(x)$ kann man diese auch als "Schachtelfunktion" schreiben: Mit der Definition $u(x) := -0.1 \cdot x^2$ kann man auch schreiben $F(x) = 1 - e^{u(x)}$, und dann ergibt sich die Ableitung nach der bekannten Ableitungsregel für Schachtelfunktionen $df(g(x))/dx = df/dg \cdot dg/dx$. (Die erste Ableitung von e^x ist ja e^x .)

⁴⁴Mit der Wahl der Bezeichnungen a und b für die Intervallgrenzen befolgen wir eine Regel aus der Integralrechnung, nach der es "verboten" ist, für die Grenzen des Integrationsintervalls dasselbe Symbol zu verwenden wie für die Variable, über die integriert wird.

⁴⁵Das erste Gleichheitszeichen beruht auf der Definition der Funktion $F(x)$, und das zweite ergibt sich aus einer Variante des Hauptsatzes der Differential- und Integralrechnung.

möglichen Werte x einer Zufallsvariablen zu summieren ist, die kleiner oder gleich b sind):

$$F(b) = \sum_{x \leq b} f(x)$$

(Ein Integral ist ja im Grunde nichts anderes als eine Summe von unendlich vielen infinitesimal kleinen Größen.)

Aus der Gleichung für kumulative Wahrscheinlichkeiten ergibt sich schließlich das zweite Gleichheitszeichen in der folgenden Gleichung⁴⁶:

$$\int_{-\infty}^{+\infty} f(x) dx = \lim_{b \rightarrow +\infty} \int_{-\infty}^b f(x) dx = \lim_{b \rightarrow +\infty} F(b) = 1$$

Das erste Integral in dieser Gleichung ist aber die gesamte Fläche unter der Dichtekurve. Das heißt: Wenn wir Flächen unter der Dichtekurve mit Integralen angeben, dann befolgen wir die Forderung der Flächenregel für Dichtekurven, daß alle Flächen als Anteile der Gesamtfläche unter der Dichtekurve anzugeben sind. Der mathematische Zusammenhang ist natürlich umgekehrt: Zu fordern ist, daß Flächen in der Weise zu bestimmen sind, wie es Integrale tun. (Das folgt unmittelbar aus den obigen Gleichungen.) Die @letzte Gleichung zeigt dann, daß man diese Forderung auch ohne Bezugnahme auf Integrale in der Weise formulieren kann, wie es die Flächenregel tut: Alle Flächen sind als Anteile der Gesamtfläche unter der Dichtekurve anzugeben.

>>

Ein Hinweis mag noch dazu beitragen, Mißverständnisse zu vermeiden. Das Feststellen von Ähnlichkeiten zwischen den Begriffen der Wahrscheinlichkeitsdichte in einem Skalenpunkt und der Geschwindigkeit in einem Zeitpunkt wurde dadurch etwas erleichtert, daß es bei der in unserer Beispielverteilung zugrundegelegten Zufallsgröße X ebenfalls um Zeiten ging - die Antwortzeiten. Das darf aber nicht zu der Annahme verleiten, daß das Konzept der Wahrscheinlichkeitsdichte ähnlich eng an die Zeit gebunden ist, wie es bei der Geschwindigkeit der Fall ist. Auch andere kontinuierliche Größen können die Rolle übernehmen, die in der Beispielverteilung der Antwortzeit zukam. Es trifft zwar zu, daß in der Psychologie Antwortzeiten die am intensivsten untersuchten kontinuierlichen Zufallsgrößen sind. Aber von den mathematischen Möglichkeiten her ist es genau so gut denkbar, daß man beispielsweise für die Gewichtsabnahme in einer Übergewichtstherapie oder für physiologische Meßwerte wie den Hautwiderstand Ogiven und Dichtekurven betrachtet. Dann wird die Dichte nicht als "Wahrscheinlichkeit pro Sekunde", sondern als "Wahrscheinlichkeit pro kg" bzw. "Wahrscheinlichkeit pro Ohm" angegeben.

⁴⁶Das erste Gleichheitszeichen ergibt sich aus der Definition eines Integrals mit $+\infty$ als oberer Integrationsgrenze.

d) Schlußbemerkung zur Vorgehensweise

<<

In den vorangehenden Abschnitten haben wir wesentliche Eigenschaften kontinuierlicher Wahrscheinlichkeitsverteilungen in der Weise dargestellt, daß sie sich aus der Ogive und der von ihr dargestellten kumulativen Verteilungsfunktion $F(x)$ ergeben. Das entspricht auch der heute als "Stand der Wissenschaft" anzusehenden Herangehensweise der mathematischen Wahrscheinlichkeitstheorie: Die kumulativen Wahrscheinlichkeiten und die Wahrscheinlichkeiten, die sich nach den Axiomen der Wahrscheinlichkeitstheorie aus den kumulativen Wahrscheinlichkeiten ableiten lassen, bilden den Ausgangspunkt aller Überlegungen, und daraus werden Konzepte wie das der Wahrscheinlichkeitsdichte abgeleitet.

Vor der Entwicklung der heutigen, auf den bekannten Axiomen aufbauenden Wahrscheinlichkeitstheorie war eine andere Herangehensweise eher verbreitet. Man begann mit der Definition einer Dichtefunktion, die etwa denselben Inhalt hatte, wie unsere entsprechende Definitionsgleichung; der logische Stellenwert war aber ein völlig anderer: Während in unserer Darstellung die Definition der Wahrscheinlichkeitsdichte eine Regel ist, die angibt, wie man von den (auf der kumulativen Verteilungsfunktion beruhenden) Intervallwahrscheinlichkeiten zur Wahrscheinlichkeitsdichte kommt, ging man früher eher davon aus, daß der Grenzwert des Quotienten aus Intervallwahrscheinlichkeit und Intervallbreite das erste ist, was man über eine kontinuierliche Wahrscheinlichkeitsverteilung weiß, so daß Intervallwahrscheinlichkeiten und kumulative Wahrscheinlichkeiten sich aus der Dichtefunktion ergeben.

Das hatte systematische und historische Gründe. Unter der systematischen Perspektive ist vor allem daran zu erinnern, daß es ja bei den diskreten Wahrscheinlichkeitsverteilungen viel natürlicher ist, von den Wahrscheinlichkeiten einzelner Skalenwerte durch Summenbildung zu den kumulativen Wahrscheinlichkeiten zu kommen. Eine derartige Vorgehensweise auch für kontinuierliche Wahrscheinlichkeitsverteilungen anzustreben hieß dann, mit der Dichtefunktion anzufangen und daraus die kumulative Verteilungsfunktion durch Integration herzuleiten.

Historisch ist darauf hinzuweisen, daß eine der ersten kontinuierlichen Wahrscheinlichkeitsverteilungen, die gründlich untersucht wurden, die Normalverteilung war, die wir noch genauer behandeln werden. Bei dieser Verteilung kennt man tatsächlich zunächst nur die Dichtefunktion und muß daraus die kumulative Verteilungsfunktion durch Integration herleiten. Nach heutiger Sicht, die auch der obigen Darstellung zugrundeliegt, ist aber zwischen zwei Aspekten zu unterscheiden:

- Bei der *allgemeinen Begriffsbildung* (die ja das Ziel der vorangehenden Abschnitte war) ist es günstiger, von den kumulativen Wahrscheinlichkeiten auszugehen und daraus die Dichtefunktion abzuleiten.
- Das schließt nicht aus, daß man bei der *Untersuchung einzelner Wahrscheinlichkeitsverteilungen* zunächst nur die Dichtefunktion kennt und von da aus erschließen muß, zu welcher kumulativen Verteilungsfunktion diese Dichtefunktion gehört.

>>

2) *Definition von Kennziffern kontinuierlicher Wahrscheinlichkeitsverteilungen aufgrund eines verallgemeinerten Erwartungswert-Begriffs*

a) Annäherung an das Problem

Die von uns behandelten Kennziffern von Häufigkeitsverteilungen und diskreten Wahrscheinlichkeitsverteilungen lassen sich in zwei Gruppen einteilen: Sie beruhen entweder unmittelbar auf der kumulativen Verteilung (Prozentränge, Centile und die daraus abgeleiteten Kennziffern wie z.B. QD) oder aber auf dem arithmetischen Mittel bzw. Erwartungswert von X oder einer von X abgeleiteten Größe $g(X)$. Für die erste Gruppe können die bisherigen Definitionen unmittelbar übernommen werden. (Beispiel: Median := derjenige x -Wert, für den $F(x) = 0.50$). Zu ergänzen ist noch, daß nach der Flächenregel der Median auch derjenige x -Wert ist, bei dem die Gesamtfläche unter der Dichtekurve durch eine senkrechte Linie halbiert wird.

Die auf einem Erwartungswert beruhenden Kennziffern werden im folgenden als "Erwartungswert-basierte Kennziffern" bezeichnet. Auch für diese Kennziffern kann das Neue wieder durch Anknüpfen an Bekanntes vorbereitet werden. In @Statistik I wurde auf eine Beziehung des arithmetischen Mittels einer Häufigkeitsverteilung zum Histogramm hingewiesen: Wenn man sich vorstellt, ein solches Histogramm aus gleichmäßiger Pappe auszuschneiden, dann ist das arithmetische Mittel derjenige Punkt der x -Achse, der unter dem Schwerpunkt dieser Figur liegt. Ebenso wie die Flächenregel kann auch diese Regel sinngemäß vom Histogramm auf die Dichtekurve übertragen werden:

Schwerpunktregel: Der Erwartungswert einer Zufallsgröße mit kontinuierlicher Wahrscheinlichkeitsverteilung ist der Abszissenwert des Schwerpunkts der Fläche unter der Dichtekurve.

Wollen Sie schätzen, über welchem Abszissen-Wert der Schwerpunkt der Fläche unter dem in Abb. @2c dargestellten Teil der Dichtekurve liegt?

Die Frage bezieht sich nur auf die Fläche unter dem in Abb. @2c dargestellten Teil der Dichtekurve, da die Dichtekurve ja rechts weitergeht und sich nur asymptotisch der Abszissenachse nähert. Zunächst einmal kann der Abszissenwert des Schwerpunkts der abgebildeten Teil-Fläche mathematisch bestimmt werden: Er liegt bei $x \approx 2.79$.

<<

Angesichts der Tatsache, daß man für die gesamte Fläche unter der Dichtekurve rechts noch einen unendlich langen Schwanz der Verteilung anhängen muß, ergibt sich die Frage, ob es dann überhaupt noch einen "Schwerpunkt" gibt. Müßte nicht bei jedem "Aufhängungspunkt" damit gerechnet werden, daß die zwar geringe, aber bis in's Unendliche reichende Masse des Verteilungsschwanzes nach Hebelgesetzen stärker nach rechts dreht, als die nur endlich weit entfernte Masse links vom Aufhängungspunkt nach links dreht? Diese Frage ist durchaus berechtigt. Tatsächlich gibt es Verteilungen, bei denen das eintreten würde - darauf wird später zurückzukommen sein. Im Augenblick mag es genügen, festzuhalten, daß eine solche Situation nicht entsteht, wenn sich nach dem im folgenden dargestellten Verfahren ein endlicher Erwartungswert ergibt, und das ist bei unserer Beispielverteilung der Fall.

>>

Ein Ergebnis der folgenden Überlegungen soll vorweggenommen werden: Wenn man den in

Abb. @2c unberücksichtigten Teil der Fläche unter der Dichtekurve hinzunimmt, dann hat die Gesamtfläche ihren Schwerpunkt bei einem Abszissenwert von ca.⁴⁷ $x = 2.8025$. Das ist also der Erwartungswert der Antwortzeit, wenn deren Wahrscheinlichkeitsverteilung die bisher untersuchte Beispielveilung ist.

In Abschnitt @B.IV.1.a wurde festgestellt, daß der Median der Beispielveilung zwischen 2.6 und 2.7 liegen muß; also ist der Erwartungswert etwas größer als der Median. Können Sie diese Tatsache in Zusammenhang mit der Form der Dichtekurve bringen?

In Statistik I wurde festgestellt, daß der Median bei positiver Schiefe kleiner als der Durchschnitt ist, und das gilt auch bei Wahrscheinlichkeitsverteilungen. (Links vom Median liegt die Hälfte der Fläche unter der Dichtekurve, und die Fläche rechts vom Median würde an einem längeren Hebel ziehen, wenn wir die ausgeschnittenen Fläche unter der Dichtekurve beim Median aufhängen würden. Also muß der Aufhängungspunkt rechts vom Median liegen, damit die Figur in's Gleichgewicht kommt..)

Dieses Schwerpunktmodell veranschaulicht, was der Erwartungswert einer Zufallsgröße mit kontinuierlicher Wahrscheinlichkeitsverteilung ist. Diesen Erwartungswert nach der bisherigen Definition als Zahl zu bestimmen, würde bedeuten, alle möglichen Werte x der Zufallsvariablen X mit der entsprechenden Wahrscheinlichkeit zu multiplizieren und die Summe dieser Produkte zu bilden. Das stößt jedoch auf zwei Probleme: Einerseits sind diese Produkte alle 0 (da ja die Wahrscheinlichkeit jedes einzelnen Wertes x in einer kontinuierlichen Wahrscheinlichkeitsverteilung 0 ist). Andererseits ist aber die Zahl dieser Summanden unendlich. Offenbar ist der bisher eingeführte Begriff des Erwartungswerts zu speziell auf diskrete Wahrscheinlichkeitsverteilungen zugeschnitten, als daß man ihn unmittelbar auf kontinuierliche Zufallsvariablen anwenden könnte.

Bei der Behandlung dieses Problems zeigt sich leider, wie lange es braucht, bis Fortschritte der Mathematik sich in Statistik-Lehrbüchern für Psychologen niederschlagen. In einigen wird für kontinuierliche Zufallsvariablen immer noch eine aus dem 19. Jahrhundert stammende Definition des Erwartungswerts weitergegeben, die genau so speziell auf kontinuierliche Wahrscheinlichkeitsverteilungen zugeschnitten ist, wie die bisher eingeführte Summenformel nur für diskrete anwendbar ist.

<<

Die entsprechende Definitionsgleichung, in der $f(x)$ die Dichtefunktion ist, lautet

$$E(X) := \int_{-\infty}^{+\infty} x \cdot f(x) \, dx$$

>>

⁴⁷Es gehört schon einige mathematische Versiertheit dazu, auch noch zu beweisen, daß dieser Schwerpunkt genau bei der Quadratwurzel aus $2.5 \cdot \pi$ liegt (wobei π die als "Kreiskonstante" bekannte Zahl 3.14... ist).

Ganz abgesehen davon, daß dadurch Leserinnen und Leser unnötig abgehängt werden, denen die Integralrechnung nicht mehr ganz geläufig ist, hat diese Vorgehensweise auch sachlich zwei erhebliche Nachteile:

- Da es zwei getrennte Definitionen des Erwartungswerts für diskrete und kontinuierliche Wahrscheinlichkeitsverteilungen gibt, müssen viele Formeln, Regeln usw. getrennt für beide Verteilungstypen formuliert werden - in Summenform für diskrete Wahrscheinlichkeitsverteilungen und in Integralform für kontinuierliche.
- Es gibt auch Wahrscheinlichkeitsverteilungen, die weder diskret noch kontinuierlich, sondern aus beiden Typen gemischt sind. Für den Erwartungswert von Zufallsvariablen mit einer solchen Wahrscheinlichkeitsverteilung müssen dann gesonderte "Mischverteilungsmodelle" herangezogen werden.

Es gibt zwei Arten von Fortschritten der Mathematik: Solche, bei denen alles komplizierter wird, so daß sie nur noch von Insidern nachzuvollziehen sind, und andererseits Fortschritte, die Mühsames durch einen neuen Ansatz einfacher machen.

Bekannt ist die Anekdote vom "kleinen Gauß", dessen Lehrer seine Schüler beschäftigen wollte und ihnen die Aufgabe stellte, die Zahlen von 1 bis 100 zu addieren. Leider machte ihm der spätere Mathematiker einen Strich durch die Rechnung, indem er nach wenigen Minuten die Lösung hatte: 5050. Was er gerechnet hatte, läßt sich mit Klammern (die er natürlich noch nicht kannte) folgendermaßen darstellen:

$(1+100) + (2+99) + (3+97)$ usw., das sind 50 Zahlenpaare, die zusammen immer 101 ergeben. Also: $50 \cdot 101 = 5050$.

Erfreulicherweise gehört auch der in der heutigen Mathematik übliche Begriff des Erwartungswerts zu den Fortschritten durch Vereinfachung. Da sich die Notwendigkeit, über die bisherige Erwartungswert-Definition hinauszugehen, bei den kontinuierlichen Wahrscheinlichkeitsverteilungen ergab, wird der neue Ansatz in den folgenden Abschnitten am Beispiel der bisher behandelten Wahrscheinlichkeitsverteilung von Antwortzeiten dargestellt. Es wird sich dann aber (am Ende von Abschnitt @B.IV.2.b) zeigen, daß dieser Begriff auch den Erwartungswert einer diskreten Wahrscheinlichkeitsverteilung als Spezialfall abdeckt und darüber hinaus auch auf die aus beiden Typen gemischten Wahrscheinlichkeitsverteilungen anwendbar ist.

Ein weiterer Vorteil des allgemeineren Erwartungswert-Begriffs ist vor allem für die Psychologie von Bedeutung. Der aus dem 19. Jahrhundert stammende Begriff ging mit großer Selbstverständlichkeit davon aus, daß es bei Zufallsvariablen immer um Zahlen geht⁴⁸. Für Anwendungen in der Psychologie ist diese stillschweigende Voraussetzung aber ein unnötiger Klotz am Bein. Im Anschluß an die Definition des Erwartungswerts einer von der Zufallsvariablen

⁴⁸Um kein falsches Bild zu erwecken, ist allerdings zu erwähnen, daß Zufallsvariablen in dieser traditionellen Wahrscheinlichkeitstheorie auch aus mehreren Zahlen bestehen konnten. Zusätzlich zur "Antwortzeit" könnte man z.B. auch einen Aspekt des Inhalts der Antwort in einer Zahl ausdrücken (z.B. eine Punktzahl für die Richtigkeit der Antwort vergeben) und sich für die gemeinsame Wahrscheinlichkeitsverteilung beider Zahlen interessieren.

X abgeleiteten Größe $g(X)$ wird in Abschnitt @B.IV.2.d darauf einzugehen sein, daß dieser Begriff auch dann noch anwendbar ist, wenn die Werte der Zufallsvariablen X keine Zahlen sind.

Um die Aufgabe nicht unnötig schwer zu machen, sind bei all diesen Überlegungen zwei Aspekte zu unterscheiden:

- Da in späteren Abschnitten die Begriffe des Erwartungswerts und der Varianz auch auf kontinuierliche Zufallsvariablen angewandt werden, ist es wichtig, wie diese Begriffe zu verstehen sind, was damit gemeint ist. Deshalb sollte man ihre Definitionen verstehen.
- Dagegen ist es für die weitere Arbeit nicht erforderlich, solche Erwartungswerte auch berechnen zu können. Dafür werden Methoden der Integralrechnung verwendet, deren Ansatz in einem gesonderten Abschnitt "für Spezialisten" (dem Abschnitt @f) kurz dargestellt wird.

Die verallgemeinerte Erwartungswert-Definition, die in den folgenden Abschnitten entwickelt wird, kommt nicht nur bei der Definition einer einzigen Kennziffer zur Anwendung. Bei Häufigkeitsverteilungen und diskreten Wahrscheinlichkeitsverteilungen konnten viele Kennziffern aufgrund des arithmetischen Mittels bzw. des Erwartungswerts einer von X abgeleiteten Größe $g(X)$ definiert bzw. berechnet werden. Dazu gilt das folgende Prinzip, das im Moment etwas abstrakt erscheinen mag, dessen Bedeutung aber durch wiederholte Anwendungen in den folgenden Abschnitten demonstriert wird:

Prinzip für Erwartungswert-basierte Kennziffern: Alle Definitionen von Kennziffern, die auf dem Erwartungswert einer Zufallsvariablen X oder einer von X abgeleiteten Größe $g(X)$ beruhen, und alle Formeln zur vereinfachten Berechnung solcher Kennziffern können sinngemäß aus den diskreten Wahrscheinlichkeitsverteilungen für andere Wahrscheinlichkeitsverteilungen übernommen werden, insbesondere auch für kontinuierliche. "Sinngemäß" heißt dabei, daß der Teil der Definition, der nur die Erwartungswerte benutzt, übernommen werden kann; statt der Darstellung des Erwartungswerts über eine Summe ist aber der nach der allgemeineren Erwartungswert-Definition ermittelte Erwartungswert einzusetzen.

Für die Varianz bedeutet das beispielsweise: Die Definitionsgleichung $\sigma_x^2 := E[(X - \mu_x)^2]$ ist ebenso wie die vereinfachte Berechnung über die Gleichung $\sigma_x^2 = E(X^2) - [E(X)]^2$ auf alle Wahrscheinlichkeitsverteilungen übertragbar; die in diesen Gleichungen auftauchenden Erwartungswerte sind aber ggf. nach der verallgemeinerten Definition zu bestimmen.

Die Demonstration des Erwartungswertbegriffs anhand unserer Beispielverteilung aus den vorangehenden Abschnitten wird dadurch erleichtert, daß dort die Wahrscheinlichkeit negativer Werte der Zufallsgröße null ist; denn das ermöglicht einen relativ einfachen Einstieg, der in Abschnitt @b) erfolgt: Ähnlich wie bei der Herleitung der Dichtekurve als Grenzfall von Histogrammen kann auch der Erwartungswert der Antwortzeit X als Grenzfall des

Erwartungswerts "abgerundeter Antwortzeiten" bestimmt werden. Dieser Ansatz wird in Abschnitt @d) (nach einigen ergänzenden Hinweisen "für Spezialisten" in Abschnitt @c) verallgemeinert, um auch den Erwartungswert einer von der Zufallsgröße X abgeleiteten Größe $g(X)$ für Situationen einzuführen, in denen diese abgeleitete Größe $g(X)$ mit Wahrscheinlichkeit I größer oder gleich 0 ist. Schließlich wird in Abschnitt @e) ein einfacher Trick gezeigt, mit dem der Ansatz auf Zufallsvariablen und abgeleitete Größen ausgedehnt werden kann, bei denen auch negative Zahlen mit einer Wahrscheinlichkeit vorkommen können, die größer als null ist.

Vorab noch eine terminologische Klärung, die zum Verständnis der Überschrift des nächsten Abschnitts beiträgt. Es ist allgemein üblich, Zahlen genau dann als negativ zu bezeichnen, wenn sie kleiner als 0 sind. Es gibt aber keine klare Sprachregelung, ob die Zahl 0 eine positive Zahl ist. In den meisten Büchern von "professionellen" Mathematikern wird die Bezeichnung "positiv" für Zahlen reserviert, die größer als 0 sind, und wenn auch die 0 mitgemeint sein soll, wird von "nichtnegativen" Zahlen gesprochen. In einigen Büchern wird aber auch die Zahl 0 zu den positiven Zahlen gezählt, und wenn man nur die Zahlen meint, die größer als 0 sind, spricht man von "streng positiven" Zahlen. Am eindeutigsten ist es, in dieser Situation das Wort "positiv" ohne Zusatz ganz zu vermeiden und aus beiden Begriffssystemen diejenigen Bezeichnungen zu verwenden, die unmißverständlich sind. Daraus ergibt sich die folgende

Sprachregelung: Eine Zahl ist

- *negativ genau dann, wenn sie kleiner als 0 ist,*
- *nichtnegativ genau dann, wenn sie größer oder gleich 0 ist, und*
- *streng positiv, wenn sie größer als 0 ist.*

Das bedeutet auch, daß der Erwartungswert einer Zufallsvariablen auch dann nach den Regeln des folgenden Abschnitts bestimmt werden kann, wenn (anders als in unserer Beispielverteilung) die Punktwahrscheinlichkeit eines Werts von $x = 0$ größer als 0 ist.

b) Der Erwartungswert nichtnegativer Zufallsvariablen

Ähnlich, wie in Abschnitt @B.IV.1.b die Dichtekurve als Grenzfall von Histogrammen für künstlich diskrete Skalen hergeleitet wurde, soll jetzt anhand der Beispielverteilung von Antwortzeiten aus Abschnitt @IV.B.1 ein Konzept des Erwartungswerts entwickelt werden, welches sich als Grenzfall des Erwartungswerts einer Zufallsvariablen mit Werten auf einer künstlich diskreten Skala verstehen läßt. Um zu einer derartigen Annäherung zu kommen, können wir an die bei der Herleitung der Dichtekurve eingeführte Vorstellung anknüpfen, daß die Antwortzeit mit einer digitalen Stoppuhr gemessen wird.

Für die Annäherung an den Erwartungswert werden nun zwei Details solcher digitaler Stoppuhren wichtig, die beim Histogramm nur begrenzt bedeutsam waren.

- Da die Stoppuhr in dem Augenblick gestartet wird, in dem die Antwortzeit beginnt, und in dem Augenblick angehalten wird, wo die Antwortzeit endet, werden die kontinuierlichen Reaktionszeiten niemals aufgerundet, sondern abgerundet, wenn sie kein ganzzahliges Vielfaches der digitalen Maßeinheit sind. Wenn die Stoppuhr z.B. Zehntelsekunden mißt, dann werden alle kontinuierlichen Antwortzeiten, die selbst kein ganzzahliges Vielfaches von 0.1 sec sind, auf das nächstniedrigere solchen Vielfache (also die nächstniedrigere Dezimalzahl mit einer Stelle hinter dem Komma) abgerundet.
- Digitaluhren haben immer eine begrenzte Stellenzahl, und wenn diese überschritten wird, fangen viele wieder bei 0 an. Es soll aber im folgenden angenommen werden, daß die Digitaluhr bei einem solchen "Überlauf" mit blinkenden Ziffern stehenbleibt. Dann weiß man wenigstens, daß die Antwortzeit mindestens so hoch war wie die "Überlaufzeit".⁴⁹

Als Meßwert wird also normalerweise die von der Uhr abgelesene Zeit verwendet, und wenn die Uhr mit blinkenden Ziffern stehenbleibt, wird die Überlaufzeit als Meßwert verwendet. Auch wenn letzteres kaum noch als "Abrundung" im üblichen Sinn des Wortes bezeichnet werden kann, sollen im folgenden der Einfachheit halber beide Arten des Ersetzens kontinuierlicher Zeit durch einen meist niedrigeren und niemals höheren digitalen Meßwert als "Abrundung" bezeichnet werden, und die so ermittelten Meßwerte bilden die "abgerundete Antwortzeit" X' . Zur Unterscheidung davon wird im folgenden auch von der "kontinuierlichen Antwortzeit" X gesprochen, wenn die nicht abgerundeten Antwortzeiten gemeint sind, auf die sich die Ogive und die Dichtekurve beziehen. Erst am Schluß der Überlegungen wird sich zeigen, welchen Vorteil es hat, zu wissen, daß der digitale Meßwert X' niemals größer ist als die kontinuierliche Antwortzeit X .

Um die Berechnung des Erwartungswerts derartiger abgerundeter Antwortzeiten X' mit Zahlen aus Tabelle @6b zu demonstrieren, soll zunächst einmal angenommen werden, daß die Digitaluhr nur ganze Sekunden mißt und schon nach 8 Sekunden mit blinkender Ziffer 7 stehenbleibt. Die entsprechenden Berechnungen sind in der folgenden Tabelle @6d zusammengefaßt.

⁴⁹Man könnte fragen, warum denn der Versuchsleiter nicht einfach aufpaßt, ob (und ggf. wie oft) die Uhr überläuft, bzw. warum er sich nicht gleich eine Uhr leistet, die eine hinreichende Stellenzahl hat, um jede vernünftigerweise zu erwartende Antwortzeit ohne Überlauf messen zu können. Augenzwinker: Das hat didaktische Gründe. Die Uhr, die beim Überlauf mit blinkenden Ziffern stehenbleibt, ist am ehesten geeignet, die Herangehensweise bei der Bestimmung des Erwartungswert einer Zufallsgröße plausibel zu machen, bei der die kumulative Wahrscheinlichkeit jeder noch so großen endlichen Zahl kleiner als 1 ist - und das ist ja bei unserer Beispielerverteilung der Fall.

x'_k	entsprechendes X -Intervall	p_k	$p_k \cdot x'_k$
0	$0 \leq X < 1$	0.095163	0.000000
1	$1 \leq X < 2$	0.234517	0.234517
2	$2 \leq X < 3$	0.263750	0.527501
3	$3 \leq X < 4$	0.204673	0.614019
4	$4 \leq X < 5$	0.119812	0.479246
5	$5 \leq X < 6$	0.054761	0.273806
6	$6 \leq X < 7$	0.019877	0.119263
7	$7 \leq X < 8$	0.005785	0.040495
8	$X \geq 8$	0.001662	0.013292
Summe:			2.302141

Tabelle 6d: Berechnung des Erwartungswerts von digital in Sekunden gemessenen "abgerundeten Antwortzeiten" X' mit Überlaufzeit 8 Sekunden.

Die möglichen Werte der abgerundeten Antwortzeit stehen in der ersten Spalte der Tabelle. Die Zeile für $x'_k = 1$ in dieser Tabelle ist folgendermaßen zu lesen: Eine abgerundete Antwortzeit $X' = 1$ wird von der Digitaluhr bei einer tatsächliche Antwortzeit X registriert, die größer oder gleich 1 und kleiner als 2 ist. Die (nach der Differenzenregel bestimmte) Wahrscheinlichkeit, daß die tatsächliche Antwortzeit X in dieses Intervall fällt, ist 0.234517 (wollen Sie das übungshalber nachrechnen?), und das Produkt aus dieser Klassen-Wahrscheinlichkeit p_k und der gemessenen Zeit x'_k ist 0.234517. Die übrigen Zeilen sind genau so zu lesen, wobei für die Berechnung zu $x'_k = 0$ und $x'_k = 8$ kleinere Besonderheiten zu beachten sind:

- Bei $x'_k = 0$ ist die Wahrscheinlichkeit einer tatsächlichen Antwortzeit von (einschließlich) 0 bis (ausschließlich) 1 gleich der (direkt aus der Tabelle ablesbaren) kumulativen Wahrscheinlichkeit $F(1)$. (Nach der Definition von $F(x)$ geht es bei dieser Funktion zwar um die Wahrscheinlichkeit eines Meßwerts bis einschließlich x , aber genau wie bei der Differenzenregel ist es natürlich auch hier gleichgültig, ob wir die Obergrenze einbeziehen oder nicht: Die Wahrscheinlichkeit, daß die Antwortzeit X genau 1 ist, beträgt 0.)
- Als Meßwert $X' = 8$ wird das Stehenbleiben der Uhr mit blinkender Ziffer 7 interpretiert. Das kommt dann vor, wenn die kontinuierliche Antwortzeit X größer oder gleich 8 Sekunden ist. Da die Wahrscheinlichkeit $p(X=8)$ null ist, können wir ersatzweise die Wahrscheinlichkeit $p(X > 8)$ bestimmen, und für diese gilt nach dem Negationssatz

$$p(X > 8) = 1 - p(X \leq 8) = 1 - F(8) = 1 - 0.998338 = 0.001662.$$

In der letzten Spalte von Tabelle @6d wird der Erwartungswert der abgerundeten

Antwortzeiten X' wie üblich bestimmt: Wir multiplizieren alle möglichen Werte von X' (also die Zahlen aus der ersten Spalte) mit ihrer Wahrscheinlichkeit (dritte Spalte), und die Summe dieser Produkte, die hier 2.302141 beträgt, ist der Erwartungswert der abgerundeten Antwortzeit X' .

Da ähnliche Berechnungen im folgenden auch für andere Digitaluhren angestellt werden sollen, ist es hilfreich, für diese Berechnung des Erwartungswerts der abgerundeten Antwortzeit X' eine Formel anzugeben, die auf andere Situationen übertragbar ist. Dazu werden (wie in den Spaltenköpfen von Tabelle @6d angedeutet) die möglichen Werte der abgerundeten Antwortzeit in aufsteigender Reihenfolge als x'_1, x'_2 usw. "durchnummeriert". Die Gesamtzahl dieser möglichen Abrundungswerte wird als m bezeichnet, so daß also x'_m die "Überlaufzeit" ist. Unter Verwendung dieser Bezeichnungen besagt die erste Zeile der folgenden Gleichung, daß der Erwartungswert von X' sich - wie üblich - ergibt, indem man alle möglichen Werte mit der entsprechenden Wahrscheinlichkeit multipliziert und diese Produkte aufsummiert.

$$\begin{aligned} E(X') &= \sum_{k=1}^m x'_k \cdot p(X' = x'_k) \\ &= \sum_{k=1}^{m-1} x'_k \cdot p(x'_k \leq X < x'_{k+1}) + x'_m \cdot p(X \geq x'_m) \end{aligned}$$

In der zweiten Zeile der Gleichung wird dann das Ereignis $X' = x'_k$ durch ein entsprechendes, in der zweiten Spalte von Tabelle @6d aufgeführtes Ereignis ersetzt:

- Für k -Werte von 1 bis $m - 1$ ist die abgerundete Antwortzeit X' dann gleich x'_k , wenn die kontinuierliche Antwortzeit X größer oder gleich x'_k und kleiner als der nächsthöhere Abrundungswert x'_{k+1} ist.⁵⁰ Dementsprechend ist für diese Abrundungswerte x'_k die Wahrscheinlichkeit $p(X' = x'_k)$ aus der ersten Zeile der Gleichung in der zweiten Zeile durch die Wahrscheinlichkeit $p(x'_k \leq X < x'_{k+1})$ ersetzt. Da dies aber nur für die ersten $m - 1$ Abrundungswerte gilt, läuft der Summations-Index k in der zweiten Zeile der Gleichung nur bis $m - 1$.
- Die Überlaufzeit x'_m ist dann als Wert der abgerundeten Antwortzeit X' anzusetzen, wenn die kontinuierliche Antwortzeit X mindestens gleich dieser Überlaufzeit ist. Das Produkt aus dieser Überlaufzeit und der entsprechenden Wahrscheinlichkeit $p(X \geq x'_m)$ paßt nicht in die mit Summenzeichen geschriebene Summe in der zweiten Zeile der Gleichung und wird deshalb gesondert addiert.

<<

Man kann die abgerundete Antwortzeit X' auch als eine von X abgeleitete Größe $g(X)$ betrachten.

⁵⁰Wenn die Gleichung nur für Tabelle @6d gebraucht würde, könnte man statt x'_{k+1} auch $x'_k + 1$ schreiben. Das Symbol x'_{k+1} für den nächsthöheren Abrundungswert hält die Gleichung aber offen für Digitaluhren mit anderen Maßeinheiten, die im weiteren Verlauf betrachtet werden sollen.

Stellt das in Tabelle @6d demonstrierte Verfahren eine exakte Anwendung unserer bisherigen Regeln für den Erwartungswert einer von X abgeleiteten Größe $g(X)$ dar?

Bei unserer bisherigen Regel wurde davon ausgegangen, daß für jeden möglichen Wert x_k der Zufallsvariablen X seine Wahrscheinlichkeit $p_k = p(X = x_k)$ mit dem entsprechenden Wert der abgeleiteten Größe multipliziert wird (also mit $g(x_k)$), und die Summe dieser Produkte ergibt den Erwartungswert von $g(X)$. Würden wir genau in dieser Weise verfahren, dann hieße das, für jede nichtnegative Zahl x das Produkt $g(x) \cdot p(X = x)$ zu bilden und diese Produkte aufzusummieren. Damit hätten wir aber wieder genau dasselbe Problem, das in Abschnitt @B.IV.2.a auftaucht: Für jede nichtnegative Zahl x ist die Wahrscheinlichkeit $p(X = x)$ null, und wir hätten unendlich viele Produkte dieser Art aufzusummieren.

Die Lösung dieses Problems ergibt sich aber aus einem einfachen Gedankengang: Der Erwartungswert der abgerundeten Antwortzeit X' darf nicht davon abhängen, ob wir mit der Schreibweise $g(X)$ daran erinnern, daß die künstlich diskrete Zufallsvariable X' durch die "Abrundungsregel" g aus der kontinuierlichen Zufallsvariablen X abgeleitet wurde, oder ob wir für die abgerundete Antwortzeit ein neues Symbol wie X' verwenden.

Um dieses Prinzip zu erfüllen, ist die Definition des Erwartungswerts einer aus der Zufallsgröße X abgeleiteten Größe $g(X)$ so zu fassen, daß sie genau dieses Prinzip zur Regel erklärt: Der Erwartungswert einer aus X abgeleiteten Größe $g(X)$ ist gleich dem Erwartungswert einer Zufallsvariablen X' , die dieselbe Verteilung hat wie $g(X)$.

Damit brauchen wir unsere bisherige Regel nicht zu korrigieren, da sie sich nur auf diskrete Wahrscheinlichkeitsverteilungen bezog, und dort führt die soeben eingeführte Definition zum gleichen Ergebnis wie die früher dargestellte. Man kann das abstrakt beweisen, aber ein Beispiel mag genügen. Wenn eine diskrete Zufallsvariable X als Werte die ganzen Zahlen von -5 bis $+5$ annehmen kann und es um den Erwartungswert von X^2 geht, dann würden beispielsweise die Ereignisse $X = -3$ und $X = 3$ etwas unterschiedlich in die Berechnung eingehen, je nach dem, ob wir die bisherige oder die neue Regel anwenden; am Ergebnis - dem Erwartungswert von X^2 - würde das unter dem Strich nichts ändern:

- Nach der bisherigen Regel sind die Produkte $p(X = -3) \cdot (-3)^2$ und $p(X = 3) \cdot 3^2$ getrennt in die Summe $\sum_k p_k \cdot x_k^2$ aufzunehmen.

- Nach der neuen Regel haben wir eine neue Zufallsvariable $X' = X^2$ zu bilden, und bei der Bildung des Erwartungswerts von X' ist das Produkt $p(X' = 9) \cdot 9$ in die Summe aufzunehmen. Für die Wahrscheinlichkeit des Ereignisses $X' = 9$ gilt dabei:

$$p(X' = 9) = p(X = -3 \vee X = 3) = p(X = -3) + p(X = 3).$$

(Das letzte Gleichheitszeichen folgt aus dem einfachen Additionssatz, da die Ereignisse $X = -3$ und $X = 3$ sich gegenseitig ausschließen.).

Unter dem Strich macht es aber offenbar keinen Unterschied, ob wir die Produkte $p(X = -3) \cdot (-3)^2$ und $p(X = 3) \cdot 3^2$ getrennt in die Summe $\sum_k p_k \cdot x_k^2$ eingehen lassen (bisherige Regel) oder aber einmal das Produkt $(p(X = -3) + p(X = 3)) \cdot 9$.

>>

Was wissen wir nun über $E(X)$ - den Erwartungswert der kontinuierlichen Antwortzeit -, wenn wir für die Digitaluhr mit Maßeinheit 1 sec und Überlaufzeit 8 sec den Erwartungswert der abgerundeten Antwortzeit als $E(X') = 2.302141$ bestimmt haben? Da die abgerundete Antwortzeit X' kleiner, aber niemals größer als die kontinuierliche Antwortzeit X sein kann, ist naheliegenderweise davon auszugehen, daß $E(X')$ zwar kleiner, aber nicht größer als $E(X)$ sein

kann⁵¹. Es gilt also

$$E(X') \leq E(X).$$

Das Ausmaß, in dem der Erwartungswert der abgerundeten Antwortzeit X' den Erwartungswert der kontinuierlichen Antwortzeit X unterschätzt, können wir aber durch die Überlegung in den Griff bekommen, wie sich größere Meßgenauigkeit und größere Überlaufzeiten der Digitaluhr auf den Erwartungswert der abgerundeten Antwortzeit auswirken. Die Ergebnisse entsprechender Berechnungen sind in der folgenden Tabelle @6e zusammengefaßt.

Maßeinheit	Überlaufzeit	$E(X')$
0.10000000	10	2.7524761
0.01000000	100	2.7974956
0.00100000	1000	2.8019956
0.00010000	10000	2.8024456
0.00001000	100000	2.8024906
0.00000100	1000000	2.8024951
0.00000010	10000000	2.8024956
0.00000001	100000000	2.8024956

Tabelle 6e: Die Reaktion des Erwartungswerts der abgerundeten Antwortzeit X' auf immer feinere Maßeinheiten und immer größere Überlaufzeiten.

Die erste Zeile dieser Tabelle bedeutet z.B.: Wenn die abgerundete Antwortzeit X' in Einheiten von 0.1 sec gemessen wird und die Überlaufzeit 10 sec beträgt⁵², dann beträgt der Erwartungswert der abgerundeten Antwortzeit 2.7524761. Dieser Erwartungswert ist nach dem gleichen Verfahren wie in Tabelle @6d berechnet, also nach der @kurz hinter dieser Tabelle angegebenen Formel für $E(X')$.

⁵¹Das Wort "naheliegenderweise" in einem mathematischen Text mag überraschen. Daher als Erinnerung: Es geht um die Definition eines Begriffs des Erwartungswerts, der für unsere Beispielverteilung anwendbar ist. Definitionen können nicht wahr oder falsch sein, aber mehr oder weniger sinnvoll. In diesem Kontext könnte man ausführlicher sagen: Um zu einer sinnvollen Definition zu kommen, ist naheliegenderweise vorzumerken: Die Definition des Erwartungswerts *soll* so sein, daß der Erwartungswert von X' nicht größer sein kann als der Erwartungswert von X , da X' nicht größer sondern höchstens kleiner sein kann als X .

⁵²Bei einer Maßeinheit von 0.1 sec bedeutet eine Überlaufzeit von 10 sec, daß die höchste von der Uhr angezeigte Zeit 9.9 sec ist. Nach Ablauf des Zeitintervalls, für das der Meßwert 9.9 sec mit nicht blinkenden Ziffern angezeigt wird, bleibt die Uhr mit blinkenden Ziffern stehen.

Wie sind für diese Zeile die Zahl m sowie die "Abrundungswerte" x'_1 , x'_2 und x'_m anzugeben, wenn der Erwartungswert von X' nach der Formel für $E(X')$ berechnet werden soll?

Bei einer Maßeinheit von 0.1 sec und einer Überlaufzeit von 10 sec gibt es insgesamt $m=101$ mögliche Werte der abgerundeten Antwortzeit, nämlich $x'_1 = 0.0$, $x'_2 = 0.1 \text{ sec}$ usw. bis zur Überlaufzeit von $x'_m = 10.0 \text{ sec}$.

Wie man sieht, wird der Erwartungswert der abgerundeten Antwortzeit immer größer, wenn die Abrundungsfehler kleiner werden; aber die Zunahme wird immer kleiner, und von der vorletzten zur letzten Zeile ist diese Zunahme schon so klein, daß man mehr als sieben Nachkommastellen bräuchte, um sie noch zu sehen.

Die Tabelle legt eine Vermutung nahe, die sich auch mathematisch beweisen läßt: Auch wenn wir die Tabelle bis ins Unendliche fortsetzen würden, würde $E(X')$ - der Erwartungswert der abgerundeten Antwortzeiten - niemals die Zahl 2.8024957 überschreiten, und wenn wir mit noch mehr Nachkommastellen rechnen würden, würde sich abzeichnen, daß $E(X')$ auch niemals größer als 2.802495609 wird. Der Erwartungswert der kontinuierlichen Antwortzeit X ist nun so definiert, daß er die kleinste derartige Zahl ist:

Der Erwartungswert der kontinuierlichen Antwortzeit X in unserer Beispielveilung ist die kleinste Zahl x , die bei jeder noch so feinen Maßeinheit und jeder noch so großen Überlaufzeit vom Erwartungswert der abgerundeten Antwortzeit X' niemals überschritten wird.

Mit mathematischen Methoden, die hier nicht weiter dargestellt zu werden brauchen, läßt sich zeigen, daß diese kleinste Zahl - also der Erwartungswert $E(X)$ - die Quadratwurzel aus $2.5 \cdot \pi$ ist, wobei π die vor allem als "Kreiskonstante" bekannte Zahl 3.14... ist. Eine recht genaue Dezimalzahl für diesen Erwartungswert ist 2.802495608. Auch hier wird wie bei diskreten Wahrscheinlichkeitsverteilungen anstelle der Schreibweise $E(X)$ oft das Symbol μ_x verwendet.

<<

Wenn man sich die Tabelle @6e unendlich fortgesetzt denkt, könnte man auch sagen, daß der Erwartungswert $E(X)$ der Grenzwert ist, gegen den der Erwartungswert $E(X')$ bei immer feinerer Meßgenauigkeit und immer größerer Überlaufzeit strebt. Mit einer derartigen Definition würde man eher an Bekanntes anknüpfen, aber die Definition als die kleinste Zahl x mit der o.g. Eigenschaft läßt sich einfacher verallgemeinern.

Damit hängt es auch zusammen, daß ein Rundungsverfahren gewählt wurde, bei dem die gerundete Antwortzeit X' niemals größer als die kontinuierliche Antwortzeit X sein kann. Anderenfalls hätte nämlich der Erwartungswert $E(X')$ auch größer als $E(X)$ werden können, und damit wäre eine Definition des Erwartungswerts $E(X)$ als kleinste Zahl x mit der o.g. Eigenschaft nicht mehr möglich gewesen.

>>

Die obige Definition des Erwartungswerts der kontinuierlichen Antwortzeit X war sehr

speziell auf dieses Demonstrationsbeispiel zugeschnitten. Es ist daher erforderlich, diese Definition so zu verallgemeinern, daß sie auch auf andere Zufallsvariablen anwendbar ist. Dazu ist zunächst darauf hinzuweisen, das es auch Wahrscheinlichkeitsverteilungen gibt, in denen es keine endliche Zahl x mit der verlangten Eigenschaft gibt. Das heißt: Für jede noch so große endliche Zahl x gibt es eine "Abrundungsregel", bei deren Anwendung der Erwartungswert der abgerundeten Zufallsgröße X' größer als x ist. Im Augenblick genügt es, darauf hinzuweisen, daß es solche Wahrscheinlichkeitsverteilungen gibt, um zu motivieren, daß für solche Fälle in der folgenden Definition eine Sonderregelung getroffen wird. Was es bedeutet, daß dann kein "eigentlicher Erwartungswert" existiert und daß "der uneigentliche Erwartungswert $+\infty$ beträgt", wird in Abschnitt @B.IV.2.c ("für Spezialisten") zu klären sein.

Eine mathematisch präzise Definition, die den wesentlichen Gedankengang bei der Herleitung des Erwartungswerts über Tabelle @6e zusammenfaßt⁵³, lautet:

Definition: Der Erwartungswert einer nichtnegativen Zufallsvariablen X ist die kleinste endliche Zahl x , bei der die Ungleichung

$$x \geq \sum_{k=1}^{m-1} x_k \cdot p(x_k \leq X < x_{k+1}) + x_m \cdot p(X \geq x_m)$$

für jede endliche, streng monotone Zahlenfolge $\{x_1, \dots, x_k, \dots, x_m\}$ mit $x_1 = 0$ gilt. Existiert keine endliche Zahl x mit dieser Eigenschaft, dann existiert auch kein "eigentlicher Erwartungswert" der Zufallsvariablen X , und der "uneigentliche Erwartungswert" von X beträgt $+\infty$.

Dazu sollte zunächst geklärt werden, was mit einer "streng monotonen, endlichen Zahlenfolge $\{x_1, \dots, x_k, \dots, x_m\}$ " gemeint ist. Auf das Beispiel der Antwortzeit bezogen, handelt es sich um diejenigen Zahlen, auf die die kontinuierliche Antwortzeit X (oder allgemein: die kontinuierliche Zufallsgröße X) "abgerundet" wird, um zu einer künstlich diskreten Zufallsvariablen X' zu kommen. Diese Zahlen bilden eine "streng monotone, endliche Zahlenfolge $\{x_1, \dots, x_k, \dots, x_m\}$ ", wenn es sich um m verschiedene Zahlen handelt, die in aufsteigender Reihenfolge geordnet sind. (Eine Zahlenfolge ist endlich, wenn sie endlich viele "Elemente" hat, und streng monoton, wenn jedes Element ab dem zweiten größer ist als das vorhergehende. Eine

⁵³Um den Bezug zu dieser Herleitung genauer nachzuprüfen, kann man auf die Gleichung @ für $E(X')$ - den Erwartungswert der abgerundeten Antwortzeit X' - von S. 118 zurückgreifen. Dort sind die möglichen Abrundungsergebnisse als x'_k bezeichnet, um den Bezug zu den abgerundeten Antwortzeiten X' hervorzuheben. Da X' in der Definition gar nicht mehr auftaucht, können die möglichen Abrundungsergebnisse auch als x_k bezeichnet werden, was natürlich keinen sachlichen Unterschied macht. Weitere kleinere Abweichungen der Definition von der Herleitung des Erwartungswerts von X über Tabelle @6e werden ebenso wie eine noch etwas einfachere Definition in Abschnitt @B.IV.1.c behandelt.

derartige symbolische Aufzählung der Elemente einer Zahlenfolge wird oft in geschweifte Klammern gesetzt.)

Eine Detailfrage zu der Definition ist schon fast "für Spezialisten"; aber da sie sich auch anderen stellen kann und dann das Verständnis der Definition erschwert, soll sie wenigstens erwähnt werden, bevor sie "für Spezialisten" näher behandelt wird. Bei der Herleitung des Erwartungswerts über Tabelle @6e wurden nur "Abrundungsregeln" in Betracht gezogen, bei denen die Überlaufzeit eine ganzzahlige Zehnerpotenz ist und bei denen die Abstände der möglichen Abrundungswerte gleich dem Kehrwert der Überlaufzeit ist. Die Definition bezieht dagegen wesentlich mehr Abrundungsmöglichkeiten ein: Die Ungleichung $x \geq \dots$ soll für *alle* "streng monotonen, endlichen Zahlenfolgen $\{x_1, \dots, x_k, \dots, x_m\}$ " gelten. Es läßt sich aber mathematisch beweisen, daß dieser Unterschied vom Ergebnis her unerheblich ist. Die in Tabelle @6e aufgeführten Abrundungsregeln stellten nur eine willkürliche Auswahl dar. (Es wäre auch höchst fragwürdig, wenn der Begriff des Erwartungswerts zu eng an die Tatsache gekoppelt wäre, daß Menschen normalerweise zehn Finger zum Rechnen haben, worauf das Dezimalsystem ja beruht.)

<<

Auf den ersten Blick scheint die umfassendere Menge von Abrundungsregeln in der Definition eine schärfere Anforderung darzustellen. Dieses Bedenken läßt sich aber entkräften. Mathematisch beweisen läßt sich: Jede Zahl x , die bei irgendeiner anderen Folge von Abrundungswerten kleiner als der Erwartungswert der abgerundeten Zufallsvariablen X' ist, wird auch in einer unendlich fortgesetzten Tabelle @6e irgendwann überschritten. Daher macht es vom Ergebnis her keinen Unterschied, ob man nur für die nach Art von Tabelle @6e gebildeten Folgen von Abrundungswerten oder (wie in der Definition) für jede endliche, streng monotone Zahlenfolge $\{x_1, \dots, x_k, \dots, x_m\}$ mit $x_1 = 0$ die Gültigkeit der Ungleichung aus der Definition verlangt: Die kleinste Zahl, welche die Anforderung erfüllt, ist dieselbe. Da dies aber so ist, würde die Definition unnötig kompliziert und willkürlich, wenn man die Beschreibung eines bestimmten Typs von Abrundungsregeln herausgreifen und in die Definition aufnehmen würde. In Abschnitt @B.IV.2.c wird übrigens gezeigt, wie sich die Definition noch weiter vereinfachen läßt.

>>

Wenn man die Definition verstanden hat, kann man sich fragen, wie man denn die kleinste Zahl x mit der dort verlangten Eigenschaft bestimmen kann. Dazu ist daran zu erinnern, daß es bei vielen Kennziffern hilfreich war, zwischen Definition und Rechenformeln zu unterscheiden. Dieser Unterschied ist bei dem oben eingeführten Erwartungswert-Begriff besonders ausgeprägt:

- Hat man den Weg zu der Definition nachvollzogen, dann bringt sie zwar klar zum Ausdruck, was der Begriff des Erwartungswerts meint; aber nur in einfachen Fällen ist es ohne zusätzliche "Rechenregeln" möglich, die Zahl x mit den in der Definition geforderten Eigenschaften zu bestimmen. (Bei der Varianz war es etwas anders: Es war zwar unnötig kompliziert, aber doch möglich, sie aufgrund der Definitionsformel zu berechnen.)
- Die "Rechenregeln" für den Erwartungswert beschränken sich entweder auf Spezialfälle (z.B. die Summenformel für diskrete Verteilungen oder die in Abschnitt @B.IV.2.a erwähnte Integralformel für kontinuierliche Verteilungen), oder sie benutzen einen Integralbegriff, der erst später erläutert wird.

Aus den in Abschnitt @B.IV.2.a genannten Gründen ist hier vor allem ein Grundverständnis des Begriffes (also der Definition) erforderlich, während die Rechenregeln und der zugrundeliegende Integral-Begriff nur "für Spezialisten" skizziert zu werden brauchen (was in Abschnitt @B.IV.2.f geschieht).

Die Definition auf S. 122 ist recht präzise; aber das Grundverständnis, um das es geht, läßt sich auch mit etwas mehr Bezug zum anschaulichen Modell der Digitaluhren formulieren. Wir brauchen nur die Worte "Überlaufzeit" und "Maßeinheit", die aus dem Digitaluhr-Modell stammen, zu vermeiden.

Der allgemeine Begriff des Erwartungswerts einer nichtnegativen Zufallsvariablen geht von Abrundungen der Zufallsvariablen aus, bei denen sich eine künstlich diskrete Skala mit endlich vielen Klassen (Intervallen) ergibt. Das letzte Intervall reicht von der Obergrenze des vorhergehenden Intervalls bis ins Unendliche. Bei der Abrundung wird jeder Wert der Zufallsvariablen, der nicht selbst auf eine Intervallgrenze fällt, durch die nächstniedrigere Intervallgrenze ersetzt. Dann ist der Erwartungswert einer Zufallsvariablen die kleinste Zahl x , die bei keiner solchen Abrundungsregel vom Erwartungswert der auf die künstlich diskrete Skala abgerundeten Werte überschritten wird. Gibt es keine solche Zahl x , dann existiert kein "eigentlicher Erwartungswert", und der "uneigentliche Erwartungswert" beträgt $+\infty$.

Zu dem Grundverständnis, um das es hier geht, gehört aber auch das Wissen, daß dieser Erwartungswert-Begriff nicht nur auf kontinuierliche Zufallsvariablen anwendbar ist, auch wenn er in diesem Zusammenhang eingeführt wurde. Es läßt sich nämlich leicht demonstrieren, daß bei einer Anwendung dieser Definition auf diskrete Wahrscheinlichkeitsverteilungen genau der nach den bisherigen Regeln bestimmte Erwartungswert herauskommt.

<<

Für eine solche Demonstration kann auf die Beispielverteilung aus Abschnitt @B.III.1 ("Würfelfest") zurückgegriffen werden. Dazu kann man zunächst feststellen: Verwendet man als "endliche, streng monotone Zahlenfolge" die möglichen Werte der Zufallsvariablen X (also die ganzen Zahlen von 0 bis 5) in aufsteigender Reihenfolge numeriert (also $x_1 = 0$ bis $x_6 = 5$), dann gilt für jedes k bis einschließlich $k = 5$ die Gleichung

$$p(x_k \leq X < x_{k+1}) = p(X = x_k) = p_k,$$

und für $k = 6 = m$ gilt

$$p(X \geq x_6) = p(X = x_6) = p_6.$$

Damit ist die Summe hinter dem \geq -Zeichen in der Ungleichung aus unserer Definition aber genau der nach der bisherigen Regel bestimmte Erwartungswert von X . Außerdem kann man leicht sehen, daß diese Summe bei keiner anderen Folge von Abrundungswerten einen größeren Wert annehmen kann; denn wenn die Wahrscheinlichkeit irgendeines möglichen Werts der Zufallsvariablen X mit einer anderen Zahl als diesem "Wert vor Rundung" multipliziert wird, dann kann es sich nur um eine Abrundung und nie um eine Aufrundung handeln. Wenn aber die Summe hinter dem \geq -Zeichen in der Ungleichung aus der Definition gleich dem "konventionellen" Erwartungswert und niemals größer werden kann, dann bedeutet das: Die kleinste Zahl x , bei der

diese Ungleichung für *alle* in der Definition vorgesehenen Abrundungsregeln gilt, ist der "konventionelle" Erwartungswert.

>>

c) Mathematische Details zur Definition des Erwartungswerts nichtnegativer Zufallsvariablen

<<

In diesem Abschnitt sollen "für Spezialisten" einige Details nachgetragen werden, die in den vorangehenden Abschnitten zurückgestellt wurden.

Zunächst ist zu erläutern, was es mit "eigentlichem" und "uneigentlichem" Erwartungswert auf sich hat, wenn es keine endliche Zahl x mit den in der Definition geforderten Eigenschaften gibt. Damit die Überlegungen nicht zu abstrakt werden, soll eine derartige Wahrscheinlichkeitsverteilung kurz eingeführt werden. Man kann wieder an Antwortzeiten als Zufallsvariable X denken, für deren Wahrscheinlichkeitsverteilung die kumulative Verteilungsfunktion durch die Gleichung

$$F(x) = \frac{x}{x+1}$$

angegeben wird. Vergleicht man die in Abb. @2g wiedergegebene Ogive mit der Ogive der bisherigen Beispielverteilung (Abb. @2a), dann stellt man fest: In der hiesigen Wahrscheinlichkeitsverteilung wächst die kumulative Wahrscheinlichkeit relativ schnell (der Median liegt bereits bei 1 sec), aber dann flacht die Ogive erheblich ab, und $F(x)$ strebt nur sehr langsam gegen 1. Das bedeutet aber, daß - anders als bei der bisherigen Beispielverteilung - auch bei großen Zahlen noch Spielraum für Zuwächse von $F(x)$ bleibt und damit für nennenswerte Intervallwahrscheinlichkeiten.

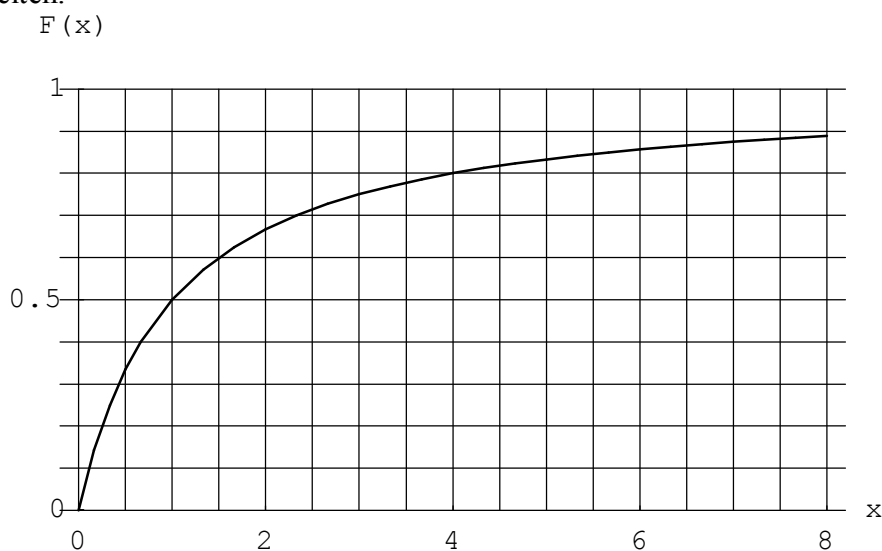


Abb. 2g: Ogive zu einer Wahrscheinlichkeitsverteilung von Antwortzeiten, bei der es keinen "eigentlichen Erwartungswert" gibt.

Das hat bei dieser Wahrscheinlichkeitsverteilung eine Konsequenz für den Erwartungswert, die aus Tabelle @6f hervorgeht. Dort sind - genau wie in Tabelle @6e für die bisherige Beispielverteilung - für die neue Verteilung die Erwartungswerte der abgerundeten Antwortzeiten X' bei immer kleineren Maßeinheiten und immer größeren Überlaufzeiten zusammengestellt.

Maßeinheit	Überlaufzeit	$E(X')$
0.10000000	10	2.3532663
0.01000000	100	4.6101784
0.00100000	1000	6.9082554
0.00010000	10000	9.2103904
0.00001000	100000	11.5129305
0.00000100	1000000	13.8155111
0.00000010	10000000	16.1180957
0.00000001	100000000	18.4206807

Tabelle 6f: Ein Beispiel, bei dem der Erwartungswert der abgerundeten Antwortzeit X' mit feineren Maßeinheiten und größeren Überlaufzeiten jede endliche Zahl x übersteigt.

Vergleicht man die letzte Spalte von Tabelle @6f mit derjenigen von Tabelle @6e, dann zeichnet sich recht deutlich ein Unterschied ab, auf den es vor allem ankommt. Bei der Behandlung von Tabelle @6e wurde festgestellt, daß es Zahlen x gibt, welche der in der letzten Spalte angegebene Erwartungswert der abgerundeten Antwortzeit auch in einer unendlich fortgesetzten Tabelle niemals überschreiten würde, und die kleinste solche Zahl ist dann der Erwartungswert der kontinuierlichen Antwortzeit. In Tabelle @6f nimmt dagegen der Erwartungswert der abgerundeten Antwortzeit von Zeile zu Zeile um mehr als 2 sec zu. Nun könnte man fragen, ob vielleicht der Punkt, an dem dann doch eine Obergrenze erreicht wird, lediglich viel später liegt, da ja die kumulative Wahrscheinlichkeit $F(x)$ auch viel langsamer gegen 1 strebt als in der alten Beispielverteilung. Es läßt sich aber (mit Mitteln der Integralrechnung) beweisen, daß dies nicht der Fall ist und daß der Zuwachs des Erwartungswerts $E(X')$ von Zeile zu Zeile auch bis ins Unendliche immer größer als 2 sec bleibt. Das bedeutet aber, daß es keine endliche Zahl x gibt, die der Erwartungswert $E(X')$ in einer unendlich fortgesetzten Tabelle nicht irgendwann überschreiten würde.⁵⁴ Konsequenz: Es gibt überhaupt keine endliche Zahl x mit der in der Definition des Erwartungswerts geforderten Eigenschaft, und damit auch keine kleinste endliche Zahl mit dieser Eigenschaft.

Auf solche Situationen geht der letzte Satz der Definition des Erwartungswerts in Abschnitt @B.IV.2.b ein. Die etwas umständliche Regelung, daß dann "kein 'eigentlicher Erwartungswert' der Zufallsvariablen X existiert und daß der 'uneigentliche Erwartungswert' $+\infty$ beträgt", liegt daran, daß $-\infty$ und $+\infty$ keine Zahlen sind, mit denen man beliebig rechnen kann. Man spricht auch von "uneigentlichen Zahlen". Auch unter professionellen Mathematikern gibt es unterschiedliche "Sprachregelungen" zum Umgang mit dieser Situation. Das zu wissen, ist wichtig, wenn man mit mathematischen Büchern arbeitet.

- Einige Mathematiker sind großzügig und lassen das Rechnen mit $-\infty$ und $+\infty$ als Zahlen zu, wobei es dann Sonderregelungen gibt, welche Rechnungen "verboten" sind. (Solche Sonderregelungen spielen eine ähnliche Rolle wie das "Verbot" einer Division durch die Zahl 0, die deshalb ja auch nicht gleich als "uneigentliche Zahl" bezeichnet wird.) Wenn man diesen Standpunkt einnimmt, kann man in der Definition des Erwartungswerts auf den zweiten Satz

⁵⁴Genauer: Für jede endliche Zahl x gilt, daß der Erwartungswert $E(X')$ zumindest in jeder Zeile der Tabelle @6f größer als x wird, deren Reihungsnummer größer als $x/2$ ist, und bei großem x auch schon früher.

ganz verzichten und im ersten Satz statt von der kleinsten *endlichen* Zahl x von der kleinsten (endlichen oder unendlichen) Zahl x mit der danach beschriebenen Eigenschaft sprechen. Konsequenz: In der problematischen Situation, auf die Tabelle @6f aufmerksam gemacht hat, ist $x = +\infty$ die einzige und damit auch die kleinste Zahl mit der in der Definition geforderten Eigenschaft.

- Andere Mathematiker sind radikal und sagen, daß in dieser kritischen Situation überhaupt kein Erwartungswert der Zufallsvariablen X existiert.

Das alles ist aber nicht nur Frage einer Sprachregelung. Grund für die Probleme ist vor allem die Tatsache, daß bestimmte Regeln ("Sätze") über Erwartungswerte nur eingeschränkt gelten, wenn "der Erwartungswert $+\infty$ vorliegt" bzw. - nach der anderen Terminologie - "kein Erwartungswert existiert". Daher ist es günstig, mit der Bezeichnung "uneigentlicher Erwartungswert" zum Ausdruck zu bringen, daß in einer solchen Situation die Größe $+\infty$ einige, aber nicht alle Eigenschaften eines Erwartungswerts hat⁵⁵. Hat man dies einmal definitorisch festgelegt, dann kann man in Formeln usw. beispielsweise $E(X) = +\infty$ schreiben; denn $+\infty$ kann nur ein uneigentlicher Erwartungswert sein.

Im Zusammenhang mit der Schwerpunktregel für den Erwartungswert blieb in Abschnitt @B.IV.2.a die Frage offen, ob nicht zumindest in einigen Wahrscheinlichkeitsverteilungen bei jedem "Aufhängungspunkt" damit gerechnet werden müßte, daß die zwar geringe, aber bis in's Unendliche reichende Masse des Verteilungsschwanzes nach Hebelgesetzen stärker nach rechts dreht, als die nur endlich weit entfernte Masse links vom Aufhängungspunkt nach links dreht. Diese Frage kann nun klar beantwortet werden: Dies trifft bei nichtnegativen Zufallsvariablen genau dann zu, wenn sie keinen eigentlichen Erwartungswert haben.

Man kann sich fragen, ob Wahrscheinlichkeitsverteilungen mit solchen Eigenschaften für die Psychologie überhaupt von Bedeutung sind. Das Beispiel von Antwortzeiten ist auch in dieser Frage aufschlußreich. In Abschnitt @B.IV.1.a wurde festgestellt, daß in der Beispielverteilung die Wahrscheinlichkeit einer Antwortzeit von über 200 Jahren noch minimal größer als 0 ist und daß man diesen offenkundigen Mangel an Realitätsbezug übergehen könne. Das gilt aber nicht mehr, wenn es nicht nur um die Wahrscheinlichkeit extremer Antwortzeiten geht, sondern auch um ihren Beitrag zum Erwartungswert. In der Beispielverteilung der vorangehenden Abschnitte war auch dieser Beitrag vernachlässigenswert⁵⁶; in der Verteilung aus Abb. @2g war dagegen die Tatsache, daß nur ein uneigentlicher Erwartungswert von $+\infty$ existiert, darauf zurückzuführen, daß die Wahrscheinlichkeitsdichte zu langsam gegen 0 strebt, wenn x ins Unendliche geht.

Fazit: Die Tatsache, daß Wahrscheinlichkeitsverteilungen mit unendlichem uneigentlichem Erwartungswert höchstens in Ausnahmefällen adäquate Modelle psychischer Prozesse sind,

⁵⁵Manche Bücher sprechen auch von einem "Quasi-Erwartungswert".

⁵⁶Ein Jahr hat "bekanntlich" 31536000 Sekunden. Laut Tabelle @6e ist aber bei einer Überlaufzeit von 10000000 sec die Differenz zwischen den Erwartungswerten der abgerundeten und der kontinuierlichen Antwortzeit bereits so minimal, daß sie sich erst in der 8. Nachkommastelle bemerkbar machen würde. Das bedeutet aber: Wenn man die Wahrscheinlichkeit, die auf unrealistische Antwortzeiten von Jahren und mehr entfällt, weiter nach unten verlagert, ändert sich am Erwartungswert kaum etwas. Genauere Rechnungen zeigen übrigens, daß der Zuwachs des Erwartungswerts $E(X')$ in den letzten Zeilen der Tabelle faktisch nur noch auf die abnehmenden Abrundungsfehler unterhalb der Überlaufzeit zurückgeht.

bedeutet keineswegs, daß es sich bei diesem Problem um eine für die Psychologie irrelevante mathematische Tüftelei handelt. Man sollte daraus die Konsequenz ziehen, bei Verteilungstypen ("Familien"), die zur Modellierung in Betracht gezogen werden, in einem frühen Stadium zu überprüfen, ob bei ihnen nicht der Erwartungswert erheblich von Eigenschaften beeinflusst wird, für die kaum ein Bezug zur Realität angenommen werden kann. Das gilt insbesondere dann, wenn der Erwartungswert vor allem von der Langsamkeit geprägt ist, mit der minimale Wahrscheinlichkeiten in extremen Skalenbereichen gegen 0 gehen. Und das gilt nicht nur für den Erwartungswert der Zufallsvariablen X , sondern auch für den Erwartungswert $E[g(X)]$ modellrelevanter, von X abgeleiteter Größen $g(X)$.

Schließlich bleibt die Ankündigung von S. @ einzulösen, daß die Definition des Erwartungswerts sich vereinfachen läßt. Zunächst kann man auf die Forderung verzichten, daß in der Zahlenfolge $\{x_1, \dots, x_k, \dots, x_m\}$ die Gleichung $x_1 = 0$ gelten soll. Ist nämlich $x_1 > 0$, dann bedeutet das lediglich, daß "versäumt" wird, das Produkt $0 \cdot p(0 \leq X < x_1)$ in die Summe aufzunehmen, und da dieses Produkt natürlich immer 0 ist, kann man darauf auch verzichten. Kommen andererseits auch negative Zahlen in der Folge vor, dann läßt sich leicht zeigen, daß der nach der Definition bestimmte Erwartungswert der kontinuierlichen Zufallsvariablen X sich nicht verändern kann, wenn auch solche Folgen von Abrundungswerten zugelassen werden: Da der Erwartungswert die kleinste Zahl x ist, bei der die Ungleichung aus der Definition für *alle* Folgen von Abrundungswerten zutrifft, ändert sich natürlich an seinem Wert nichts, wenn man auch noch negative Zahlen als Abrundungswerte zuläßt: Der Erwartungswert der abgerundeten Zufallsvariablen X' wird höchstens kleiner, wenn man unterhalb der kleinsten streng positiven Zahl x_k statt auf 0 auf eine negative Zahl abrundet. Es ändert aber offenbar nichts an der kleinsten Zahl x , die mindestens so groß ist wie *alle* Erwartungswerte abgerundeter Zufallsvariablen X' , wenn man *zusätzlich* zu den Folgen von Abrundungswerten mit $x_1 = 0$ auch noch solche einbezieht, bei denen sich durch negative Abrundungswerte ein niedrigerer Erwartungswert der abgerundeten Zufallsvariablen X' ergibt.

Formal noch einfacher wird die Definition, wenn man in der darin enthaltenen Ungleichung das Produkt $x_m \cdot p(X \geq x_m)$ streicht und lediglich fordert, daß die Ungleichung

$$x \geq \sum_{k=1}^{m-1} x_k \cdot p(x_k \leq X < x_{k+1})$$

für jede endliche, streng monotone Zahlenfolge $\{x_1, \dots, x_k, \dots, x_m\}$ gilt. Auch hier läßt sich beweisen, daß diese Variante der Definition zur gleichen Zahl x als Erwartungswert führt wie die Definition auf S. 122 mit der etwas längeren Ungleichung.⁵⁷

Für eine letzte Vereinfachungsmöglichkeit, die noch erwähnt werden soll, sind zunächst zwei Begriffe aus der mengentheoretisch orientierten Mathematik zu erläutern.

- Ist A eine beliebige Zahlenmenge, dann ist eine Zahl x eine "obere Schranke" dieser Zahlenmenge, wenn die Ungleichung $x \geq a$ für jedes Element a der Menge A gilt.

⁵⁷Beweisansatz: Jede endliche, streng monotone Zahlenfolge $\{x_1, \dots, x_k, \dots, x_m\}$ läßt sich durch Anfügen einer weiteren Zahl x_{m+1} so verlängern, daß die Wahrscheinlichkeit $p(x_m \leq X < x_m)$ beliebig nahe an die Wahrscheinlichkeit $p(X \geq x_m)$ heranrückt. Damit kann aber die Summe der Produkte $x_k \cdot p(x_k \leq X < x_{k+1})$ ohne das "Restglied" $x_m \cdot p(X \geq x_m)$ in der verlängerten Folge beliebig nahe an die Summe mit Restglied in der Ausgangsfolge herankommen.

- Das "Supremum" einer Zahlenmenge A ist die kleinste obere Schranke von A .

Ein Rückgriff auf diese Begriffe hat zunächst den Vorteil, daß damit eine Frage explizit geklärt wird, die bisher noch nicht angeschnitten wurde. Es könnte leichtfertig erscheinen, daß in der Definition des Erwartungswerts einfach von "der kleinsten Zahl x mit der Eigenschaft ..." gesprochen wird; denn es gibt auch Eigenschaften, bei denen es keine kleinste Zahl mit dieser Eigenschaft gibt. (Beispielsweise gibt es keine kleinste Zahl x mit der Eigenschaft $x > 0$.) Der Rückgriff auf den Begriff des Supremums würde aber zumindest alle Kenner der damit verbundenen Eigenschaften daran erinnern, daß jede Zahlenmenge A , die eine endliche obere Schranke hat, auch ein endliches Supremum besitzt. Eine Definition, die auf diesen Zusammenhang zurückgreift und auch die anderen Vereinfachungen aufnimmt, würde lauten: Eine nichtnegative Zufallsvariable X hat einen "eigentlichen Erwartungswert" genau dann, wenn die Menge aller Summen

$$\sum_{k=1}^{m-1} x_k \cdot p(x_k \leq X < x_{k+1})$$

bei denen $\{x_1, \dots, x_k, \dots, x_m\}$ eine endliche, streng monotone Zahlenfolge ist, ein endliches Supremum hat. In diesem Fall ist dieses Supremum der (eigentliche) Erwartungswert von X . Anderenfalls beträgt der "uneigentliche Erwartungswert von X " $+\infty$.

Der Gewinn der in dieser Definition zusammengefaßten Varianten bestünde in einer formal einfacheren Definition; der Verlust an unmittelbar erkennbarem Bezug zu den vorangehenden Überlegungen ist aber in einer Einführung für Anfänger in Wahrscheinlichkeitstheorie ebenfalls zu beachten. Daher der hier gewählte Weg, die Möglichkeit weiterer Vereinfachungen nur "für Spezialisten" zu erwähnen.

Zu ergänzen ist schließlich, daß in der heutigen mathematischen Wahrscheinlichkeitstheorie der Begriff des Erwartungswerts lediglich ein Spezialfall eines allgemeinen Integralbegriffs ist, der in Abschnitt @B.IV.2.f näher dargestellt wird.

>>

d) Der Erwartungswert einer nichtnegativen, von einer Zufallsvariablen X abgeleiteten Größe $g(X)$

Das in Abschnitt @B.IV.2.b eingeführte allgemeine Erwartungswert-Konzept wurde bisher nur für den Erwartungswert einer nichtnegativen Zufallsvariablen definiert. Es läge nahe, als nächstes auf Zufallsvariablen einzugehen, die auch negative Werte haben können. Der Erwartungswert solcher Zufallsvariablen wird aber erst im nächsten Abschnitt eingeführt, da dazu das Ergebnis des hiesigen Abschnitts benutzt wird: Der Erwartungswert einer nichtnegativen, von einer Zufallsvariablen X abgeleiteten Größe $g(X)$.

Als Beispiel zur Demonstration eignet sich wieder wie in früheren Verteilungen die "abgeleitete Größe" $g(X) = X^2$. Dann kann nämlich der Erwartungswert von X^2 zusammen mit dem bereits in Abschnitt @b) ermittelten Erwartungswert von X zur Varianzberechnung nach der bekannten Formel $\sigma_x^2 = E(X^2) - [E(X)]^2$ verwendet werden.

Eine zahlenmäßige Bestimmung des Erwartungswerts von X^2 verläuft ähnlich wie beim Erwartungswert von X . Man kann sich wieder vorstellen, daß die Größe X^2 abgerundet wird, und

diese abgerundeten Werte von X^2 bilden diesmal die Zufallsvariable X' . Man könnte auch wieder an ein digitales Meßgerät denken, das die quadrierte Antwortzeit digital abrundet und bei einem "Überlaufpunkt" mit blinkenden Ziffern stehenbleibt. Solche konkrete Vorstellungen helfen beim ersten Kennenlernen eines Gedankens, und jeder kann selbst sehen, wie lange sie noch hilfreich sind. Um das Festhalten an der Vorstellung nicht unnötig zu verlängern, sollen jedenfalls die bisher verwendeten Begriffe "Maßeinheit" und "Überlaufzeit" im folgenden durch "Intervallbreite" und "Kappungsgrenze" ersetzt werden. Dann kann jeder entscheiden, wie lange es hilft, bei diesen Begriffen an ein digitales Meßgerät zu denken.

Mit den neuen Begriffen läßt sich die in der folgenden Tabelle @6g demonstrierte Berechnung folgendermaßen darstellen: Die quadrierte Antwortzeit X^2 wird auf ein ganzzahliges Vielfaches der Intervallbreite 5 abgerundet; ab der Kappungsgrenze 50 (also bei allen Werten von X^2 ab 50) wird die quadrierte Antwortzeit jedoch auf 50 abgerundet. Die so ermittelten "abgerundeten quadrierten Antwortzeiten" bilden die Zufallsvariable X' , deren Erwartungswert in der Tabelle bestimmt wird.

x'_k	entsprechendes X^2 -Intervall	P_k	$P_k \cdot x'_k$
0	$0 \leq X^2 < 5$	0.393469	0.000000
5	$5 \leq X^2 < 10$	0.238651	1.193256
10	$10 \leq X^2 < 15$	0.144749	1.447493
15	$15 \leq X^2 < 20$	0.087795	1.316923
20	$20 \leq X^2 < 25$	0.053250	1.065006
25	$25 \leq X^2 < 30$	0.032298	0.807448
30	$30 \leq X^2 < 35$	0.019590	0.587691
35	$35 \leq X^2 < 40$	0.011882	0.415861
40	$40 \leq X^2 < 45$	0.007207	0.288266
45	$45 \leq X^2 < 50$	0.004371	0.196697
50	$X^2 \geq 50$	0.006738	0.336897
Summe:			7.655538

Tabelle 6g: Berechnung des Erwartungswerts der abgerundeten quadrierten Antwortzeit X' mit Intervallbreite 5 und Kappungsgrenze bei $X^2 = 50$.

Die Eintragungen in der zweiten Zeile für $x'_k = 5$ sind beispielsweise folgendermaßen zu lesen: Ein X' von 5 - also eine abgerundete quadrierte Antwortzeit von 5 - ergibt sich, wenn die quadrierte Antwortzeit X^2 mindestens 5 und weniger als 10 beträgt (2. Spalte). Um die entsprechende Wahrscheinlichkeit (3. Spalte) zu berechnen, müssen wir diese Aussage über die

quadrierte Antwortzeit X^2 zunächst in eine Aussage über X übersetzen. Da X in unserem Beispiel nur nichtnegative Werte haben kann, ist die Angabe $5 \leq X^2 < 10$ gleichbedeutend mit $\sqrt{5} \leq X < \sqrt{10}$. Die entsprechende Wahrscheinlichkeit ist natürlich nach der Differenzenregel zu bestimmen. Für die kumulativen Wahrscheinlichkeiten der Zahlen $\sqrt{5} \approx 2.236$ und $\sqrt{10} \approx 3.162$ enthält Tabelle @6b keine genauen Einträge; aber wenn man etwas "über den Daumen peilt", dann kann man feststellen: Für das Intervall von 2.2 bis 3.2 ergäbe sich aus der Tabelle eine Wahrscheinlichkeit von $0.640845 - 0.383687 = 0.257158$, und damit ist die in der dritten Spalte der Tabelle angegebene Intervallwahrscheinlichkeit von 0.238651 nur leicht überschritten. (Das Intervall von 2.2 bis 3.2 ist ja auch etwas zu groß. Die Intervallwahrscheinlichkeit ist vom Computer berechnet und sollte eigentlich stimmen.⁵⁸)

Worauf es ankommt: Um die Wahrscheinlichkeit eines X^2 in dem in der zweiten Spalte angegebenen Bereich zu berechnen, muß diese Bereichsangabe zunächst in ein Intervall von X übersetzt werden, und dann kann die entsprechende Intervallwahrscheinlichkeit nach der Differenzenregel bestimmt werden. (Für die letzte Zeile ist natürlich die Angabe $X^2 \geq 50$ aus der zweiten Spalte zu übersetzen in $X \geq \sqrt{50}$, und da die Wurzel aus 50 etwa 7.071 ist, läßt sich die Herleitung der Eintragung in der 3. Spalte für diese Zeile folgendermaßen zusammenfassen:

$$p(X^2 \geq 50) = p(X \geq \sqrt{50}) = 1 - p(X < \sqrt{50}) = 1 - F(\sqrt{50}) \approx 1 - F(7.071) \approx 0.006738.$$

Die in der dritten Spalte angegebenen Wahrscheinlichkeiten sind natürlich für die Berechnung des Erwartungswerts der abgerundeten quadrierten Antwortzeit die Klassenwahrscheinlichkeiten p_k . Diese werden dann in der letzten Spalte mit den entsprechenden Werten von x'_k multipliziert, und die Summe dieser Produkte ergibt den Erwartungswert $E(X') = 7.655538$.

Ganz entsprechend wie bei der Annäherung an den Erwartungswert von X über Tabelle @6e wird nun in Tabelle @6h demonstriert, wie der Erwartungswert der abgerundeten quadrierten Antwortzeit auf immer kleinere Intervallbreiten und immer größere Kappungsgrenzen reagiert.

⁵⁸Natürlich kann man das, was der Computer berechnet hat, auf einem Taschenrechner mit genügend mathematischen Funktionen nachvollziehen: Die Intervallwahrscheinlichkeit von $\sqrt{5}$ bis $\sqrt{10}$ ist ja nach der Differenzenregel $F(\sqrt{10}) - F(\sqrt{5})$, und dafür kann man die kurz hinter Tabelle @6b angegebene Formel für $F(x)$ verwenden.

Intervallbreite	Kappungsgrenze	$E(X')$
0.10000000	10	6.28965224
0.01000000	100	9.99454706
0.00100000	1000	9.99950001
0.00010000	10000	9.99995000
0.00001000	100000	9.99999500
0.00000100	1000000	9.99999950
0.00000010	10000000	9.99999995
0.00000001	100000000	9.99999999

Tabelle 6h: Die Reaktion des Erwartungswerts der abgerundeten quadrierten Antwortzeit auf immer feinere Intervallbreiten und immer größere Kappungsgrenzen.

Es zeichnet sich ab, daß der Erwartungswert der abgerundeten quadrierten Antwortzeit immer näher an die Zahl 10 herangeht, und es läßt sich auch beweisen, das die Zahl 10 die kleinste Zahl ist, die auch bei noch so kleiner Intervallbreite und noch so großer Kappungsgrenze niemals überschritten wird. Daher ist diese Zahl 10 der Erwartungswert der quadrierten Antwortzeit X^2 . Als Gleichung formuliert: $E(X^2) = 10$.

Bevor der Gedankengang verallgemeinert wird, sollte das Ergebnis genutzt werden, um Varianz und Standardabweichung der Antwortzeit X für unsere Beispielverteilung zu bestimmen. Wie bei der Einführung des Prinzips für Erwartungswert-basierte Kennziffern (vgl. S. 114) angekündigt wurde, ist die Gleichung $\sigma_x^2 = E(X^2) - [E(X)]^2$ "sinngemäß" übertragbar, und nach den vorangehenden Demonstrationen sollte deutlich sein, was das bedeutet: Es sind einfach die nach den neuen Verfahren bestimmten Erwartungswerte in diese Gleichung einzusetzen. Daher ergibt sich für die Beispielverteilung eine Varianz der Zufallsvariablen "Antwortzeit" von

$$\sigma_x^2 = 10 - 2.8025^2 \approx 2.1460.$$

Natürlich ist die Standardabweichung wieder die Wurzel aus der Varianz:

$$\sigma_x = \sqrt{2.1460} \approx 1.4649$$

Da Varianz und Standardabweichung nicht die einzigen Kennziffern sind, die auf dem Erwartungswert einer von der Zufallsvariablen X abgeleiteten Größe $g(X)$ beruhen, soll jetzt der Gedankengang, der zum Erwartungswert von X^2 führte, so verallgemeinert werden, daß er auch auf andere von X abgeleitete Größen $g(X)$ anwendbar ist. Dabei bleibt jedoch einstweilen die Beschränkung aufrechterhalten, daß es sich um eine abgeleitete Größe handeln muß, die mit Wahrscheinlichkeit 1 nichtnegativ ist.

Es wird kaum noch überraschen, daß der Grundgedanke derselbe bleibt: Man betrachtet zunächst den Erwartungswert einer Zufallsvariablen X' , die aus der abgeleiteten Größe $g(X)$ durch "Abrundung" entsteht. Um für diesen Erwartungswert $E(X')$ eine Formel anzugeben, die möglichst eng an die Gleichung @ für den Erwartungswert $E(X')$ von S. 118 angelehnt ist, sollen die möglichen Abrundungswerte zunächst wieder in aufsteigender Reihenfolge als x'_1, x'_2 usw. bezeichnet werden, und die Zahl dieser Abrundungswerte ist wieder m , so daß die Kappungsgrenze x'_m ist. Wie bei der Gleichung @ von S. 118 wird auch bei der folgenden in der ersten Zeile lediglich festgehalten, daß zur Bestimmung des Erwartungswerts $E(X')$ alle möglichen Abrundungswerte x'_k mit der entsprechenden Wahrscheinlichkeit zu multiplizieren sind, und die Summe dieser Produkte ist dann der Erwartungswert von X' .

$$\begin{aligned} E(X') &= \sum_{k=1}^m x'_k \cdot p(X' = x'_k) \\ &= \sum_{k=1}^{m-1} x'_k \cdot p(x'_k \leq g(X) < x'_{k+1}) + x'_m \cdot p(g(X) \geq x'_m) \end{aligned}$$

Die zweite Zeile der Gleichung weicht dagegen geringfügig von der Gleichung @ von S. 118 ab, da X' im hiesigen Kontext nicht mehr durch Abrundung der kontinuierlichen Zufallsvariablen X entsteht, sondern durch Abrundung der aus X abgeleiteten Größe $g(X)$. Daher gilt für jeden Abrundungswert x'_k :

- Ist $k < m$ (handelt es sich also nicht um die Kappungsgrenze x'_m), dann ist X' - der abgerundete Wert von $g(X)$ - genau dann gleich x'_k , wenn die abgeleitete Größe $g(X)$ größer oder gleich x'_k und kleiner als der nächstgrößere Abrundungswert x'_{k+1} ist. Daher gilt für die Wahrscheinlichkeit, daß der abgerundete Wert von $g(X)$ gleich x'_k ist, die Gleichung

$$p(X' = x'_k) = p(x'_k \leq g(X) < x'_{k+1}).$$

- Für $k = m$ ist dagegen x'_k (bzw. x'_m) die Kappungsgrenze, und dann ist der abgerundete Wert von $g(X)$ genau dann gleich dieser Kappungsgrenze, wenn die abgeleitete Größe $g(X)$ mindestens so groß ist wie die Kappungsgrenze. Für die entsprechende Wahrscheinlichkeit gilt also:

$$p(X' = x'_m) = p(g(X) \geq x'_m).$$

<<

Die Anwendung der obigen Gleichung in Tabelle @6g und Tabelle @6h wurde dadurch erleichtert, daß jeder Vorgabe eines Intervalls für X^2 (zweite Spalte von Tabelle @6g) ein Intervall von X entsprach, dessen Grenzen sich einfach als Quadratwurzel des in der Tabelle angegebenen Intervalls für X^2 ergaben. Bei anderen von X abgeleiteten Größen kann das aber etwas komplizierter sein. Wenn man beispielsweise die Varianz σ_x^2 unmittelbar nach der Definitionsformel als Erwartungswert von $g(X) = (X - \mu_x)^2$ bestimmen will und es dabei z.B. um die Wahrscheinlichkeit $p(1 \leq g(X) < 2)$ geht, dann ist zu beachten, daß auch bei negativer Differenz $X - \mu_x$ (also bei einem X unterhalb des Erwartungswerts μ_x) das Quadrat dieser Differenz in das Intervall von (einschließlich) 1 bis (ausschließlich) 2 fallen kann. Die entsprechende Wahrscheinlichkeit kann nach dem einfachen Additionssatz bestimmt werden:

$$p(1 \leq g(X) < 2) = p(1 \leq X - \mu_x < \sqrt{2}) + p(-\sqrt{2} < X - \mu_x \leq -1).$$

Warum konnte bei der Bestimmung des Erwartungswert von X^2 für die Beispielverteilung darauf verzichtet werden, negative Werte von X entsprechend zu berücksichtigen?

In der Beispielverteilung ist die Wahrscheinlichkeit negativer Werte der Antwortzeit 0. Bei Wahrscheinlichkeitsverteilungen, in denen die Wahrscheinlichkeit $p(X < 0)$ größer als 0 ist, müßte man dagegen auch schon beim Erwartungswert von X^2 beachten, daß auch bei negativem X das Quadrat X^2 in ein Intervall zwischen zwei positiven Intervallgrenzen fallen kann.

Allgemein ist festzuhalten, daß es sehr von der Funktion g abhängt, wie die Wahrscheinlichkeiten $p(x'_k \leq g(X) < x'_{k+1})$ und $p(X \geq x'_m)$ in der @obigen Gleichung für $E(X')$ zu bestimmen sind. Es gibt sogar Funktionen g (also Regeln für die Bestimmung einer von X abgeleiteten Größe $g(X)$), bei denen diese Wahrscheinlichkeiten überhaupt nicht bestimmt werden können. Und zwar nicht, weil dazu kein adäquates Verfahren bekannt wäre, sondern weil diese Wahrscheinlichkeiten gar nicht existieren. Man spricht dann von "nicht meßbaren Funktionen (oder Abbildungen)". Dieses Problem, auf das in Abschnitt @B.IV.2.f näher einzugehen sein wird, stellt sich aber nicht bei den Funktionen, die bei den üblichen Erwartungswert-basierten Kennziffern zur Anwendung kommen. Daher genügt im Augenblick der Hinweis, daß die folgende Definition des Erwartungswerts $E[g(X)]$ nur anwendbar ist, wenn die in dieser Definition auftauchenden Wahrscheinlichkeiten existieren, und daß dies bei den abgeleiteten Größen $g(X)$, die bei den üblichen Erwartungswert-basierten Kennziffern eine Rolle spielen, immer der Fall ist.

>>

Aus der @obigen Gleichung für den Erwartungswert $E(X')$ ergibt sich nun eine allgemeine Definition des Erwartungswerts einer von X abgeleiteten nichtnegativen Größe $g(X)$ in ganz ähnlicher Weise, wie sich die Definition des Erwartungswerts von X auf S. 122 aus der @dortigen Gleichung für den Erwartungswert der abgerundeten Zufallsvariablen ergab: Es ist die kleinste endliche Zahl x , die bei keiner Abrundungsregel vom Erwartungswert der abgerundeten Größe überschritten wird. Die Parallele geht weiter: Für Situationen, in denen es keine endliche Zahl x mit dieser Eigenschaft gibt, wird eine entsprechende Sonderregelung getroffen. Die Definition lautet:

Definition: Der Erwartungswert einer von einer Zufallsvariablen X abgeleiteten nichtnegativen Größe $g(X)$ ist die kleinste endliche Zahl x , bei der die Ungleichung

$$x \geq \sum_{k=1}^{m-1} x_k \cdot p(x_k \leq g(X) < x_{k+1}) + x_m \cdot p(g(X) \geq x_m)$$

für jede endliche, streng monotone Zahlenfolge $\{x_1, \dots, x_k, \dots, x_m\}$ mit $x_1 = 0$ gilt. Existiert keine endliche Zahl x mit dieser Eigenschaft, dann existiert auch kein "eigentlicher Erwartungswert" der abgeleiteten Größe $g(X)$, und der "uneigentliche Erwartungswert" von $g(X)$ beträgt $+\infty$.

Da der in der obigen Definition zusammengefaßte Erwartungswert-Begriff bisher nur an einem spezielles Beispiel demonstriert wurde, sind einige Klarstellungen angebracht, damit man nicht zu sehr an Merkmalen dieses Demonstrationsbeispiels hängenbleibt.

- Wie beim Erwartungswert von X ist auch die Definition des Erwartungswerts einer von der Zufallsvariablen X abgeleiteten Größe $g(X)$ nicht nur auf kontinuierliche Zufallsvariablen anwendbar. Der erweiterte Begriff wurde zwar in diesem Zusammenhang erforderlich, aber er ist auch auf Fälle anwendbar, in denen die Größe $g(X)$ von einer Zufallsvariablen X mit diskreter Wahrscheinlichkeitsverteilung abgeleitet ist.
- Im Demonstrationsbeispiel war nicht nur die von der Zufallsvariablen X abgeleitete Größe $g(X)$ nichtnegativ, sondern auch die Zufallsvariable X selbst. Das ist aber nicht erforderlich, um die Definition anwenden zu können. Wenn es beispielsweise um den Erwartungswert von X^2 geht, dann ist X^2 auch bei negativem X nichtnegativ, und damit ist die Definition anwendbar.
- Bisher wurden immer nur Zufallsvariablen X behandelt, deren Werte Zahlen sind. Auch das ist aber nicht notwendig. Wie bereits in Abschnitt @B.IV.2.a angekündigt wurde, besteht ein wesentlicher Vorteil des modernen Erwartungswert-Begriffs darin, daß auch Raum für Zufallsvariablen ist, deren Werte keine Zahlen sind. Wenn es beispielsweise darum geht, das Verhalten von Personen in einem Test wahrscheinlichkeitstheoretisch zu beschreiben, könnte man auch von einer Zufallsvariablen X ausgehen, die angibt, welche Antworten gegeben wurden. In einem Fragebogen, bei dem zu jeder Frage nur die Antworten "ja", "nein" oder "weiß nicht" möglich sind, könnte beispielsweise die Angabe $X = jnjj?nnj?n$ bedeuten, daß zu den Fragen 1, 3, 4 und 8 die Antwort "ja" gegeben wurde, während die Fragen 2, 6, 7 und 10 mit "nein" beantwortet wurden und die Fragen 5 und 9 mit "weiß nicht". Bei einem anderen Test, in dem die Antwortmöglichkeiten offener sind, könnte eine Zufallsvariable X die Gesamtheit der von einer Person gegebenen Antworten wiedergeben. Für eine derartige Zufallsvariable X geben die obigen Definitionen natürlich keinen Erwartungswert her⁵⁹, aber wenn man die nach einer Auswertungsregel ermittelte "Punktzahl" als eine von der Zufallsvariablen X abgeleitete Größe $g(X)$ betrachtet, dann ist der Erwartungswert dieser Punktzahl durch die obige Definition eindeutig bestimmt, sobald eine Wahrscheinlichkeitsverteilung der Zufallsvariablen X durch ein Modell vorgegeben wird.

e) Der Erwartungswert einer von einer Zufallsvariablen X abgeleiteten Größe $g(X)$ mit negativen Werten

Die bisherige Definition des Erwartungswerts einer von einer Zufallsvariablen X abgeleiteten Größe $g(X)$ ist nur anwendbar, wenn die Werte von $g(X)$ nichtnegative Zahlen sind. In diesem Abschnitt soll die Definition so erweitert werden, daß diese Beschränkung aufgehoben werden

⁵⁹Zu erwähnen ist allerdings, daß es inzwischen auch Ansätze gibt, für solche "nichtnumerische" Zufallsvariablen einen Erwartungswert zu definieren. Diese neueste Entwicklung wird auf Methodiker-Kongressen diskutiert. Eine Zusammenfassung der wesentlichen Ergebnisse eines entsprechenden Kongreßvortrags kann im Internet unter der Adresse

<http://userpage.fu-berlin.de/~iseler/papers>
abgerufen werden. (Eine Zeitschriften-Veröffentlichung ist für 1998 vorgesehen.)

kann. Ganz nebenbei ergibt sich damit auch der Erwartungswert einer Zufallsvariablen X , die auch negative Werte haben kann.

Der Ansatz ist sehr einfach. Für den ursprünglichen, auf diskrete Wahrscheinlichkeitsverteilungen abgestellten Erwartungswert-Begriff galt eine Regel über die Differenz zweier von einer Zufallsvariablen X abgeleiteten Größen $g(X)$ und $h(X)$: Betrachte man die Differenz $g(X) - h(X)$ als eine weitere von X abgeleitete Größe, dann ist deren Erwartungswert auch die Differenz der Erwartungswerte von $g(X)$ und $h(X)$. Als Gleichung geschrieben:

$$E[g(X) - h(X)] = E[g(X)] - E[h(X)]$$

Damit der verallgemeinerte Erwartungswert-Begriff wirklich eine Verallgemeinerung des alten ist, sollte diese Beziehung natürlich erhalten bleiben. Aus dieser Forderung läßt sich nun Kapital schlagen, indem man eine von X abgeleitete Größe $g(X)$, die auch negative Werte haben kann, einfach als Differenz zweier abgeleiteter Größen darstellt, die keine negativen Werte annehmen.

Wenn etwa für die Beispielverteilung von Antwortzeiten aus den vergangenen Abschnitten die Schiefe bestimmt werden soll, dann ist dies (nach dem Prinzip für Erwartungswert-basierte Kennziffern, vgl. S. 114) wieder der Erwartungswert von z_X^3 . Die abgeleitete Größe z_X^3 kann auch negative Werte haben (nämlich für alle Werte von X , die kleiner als μ_X sind). Nun kann man aber zwei von X abgeleitete Größen $g(X)$ und $h(X)$ definieren, die beide nichtnegativ sind und deren Differenz gerade z_X^3 ist. Sie werden aus X nach folgenden Regeln abgeleitet:

- Ist $z_X^3 \geq 0$, dann ist $g(X) := z_X^3$ und $h(X) := 0$.
- Ist dagegen $z_X^3 < 0$, dann ist $g(X) := 0$ und $h(X) := |z_X^3|$ (wobei die "senkrechten" Striche natürlich für "Absolutbetrag von ..." stehen).

Es ist für das weitere Verständnis wichtig, sich an dieser Stelle die Zeit zu nehmen, sich zweierlei klar zu machen:

- Die Werte von $g(X)$ und $h(X)$ sind in jedem Fall nichtnegativ.
- Für jeden Wert der Zufallsvariablen X gilt die Gleichung

$$z_X^3 = g(X) - h(X).$$

Da $g(X)$ und $h(X)$ zwei nichtnegative, von X abgeleitete Größen sind, können ihre Erwartungswerte nach der Definition von S. 134 (z.B. mit dem in Tabelle @6h demonstrierten Annäherungsverfahren) bestimmt werden⁶⁰. Die Werte, die sich dabei ergeben⁶¹, sind $E[g(X)] = 5.53646$ und $E[h(X)] = 4.90535$. Damit ergibt sich für die Schiefe:

⁶⁰Anzumerken ist, daß die Wahrscheinlichkeitsverteilungen der abgeleiteten Größen $g(X)$ und $h(X)$ weder diskret noch kontinuierlich, sondern aus beiden Typen gemischt sind. (Bei beiden Größen gibt es eine von 0 verschiedene Punktwahrscheinlichkeit; denn wie man leicht nachprüfen kann, ist $p(g(X) = 0) = p(X \leq \mu_X)$ und $p(h(X) = 0) = p(X \geq \mu_X)$. An dieser Stelle erweist es sich also als vorteilhaft, daß die verallgemeinerte Erwartungswert-Definition nicht auf spezielle Verteilungstypen eingeschränkt ist.

⁶¹Es wäre ziemlich mühsam, das nachzurechnen. Man kann es "dem Computer glauben".

$$\text{Schiefe} = E(z_X^3) = E[g(X) - h(X)] = E[g(X)] - E[h(X)] = 5.53646 - 4.90535 = 0.63111.$$

Daß die Beispielverteilung von Antwortzeiten eine positive Schiefe hat, geht natürlich auch aus Abb. @2c hervor.

Für eine Verallgemeinerung des Gedankengangs ist eine in der Mathematik übliche Schreibweise hilfreich: Man bezeichnet die beiden "Teilfunktionen", in die $g(X)$ zerlegt wird, nicht mit besonderen Buchstaben, sondern als $g_+(X)$ und $g_-(X)$. Die folgende Zerlegungsregel enthält sowohl eine Verallgemeinerung der obigen Regeln für die Bildung der beiden Teilfunktionen als auch eine Definition des Erwartungswerts.

Zerlegungsregel: Zur Bestimmung des Erwartungswerts einer von einer Zufallsvariablen X abgeleiteten Größe $g(X)$, die auch negative Werte haben kann, wird die Funktion g nach folgenden Regeln in einen "Positiv-Teil" g_+ und einen "Negativ-Teil" g_- zerlegt:

- Ist $g(X) \geq 0$, dann ist $g_+(X) := g(X)$ und $g_-(X) := 0$.
- Ist $g(X) < 0$, dann ist $g_+(X) := 0$ und $g_-(X) := |g(X)|$.

Damit sind $g_+(X)$ und $g_-(X)$ zwei nichtnegative, von X abgeleitete Größen, deren Erwartungswerte sich aus der Definition in Abschnitt @B.IV.2.d (S. 134) ergeben.

Der Erwartungswert der von X abgeleiteten Größe $g(X)$ ergibt sich dann aus der Definitionsgleichung

$$E[g(X)] := E[g_+(X)] - E[g_-(X)],$$

wobei für "uneigentliche" Erwartungswerte sinngemäße Regelungen gelten.

<<

Die in der Zerlegungsregel erwähnten "sinngemäßen Regelungen" für uneigentliche Erwartungswerte bilden ein Beispiel für die auf S 127 erwähnten Probleme beim Rechnen mit den "uneigentlichen Zahlen" $-\infty$ und $+\infty$: Solange mindestens eine der beiden Teilfunktion $g_+(X)$ und $g_-(X)$ einen eigentlichen (also endlichen) Erwartungswert hat, gibt es keine Probleme. Sind dagegen beide Erwartungswerte $+\infty$ und damit uneigentlich, dann muß die Differenz $(+\infty) - (+\infty)$ als undefiniert betrachtet werden, so daß auch nach den großzügigsten Rechenregeln kein Erwartungswert von $g(X)$ existiert. Im einzelnen lauten die Regelungen:

- Haben $g_+(X)$ und $g_-(X)$ beide einen eigentlichen Erwartungswert, dann ist auch der nach obiger Gleichung bestimmte Erwartungswert von $g(X)$ ein eigentlicher Erwartungswert.
- Hat nur eine der Teilfunktionen $g_+(X)$ und $g_-(X)$ einen eigentlichen und die andere einen uneigentlichen Erwartungswert, dann hat auch $g(X)$ nur einen uneigentlichen Erwartungswert. Dieser beträgt $+\infty$, wenn $g_+(X)$ einen uneigentlichen Erwartungswert hat, und $-\infty$ bei uneigentlichem Erwartungswert von $g_-(X)$.
- Haben $g_+(X)$ und $g_-(X)$ beide nur einen uneigentlichen Erwartungswert, dann hat $g(X)$ überhaupt keinen Erwartungswert, auch keinen uneigentlichen.

>>

f) Bestimmung von Erwartungswerten mit Mitteln der Integralrechnung

<<

Dieser Abschnitt ist ganz "für Spezialisten" und soll irgendwann im Internet stehen. Wichtigster Inhalt: Bei einer kontinuierlichen Zufallsvariablen X läßt sich der Erwartungswert einer von X abgeleiteten Größe $g(X)$ mit der Gleichung

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) \cdot f(x) dx$$

bestimmen, wobei $f(x)$ die Dichtefunktion ist. Dabei kann das Integral auf der rechten Seite dieser Gleichung im Sinne des aus der Schule bekannten "Riemann-Integrals" interpretiert werden. Greift man auf einen moderneren Integral-Begriff (das "Lebesgue-Integral") zurück, dann gibt es eine entsprechende Gleichung, die nicht nur auf kontinuierliche, sondern auf beliebige (also auch diskrete und gemischte) Zufallsvariablen anwendbar ist. Der in den vorangehenden Abschnitten eingeführte allgemeine Begriff des Erwartungswerts ist im Grunde nichts anderes als ein solches Lebesgue-Integral, wobei lediglich auf die Integral-Schreibung verzichtet wurde.

>>

3) Die Normalverteilung

a) Allgemeines

Die wohl wichtigste kontinuierliche Verteilung ist die sogenannte Normalverteilung, die nach ihrem "Entdecker" auch häufig als Gauß'sche Verteilung bezeichnet wird. Eine Normalverteilung entsteht häufig dann, wenn ein Merkmal X durch Zusammenwirken sehr vieler Einzelfaktoren entsteht. Darauf wird @in einem späteren Abschnitt näher eingegangen.

Der Singular "die Normalverteilung" ist dabei ähnlich zu verstehen, als wenn man von "der Binomialverteilung" spricht: Es handelt sich um einen Sammelbegriff für eine Familie von kontinuierlichen Wahrscheinlichkeitsverteilungen. Die Parameter, durch die sich die Mitglieder dieser Familie unterscheiden, sind der Erwartungswert μ und die Standardabweichung σ . (Manchmal wird anstelle der Standardabweichung auch die Varianz als Parameter bezeichnet.)

<<

Die allgemeine Formel für die Dichte einer Normalverteilung mit Erwartungswert μ und Standardabweichung σ lautet:

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2}$$

Dabei ist π die bekannte "Kreiskonstante" 3.14..., während e die schon häufiger verwendete Basis der natürlichen Logarithmen ist ($e=2.718...$).

Diese Formel läßt sich nicht beweisen; sie ist Teil einer Definition: Eine Wahrscheinlichkeits-

verteilung ist definitionsgemäß eine Normalverteilung, wenn es sich um eine kontinuierliche Wahrscheinlichkeitsverteilung handelt, deren Dichtefunktion sich in der obigen Gleichung darstellen läßt. Beweisen läßt sich jedoch ein Lehrsatz, der besagt, unter welchen Bedingungen eine solche Normalverteilung zustande kommt. Dieser Lehrsatz, der sog. "Zentrale Grenzwertsatz", wird in Abschnitt @B.IV.3.d behandelt.

Für die kumulative Verteilungsfunktion gibt es keine Formel der bisher gewohnten Art. Eine Anwendung der @Integralformel von S. 108 ergibt zwar die Gleichung

$$F(b) = \int_{-\infty}^b \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{1}{2}\left(\frac{x - my}{\sigma}\right)^2} dx$$

Dieses Integral läßt sich aber nicht in der Weise "auflösen", wie man das aus der Schulmathematik bei Integralen gewohnt ist. Mit Methoden, die hier nicht weiter behandelt werden sollen, kann man jedoch numerische Werte für dieses Integral - also für die kumulative Wahrscheinlichkeit bei Normalverteilung - berechnen, und Ergebnisse derartiger Berechnungen sind in Tabellen verfügbar, deren Benutzung in den nächsten Abschnitten behandelt wird.

Viele der im folgenden dargestellten Eigenschaften der Normalverteilung lassen sich allein aus der Formel für die Dichtefunktion herleiten⁶².

>>

Allen Normalverteilungen gemeinsam ist es, daß eine Zufallsvariable mit einer solchen Verteilung prinzipiell alle negativen und positiven Zahlen als Werte haben kann. In Abb. @3a sind die Dichtekurven für verschiedene Normalverteilungen mit Erwartungswert $\mu = 100$ und verschiedenen Standardabweichungen dargestellt.

⁶² Beweise finden sich z.B. bei Hays (1973).

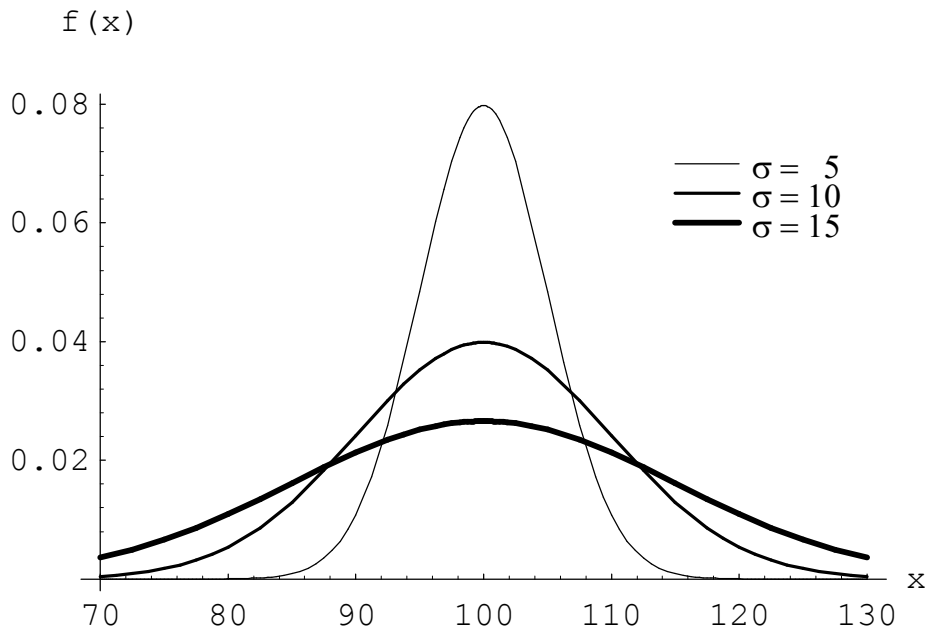


Abb. 3a: Dichtekurven von Normalverteilungen mit $\mu=100$ und unterschiedlichen Standardabweichungen

Die folgenden gemeinsamen und unterschiedlichen Merkmale gelten allgemein für Dichtekurven von Normalverteilungen:

- Dichtekurven von Normalverteilungen haben eine symmetrische "Glockenform" mit Gipfel beim Erwartungswert und Wendepunkten im Abstand von einer Standardabweichung unterhalb und oberhalb des Erwartungswerts. Das bedeutet auch: Dichtekurven von Normalverteilungen mit gleicher Standardabweichung und unterschiedlichem Erwartungswert unterscheiden sich nur durch eine Links-rechts-Verschiebung.
- Da die Gesamt-Fläche unter jeder Dichtekurve gleich sein muß (nämlich 1), ist diese Fläche bei niedriger Standardabweichung in einem engen Skalenbereich zusammengestaucht, so daß die Dichte in der Umgebung des Erwartungswerts hoch sein muß, damit noch eine Gesamt-Fläche von 1 herauskommt. Bei hoher Standardabweichung verteilt sich die Fläche dagegen über einen entsprechend breiteren Skalenbereich, so daß auch in der Umgebung des Erwartungswerts eine niedrigere Dichte vorliegt als bei niedriger Standardabweichung.
- Dieser Unterschied darf aber nicht mit den in @Statistik I im Zusammenhang mit dem Exzeß eingeführten Begriffen "steilgipflig" und "flachgipflig" verwechselt werden. Dort wurde ausdrücklich darauf hingewiesen, daß diese Merkmale immer relativ zur Gesamtverteilung zu sehen sind und daß die mit linearen Skalentransformationen verbundene Änderung der gesamten Breite einer Verteilung am Exzeß (also an der Steilgipfligkeit bzw. Flachgipfligkeit) der Verteilung nichts ändert.
- Versteht man (wie in @Statistik I begründet) unter der Form einer Verteilung die Gesamtheit all derjenigen Merkmale, die bei Lineartransformationen ohne Umpolung erhalten bleiben und sich daher auch in der Verteilung der z-Werte niederschlagen, dann kann man sogar sagen, daß

alle Normalverteilungen dieselbe Form haben⁶³. Insbesondere sind Schiefe und Exzeß⁶⁴ jeder Normalverteilung 0.

Diese allen Normalverteilungen gemeinsame Form läßt sich graphisch darstellen, indem man wie in Abb. @3b eine einzige Dichtekurve zeichnet und den erforderlichen Spielraum für Unterschiede von Erwartungswerten und Standardabweichungen durch die Beschriftung der Koordinatenachsen schafft.

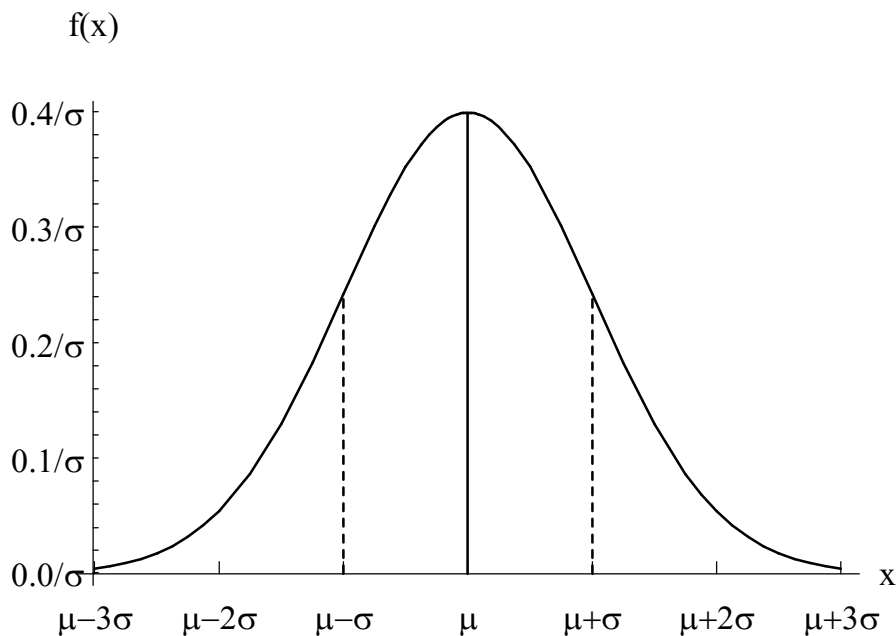


Abb. 3b: Eine Dichtekurve für alle Normalverteilungen ergibt sich durch flexible Achsenbeschriftungen.

Wie wäre z.B. für eine Normalverteilung mit einem Erwartungswert von $\mu = 100$ und einer Standardabweichung von $\sigma = 10$ die Dichte im Skalenpunkt 90 zu bestimmen. (Diese Dichte läßt sich sowohl aus Abb. @3a als auch aus Abb. @3b bestimmen!)

- In Abb. @3a ist bei $\mu = 100$ und $\sigma = 10$ die mitteldick gezeichnete Kurve zugrundezulegen, und daraus ergibt für $x = 90$ eine Dichte von etwa $f(x) = 0.024$.
- Um in Abb. @3b ablesen zu können, muß zunächst der richtige Ablesepunkt auf der

⁶³Begründung: Wie in Kürze gezeigt wird, ist die Wahrscheinlichkeitsverteilung der z-Werte bei allen Normalverteilungen gleich, nämlich eine Standardnormalverteilung.

⁶⁴Wie in @Statistik I näher begründet wurde, wird beim Exzeß häufig die Konstante 3 vom arithmetischen Mittel (bzw. vom Erwartungswert) von z^4 subtrahiert, damit das Vorzeichen der Kennziffer ausdrückt, ob der Exzeß höher oder niedriger ist als bei der Normalverteilung. Natürlich gilt die Aussage, daß der Exzeß der Normalverteilung immer 0 ist, nur für die so definierte Kennziffer.

Abszissenachse bestimmt werden. Bei den angenommenen Werten $\mu = 100$ und $\sigma = 10$ entspricht dem x -Wert von 90 der Ablesepunkt $\mu - \sigma$ auf der Abszissenachse. Für diesen Punkt kann man einen Wert von etwa $0.24/\sigma$ ablesen, und bei einer Standardabweichung von $\sigma = 10$ ergibt sich wieder eine Dichte von ca. 0.024.

Natürlich läuft die Bestimmung des Ablesepunktes auf der Abszissenachse von Abb. @3b auf die Berechnung eines z -Werts hinaus. Dann hat man bei $\mu + z \cdot \sigma$ abzulesen. (Im obigen Beispiel war der z -Wert -1 , und deshalb war bei $\mu + (-1) \cdot \sigma = \mu - \sigma$ abzulesen.

Man kann sich die Bedeutung von Abb. @3b auch so vorstellen, daß man für jede spezielle Normalverteilung eine entsprechende Achsenbeschriftung mit Zahlen anbringen kann, die sich aus den jeweiligen Parametern μ und σ ergeben. Daraus kann man nun eine weitere wichtige Eigenschaft aller Normalverteilungen herleiten:

In allen Normalverteilungen hängen Intervallwahrscheinlichkeiten und kumulative Wahrscheinlichkeiten nur von den z -Werten der Intervallgrenzen ab.

In den folgenden Anwendungsbeispielen für diese Regel geht es jeweils nur um die Feststellung, daß zwei Wahrscheinlichkeiten gleich sind. Wie sie zahlenmäßig bestimmt werden, wird in Abschnitt @B.IV.3.b behandelt.

Beispiel 1:

Welche Fläche in Abb. @3b entspricht

a) einem Wert von 40 bis 60 in einer Normalverteilung mit $\mu=50$ und $\sigma=10$ und

b) einem Wert von 85 bis 115 in einer Normalverteilung mit $\mu=100$ und $\sigma=15$?

In beiden Fällen sind die z -Werte der Intervallgrenzen -1 und $+1$; in Abb. @3b geht es also um den Bereich von $\mu - \sigma$ bis $\mu + \sigma$, also um die Fläche zwischen den beiden gestrichelten Linien. Der Anteil dieser Fläche an der Gesamtfläche unter der Dichtekurve ist natürlich unabhängig davon, welche Achsenbeschriftungen man anbringt, und deshalb sind die beiden Intervallwahrscheinlichkeiten, die den genannten Flächen entsprechen, auch gleich.

Beispiel 2:

Welche Fläche in Abb. @3b entspricht

a) einem Wert bis 70 in einer Normalverteilung mit $\mu=50$ und $\sigma=10$ und

b) einem Wert bis 130 in einer Normalverteilung mit $\mu=100$ und $\sigma=15$?

Wieder haben wir denselben z -Wert - diesmal $+2$ - für die Obergrenze; es geht also in beiden Fällen um die gesamte Fläche bis $\mu + 2\sigma$. Deshalb sind die kumulativen Wahrscheinlichkeiten, die diesen beiden Flächen entsprechen, gleich.

Da kumulative Wahrscheinlichkeiten nur von den entsprechenden z -Werten abhängen, kann man auch eine Ogive zeichnen, die mit flexibler Achsenbeschriftung auf alle Normalverteilungen

anwendbar ist. Sie ist in Abb. @3c wiedergegeben.

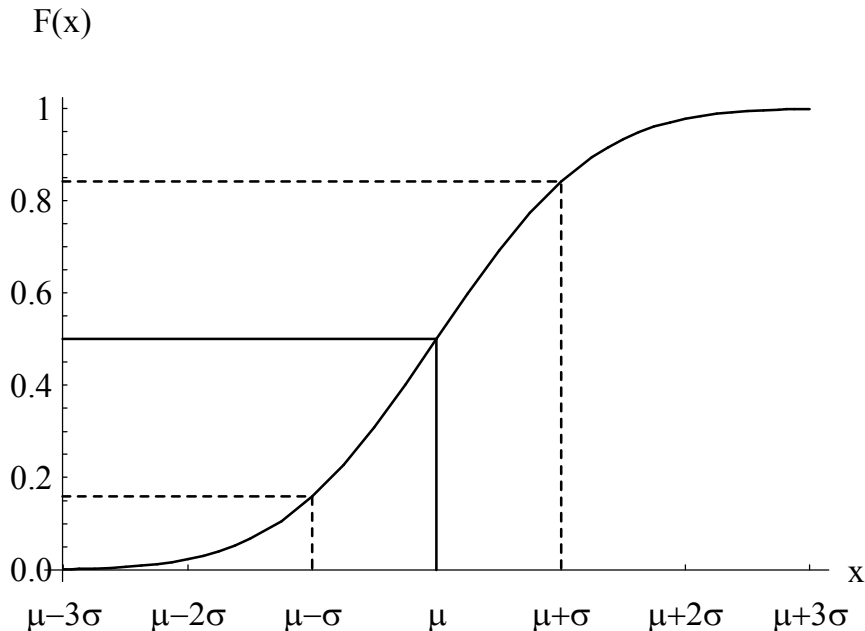


Abb. 3c: Durch flexible Beschriftung der Abszissenachse ergibt sich eine auf alle Normalverteilungen anwendbare Ogive.

Im Unterschied zu Abb. @3b hat hier die Ordinatenachse eine Beschriftung, die nicht von den Parametern abhängt; denn nach der obigen Regel muß die kumulative Wahrscheinlichkeit für gleiche z -Werte (also gleiche Ablesepunkte auf der Abszissenachse) in allen Normalverteilungen gleich sein. Die eingezeichneten Hilfslinien ermöglichen die folgenden Feststellungen:

- In jeder Normalverteilung ist der Median (also der x -Wert mit kumulativer Wahrscheinlichkeit 0.5) gleich dem Erwartungswert μ . (Das ergab sich auch schon aus Abb. @3b: Da die Dichtekurve symmetrisch ist, wird die Fläche unter dieser Kurve von der durchgezogenen senkrechten Linie halbiert.)
- In Abschnitt @B.IV.1.a wurde als graphischer Aspekt der Differenzenregel für Intervallwahrscheinlichkeiten festgestellt, daß bei einer kontinuierlichen Wahrscheinlichkeitsverteilung der Höhenunterschied der Ogive zwischen zwei Skalenpunkten gleich der Intervallwahrscheinlichkeit für das von diesen Skalenpunkten begrenzten Intervalls ist. Die gestrichelten Hilfslinien ermöglichen es nun, von $\mu - \sigma$ bis $\mu + \sigma$ eine Zunahme der kumulativen Wahrscheinlichkeit von ca. 0.16 auf ca. 0.84 festzustellen⁶⁵; der Höhenunterschied und damit die Intervallwahrscheinlichkeit für das entsprechend Intervall beträgt also etwa $0.84 - 0.16 = 0.68$

Schon in @Statistik I wurde ein Näherungswert für die Intervallwahrscheinlichkeit des Intervalls von $\mu - \sigma$ bis $\mu + \sigma$ angegeben. Erinnern Sie sich?

Es wurde festgestellt, daß bei Häufigkeitsverteilungen mit mittlerem Exzeß etwa zwei Drittel der Meßwerte um höchstens eine Standardabweichung vom Durchschnitt entfernt liegen, und das

⁶⁵Etwas genauere Zahlen werden in Abschnitt @B.IV.3.b ermittelt.

bedeutet ja (auf Normalverteilungen übertragen) von $\mu - \sigma$ bis $\mu + \sigma$. Die Zahl 0.68 ist exakter und für Normalverteilungen zutreffend; aber die gröbere Angabe "etwa zwei Drittel" ist bei Häufigkeitsverteilungen angebrachter, da diese natürlich nur selten so exakt einer Normalverteilung entsprechen, daß eine Angabe mit zwei Nachkommastellen gerechtfertigt wäre.

Die Bestimmung von Intervallwahrscheinlichkeiten anhand der Ogive ist natürlich recht ungenau. Für genauere Zahlen kann man aber ebenfalls die Tatsache ausnutzen, daß für Intervalle mit gleichen z -Werten der Intervallgrenzen die entsprechenden Intervallwahrscheinlichkeiten bei allen Normalverteilungen gleich sind: Hat man für eine Normalverteilung die kumulativen Wahrscheinlichkeiten (und damit natürlich auch die Intervallwahrscheinlichkeiten), dann hat man sie auch für alle übrigen Normalverteilungen. Als solche "Standardverteilung" ist eine Normalverteilung mit $\mu = 0$ und $\sigma = 1$ besonders geeignet, weil sich bei ihr die Umrechnung in z -Werte erübrigt; denn für jede Zahl x gilt dann $z_x = (x - 0) / 1 = x$. Diese Normalverteilung heißt deshalb die "Standardnormalverteilung". Sie ist Gegenstand des folgenden Abschnitts.

b) Die Standardnormalverteilung

Die Bedeutung der Standardnormalverteilung - also der Normalverteilung mit Erwartungswert $\mu = 0$ und Standardabweichung $\sigma = 1$ - hat dazu geführt, daß für die kumulative Verteilungsfunktion dieser Wahrscheinlichkeitsverteilung ein besonderes Symbol eingeführt wurde: Anstelle des sonst üblichen $F(x)$ schreibt man $\Phi(z)$. Die Verwendung des Buchstaben z drückt dabei aus, daß in dieser Verteilung alle x -Werte auch z -Werte sind. Die aus Abb. @3c zu entnehmende Feststellung, daß die kumulative Wahrscheinlichkeit bei $\mu + \sigma$ etwa 0.84 beträgt, läßt sich also einfach als $\Phi(1) = 0.84$ schreiben.

Es gibt keine einfache Formel, mit der sich diese kumulativen Wahrscheinlichkeiten berechnen lassen. Daher gibt es in fast allen Statistikbüchern Tabellen, aus denen sich die kumulativen Wahrscheinlichkeiten der Standardnormalverteilung direkt ablesen oder aber mit Hilfe der folgenden Eigenschaften herleiten lassen:

Grundeigenschaften der Standardnormalverteilung:

- 1. Die Dichtekurve der Standardnormalverteilung ist symmetrisch um den 0-Punkt (denn jede Dichtekurve einer Normalverteilung ist symmetrisch um ihren Erwartungswert).*
- 2. Die gesamte Fläche unter der Dichtekurve ist (wie bei jeder kontinuierlichen Wahrscheinlichkeitsverteilung) gleich 1.*

Verschiedene derartige Tabellen sind unterschiedlich angelegt. In einigen sind kumulative Wahrscheinlichkeiten direkt angegeben, in anderen Intervallwahrscheinlichkeiten oder die sog. "zweiseitigen Überschreitungswahrscheinlichkeiten", die auch hier demnächst eingeführt werden.

Wenn man bei der Behandlung der kontinuierlichen Wahrscheinlichkeitsverteilungen von der Ogive ausgegangen ist und schon eine sehr allgemeine Differenzenregel für Intervallwahrscheinlichkeiten hat, die mit kumulativen Wahrscheinlichkeiten arbeitet, liegt es nahe, mit einer Tabelle zu arbeiten, die solche kumulativen Wahrscheinlichkeiten angibt. Eine solche Tabelle ist im Tabellen-Anhang dieses Skripts als Tabelle @1 enthalten.

Ablesebeispiel: Wie groß ist in einer Standardnormalverteilung die kumulative Wahrscheinlichkeit eines x -Wertes von 1.00?

In einer Standardnormalverteilung ist jeder x -Wert bereits ein z -Wert. Für einen z -Wert von 1.00 finden wir in der Tabelle eine kumulative Wahrscheinlichkeit von $\Phi(z) = 0.841$ angegeben.⁶⁶ Auf Grund dieser Angabe können wir nun unter Verwendung unserer oben genannten Eigenschaften auch weitere Fragen beantworten, z.B. die folgende:

Wie groß ist die Wahrscheinlichkeit, daß eine standardnormalverteilte Zufallsvariable größer als 1.00 wird?

Um solche weiterführenden Fragen zu beantworten, legt man am besten eine Skizze der Standardnormalverteilung an und schraffiert die Fläche, auf die es einem ankommt. Für eine solche Skizze ist keine genaue Dichtekurve erforderlich; es genügt, eine ungefähre "Glockenkurve" mit Symmetrieachse im Punkt $\mu = 0$ und Wendepunkten bei $z = -1$ und $z = +1$ zu zeichnen, welche die Abszissenachse etwa bei $z = -3$ und $z = +3$ nahezu erreicht. In unserem Falle würde diese Skizze etwa folgendermaßen aussehen:

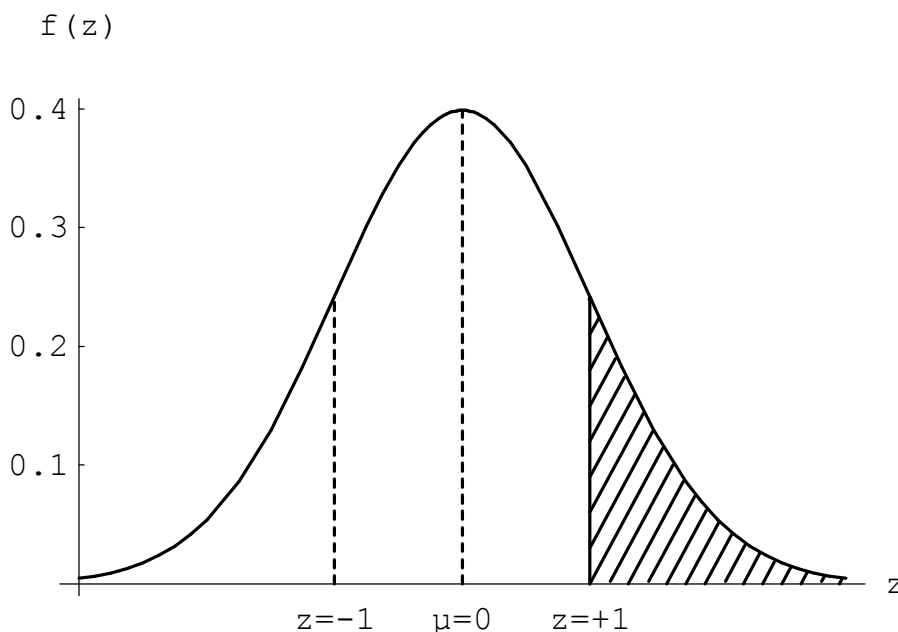


Abb. 4: Planskizze zur Bestimmung der "einseitigen Überschreitungswahrscheinlichkeit" eines z -Wertes von 1 bei Standardnormalverteilung.

⁶⁶Hier und bei den folgenden Beispielen soll eine Rechengenauigkeit von drei Stellen hinter dem Komma genügen.

Wie sich in dem vorhergehenden Ablesebeispiel zeigte, ist die kumulative Wahrscheinlichkeit eines z -Wertes von $+1$ (also der nicht-schraffierte Teil der Fläche unter der Kurve in Abb. @4) gleich 0.841 . Da sich dieser Teil der Kurve zusammen mit dem schraffierten Teil zu 1 ergänzen muß (vgl. oben Grundeigenschaft 2) können wir angeben:

Schraffierte Fläche in Abb. 4

$$= 1 - \text{nicht schraffierte Fläche in Abb. 4}$$

$$= 1 - 0.841 = 0.159$$

Die gesuchte Wahrscheinlichkeit ist also 0.159 .

Solche Wahrscheinlichkeiten, daß ein bestimmter z -Wert überschritten wird, nennt man auch "Überschreitungswahrscheinlichkeiten". Genauer unterscheidet man zwischen "einseitigen" und "zweiseitigen" Überschreitungswahrscheinlichkeiten. Die soeben berechnete Wahrscheinlichkeit ist die einseitige Überschreitungswahrscheinlichkeit eines z -Werts von 1.00 . Bei der zweiseitigen Überschreitungswahrscheinlichkeit dieses z -Werts geht es dagegen um die Wahrscheinlichkeit eines z -Werts von über 1.00 oder unter -1.00 . Diese wird später ermittelt.

Man kann das Ergebnis der obigen Berechnung als Regel formulieren: Die einseitige Überschreitungswahrscheinlichkeit jedes positiven z -Werts ist $1 - \Phi(z)$. Ähnlich könnte man aus allen Ergebnissen der folgenden Beispiele allgemeine Regeln machen. Statt sein Gedächtnis mit vielen solchen Regeln zu belasten, ist es aber günstiger, vor allem die Technik der Bestimmung von Wahrscheinlichkeiten in der Standardnormalverteilung anhand von Planskizzen unter Anwendung der obigen Grundeigenschaften zu üben. Hat man das einige Male gemacht, dann kann man es bald "im Kopf", ohne sich eine Planskizze anzulegen.

Auch die Antwort auf die folgende Frage läßt sich leicht aus bisherigen Ergebnissen und einer der beiden Grundeigenschaften herleiten:

Wie groß ist in einer Standardnormalverteilung die kumulative Wahrscheinlichkeit eines z -Wertes von -1 ?

Wenn wir uns wieder eine Planskizze anlegen würden, dann hätten wir den Teil der Fläche zu schraffieren, der links von der Linie für $z = -1$ in Abb. @4 liegt. Wir können uns nun die Tatsache zunutze machen, daß jede Standardnormalverteilung symmetrisch ist (Grundeigenschaft 1). Die Fläche links von der Linie für $z = -1$ ist demnach symmetrisch (und damit flächengleich) der schraffierten Fläche rechts von der gestrichelten Linie, deren Größe wir schon als 0.159 berechnet haben, und dies ist deshalb auch die gesuchte Wahrscheinlichkeit.

In ganz ähnlicher Weise kann man die kumulative Wahrscheinlichkeit aller negativen z -Werte in der Standardnormalverteilung angeben: Es gilt die Formel

$$\Phi(-z) = 1 - \Phi(z),$$

und das ist so ziemlich die einzige Verallgemeinerung eines Rechenergebnisses zur Standardnormalverteilung, die zu merken sich lohnt.

Auch diese Wahrscheinlichkeit würde man als "einseitige" Überschreitungswahrscheinlichkeit

(des z -Werts von -1.00) bezeichnen, und damit sollte auch noch klarer sein, was mit "Überschreitungswahrscheinlichkeiten" gemeint ist: Es geht immer um die Wahrscheinlichkeit, einen Wert zu erzielen der weiter⁶⁷ vom Erwartungswert $\mu = 0$ entfernt ist als der z -Wert, um dessen Überschreitungswahrscheinlichkeit es geht. Bei "einseitigen" Überschreitungswahrscheinlichkeiten geht es dabei nur um Abweichungen in der gleichen Richtung wie bei dem vorgegebenen z -Wert, während es bei zweiseitigen Überschreitungswahrscheinlichkeiten um Abweichungen in beiden Richtungen geht. Diese zweiseitige Überschreitungswahrscheinlichkeit kann jetzt ebenfalls bestimmt werden.

Wie groß ist die "zweiseitige Überschreitungswahrscheinlichkeit" eines z -Werts von 1.00 , also die Wahrscheinlichkeit, daß eine standardnormalverteilte Zufallsvariable kleiner als -1 oder größer als $+1$ wird?

Für eine Planskizze wären in Abb. @4 die beiden Flächen links von $z = -1$ und rechts von $z = +1$ zu schraffieren. Die Größe der gesamten schraffierten Fläche ist dann natürlich die Summe der Größen der beiden Teilflächen, die bereits als 0.159 bestimmt wurden. Also ist die gesuchte Wahrscheinlichkeit $0.159 + 0.159 = 0.318$.

Anhand der "allgemeinen Ogive" (Abb. @3c) konnte eine Entsprechung zu der Regel aus @Statistik I gefunden werden, nach der in Häufigkeitsverteilungen mit mittlerem Exzeß etwa zwei Drittel der Meßwerte um höchstens eine Standardabweichung vom Durchschnitt entfernt liegen: Die Intervallwahrscheinlichkeit für das Intervall von $\mu - \sigma$ bis $\mu + \sigma$ beträgt ca. 0.68 .

Wie kann diese Wahrscheinlichkeit anhand der bisherigen Ergebnisse genauer angegeben werden?

Die entsprechende Fläche in Abb. @4 liegt zwischen den senkrechten Linien für $z = -1$ und $z = +1$. Die Fläche außerhalb dieser Linien haben wir bereits als 0.318 bestimmt. Da die gesamte Fläche unter der Kurve 1 ist, ergibt sich die Fläche zwischen den Linien als $1 - 0.318 = 0.682$.

Die Intervallwahrscheinlichkeit für das Intervall von $z = -1$ bis $z = +1$ könnte man auch nach der Differenzenregel für Intervallwahrscheinlichkeiten bestimmen. Wie würde das gehen?

Die kumulativen Wahrscheinlichkeiten der beiden Intervallgrenzen wurden bereits als 0.159 und 0.841 bestimmt. Die Intervallwahrscheinlichkeit beträgt also $0.841 - 0.159 = 0.682$, und daß dieser Wert mit dem zuvor auf anderem Weg ermittelten übereinstimmt, ist natürlich kein Zufall.

Dieser Wert ist zwar nicht viel genauer als der aus der Ogive abgelesene Wert; aber mit Hilfe der Standardnormalverteilungs-Tabelle im Anhang könnte man diese Fläche auch mit mehr Nachkommastellen bestimmen. (Wer das Verständnis des bisherigen Gedankengangs überprüfen

⁶⁷Genau wie bei Intervallwahrscheinlichkeiten macht es auch bei Überschreitungswahrscheinlichkeiten in einer kontinuierlichen Wahrscheinlichkeitsverteilung keinen Unterschied, ob man die Grenze einbezieht oder nicht - ob es also um die Wahrscheinlichkeit von Werten geht, die "weiter" oder "mindestens so weit" vom Erwartungswert entfernt sind wie ein vorgegebener z -Wert.

oder vertiefen will, kann als Übungsaufgabe die Fläche zwischen den Linien für $z = -1$ und $z = +1$ mit 5 Nachkommastellen bestimmen.)

Nach der o.g. "Daumenregel" aus @Statistik I liegen nur ca. 5% der Meßwerte um 2 Standardabweichungen oder mehr vom Durchschnitt entfernt. Wie lautet die entsprechende Wahrscheinlichkeit in der Standardnormalverteilung?

Es geht offenbar um die zweiseitige Überschreitungswahrscheinlichkeit eines z -Werts von 2.00. Wiederholt man die bisher für $z = 1$ durchgeführten Berechnungen für $z = 2$, dann ergibt sich: Bis $z = 2.00$ liegt lt. Tabelle eine Fläche von 0.977; also liegt oberhalb von $z = 2.00$ eine Fläche von $1 - 0.977 = 0.023$. Eine ebenso große Fläche liegt auch unterhalb von $z = -2.00$; also liegt insgesamt eine Fläche von $2 \cdot 0.023 = 0.046$ um 2 Standardabweichungen oder mehr vom Erwartungswert $\mu = 0$ entfernt. Die zweiseitige Überschreitungswahrscheinlichkeit eines z -Werts von 2.00 beträgt also 0.046.

Die Angaben "etwa zwei Drittel" bzw. "ca. 5%" stimmen also ziemlich genau. Man kann dieses Ergebnis aber zum Anlaß nehmen, den z -Wert genauer zu bestimmen, dessen zweiseitige Überschreitungswahrscheinlichkeit 5% (also 0.05) beträgt. Eine Planskizze dafür ist Abb. @4a.

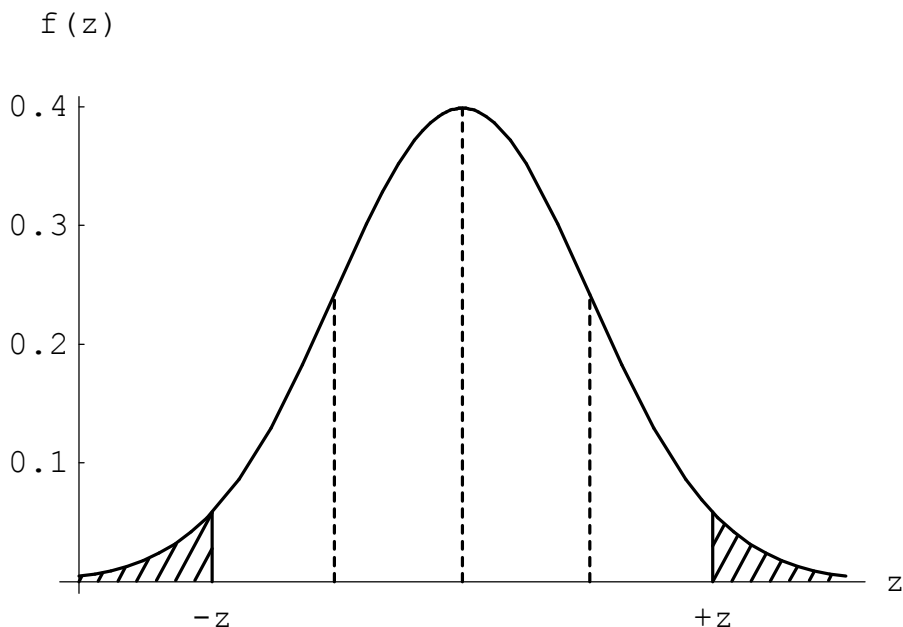


Abb. 4a: Planskizze zur Bestimmung des z -Werts mit einer zweiseitigen Überschreitungswahrscheinlichkeit von 0.05.

Anhand dieser Planskizze läßt sich der gesuchte z -Wert mit folgenden Überlegungen bestimmen:

- Die Forderung, daß die zweiseitige Überschreitungswahrscheinlichkeit des gesuchten z -Werts 5% betragen soll, bedeutet, daß die Größe der beiden schraffierten Flächen zusammen 0.05 betragen soll.
- Wegen der Symmetrie der Standardnormalverteilung müssen die beiden schraffierten Flächen gleich groß sein; also muß die Größe jeder einzelnen $0.05 / 2 = 0.025$ sein.

- Wenn die Größe der rechten schraffierten Fläche 0.025 sein soll, dann muß die Größe der Fläche bis $+z$ natürlich $1 - 0.025 = 0.975$ betragen. Gesucht ist also der z -Wert mit einer kumulativen Wahrscheinlichkeit von 0.975 .
- Dieser z -Wert ergibt sich, indem man die Standardnormalverteilung-Tabelle "rückwärts liest". Man sucht also, bei welchem z -Wert die in der Tabelle angegebene kumulative Wahrscheinlichkeit am nächsten bei 0.975 liegt. Das ist bei $z = 1.96$ der Fall.

Dieser z -Wert ist für die praktische Anwendung der Wahrscheinlichkeitstheorie in der Inferenzstatistik von einiger Bedeutung. Natürlich kann man nicht die ganze z -Tabelle "vorwärts und rückwärts" auswendig lernen, aber wenigstens zwei Zahlen sollte man kennen:

Wichtigste Zahlen zur Standardnormalverteilung:

- Zwischen $z = -1$ und $z = +1$ liegt eine Wahrscheinlichkeit von rund 68%
- Der z -Wert mit einer zweiseitigen Überschreitungswahrscheinlichkeit von 5% beträgt 1.96.

Zwei weitere z -Werte sind auch noch wichtig, so daß es nützlich ist, auch diese zu kennen:

- Der z -Wert mit einer zweiseitigen Überschreitungswahrscheinlichkeit von 1% beträgt 2.58.
- Der z -Wert mit einer einseitigen Überschreitungswahrscheinlichkeit von 5% ist 1.64.

(Wollen Sie beide Angaben als Übungsaufgabe überprüfen?)

Als letzte Aufgabe soll die Wahrscheinlichkeit eines z -Wertes bestimmt werden, der größer als -0.6 , aber nicht größer als 0.3 ist. Das kann man mit Hilfe der Differenzenregel für Intervallwahrscheinlichkeiten oder (übungshalber) mit einer Planskizze zur Standardnormalverteilung.

Zur Anwendung der Differenzenregel für Intervallwahrscheinlichkeiten braucht man die beiden kumulativen Wahrscheinlichkeiten. Diejenige für $z = 0.3$ kann unmittelbar der Standardnormalverteilungstabelle entnommen werden: Sie beträgt 0.618 . Für die kumulative Wahrscheinlichkeit von $z = -0.6$ kann man auf die Regel $\Phi(-z) = 1 - \Phi(z)$ zurückgreifen, und da aus der Standardnormalverteilungstabelle für $z = 0.6$ eine kumulative Wahrscheinlichkeit von 0.726 abzulesen ist, ergibt sich

$$\Phi(-0.6) = 1 - \Phi(0.6) = 1 - 0.726 = 0.274.$$

Nach der Differenzenregel für Intervallwahrscheinlichkeiten ergibt sich damit die gesuchte Wahrscheinlichkeit als $0.618 - 0.274 = 0.344$.

Eine entsprechende Planskizze ist Abb. @5.

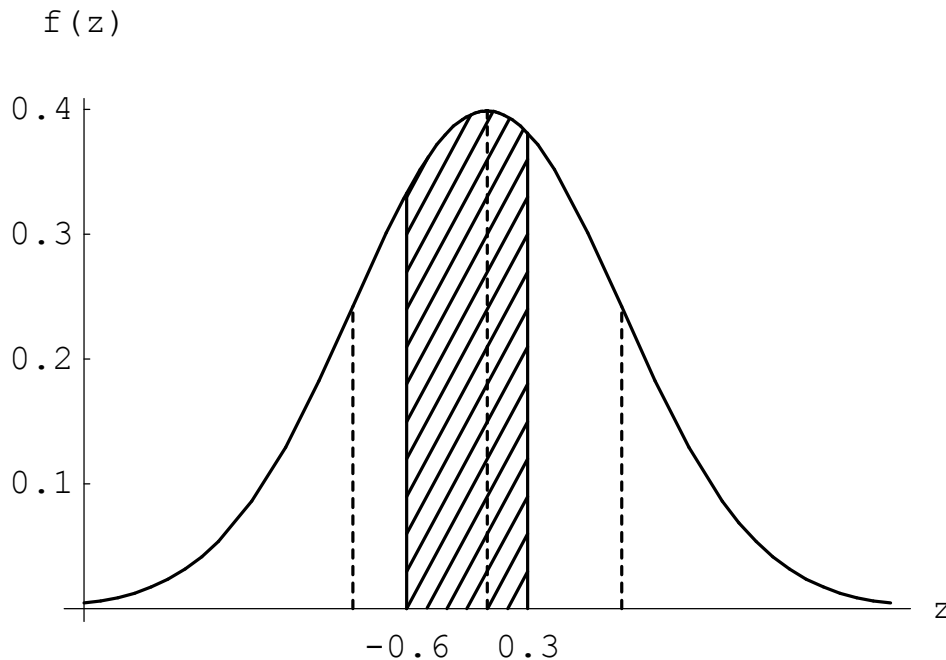


Abb. 5: Planskizze zur Bestimmung der Intervallwahrscheinlichkeit von $z = -0.6$ bis $z = 0.3$.

Die Fläche links von $z = -0.6$ muß zusammen mit der schraffierten Fläche die Fläche bis $z = 0.3$ ergeben. Anders ausgedrückt: Die schraffierte Fläche ergibt sich, indem wir die Fläche bis $z = -0.6$ von der Fläche bis $z = 0.3$ subtrahieren. Das ist aber genau das, was bereits aufgrund der Differenzenregel gerechnet wurde. Die Planskizze zeigt den Bezug zwischen dem im hiesigen Abschnitt dargestellten Verfahren zur Bestimmung von Wahrscheinlichkeiten in der Standardnormalverteilung und der Differenzenregel für Intervallwahrscheinlichkeiten.

c) Weitere Normalverteilungs-Skalen

Zu vielen psychologischen Tests gibt es im zugehörigen Testhandbuch Tabellen, mit denen man aus "vorläufigen Testwerten" (z.B. aus der Zahl richtig gelöster Aufgaben) einen Wert ablesen kann, der in einer "Population" normalverteilt ist, für welche der Test gedacht ist. Wenn aber der Erwartungswert und die Standardabweichung dieser Normalverteilung bei jedem Test verschieden wären, müßte man diese Kennziffern für jeden Test kennen, um Ergebnisse interpretieren zu können. Um das zu vereinfachen, gibt es einige wenige "Normalverteilungs-Skalen", auf denen die Testwerte angegeben werden können. Die älteste solche Skala ist die IQ-Skala mit einem Erwartungswert von $\mu = 100$ und einer Standardabweichung von $\sigma = 15$. Wenn man dann als Testergebnis z.B. einen IQ von 130 mitgeteilt bekommt, kann man diesen Wert interpretieren.

Aus verschiedenen Erwägungen heraus wurden weitere Normalverteilungs-Skalen eingeführt, die in der folgenden Tabelle zusammengefaßt sind.

Name der Normal- verteilungs-Skala	μ	σ
z-Skala	0	1
IQ-Skala	100	15
SW-Skala	100	10
T-Skala	50	10
C-Skala	5	2

Mit Hilfe dieser Angaben kann man für jeden Wert auf einer der genannten Skalen den zugehörigen z-Wert berechnen.

Beispiel: Der o.g. IQ-Wert von 130 entspricht einem z-Wert von 2.0; ein SW von 85 entspricht einem z-Wert von -1.5 usw.. Mit Hilfe dieser Transformation in z-Werte kann man auch die verschiedenen Skalen miteinander vergleichen. (Beispiel: Ein IQ-Wert von 85 entspricht einem T-Wert von 40, denn beide entsprechen einem z-Wert von -1).

Man kann nun die kumulativen Wahrscheinlichkeiten, die Überschreitungswahrscheinlichkeiten und die Wahrscheinlichkeiten von Intervallen auf allen Normalverteilungsskalen angeben, indem man die Skalenwerte zunächst in z-Werte transformiert und dann aufgrund der Standardnormalverteilungstabelle die entsprechenden Wahrscheinlichkeiten berechnet.

Beispiel: Wie groß ist die kumulative Wahrscheinlichkeit eines IQ-Wertes von 95.500...?

Wir transformieren in einen z-Wert:

$$z = \frac{95.5 - 100}{15} = -0.3$$

Die Wahrscheinlichkeit (d.h. die relative Populationshäufigkeit) eines IQ-Wertes von höchstens 95.5000 ist dann gleich der Wahrscheinlichkeit eines z-Wertes von höchstens -0.3:

$$\begin{aligned} p(\text{IQ} \leq 95.500\dots) &= p(z < -0.3) = \Phi(-0.3) = 1 - \Phi(0.3) \\ &= 1 - 0.618 = 0.382 \end{aligned}$$

Bei der Anwendung dieses Verfahrens ist jedoch zu berücksichtigen, daß die oben genannten Skalen meistens als künstlich diskrete Skalen verwendet werden: Kommastellen werden auf- oder abgerundet. Konsequenz: Bevor man die Skalenwerte in z-Werte übersetzt, muß man die auf der Basis der künstlich diskreten Skala formulierte Frage in eine kontinuierliche Skala übersetzen.

Beispiel: Wie groß ist die Wahrscheinlichkeit eines Standardwerts (SW) von mehr als 92 und weniger als 103?

Da bei der Rundung alle Standardwerte von 91.5 bis 92.5 zu der Klasse 92 geschlagen werden, beginnt das hier (auf der künstlich diskreten Skala angegebene!) Intervall auf der kontinuierlichen Skala erst bei 92.5; aus entsprechenden Gründen endet es schon bei 102.5. Es ergibt sich dann:

$$p(92.500 \leq \text{SW} \leq 102.5) = p(-0.75 \leq z < 0.25) = \Phi(0.25) - \Phi(-0.75) = 0.372.$$

Ähnlich könnte man sich fragen, wie groß die Wahrscheinlichkeit (relative Populationshäufigkeit)

eines IQ-Wertes von 100 ist. Das ist die Wahrscheinlichkeit des Intervalls von 99.5 bis 100.5; wir können also angeben:

$$p(99.5 \leq IQ \leq 100.5) = p(-0.0333 \leq z < 0.0333) = \Phi(0.0333) - \Phi(-0.0333) = 0.024$$

Nur 2.4% der Gesamtbevölkerung haben also einen IQ von 100. Noch weiter geht die Rundung bei der sogenannten Stanine-Skala: Es handelt sich grundsätzlich um eine C-Skala, bei der jedoch alle Werte, die kleiner als 1 sind, zur 1 und alle Werte, die größer als 9 sind, zur 9 geschlagen werden. Die Skala hat also insgesamt nur 9 Klassen.⁶⁸

Wie groß ist demnach die Wahrscheinlichkeit eines Stanine-Wertes von 1?

Umgewandelt in eine kontinuierliche Skala: Wie groß ist die Wahrscheinlichkeit eines Stanine-Wertes von weniger als 1.5? Der Wert 1.5 entspricht bei einem $\mu = 5$ und einem $\sigma = 2$ ein $z = -1.75$; dessen kumulative Wahrscheinlichkeit wir nach der üblichen Methode für negative z -Werte angeben:

$$\Phi(-1.75) = 1 - \Phi(1.75) = 0.040.$$

Die "Korrektur", die man vornimmt, um eine auf einer künstlich diskreten Skala (also mit gerundeten Skalenwerten) formulierte Fragestellung in ein kontinuierliche Skala zu übersetzen, nennt man auch die "*Kontinuitätskorrektur*". Sie wird uns auch noch in anderen Zusammenhängen begegnen.

<<

Das Verfahren, mit dem man eine Tabelle zur Umrechnung "vorläufiger Testwerte" (z.B. Zahl "richtiger Lösungen") in Werte auf einer der o.g. Skalen erstellt, ist übrigens recht einfach: Man bestimmt die (interpolierten) Prozentränge aller vorläufigen Testwerte, ermittelt aufgrund der Standardnormalverteilungs-Tabelle die entsprechenden z -Werte und rechnet diese dann nach der Formel $x = \mu + \sigma \cdot z$ in Werte auf der angestrebten Skala um.

>>

Bereits in @Statistik I wurde festgestellt, daß aus der Konstanz der Verteilungsform bei Lineartransformationen und ihrer Änderung bei monotonen, nichtlinearen Skalentransformationen auch folgt, daß bei Normalverteilung des gemessenen Merkmals eine Ordinalskala genau dann eine Intervallskala ist, wenn die Verteilung der Meßwerte ebenfalls eine Normalverteilung ist, und daß darauf die Praxis beruht, Testwerte auf einer Normalverteilungs-Skala anzugeben.

d) Der Zentrale Grenzwertsatz

Der "Zentrale Grenzwertsatz" behandelt eine Eigenschaft, die sich in unseren weiteren Überlegungen als sehr wichtig herausstellen wird. Eine mathematisch präzise Formulierung dieses Satzes fällt selbst professionellen Mathematikern schwer. Es geht darum, daß die Wahrscheinlich-

⁶⁸Da auch diese Skala als eine Art Standard-Skala mit 9 Klassen konzipiert ist, erklärt sich der Name "Stanine-Skala" als Kontraktion von "standard-nine"-Skala.

keitsverteilung von Größen, die sich als Summe "vieler" einzelner Zufallsgrößen darstellen lassen, oft eine Normalverteilung ist. Mathematisch etwas präziser ist die folgende Formulierung:

Bilden wir eine Summe von vielen stochastisch unabhängigen Zufallsgrößen, deren Wahrscheinlichkeitsverteilungen alle die gleiche Varianz haben, so ist die Wahrscheinlichkeitsverteilung dieser Summe approximativ eine Normalverteilung.

Diese Approximation gilt mit umso größerer Genauigkeit, je größer die Zahl der Summanden und je kleiner die Abweichungen der Einzelverteilungen von der Normalverteilung sind.

Für den Beweis dieses Satzes benötigt man noch einige weitere Voraussetzungen, die hier nicht diskutiert werden, weil sie bei in der Psychologie verwendeten Wahrscheinlichkeitsverteilungen praktisch immer erfüllt sind.

Wichtig ist, daß es nicht erforderlich ist, daß die summierten zufälligen Größen dieselben Verteilungen haben müssen. Es genügt, wenn sie alle die gleiche Varianz haben - und selbst diese Voraussetzung läßt sich noch abschwächen. Der Name "Grenzwertsatz" läßt sich damit erklären, daß die exakte Normalverteilung gewissermaßen als Grenzfall einer solchen Summenverteilung mit unendlich vielen Summanden entsteht.

Wichtiger Anwendungsfall: Die zufällige Größe X in der *Binomialverteilung* kann man als Summe von n stochastisch unabhängigen Größen X_1, X_2, \dots, X_n ansehen, die angeben, ob im 1., 2. usw. Versuch ein "kritisches Ereignis" auftritt (zufällige Größe = 1) oder nicht (zufällige Größe = 0). Solche zufälligen Größen, die in dieser Weise mit Zahlenwerten von 1 bzw. 0 angeben, ob ein Ereignis eintritt oder nicht, nennt man "Indikatorvariablen". Offenbar gilt dann für die Zufallsvariable X der Binomialverteilung:

$$X = \sum_{j=1}^n X_j$$

Da diese zufälligen Größen stochastisch unabhängig sind und alle die gleiche Varianz haben, ist die Wahrscheinlichkeitsverteilung der Summe (also des X der Binomialverteilung) approximativ eine Normalverteilung, wenn n nicht zu klein und p nicht zu weit von 0.5 entfernt ist. Wann diese Bedingungen erfüllt sind, ist z.T. eine Ermessensfrage; es hängt davon ab, welchen Genauigkeitsgrad man für eine solche Approximation verlangt. Verbreitet ist die folgende

Faustregel: Die Binomialverteilung läßt sich mit einer Normalverteilung approximieren, wenn die Produkte $n \cdot p$ und $n \cdot q$ beide größer als 10 sind. μ und σ werden nach den entsprechenden Formeln (S. 73) bestimmt.

Man muß dabei allerdings berücksichtigen, daß die Binomialverteilung eine diskrete Verteilung

ist, die daher nur mit einer "künstlich diskreten Normalverteilung" approximiert werden kann, wie wir es im vorigen Abschnitt bei den Normalverteilungsskalen kennenlernten ("Kontinuitätskorrektur"). Man muß also auch hier jede auf die diskrete Skala bezogene Frage nach Wahrscheinlichkeiten zunächst in eine Frage für eine kontinuierliche Skala umformulieren.

Beispiel: Wir würfeln mit einem ungefälschten Würfel 180 mal. Wie groß ist dabei die Wahrscheinlichkeit, weniger als 25 Sechser zu würfeln?

Die zufällige Größe X (= Zahl der Sechser) hat offenbar eine Binomialverteilung mit $n=180$, $p=1/6$ und $q=5/6$. Da die Produkte $n \cdot p = 30$ und $n \cdot q = 150$ beide größer als 10 sind, können wir die Binomialverteilung mit einer Normalverteilung approximieren. Erwartungswert und Standardabweichung dieser Normalverteilung können wir mit den Formeln der Binomialverteilung bestimmen:

$$\mu_x = n \cdot p = 180 \cdot \frac{1}{6} = 30$$

$$\sigma_x = \sqrt{n \cdot p \cdot q} = \sqrt{180 \cdot \frac{1}{6} \cdot \frac{5}{6}} = 5.00$$

Um einen z-Wert ausrechnen zu können, müssen wir zunächst unsere Frage für eine kontinuierliche Skala umformulieren. Sie lautet dann: Wie groß ist die Wahrscheinlichkeit, daß eine mit Erwartungswert 30 und Standardabweichung 5.00 kontinuierlich normalverteilte zufällige Größe kleiner als 24.5 wird. Jetzt können wir den z-Wert bestimmen:

$$z = \frac{24.5 - 30}{5.00} = -1.10$$

Die Wahrscheinlichkeit eines z-Wertes von weniger als -1.10 erhalten wir als

$$\Phi(-1.10) = 1 - \Phi(1.10) = 1 - 0.864 = 0.136$$

Die gesuchte Wahrscheinlichkeit beträgt also 0.136.

e) Die Interpretation von Verteilungen aufgrund des Zentralen Grenzwertsatzes

Wenn wir in der Natur häufig Merkmale vorfinden, deren Verteilung in etwa eine Normalverteilung ist, dann *kann* das daran liegen, daß an der Bildung dieses Merkmals viele Faktoren (viele Gene und/oder viele Umwelteinflüsse) stochastisch unabhängig additiv zusammenwirken. Es gibt z.B. Psychologen, die aus derartigen Überlegungen heraus eine Normalverteilung der Intelligenz postulieren. Es ist allerdings zu fragen, inwieweit die Voraussetzungen des Zentralen Grenzwertsatzes (Additivität, Unabhängigkeit und gleiche Varianz der Faktoren) bei der Intelligenz zutreffen.

Daneben kann man auch Abweichungen von der Normalverteilung auf dem Hintergrund des Zentralen Grenzwertsatzes interpretieren, indem man aufzeigt, welche Verletzungen der Voraussetzungen des Zentralen Grenzwertsatzes für die aufgefundenen Abweichungen verantwortlich sein können.

Beispiel 1: Führt man einen Test, der Maskulinität (hohe X-Werte) bzw. Femininität (niedrige X-Werte) messen soll, an einer aus Männern und Frauen gemischten Stichprobe durch, dann ergibt sich verständlicherweise eine bimodale Verteilung. Man könnte dies auf zwei Weisen interpretieren: Entweder ist das biologische Geschlecht für den größten Teil der Varianz der X-Werte direkt verantwortlich, dann läge eine Verletzung der beim Zentralen Grenzwertsatz erforderlichen Voraussetzung gleicher Varianzbeiträge von vielen Faktoren vor. Es könnte aber auch sein, daß viele Umweltfaktoren (z.B. Rollenerwartungen der Gesellschaft) an der Hervorbringung der vom Test gemessenen Maskulinität und Femininität ebenso stark beteiligt sind wie das biologische Geschlecht. Da diese Faktoren aber offenbar in Abhängigkeit vom biologischen Geschlecht vorkommen werden, läge eine Verletzung der beim Zentralen Grenzwertsatz wichtigen Voraussetzung der stochastischen Unabhängigkeit der Faktoren vor.

Beispiel 2: Die Reaktionszeit eines Menschen auf bestimmte Aufgaben ist eine zufällige Größe, deren Wahrscheinlichkeitsverteilung eine starke Linksschiefe aufweist. Transformiert man jedoch die Reaktionszeiten in Logarithmen, dann haben diese zumindest näherungsweise eine Normalverteilung.

Interpretation: Es könnte sein, daß an dieser Wahrscheinlichkeitsverteilung der Lösungszeiten viele unabhängige Faktoren beteiligt sind, jedoch nicht additiv sondern multiplikativ. Das würde nämlich bedeuten, daß dieselben Faktoren nach logarithmischer Transformation additiv den Logarithmus der Lösungszeit ergeben. Die Voraussetzungen des Zentralen Grenzwertsatzes wären also für die Logarithmen der Lösungszeit besser erfüllt als für die Lösungszeiten selbst, und daher wäre es verständlich, wenn nicht die Lösungszeiten sondern ihre Logarithmen normalverteilt wären.

Zusammengefaßt: Sowohl das Vorliegen wie auch das Nichtvorliegen einer Normalverteilung eines Merkmals kann man auf dem Hintergrund des Zentralen Grenzwertsatzes sinnvoll interpretieren.

Natürlich können solche Interpretationen nur den Sinn vorsichtiger Hypothesen haben; denn für Abweichungen von der Normalverteilung gibt es in der Regel sehr viele mögliche Interpretationen.

f) Das Verhalten einer Normalverteilung bei Skalentransformationen

Wir haben in der beschreibenden Statistik gesehen, daß die Kennziffern für die Form einer Verteilung sich nicht ändern, wenn wir die X-Werte einer Lineartransformation unterwerfen. Höchstens das Vorzeichen der Schiefe ändert sich (nämlich bei linearer Umpolung). Da aber Normalverteilungen symmetrisch sind und daher eine Schiefe von 0 haben, wäre auch eine solche Vorzeichenänderung wirkungslos. Andererseits ändert sich die Form bei monotonen, aber

nichtlinearen Skalentransformationen. Entsprechend gilt der

Lehrsatz: Hat eine Zufallsvariable X eine Normalverteilung, dann hat auch jede Lineartransformation dieser Zufallsvariablen eine Normalverteilung. Umgekehrt geht die Normalverteilung bei monotonen, aber nichtlinearen Skalentransformationen verloren.

4) Mehrdimensionale kontinuierliche Verteilungen

Auch für mehrdimensionale kontinuierliche Wahrscheinlichkeitsverteilungen eignen sich Antwortzeiten als Demonstrationsbeispiel, und auch der Einstieg über eine kumulative Wahrscheinlichkeitsverteilung ist wieder am einfachsten: Sind die Zufallsvariablen X und Y die Antwortzeiten einer Person bei zwei Aufgaben, dann ist die kumulative Verteilungsfunktion $F(x, y)$ zu interpretieren als

$$F(x, y) := p(X \leq x \wedge Y \leq y).$$

Wenn es außerdem noch eine Zufallsvariable Z die Antwortzeit in einer dritten Aufgabe angibt, dann ist $F(x, y, z)$ entsprechend zu interpretieren. Die folgende Darstellung beschränkt sich aber überwiegend auf zweidimensionale Verteilungen.

Für mehrdimensionale Intervallwahrscheinlichkeiten (z.B. für die Wahrscheinlichkeit, daß die Antwortzeit X in das Intervall von 3 bis 4 Sekunden fällt und gleichzeitig Y in das Intervall von 2 bis 5 Sekunden) gibt es wieder Differenzenregeln, die allerdings etwas komplexer sind als bei eindimensionalen kontinuierlichen Wahrscheinlichkeitsverteilungen.

<<

Der Grund für diesen Unterschied besteht darin, daß man bei den eindimensionalen Verteilungen das Ereignis $X \leq x_{IE}$ als Oder-Verknüpfung der einander ausschließenden Ereignisse $X \leq x_{IA}$ und $x_{IA} < X \leq x_{IE}$ auffassen und dann den einfachen Additionssatz anwenden konnte. Haben wir dagegen Intervallgrenzen x_{IA} und x_{IE} für X sowie y_{IA} und y_{IE} für Y , dann muß man die verschiedenen Kombinationen der auf X bezogenen Ereignisse $X \leq x_{IA}$ und $x_{IA} < X \leq x_{IE}$ sowie der auf Y bezogenen Ereignisse $Y \leq y_{IA}$ und $y_{IA} < Y \leq y_{IE}$ betrachten und kommt zu dem Ergebnis

$$p(x_{IA} < X \leq x_{IE} \wedge y_{IA} < Y \leq y_{IE}) = F(x_{IE}, y_{IE}) + F(x_{IA}, y_{IA}) - F(x_{IA}, y_{IE}) - F(x_{IE}, y_{IA}).$$

Wollen Sie versuchen, diese Gleichung selbst herzuleiten?

>>

Im folgenden sollen aber die bei den eindimensionalen kontinuierlichen Wahrscheinlichkeitsverteilungen im Detail demonstrierten Überlegungen nicht genau so ausführlich wiederholt werden; es genügt, die Grundzüge einer sinngemäßen Übertragung darzustellen.

a) Der Begriff der gemeinsamen Dichte

Im Sinne einer anschaulichen, wenn auch mathematisch etwas unpräzisen Umschreibung konnte festgestellt werden, daß die Wahrscheinlichkeitsdichte $f(x)$ bei eindimensionalen kontinuierlichen Wahrscheinlichkeitsverteilungen angibt, "wie dicht die Wahrscheinlichkeit in der unmittelbaren Umgebung von x gepackt ist". Um diese Aussage sinngemäß auf zweidimensionale kontinuierliche Wahrscheinlichkeitsverteilungen zu übertragen, sollte man sich zunächst klar machen: Was bei eindimensionalen Wahrscheinlichkeitsverteilungen die Abszissenachse ist, wird bei zweidimensionalen Wahrscheinlichkeitsverteilungen durch eine x - y -Ebene dargestellt, die etwa mit dem Koordinatensystem vergleichbar ist, in das man Meßwertpaare in einem "Streudiagramm" als "Punktwolke" einträgt. Die wesentliche Information eines solchen Diagramms besteht darin, daß in verschiedenen Bereichen der x - y -Ebene die Punkte unterschiedlich "dicht gepackt" sind. Ähnlich kann man auch sagen, daß bei zweidimensionalen kontinuierlichen Wahrscheinlichkeitsverteilungen die Wahrscheinlichkeit in unterschiedlichen Bereichen der x - y -Ebene unterschiedlich dicht gepackt ist. In einem solchen Sinn gibt die "gemeinsame Dichte" $f(x,y)$ an, wie dicht die Wahrscheinlichkeit in der Umgebung eines Punktes der x - y -Ebene mit den Koordinaten x und y gepackt ist.

Der Dichtekurve als graphischer Darstellung entspricht dann eine gewölbte Fläche, die sich über einer waagerechten x - y -Ebene erhebt. Ihre Höhe über einem Punkt der x - y -Ebene gibt dann die Wahrscheinlichkeit in diesem Punkt an. Da das Wort "Kurve" für Linien reserviert ist, kann man von einer "Dichtefläche" sprechen. Eine Darstellung eines solchen dreidimensionalen Gebildes auf zweidimensionalem Papier ist natürlich nur in Grenzen möglich, aber die folgende Abb. @5a deutet zumindest an, worum es dabei geht. Man kann wieder an Antwortzeiten in Sekunden denken.

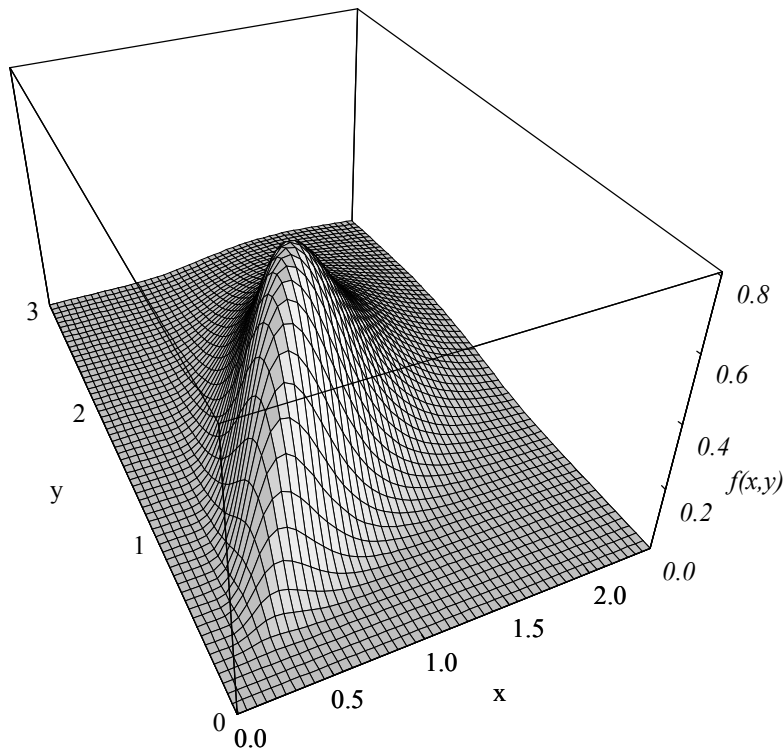


Abb. 5a: Dichtefläche einer zweidimensionalen kontinuierlichen Wahrscheinlichkeitsverteilung von Antwortzeiten.

Der Eindruck eines Berges, der morgens schon von der Sonne angestrahlt wird, während das Tal noch im Dämmerlicht liegt, unterstützt die dreidimensionale Vorstellung, und es ist wohl auch nicht schwer, zu erraten, wo die Sonne steht. (Bei moderner Computergrafik zur Darstellung solcher Gebilde kann man derartige Effekte einprogrammieren.) Aber es ist hier natürlich - nicht zuletzt wegen perspektivischer Verzerrung - viel schwieriger als bei einer Dichtekurve, irgendwelche Werte abzulesen. Trotzdem dürfte aber zu erkennen sein, daß es sich um eine Art "Höhenrücken" handelt, der sich von der vorderen zur hinteren Ecke erstreckt (und nicht von der linken zur rechten). Um von dieser bildhaften Betrachtungsweise zur mathematischen Bedeutung zurückzufinden, kann man sich (unter Beachtung der Achsenbeschriftungen) fragen:

Was bedeutet dieser Verlauf des "Höhenrückens"?

Offenbar liegt eine recht enge positive Korrelation zwischen den zwei Antwortzeiten vor.

Zur Wahrscheinlichkeitsdichte in eindimensionalen kontinuierlichen Wahrscheinlichkeitsverteilungen gab es zwei Zugänge: Den graphischen, bei dem aus Histogrammen für immer feinere künstlich diskrete Skalen als Grenzfall eine Dichtekurve wurde, und den algebraischen, bei dem sich die Dichte als Grenzwert des Quotienten aus Intervallwahrscheinlichkeit und Intervallbreite für immer feinere Intervallbreiten ergab. Ähnliche Zugänge sind auch für zweidimensionale

kontinuierliche Wahrscheinlichkeitsverteilungen möglich.

Für den graphischen Weg ist zunächst zu klären, wie ein Histogramm für eine zweidimensionale Häufigkeitsverteilung aussehen würde. Es wäre ein dreidimensionaler Körper. An die Stelle der Abszissenachse würde (wie in Abb. @5a) eine waagerechte Ebene treten, in der jeder zweidimensionalen Klasse ein rechteckiges Feld entspricht. Auf jedem solchen Feld steht dann eine rechteckige Säule, deren Höhe die Besetzungshäufigkeit der entsprechenden zweidimensionalen Klasse angibt. Derartige Histogramme lassen sich natürlich auch für diskrete Wahrscheinlichkeitsverteilungen erstellen. Wenn man nun beispielsweise die beiden Antwortzeiten X und Y zunächst auf volle Sekunden abrundet, dann auf Zehntelsekunden, Hundertstelsekunden usw. und jeweils ein dreidimensionales Histogramm zeichnet⁶⁹, dann entsteht als Grenzfall eine Dichtefläche wie die in Abb. @5a.

Es dürfte kaum noch überraschen, daß es auch zur Flächenregel für Intervallwahrscheinlichkeiten eine Entsprechung gibt. Natürlich übernimmt jetzt das Volumen zwischen der Dichtefläche und der x - y -Ebene die Rolle, die vorher die Fläche zwischen der Dichtekurve und der Abszissenachse hatte. Gibt man dieses Volumen so an, daß das Gesamtvolumen unter der Dichtefläche I ist, dann ist derjenige Anteil dieses Volumens, der über einem bestimmten Ausschnitt der x - y -Ebene (z.B. einem Rechteck oder auch einem Dreieck oder einer Kreisfläche) liegt, gleich der Wahrscheinlichkeit eines Wertepaars in diesem Ausschnitt.

Für den algebraischen Zugang kann man zunächst wieder von der Beziehung zu einem Streudiagramm ausgehen. Dort kann man angeben, wie viele Punkte auf eine bestimmte Fläche fallen, und entsprechend kann man in einer kontinuierlichen 2-dimensionalen Wahrscheinlichkeitsverteilung angeben, mit welcher Wahrscheinlichkeit ein zufälliges Meßwertpaar (X, Y) (das z.B. aus zwei Antwortzeiten besteht) in eine bestimmte Teilfläche der x - y -Ebene fällt. Läßt man diese Teilfläche nun immer kleiner werden und dividiert jeweils die verbleibende Wahrscheinlichkeit eines X - Y -Meßwertpaars in der Teilfläche durch die Größe der Teilfläche, so entsteht im "Grenzfall" (Teilfläche zu einem Punkt der x - y -Ebene zusammengeschrumpft) das, was man die "gemeinsame Dichte der zufälligen Größen X und Y im diesem Punkt der x - y -Ebene" nennt. Entsprechend könnte man auch in Analogie zu einer 3-dimensionalen Punktwolke die gemeinsame Dichte von 3 zufälligen Variablen X , Y und Z definieren. Mit mathematischer Abstraktion sind ähnliche Definitionen auch noch bei mehr als 3 zufälligen Größen mit kontinuierlicher Verteilung möglich.

Beschränken wir uns wieder auf zwei Zufallsvariablen X und Y , dann können wir für die gemeinsame Dichte $f(x,y)$ wieder eine ganz ähnliche Formel angeben wie bei eindimensionalen kontinuierlichen Wahrscheinlichkeitsverteilungen:

⁶⁹Natürlich müßte man - ähnlich wie in Abb. @2b - die Wahrscheinlichkeiten auf der senkrechten Achse jeweils durch die "Intervallbreite" dividieren, um zu vergleichbaren Maßstäben zu kommen. Intervallbreite wäre hier die Fläche des Rechtecks, auf der jede Säule steht.

$$f(x,y) := \lim_{d \rightarrow 0, d > 0} \frac{p(x \leq X \leq x+d \wedge y \leq Y \leq y+d)}{d^2}$$

<<

Falls es jemanden interessiert, welche Dichtefunktion der Abb. @5a zugrundeliegt: Aus der Annahme, daß die natürlichen Logarithmen der Antwortzeiten X und Y eine bivariate Normalverteilung mit den Parametern $\mu_x = 0.2$, $\sigma_x = 0.2$, $\mu_y = 0.4$, $\sigma_y = 0.5$ und $\rho_{xy} = 0.7$ haben, ergibt sich eine etwas komplexe Dichtefunktion, für die man am besten zunächst eine Hilfsfunktion einführt (in welcher \ln für "natürlicher Logarithmus" steht):

$$h(x,y) := \frac{\left(\frac{\ln(x)-\mu_x}{\sigma_x}\right)^2 + \left(\frac{\ln(y)-\mu_y}{\sigma_y}\right)^2 - 2\rho_{xy} \cdot \frac{\ln(x)-\mu_x}{\sigma_x} \cdot \frac{\ln(y)-\mu_y}{\sigma_y}}{2 \cdot (1 - \rho_{xy}^2)}$$

Mit dieser Hilfsfunktion läßt sich dann die der Abb. @5a zugrundeliegende Dichtefunktion angeben als

$$f(x,y) = \frac{1}{2\pi \cdot \sqrt{1 - \rho_{xy}^2} \cdot \sigma_x \cdot \sigma_y \cdot x \cdot y} \cdot e^{-h(x,y)}$$

>>

b) Randverteilung und bedingte Verteilungen

Genauso wie aus einer diskreten kann man auch aus einer kontinuierlichen 2-dimensionalen Verteilung zwei Randverteilungen herausziehen. Wie bei Häufigkeitsverteilungen und bei diskreten Wahrscheinlichkeitsverteilungen stellen diese Randverteilungen die Wahrscheinlichkeitsverteilungen von X bzw. Y unter Zusammenfassung aller Werte der jeweils anderen Variablen dar.⁷⁰ In der folgenden Abb. @5b sind Dichtekurven für die Randverteilungen von X und Y dargestellt.

⁷⁰Für Spezialisten: Es ist fast selbstverständlich, daß bei der Bildung der Randverteilung von X an die Stelle der Summation über alle y -Werte ein Integral tritt. Bezeichnet man die gemeinsame Dichte im Punkt (x,y) als $f(x,y)$, dann läßt sich die Randverteilung von X angeben als:

$$f(x) = \int_{-\infty}^{+\infty} f(x,y) dy$$

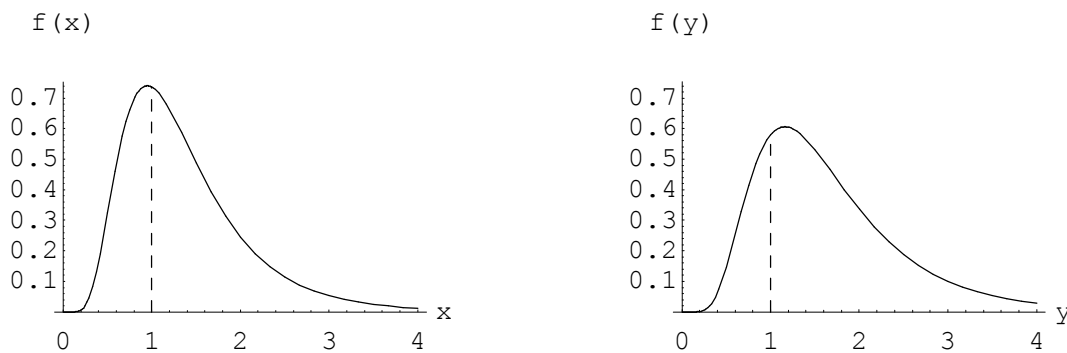


Abb. @5b: Randverteilungen zu der gemeinsamen Verteilung von Antwortzeiten aus Abb. @5a.

Die senkrechten gestrichelten Hilfslinien erleichtern die Feststellung, daß Antwortzeit X tendentiell kürzer ist als Antwortzeit Y . Dies ergibt sich auch aus den Erwartungswerten von $\mu_x = 1.38403$ und $\mu_y = 1.69046$. Die dargestellten Dichten nennt man auch "Randdichten".

Eine Übertragung des Begriffs der bedingten Wahrscheinlichkeitsverteilung auf die hiesige Verteilung ist dagegen mit einem Problem verbunden. Wenn wir für zwei kontinuierliche Zufallsvariablen X und Y etwa die bedingte Wahrscheinlichkeit bestimmen wollen, daß Y kleiner als 1 ist, gegeben $X=2$, dann würde sich aus der Definition der bedingten Wahrscheinlichkeit ergeben:

$$p(Y < 1 / X = 2) = \frac{p(Y < 1 \wedge X = 2)}{p(X = 2)}$$

Bei dem Bruch auf der rechten Seite wird dann aber null durch null dividiert, und das macht keinen Sinn. Man kann aber die in Abschnitt @ B III 4 b für zweidimensionale diskrete Wahrscheinlichkeitsverteilungen eingeführte Formel

$$f(x/Y=y) = \frac{f(x,y)}{f(y)}$$

sinngemäß übertragen, indem man die darin auftretenden Ausdrücke uminterpretiert:

- $f(x/Y=y)$ ist eine "bedingte Dichte"
- $f(x,y)$ ist die gemeinsame Dichte
- $f(y)$ ist die Dichte der Randverteilung von Y im Punkt y .

Die Bedeutung dieser Formel für die bedingte Dichte läßt sich auch gut an Abb. @5a demonstrieren. Dort sieht man ein Gitter von Linien, die "den Berg hinauf und dann wieder hinab" verlaufen. Einige davon verlaufen etwa parallel zur x -Achse ("von links nach rechts"); diese stellen die Funktion $f(x, y)$ für einen bestimmten Wert von y dar. Andere, die diese Funktion für einen bestimmten Wert von x darstellen, verlaufen etwa parallel zur y -Achse ("von hinten nach vorn"). Für welchen Wert von y bzw. x die jeweilige Linie gilt, kann man an der Stelle ablesen, wo die Linie die entsprechende Achse erreicht. Vergleicht man nun die Linien, die aus dem

Bereich um $y = 0.5$ kommen, mit denen aus dem Bereich um $y = 1$, dann sieht man leicht, daß die letzteren wesentlich höher steigen, und entsprechend liegt zwischen jeder Linie und der x - y -Ebene eine unterschiedliche Fläche. Das liegt daran, daß y -Werte um 1 wahrscheinlicher sind als y -Werte um 0.5 (vgl. die Dichtekurve der Randverteilung für Y in Abb. @5b). Die Division von $f(x,y)$ durch $f(y)$ kann man nun so deuten, daß jede der betrachteten Linien umso mehr "in ihrem Gipfelpunkt angehoben" wird, je kleiner $f(y)$. Dann entstehen Dichtekurven für die bedingte Dichte $f(x/Y = y)$, von denen einige im Bereich von $y = 0.5$ bis $y = 1.5$ in Abb. @5c dargestellt sind.

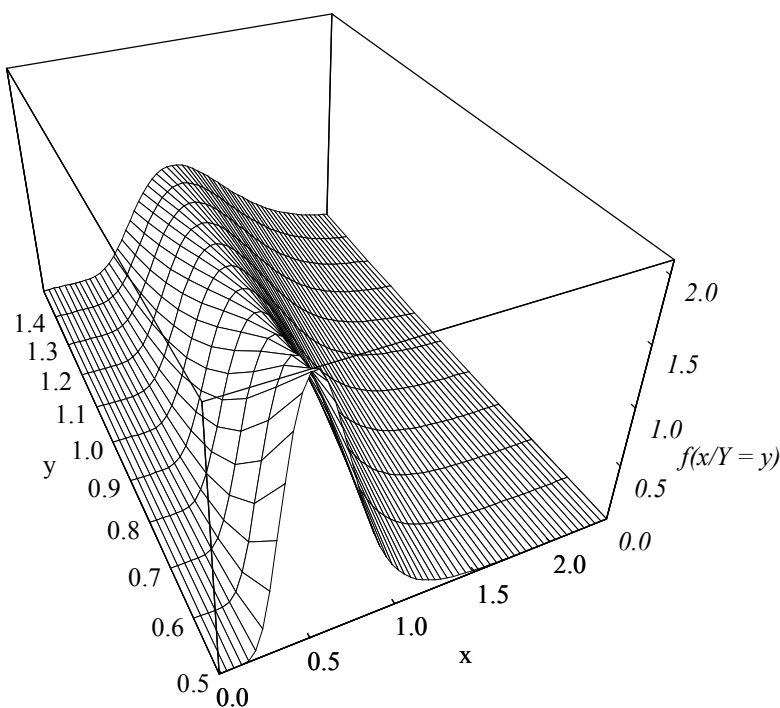


Abb 5c: Kurven für die bedingte Dichte $f(x/Y = y)$ bei verschiedenen y -Werten.

Man kann aufgrund der Abbildung zwar nicht genau nachprüfen, daß unter allen diesen Kurven bis zur x - y -Ebene jeweils eine Fläche von 1 ist; aber es ist doch deutlich, daß die Anhebung gegenüber Abb. @5a umso größer ist, je geringer der in Abb. @5b ablesbare Wert der "Randdichte" $f(y)$. (Auf den ersten Blick könnte man sogar den Eindruck bekommen, daß die Kurve im Bereich um $y=1$ gesunken ist, obwohl die Division durch $f(1) = 0.6$ doch auch eine Anhebung bewirken müßte; aber wenn man die Zahlen an der "vertikalen" Achse beider Abbildungen beachten, löst sich dieser scheinbare Widerspruch auf.)

Die Erwartungswerte der bedingten Verteilungen nennt man - genau wie im diskreten Fall - bedingte Erwartungswerte.

<<

Die mathematische Rechtfertigung der Gleichung für die bedingte Dichte erfolgt natürlich nicht nur über eine Analogie zu diskreten Verteilungen. Man kann zunächst eine bedingte kumulative Verteilungsfunktion $F(x | Y = y)$ als Grenzfall für bedingte Wahrscheinlichkeiten definieren, deren Bedingung sich auf ein immer kleineres y -Intervall bezieht, also durch die Gleichung

$$F(x | Y=y) = \lim_{d \downarrow 0} p(X \leq x | y \leq Y \leq y+d) = \lim_{d \downarrow 0} \frac{p(X \leq x \wedge y \leq Y \leq y+d)}{p(y \leq Y \leq y+d)}$$

Aus dieser bedingten kumulativen Verteilung läßt sich dann die Gleichung für die bedingte Dichte von @S. 161 herleiten, indem man die übliche Definition der Dichte sinngemäß überträgt.

Bei all diesen Überlegungen zur bedingten Dichte ist aber ein Tatsache zu beachten, die den Begriff der bedingten Verteilung bei kontinuierlichen Zufallsgrößen etwas schwierig macht. Wenn wir zu einer gemeinsamen Verteilung zweier kontinuierlicher Zufallsvariablen alle diese bedingten Dichten haben und dann für *einen* bestimmten y -Wert eine andere bedingte Verteilung der Zufallsgröße X annehmen, ändert sich die gemeinsame Verteilung von X und Y nicht. Grund: Die Wahrscheinlichkeit, daß genau der y -Wert auftritt, für den wir die bedingte Verteilung geändert haben, ist null, weil Y eine kontinuierliche Zufallsgröße ist. Daher gibt es zu einer kontinuierlichen zweidimensionalen Verteilung mehrere "Versionen" der bedingten Verteilung. Entsprechend gibt es dann auch mehrere Versionen der bedingten Erwartungswerte.

>>

c) Kennziffern in kontinuierlichen 2-dimensionalen Verteilungen

Die Definitionen der Kennziffern σ_{xy} , ρ_{xy} usw. können aus den diskreten Verteilungen wörtlich übernommen werden, solange man sich auf die Definitionen mit Hilfe von Erwartungswerten bezieht.

Dabei wird natürlich der Erwartungswerts einer von X und Y abgeleiteten Größe $g(X, Y)$ - z.B. des Produkts $X \cdot Y$ für die Kovarianzformel $\sigma_{xy} := E(X \cdot Y) - E(X) \cdot E(Y)$ - wieder ganz entsprechend wie in den eindimensionalen kontinuierlichen Wahrscheinlichkeitsverteilungen durch immer feinere Abrundungen und immer größere Kappungsgrenzen bestimmt. Darauf soll aber nicht im Detail eingegangen werden.

<<

Man kann sogar die Definition des Erwartungswerts einer von X abgeleiteten Größe wörtlich übernehmen und braucht lediglich überall $g(X, Y)$ anstelle von $g(X)$ zu schreiben. Entsprechend arbeitet man auch mit einer Zerlegung in einen Positivteil $g_+(X, Y)$ und einen Negativteil $g_-(X, Y)$, falls auch negative Werte von $g(X, Y)$ vorkommen.

Wenn man diesen Erwartungswert mit Integralen bestimmen will, kann man eine Formel sinngemäß übertragen, die auf S. 82 für zweidimensionale diskrete Wahrscheinlichkeitsverteilungen ein geführt wurde. Dort wurde der Erwartungswert einer aus zwei Zufallsgrößen X und Y abgeleiteten Größe $g(X, Y)$ durch die Gleichung $E[g(X, Y)] = \sum_x \sum_y f(x, y) \cdot g(x, y)$ bestimmt.

Ersetzt man in dieser Formel die Summen über alle möglichen Werte x und y der beiden Zufallsgrößen durch entsprechende Integrale, dann erhält man die Gleichung

$$E[g(X,Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y) \cdot f(x,y) \, dx \, dy$$

Für den Erwartungswert des Produkts $X \cdot Y$ (der wie bei diskreten Verteilungen im Zusammenhang mit der Kovarianz verwendet wird) gilt z.B.

$$E(X \cdot Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x \cdot y \cdot f(x,y) \, dx \, dy$$

>>

d) Bivariate und Multivariate Normalverteilungen

Eine für die Psychologie wichtige Familie mehrdimensionaler kontinuierlicher Verteilungen sind die multivariaten Normalverteilungen. Man nennt sie auch mehrdimensionale Normalverteilungen. Viele "multivariate Verfahren" (also Verfahren, bei denen mehrere Variablen pro Person oder sonstiger Beobachtungseinheit erhoben werden) setzen voraus, daß die Verteilung der Variablen "in der Population" eine solche multivariate Normalverteilung ist. Daher soll hier zumindest erläutert werden, worum es bei diesen Verteilungen geht.

Das Wesentlichste läßt sich schon an den einfachsten Mitgliedern dieser Familie aufzeigen: Den zweidimensionalen oder bivariaten Normalverteilungen, also diejenigen, bei denen es um die gemeinsame Verteilung zweier Zufallsvariablen geht. Diese werden im folgenden wie üblich als X und Y bezeichnet. Wie der Name schon vermuten läßt, müssen die Randverteilungen von X und Y Normalverteilungen sein. Dieses Merkmal reicht aber noch nicht aus, um zu gewährleisten, daß die gemeinsame Verteilung eine bivariate Normalverteilung ist. Was sonst noch gegeben sein muß, läßt sich wieder mit Darstellungen von Dichteflächen und bedingten Dichten klären. Eine Dichtefläche zu einer bivariaten Normalverteilung ist in Abb. 5d wiedergegeben. Wie in Abb. @5a liegt ein "Höhenrücken" vor, dessen Verlauf auf eine positive Korrelation von X und Y hinweist.

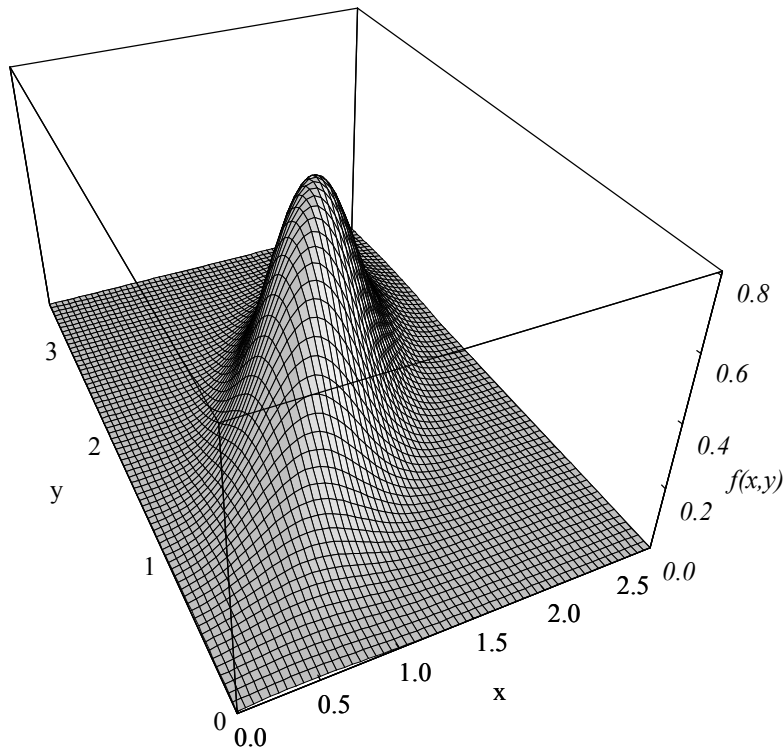


Abb. 5d: Dichtefläche zu einer bivariaten Normalverteilung mit $\mu_x = 1.4$, $\sigma_x = 0.5$, $\mu_y = 1.7$, $\sigma_y = 0.6$ und $\rho_{xy} = 0.7$.

Ein genauerer Vergleich der beiden Dichteflächen zeigt gewisse Unterschiede, aber bei zweidimensionalen Darstellungen dreidimensionaler Gebilde kommen leicht Fehleindrücke durch perspektivische Verzerrungen vor. Wesentlich deutlicher werden Unterschiede, um die es vor allem geht, wenn man die Darstellung bedingter Dichtkurven in Abb. @5c mit der entsprechenden Abb. @5e vergleicht. Diese ist aus Abb. @5d in derselben Weise gewonnen wie Abb. @5c aus Abb. @5a: Entsprechend der Formel für die bedingte Dichte von @S. 161 sind die Werte der gemeinsamen Dichte $f(x,y)$ durch die Randdichte $f(y)$ dividiert, so daß die Teilkurven für gleiche y -Werte aus Abb. @5d umso mehr "im Gipfel angehoben" sind, je kleiner die entsprechende Randdichte ist.

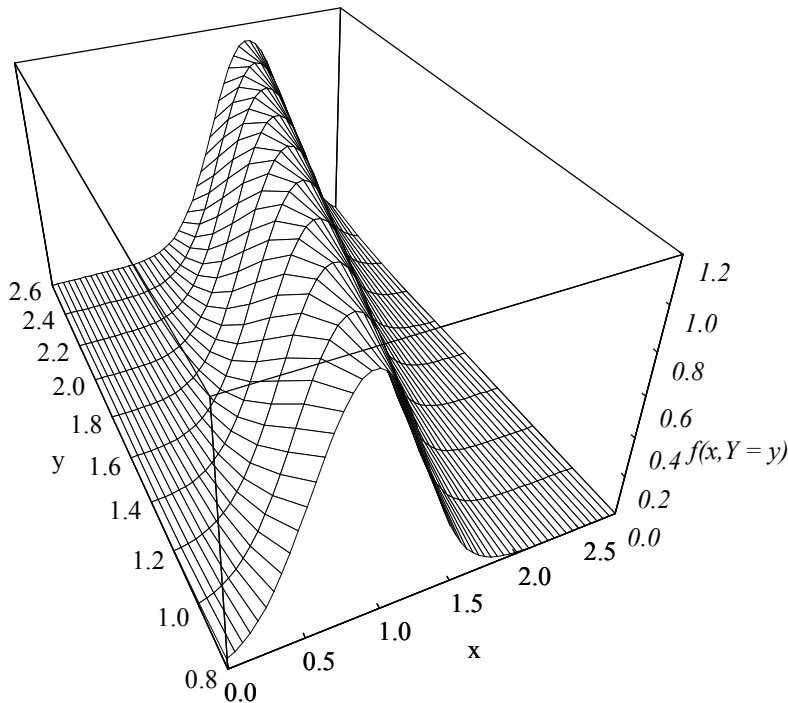


Abb. 5e: Kurven für die bedingte Dichte $f(x/Y = y)$ in einer bivariaten Normalverteilung bei verschiedenen y -Werten.

Bildlich gesprochen: Der "Tunnel" in Abb. @5c ist "vorne" (also bei niedrigen y -Werten) schmaler und höher als hinten, und außerdem macht er hinten einen leichten Schwenk nach rechts. Dagegen ist der Tunnel in Abb. @5e in seinem gesamten Verlauf gleich hoch und gleich breit und verläuft gerade. Ein weiteres Merkmal ist weniger leicht zu sehen, wenn man nicht darauf aufmerksam gemacht wird; aber vielleicht hilft eine hinweisende Frage weiter:

Bei einer der beiden Darstellungen sind die bedingten Dichtekurven Normalverteilungen, bei der anderen haben sie eine positive Schiefe. Auf welche von beiden trifft was zu?

Wahrscheinlich ist es nach dieser Frage zu sehen, daß die Normalverteilungen bei den bedingten Dichtekurven von Abb. @5e vorliegen und daß diejenigen aus Abb. @5c die positive Schiefe haben.

Wenn man auch die bildhaften Feststellungen über die beiden Tunnel etwas mathematischer formuliert und auch noch die als erstes erwähnte Eigenschaft hinzunimmt, daß die beiden Randverteilungen Normalverteilungen sind, hat man bereits die wesentlichen Merkmale, auf die es ankommt:

Definition: Die gemeinsame Wahrscheinlichkeitsverteilung zweier Zufallsvariablen X und Y ist eine bivariate Normalverteilung, wenn die folgenden Eigenschaften zutreffen:

1. Die Randverteilungen von X und Y sind Normalverteilungen.
2. Die bedingten Dichtefunktionen $f(x/ Y = y)$ ergeben für alle Werte des vorgegebenen y Normalverteilungen mit gleicher Varianz.
3. Der entsprechende bedingte Erwartungswert $E(X/ Y = y)$ ist eine lineare Funktion des vorgegebenen y ; die Beziehung ist also linear.
4. Die Eigenschaften 2 und 3 gelten entsprechend für die bedingte Dichtefunktion und den bedingten Erwartungswert von Y bei Vorgabe des Werts von X , also für $f(y/ X = x)$ und für $E(Y/ X = x)$.

Die Definition einer multivariaten Normalverteilung ist dann eine sinngemäße Ausdehnung dieser Definition auf die gemeinsame Wahrscheinlichkeitsverteilung von mehr als zwei Zufallsvariablen.

<<

Die bivariate Normalverteilung läßt sich auch über eine Formel für die Dichte definieren (und das ist natürlich die in der mathematischen Wahrscheinlichkeitstheorie übliche Definition):

Definition: Die gemeinsame Verteilung zweier zufälliger Größen X und Y heißt eine *bivariate (zweidimensionale) Normalverteilung*, wenn die Variablen eine gemeinsame Dichte gemäß der folgenden Formel haben (wobei ρ die Produkt-Moment-Korrelation der Variablen ist):

$$f(x,y) = \frac{1}{2\pi \cdot \sigma_x \cdot \sigma_y \cdot \sqrt{1-\rho^2}} \cdot e^{-\frac{z_x^2 + z_y^2 - 2\rho z_x z_y}{2(1-\rho^2)}}$$

Eine solche Definition über eine Dichteformel ist auch für den allgemeinen Fall einer multivariaten Normalverteilung mit mehr als zwei Variablen möglich. Sie ist aber noch komplexer. Man kann aber eine aus dieser Formel ableitbare Eigenschaft zur Grundlage einer Definition machen, die folgendermaßen lautet:

Definition: Die gemeinsame Verteilung mehrerer (endlich vieler) Zufallsgrößen ist eine *mehrdimensionale Normalverteilung*, wenn jede Linearkombination dieser Zufallsgrößen eine normalverteilte Zufallsvariable ergibt.

Beispiel: Zwei Intelligenztests X und Y messen in unterschiedlichem Ausmaß "kulturfreie" und "kulturabhängige" Intelligenz: Während X stärker die "kulturfreie" Intelligenz erfaßt, geht in Y vor allem die kulturabhängige Intelligenz ein. Nun kann man z.B. für jede Person den Durchschnitt ihrer Leistungen in beiden Tests berechnen, um ihre "allgemeine Intelligenz" zu bestimmen, und dieser Durchschnitt wäre identisch mit der Linearkombination $0.5 \cdot X + 0.5 \cdot Y$. Außerdem könnte man überlegen, die "kulturfreie" Intelligenz zu schätzen, indem man für jede Person auch die Linearkombination $0.8 \cdot X - 0.2 \cdot Y$ berechnet⁷¹, und für weitere Zwecke könnte man noch andere Linearkombinationen in Erwägung ziehen. Dann gilt: Die gemeinsame

⁷¹Die Koeffizienten 0.8 und 0.2 wären genauer mit Methoden der Faktorenanalyse zu ermitteln. Entscheidend ist, daß die Subtraktion eines Werts, der stärker die kulturabhängige Intelligenz erfaßt, die für die Schätzung der kulturfreien Intelligenz unerwünschte "Beimischung" von kulturabhängiger Intelligenz unterdrückt. Insofern hätte die Variable Y hier die Funktion eines "Suppressors".

Verteilung von X und Y ist eine multivariate (hier eine bivariate) Normalverteilung, wenn die Verteilung jeder solchen Linearkombination eine Normalverteilung ist.

Wie multivariate Normalverteilungen entstehen können, zeigt der folgende

Lehrsatz: Haben zwei (oder mehr) zufällige Größen jede als Randverteilung eine Normalverteilung und sind sie voneinander (vollständig) stochastisch unabhängig, so ist die gemeinsame Verteilung verschiedener Linearkombinationen dieser Zufallsgrößen eine bivariate (bzw. eine multivariate) Normalverteilung.

Beispielsweise gehen einige Verfahren der Faktorenanalyse von der Modellvorstellung aus, daß die Meßwerte in verschiedenen beobachtbaren Variablen (z.B. Tests) sich (bis auf Meßfehler) als Linearkombinationen von (zunächst unbekannt) Meßwerten für einige wenige latente Grundeigenschaften (die "Faktoren") darstellen lassen. Aus dem obigen Lehrsatz folgt dann: Sind diese Grundeigenschaften stochastisch unabhängig und normalverteilt, dann ist die gemeinsame Verteilung der beobachtbaren Variablen eine multivariate Normalverteilung (allerdings wieder mit der Einschränkung: Bis auf Meßfehler).

>>

C. Stichprobentheorie

I. Grundbegriffe und Zielsetzung

Bevor wir die Zielsetzung der Stichprobentheorie näher behandeln, sind die Begriffe "Population" und "Stichprobe" zumindest vorläufig zu klären.

Eine Population ist die Gesamtheit aller Repräsentanten einer bestimmten Klasse. (Beispiel: Population der deutschen Studenten = Gesamtheit aller Repräsentanten der Klasse "deutscher Student"). Man spricht hier auch von einer "Grundgesamtheit".

Eine Stichprobe ist eine nach einer bestimmten Regel erhobene Auswahl von Repräsentanten einer Klasse (Beispiel: Zufallsstichprobe der deutschen Studenten = eine nach dem Zufallsprinzip gezogene Auswahl von Repräsentanten der deutschen Studenten).

In späteren Abschnitten wird gezeigt, daß (und warum) diese anschaulichen Begriffe in der Stichprobentheorie durch etwas allgemeinere ersetzt werden; aber für eine erste Annäherung ist es wohl hilfreich, von diesen konkreten Begriffen auszugehen.

Es gehört heute schon fast zur Allgemeinbildung, einen Grundgedanken zu kennen, auf dem Schlüsse von einer Stichprobe auf eine Population beruhen: Stichproben können (zumah wenn sie klein sind) mit Fehlern behaftet sein; aber es ist sehr unwahrscheinlich, daß die Verteilung eines

Merkmals in einer hinreichend großen, "nach allen Regeln der Kunst" erhobenen Stichprobe erheblich von der Verteilung in einer Population abweicht, aus der die Stichprobe gezogen wurde.

So formuliert, ist der Gedanke allerdings noch recht unpräzise. Es ist wichtig, zu wissen, mit welchem Ausmaß von Abweichungen zwischen Stichprobenverteilung und Populationsverteilung man rechnen muß. Antworten auf derartige Fragen zu geben, ist Aufgabe der Stichprobentheorie.

Dabei wird vor allem auf das Urnenmodell zurückgegriffen. Bei seiner Behandlung in Abschnitt @B.I.3.a wurde darauf hingewiesen, daß man dabei zwei Arten von Fragestellungen unterscheiden kann:

- Man kann einerseits fragen, mit welcher Wahrscheinlichkeit bestimmte Zugergebnisse ("Daten") bei einer angenommenen Zusammensetzung des Urneninhalts auftreten.
- Außerdem kann man fragen, welche Schlüsse über den unbekanntes Inhalt der Urne gezogen werden können, wenn man die Ergebnisse der Ziehungen (also die "Daten") kennt.

In der Praxis der Statistikanwendung geht es natürlich vor allem um Rückschlüsse von Stichproben-Daten auf die Populationsverteilung des erhobenen Merkmals, und das entspricht der zweiten Fragestellung. Aber solche Rückschlüsse setzen (wie beim Urnenmodell) voraus, daß man zunächst einmal eine einfachere Frage beantworten kann, die der ersten Fragestellung im Urnenmodell entspricht: Mit welcher Wahrscheinlichkeit treten bestimmte Daten bei einer angenommenen Populationsverteilung des Merkmals auf? Das ist die Aufgabe der Stichprobentheorie.

Diese Herangehensweise mag auf den ersten Blick praxisfern aussehen: Wir erheben doch Stichprobendaten, weil wir die Populationsverteilung eines Merkmals nicht kennen. Was hilft es da, wenn wir etwas über die Wahrscheinlichkeitsverteilung der Daten bei einer angenommenen Populationsverteilung des Merkmals aussagen können?

Um eine erste Antwort auf diese kritische Frage zu geben, lohnt sich ein Rückgriff auf unsere Feststellung, daß die Wahrscheinlichkeit, bei der Beantwortung eines Multiple-Choice-Tests aus 5 Aufgaben mit je 6 Alternativen mit Würfeln 2 Treffer zu erzielen, 0.1608 beträgt. Wenn nun eine Person in einem solchen Test tatsächlich 2 Richtige erzielt, dann sind das zwar etwas mehr als der Erwartungswert von 0.8333; aber die Wahrscheinlichkeit, ein solches Ergebnis rein zufällig zu erzielen, ist noch zu groß, um zu sagen, daß das kaum noch Zufall sein kann. Worauf es im hiesigen Zusammenhang ankommt, ist die Feststellung: Obwohl wir nicht wissen, ob die Person tatsächlich geraten hat, kann es für die Schlußfolgerung wichtig sein, sich zu fragen, wie groß denn die Wahrscheinlichkeit von 2 Treffern unter der Annahme des reinen Rätens wäre.

Welche Bedeutung eine solche Berechnung von Wahrscheinlichkeiten unter einer angenommenen Situation beim Schluß von Stichproben auf Populationen haben kann, mag das folgende Beispiel veranschaulichen. In einer Zufallsstichprobe von 400 Ehefrauen wurde gefragt, ob die Hausarbeit zwischen ihnen und ihrem Ehemann gleichmäßig verteilt ist. Nur 4.3% beantworteten diese Frage mit "Ja". Natürlich liegt die Vermutung nahe, daß dann auch etwa 4.3% der Ehefrauen in der Ehefrauen-Population, aus der die Stichprobe stammt, die Frage mit "Ja"

beantworten würden. Um genauer eingrenzen zu können, mit welchem Fehler ein solcher Rückschluß behaftet ist, könnte man z.B. feststellen: Wenn der Prozentsatz der Ja-Antworten in der Population 10% betrüge, dann würde man in einer Zufallsstichprobe von $N=400$ mit einer Wahrscheinlichkeit von 99% einen Prozentsatz von Ja-Antworten zwischen 6.3% und 13.7% erhalten.⁷² Daraus kann man dann den Schluß ziehen, daß der tatsächlich erzielte Prozentsatz von 4.3% Ja-Antworten mit der Annahme von 10% Ja-Antworten in der Population kaum vereinbar ist. In dieser Weise kann man genauer einkreisen, mit welchen Annahmen über den Prozentsatz der Ja-Antworten in der Population das Ergebnis von 4.3% Ja-Antworten in der Stichprobe denn vereinbar ist.

Man kann diesen Gedankengang folgendermaßen verallgemeinern: Hinter allen Antworten auf die Frage, mit welchen Annahmen über die Population die Daten einer Stichprobe vereinbar sind, stehen Überlegungen, welche Daten bei Zugrundelegung dieser Annahmen mit einer bestimmten Wahrscheinlichkeit auftreten. Solche Überlegungen anzustellen, ist Aufgabe der Stichprobentheorie. Wie man aus derartigen Überlegungen dann Rückschlüsse von der Stichprobenverteilung eines Merkmals auf die Populationsverteilung ableitet, wird in den Abschnitten @D und @E behandelt.

Noch eine weitere Anwendung der Stichprobentheorie ist erwähnenswert. Es ist zwar allgemein bekannt, daß die mit einer Stichprobe verbundenen Fehler auch mit der Stichprobengröße zusammenhängen; aber wenn man diesen Zusammenhang genauer kennen will (um beispielsweise die für eine bestimmte Genauigkeit von Schätzungen erforderliche Stichprobengröße zu bestimmen), dann erhält man Antworten auf diese Frage von der Stichprobentheorie.

II. Stichprobentechniken

Man kann eine Stichprobe nach verschiedenen Verfahren ziehen, die man auch "Stichprobentechniken" nennt. Die von uns zu behandelnden statistischen Methoden gehen alle von einem Modell der Zufallsstichprobe aus, das man sich zunächst am Urnenmodell veranschaulichen kann. Ist die Population, aus der wir die Stichprobe ziehen wollen, überschaubar (z.B. Population der Studenten einer Universität), dann können wir jedem Mitglied der Population ein Los zuordnen (indem wir z.B. den Namen und das Geburtsdatum oder die Matrikelnummer draufschreiben) und diese Lose in eine Urne geben, durchmischen und so viele Lose ziehen, wie die Stichprobe

⁷²Wie man zu einem solchen Schluß kommt, ist im Augenblick unwichtig. (Es wird in Abschnitt @C.V behandelt.) Im Moment kommt es vor allem darauf an, welche Rolle eine solche Aussage beim Rückschluß auf die Populationsverteilung des Merkmals spielt.

Elemente (z.B. Vpn) haben soll. Je nachdem, ob wir dabei "mit Zurücklegen" oder "ohne Zurücklegen" verfahren, entsteht eine "Zufallsstichprobe mit Zurücklegen" oder eine "Zufallsstichprobe ohne Zurücklegen".

Die von uns zu behandelnden Verfahren der schlußfolgernden Statistik gehen im Grunde alle vom Modell der "Zufallsstichprobe mit Zurücklegen" aus, weil dieses Modell am einfachsten wahrscheinlichkeitstheoretisch zu analysieren ist. In der Praxis würde man dieses Verfahren niemals anwenden: Man würde wohl kaum einer Vp die Chance geben, mehrmals in die Stichprobe zu kommen, wenn ihr Los nach dem Zurücklegen ein zweites mal gezogen wird.

Der Fehler, den man macht, wenn man eine "Zufallsstichprobe ohne Zurücklegen" mit mathematischen Methoden verrechnet, die für "Zufallsstichprobe mit Zurücklegen" entwickelt wurden, ist jedoch nicht gravierend. Zunächst läßt sich mathematisch beweisen, daß die Wahrscheinlichkeitsverteilungen, die wir im Zusammenhang mit der Stichprobentheorie und der schlußfolgernden Statistik behandeln werden, beim Stichprobenziehen mit und ohne Zurücklegen praktisch identisch sind, solange der Umfang der Stichprobe nicht mehr als etwa 5% der Populationsgröße beträgt.⁷³ Außerdem läßt sich zeigen, daß "Zufallsstichproben ohne Zurücklegen" dann, wenn sie von einer "Zufallsstichprobe mit Zurücklegen" abweichen, eher repräsentativ sind als die letzteren. Wenn wir also die Modelle "mit Zurücklegen" verwenden, dann sind wir eher zu vorsichtig als zu unvorsichtig in unseren Schlußfolgerungen auf die Population.

Es gibt noch andere Stichprobentechniken, die hier nicht näher behandelt werden sollen. In der Psychologie werden jedoch häufig überhaupt keine nach einer Regel gezogenen Stichproben, sondern sogenannte "Stichproben aufs Geradewohl" verwendet: Man testet alle Mitglieder der zu untersuchenden Population, die gerade greifbar sind, und häufig besteht nicht einmal Klarheit darüber, welche Population von der Stichprobe erfaßt werden soll. Es ist eine Grundlagenfrage der Statistikanwendung in der Psychologie, ob es zu rechtfertigen ist, die für Zufallsstichproben "mit Zurücklegen" entwickelten statistischen Verfahren einfach auf diese "Stichproben aufs Geradewohl" anzuwenden. Wir können diese Frage hier nur stellen und nicht näher diskutieren.

III. Beobachtungen als zufällige Größen

Die gesamte Stichprobentheorie basiert auf der Vorstellung, daß die Meßwerte einer Stichprobe durch Zufall größer oder kleiner ausfallen können. Man kann sich z.B. fragen: Wenn

⁷³Ausnahme: Die Wahrscheinlichkeit, mehrere Elemente einer in der Population nur sehr selten vertretenen Klasse zu erhalten, ist beim Stichprobenziehen "ohne Zurücklegen" spürbar geringer als beim Ziehen "mit Zurücklegen". Haben wir nämlich ein solches Element gezogen, dann sinkt die relative Häufigkeit dieser Klasse in der Urne beträchtlich, wenn wir ohne Zurücklegen ziehen.

in einer Population 62% der Personen einen Meßwert bis 83.5 haben und wir aus dieser Population eine Stichprobe nach dem Urnenmodell mit Zurücklegen ziehen, wie groß ist dann die Wahrscheinlichkeit, daß der erste Meßwert unserer Stichprobe eine Zahl bis 83.5 ist? Natürlich beträgt diese Wahrscheinlichkeit 0.62 oder 62%; denn nach dem Urnenmodell ist die Wahrscheinlichkeit, eine Kugel mit einer bestimmten Eigenschaft (hier: Eine Person mit einem Meßwert bis 83.5) zu ziehen gleich dem Anteil der Kugeln mit dieser Eigenschaft am Gesamthalt der Urne. Diese Überlegung können wir natürlich nicht nur für Meßwerte bis 83.5 anstellen. Für jede Zahl x gilt ganz entsprechend: Die Wahrscheinlichkeit, daß unser erster Meßwert höchstens gleich x ist, ist gleich der relativen Häufigkeit von Personen mit Meßwert bis x in der Population. Und das gilt selbstverständlich nicht nur für den ersten Meßwert, sondern auch für jeden weiteren. Damit gilt der erste Satz des folgenden Grundprinzips. (Der zweite wird anschließend kommentiert.)

Grundprinzip der Stichprobentheorie nach dem Urnenmodell: Jede Beobachtung eines Merkmals (z.B. jeder Meßwert) in der Stichprobe kann als eine zufällige Größe betrachtet werden, deren Wahrscheinlichkeitsverteilung identisch mit der Populationsverteilung des Merkmals ist. Die verschiedenen Beobachtungen, die an den verschiedenen Elementen (z.B. V_{pn}) einer Zufallsstichprobe gemacht werden, können als stochastisch unabhängig betrachtet werden.

Der zweite Satz gilt exakt nur beim "Stichprobenziehen mit Zurücklegen"; beim "Stichprobenziehen ohne Zurücklegen" gilt er zumindest approximativ. Diese Interpretation von Stichprobendaten als unabhängige zufällige Größen mit einer Wahrscheinlichkeitsverteilung, die der Populationsverteilung des beobachteten Merkmals entspricht, ist eine wesentliche Grundlage der ganzen schlußfolgernden Statistik.

Da die Populationsverteilung in der Stichprobentheorie vor allem in Form der Wahrscheinlichkeitsverteilung der Einzelbeobachtungen auftritt, verwendet man für die Kennziffern der Populationsverteilung die gleichen griechischen Buchstaben wie bei Wahrscheinlichkeitsverteilungen. Die Symbole μ_x , σ_x und σ_x^2 werden also auch für das Populationsmittel⁷⁴, die Populations-Standardabweichung und die Populations-Varianz eines Merkmals verwendet.

IV. Der mathematische Stichprobenbegriff

Im vorangehenden Abschnitt haben wir aus dem Modell des Urnenziehens mit Zurücklegen

⁷⁴Der Begriff "Populationsmittel" bezeichnet hier und im folgenden immer das arithmetische Mittel der Populationsverteilung eines Merkmals. Präziser wäre natürlich der Ausdruck "arithmetisches Populationsmittel", aber der kürzere Ausdruck ist üblich.

die Sichtweise abgeleitet, nach der die Daten einer Stichprobe als stochastisch unabhängige Zufallsvariablen betrachtet werden, deren Wahrscheinlichkeitsverteilung mit der Populationsverteilung des Merkmals identisch ist. Der mathematischen Stichprobentheorie liegt nun eine weitere Verallgemeinerung zugrunde, die auch für das Verständnis über den Sinn der Statistik-anwendung in der Psychologie wichtig ist. Den entscheidenden Gedankenschritt könnte man etwa folgendermaßen formulieren: Unabhängige Zufallsvariablen mit gleicher Wahrscheinlichkeitsverteilung können nicht nur beim Ziehen einer Stichprobe aus einer Population nach dem Urnenmodell mit Zurücklegen entstehen. Für mathematische Schlußfolgerungen ist es gleichgültig, wie es zu dieser Situation kommt. Hauptsache ist, daß unsere Meßwerte als N unabhängige, identisch verteilte Zufallsvariablen betrachtet werden können. Wegen der zentralen Bedeutung dieser Annahme wird dafür häufig eine Abkürzung verwendet, die für die ersten drei Worte von "independent, identically distributed random variables" steht. Es handelt sich um die

i.i.d.-Annahme: Eine Stichprobe besteht aus N unabhängigen, identisch verteilten Zufallsvariablen. (Dabei bezeichnet N - wie üblich - die Stichprobengröße.)

Allein aus dieser Annahme ergeben sich bereits viele Antworten auf die (in Abschnitt @C.I herausgestellte) Grundfrage der Stichprobentheorie, mit welcher Wahrscheinlichkeit bestimmte Daten unter angenommenen Populationsverhältnissen zu erwarten sind.

<<

Diese Verallgemeinerung des Stichprobenbegriffs ist auch für die Psychologie wichtig. Bei der Behandlung von Stichproben nach dem Urnenmodell wird ja implizit vorausgesetzt, daß es für jede zu einer Population gehörende Person einen festen Meßwert gibt, den man nur festzustellen braucht, wenn die Person in die Stichprobe kommt. Heute sehen es aber viele Methodiker eher so, daß die Verhaltensdispositionen einer Person durch eine individuelle Wahrscheinlichkeitsverteilung von Meßwerten zu charakterisieren sind. Bei einigen Untersuchungen (z.B. bei Reaktionszeiten) kann man mehrere Meßwerte pro Person erheben und diese dann als eine Stichprobe im obigen mathematischen Sinn betrachten.

Aber auch in Bereichen, in denen die Erhebung mehrerer Meßwerte pro Person nicht sinnvoll ist, ist das mathematische Stichprobenkonzept für die Psychologie adäquater als das einfache Urnenmodell. In psychologischen Untersuchungen liegt meistens eine Art 2-stufiges Stichprobenverfahren vor: Zunächst ziehen wir aus einer Population von Menschen eine Stichprobe von V_{pn} , und dann ziehen wir aus der Wahrscheinlichkeitsverteilung der möglichen Verhaltensweisen der in der V_{pn} -Stichprobe aufgenommenen Menschen nochmals eine Stichprobe, indem wir unsere experimentellen Daten erheben.

>>

Die Eleganz des verallgemeinerten mathematischen Stichprobenmodells hat leider einen Preis: Die Überlegungen werden ziemlich abstrakt und erschweren dadurch eine erste Annäherung an die Stichprobentheorie. Daher wird die Stichprobentheorie im folgenden im Sinne der gängigen Vorstellung des Urnenmodells mit Zurücklegen dargestellt. Die Ergebnisse sind aber auf andere Stichproben im Sinne der allgemeinen i.i.d.-Annahme übertragbar.

V. Wahrscheinlichkeitsverteilungen von Stichprobenkennziffern

Wenn wir eine Stichprobe von Meßwerten ziehen, dann sind alle diese Meßwerte zufällige Größen, und entsprechend ist auch der Durchschnitt eine zufällige Größe und hat eine Wahrscheinlichkeitsverteilung: Er kann zufällig größer oder kleiner ausfallen.

Es ist relativ leicht zu beweisen, daß der Erwartungswert dieser Wahrscheinlichkeitsverteilung des Stichprobendurchschnitts \bar{x} angegeben werden kann als

$$E(\bar{x}) = \mu_x$$

Der Erwartungswert der Wahrscheinlichkeitsverteilung des Stichprobendurchschnitts \bar{x} ist also gleich dem Populationsmittel. Auch die Varianz dieser Wahrscheinlichkeitsverteilung kann man angeben, und zwar als

$$\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{N}$$

Dabei ist $\sigma_{\bar{x}}^2$ die Varianz der Wahrscheinlichkeitsverteilung des Stichprobendurchschnitts, während σ_x^2 die Varianz der X-Werte in der Grundgesamtheit ist.

Für die Standardabweichung derselben Wahrscheinlichkeitsverteilung gilt entsprechend die Gleichung

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}}$$

Diese Gleichung wird häufig als das "Wurzel-N-Gesetz" bezeichnet. Dieses Gesetz ist aus mehreren Gründen wichtig:

- Einerseits präzisiert das Wurzel-N-Gesetz die bekannte Tatsache, daß die zufälligen Stichprobenfehler, mit denen man rechnen muß, um so kleiner werden, je größer die Stichprobe ist. Wie jede Standardabweichung einer Zufallsvariablen, drückt auch $\sigma_{\bar{x}}$ aus, mit welchen Abweichungen vom Erwartungswert man rechnen muß. Da der Erwartungswert der Wahrscheinlichkeitsverteilung des Stichprobendurchschnitts das Populationsmittel μ_x ist (s.o.), kann man also sagen: $\sigma_{\bar{x}}$ drückt aus, mit welchen zufälligen Abweichungen des Stichprobendurchschnitts \bar{x} vom Populationsmittel μ_x man rechnen muß. Dazu ermöglicht das Wurzel-N-Gesetz z.B. die Schlußfolgerung, daß man die Stichprobengröße vervierfachen muß, um den Zufallsfehler, mit dem der Stichprobendurchschnitt behaftet ist, zu halbieren.
- Recht verbreitet ist die Vermutung, daß eine Stichprobe um so größer sein müsse, je größer die Population ist. Das Wurzel-N-Gesetz zeigt aber, daß diese Vermutung unzutreffend ist: Der Zufallsfehler, mit dem ein Stichprobendurchschnitt behaftet ist, hängt zwar von der Größe der Stichprobe ab, aber nicht von der Größe der Population.

Über die Form der Wahrscheinlichkeitsverteilung des Stichprobendurchschnitts läßt sich folgende Aussage machen: Sie ist genau dann eine Normalverteilung, wenn die Verteilung des Merkmals X in der Grundgesamtheit eine Normalverteilung ist. Sie ist aber nach dem Zentralen Grenzwertsatz, den wir auf die Summe aller X -Werte anwenden, auch dann zumindest approximativ eine Normalverteilung, wenn die Verteilung des Merkmals X in der Grundgesamtheit nicht allzu stark von einer Normalverteilung abweicht und wenn die Stichprobe hinreichend groß ist. Diese beiden Bedingungen ergänzen sich gegenseitig: Je größer die Abweichung der Verteilung der X -Werte von der Normalverteilung, desto größer muß die Stichprobe sein, damit man approximativ eine Normalverteilung des Stichprobendurchschnitts annehmen kann und umgekehrt.

Faustregel: Bei mäßigen Abweichungen der Populationsverteilung von der Normalverteilung kann bereits ab etwa $N=30$ nach dem zentralen Grenzwertsatz die Wahrscheinlichkeitsverteilung des Stichprobendurchschnitts durch eine Normalverteilung approximiert werden; und nur bei extremen Abweichungen der Populationsverteilung von der Normalverteilung muß die Stichprobengröße N größer als 100 sein, damit man die Wahrscheinlichkeitsverteilung des Stichprobendurchschnitts \bar{x} mit einer Normalverteilung approximieren kann.

Den Erwartungswert und die Varianz dieser Normalverteilungen können wir nach den obigen Formeln bestimmen, und damit ist die Wahrscheinlichkeitsverteilung des Stichprobendurchschnitts in den beiden Fällen mit Normalverteilung festgelegt.

Beispiel: Wenn man an einer Zufallsstichprobe von $N = 225$ ein Merkmal erhebt, dessen Populationsverteilung ein Mittel von $\mu_x = 70$ und eine Standardabweichung von $\sigma_x = 6$ hat, dann kann man die Wahrscheinlichkeitsverteilung des Stichprobendurchschnitts \bar{x} näherungsweise als eine Normalverteilung mit Erwartungswert

$$E(\bar{x}) = \mu_x = 70$$

und Standardabweichung

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}} = \frac{6}{\sqrt{225}} = 0.40$$

betrachten. Das bedeutet z.B., daß der Stichprobendurchschnitt in der angenommenen Situation mit einer Wahrscheinlichkeit von 95% in den Bereich von 69.216 und 70.784 fallen wird. (Dies sind nämlich in der angegebenen Normalverteilung die Zahlen mit z -Werten von -1.96 und +1.96.)

Diese Überlegungen zum Stichprobendurchschnitt können wir zumindest in dem folgenden Satz verallgemeinern:

Jede aus den Stichprobendaten berechnete Stichprobenkennziffer ist eine zufällige Größe und hat eine Wahrscheinlichkeitsverteilung, die sowohl von der Verteilung des beobachteten Merkmals in der Grundgesamtheit als auch von der Größe der Stichprobe abhängt.

Die weiteren für den Durchschnitt gezeigten Eigenschaften lassen sich jedoch nicht ohne weiteres übertragen. Insbesondere kann nicht ohne weiteres davon ausgegangen werden, daß der Erwartungswert der Wahrscheinlichkeitsverteilung einer Stichprobenkennziffer immer gleich dem Wert der entsprechenden Populations-Kennziffer ist. Für den Durchschnitt trifft das zu, wie wir gesehen haben; aber in @D.II.1 werden wir sehen, daß es bei der Varianz anders ist.

<< Nachzutragen bleibt noch die Begründung für eine im Zusammenhang mit dem Beispiel aus Abschnitt @C.I aufgestellte Behauptung. Dort wurde gesagt: Wenn der Prozentsatz der Ja-Antworten in der Population 10% betrüge, dann würde man in einer Zufallsstichprobe von $N=400$ mit einer Wahrscheinlichkeit von 99% einen Prozentsatz von Ja-Antworten zwischen 6.3% und 13.7% erhalten. Natürlich sind auch Stichproben-Prozentsätze Kennziffern, die durch Zufall größer oder kleiner ausfallen können. Es ist aber einfacher, wenn wir statt des Prozentsatzes die absolute Häufigkeit von Ja-Antworten als eine zufällige Größe X behandeln, über die wir folgende Aussagen machen können:

- Nach dem Urnenmodell mit Zurücklegen (das wir aus den in Abschnitt @C.II aufgeführten Gründen zugrundelegen können) haben wir $N=400$ unabhängige "Versuche", und die Zufallsgröße X ist die Zahl der Versuche, bei denen das "kritische Ereignis" einer Ja-Antwort auftritt. Wenn nun der Prozentsatz der Ja-Antworten in der Population 10% ist, dann ist X binomialverteilt mit $N=400$ und $p=0.10$.
- Diese Binomialverteilung können wir mit einer Normalverteilung approximieren, deren Kennziffern sich als $\mu = 400 \cdot 0.10 = 40$ und $\sigma^2 = 400 \cdot 0.10 \cdot 0.90 = 36$ angeben lassen. Die Standardabweichung dieser Normalverteilung ist also $\sigma = 6$.
- Rechnen wir die angegebenen Prozentsätze von 6.3% und 13.7% Ja-Antworten in absolute Häufigkeiten um, dann ergibt sich 25 bzw. 55.
- Zusammenfassend ergibt sich die Frage: Wie groß ist die Wahrscheinlichkeit, daß eine normalverteilte Größe mit $\mu = 40$ und $\sigma = 6$ in das Intervall von 24.5 bis 55.5 fällt (Kontinuitäts-Korrektur!). Daß diese Wahrscheinlichkeit ziemlich genau 99% beträgt, kann als Übungsaufgabe überprüft werden.

>>

D. Parameterschätzung

I. Zielsetzung und Terminologie

Es ist ein leider weit verbreiteter Irrtum, zu glauben, die Statistik ginge von der stillschwei-

genden Annahme aus, daß alle aus einer Stichprobe berechneten Kennziffern automatisch auf die Population "übertragen" werden könnten - allenfalls mit der Angabe einer "Fehlertoleranz". Tatsächlich gibt es Stichproben-Kennziffern, mit denen man so umgehen kann. In den folgenden Abschnitten werden wir z.B. erfahren, daß das für den Durchschnitt sinnvoll ist. Aber schon bei der Varianz ist es anders: Es ist vielleicht intuitiv plausibel (und läßt sich auch nachweisen), daß die Stichprobenvarianz eines Merkmals vor allem in kleinen Stichproben die Varianz dieses Merkmals in der Population tendentiell unterschätzt. Deswegen ist es sinnvoll, die Stichprobenvarianz "nach oben zu korrigieren", wenn wir die Populationsvarianz schätzen wollen. (Eine entsprechende Formel werden wir in Abschnitt @D.II.1 kennenlernen.)

Wenn aber nicht jede Stichprobenkennziffer automatisch auch der beste Schätzwert für die entsprechende Kennziffer der Populationsverteilung eines Merkmals ist, dann stellt sich für jede uns interessierende Kennziffer dieser Populationsverteilung die Frage nach einer optimalen Schätzregel. Schon der soeben angedeutete Unterschied zwischen der Brauchbarkeit von Stichprobendurchschnitt und Varianz läßt vermuten, daß diese Frage für verschiedene Kennziffern unterschiedliche Probleme aufwirft, und deshalb ist es auch in einer Einführung in die Statistik nicht möglich, zu allen in der Beschreibenden Statistik eingeführten Stichprobenkennziffern die optimalen Schätzregeln für entsprechende Populationskennziffern zu behandeln. Wir werden dies vor allem für Durchschnitt und Varianz tun. Diese zwei Beispiele sind zunächst um ihrer selbst willen wichtig.

Außerdem werden wir an diesen beiden Beispielen einige Begriffe kennenlernen, die bisher unter der intuitiven Bezeichnung "optimale Schätzregel" zusammengefaßt wurden. Schon in der Regressionsstatistik (bei der es ja auch um Schätzung geht) haben wir festgestellt, daß man grundsätzlich verschiedene "Optimalitätskriterien" aufstellen könnte: Man könnte entweder die Beträge oder die Quadrate der Abweichungen zwischen Schätzwert und tatsächlichem Wert möglichst klein machen, und je nach dem, welches dieser Kriterien man zugrundelegt, kommt man möglicherweise zu unterschiedlichen Regressionsgleichungen. Dort haben wir auch gesehen, daß sich das Prinzip möglichst kleiner Abweichungsquadrate durchgesetzt hat, weil sich darauf am besten "weiterbauen" läßt. (Die additive Zerlegung der Varianz von Y in die "determinierte Varianz" und die "Schätzfehlervarianz" wäre z.B. nicht möglich, wenn wir statt der Abweichungsquadrate die Abweichungsbeträge möglichst klein gemacht hätten.) Ähnlich ist es auch mit den Kriterien guter Schätzregeln, die wir in Abschnitt @D.II kennenlernen werden: Es ließen sich auch andere Kriterien formulieren, die auf den ersten Blick ähnlich plausibel wären wie das Kriterium möglichst kleiner Abweichungsbeträge in der Regressionsstatistik; aber die hier zu behandelnden sind diejenigen, die sich durchgesetzt haben, weil sich auf den nach entsprechenden Regeln gewonnenen Schätzwerten am besten aufbauen läßt.

Bevor wir uns dieser Aufgabe zuwenden, sollen noch einige Begriffe und Regeln für Symbole in Formeln usw. eingeführt werden. Zunächst drei Begriffe, die zu unterscheiden sind:

Parameter sind Populationskennziffern, Kennziffern von Wahrscheinlichkeitsverteilungen usw..

Statistiken sind alle Werte, die man aus den Daten einer Stichprobe berechnen kann (z.B. die aus der beschreibenden Statistik bekannten Kennziffern, aber z.B. auch die Summe der Meßwerte oder die Summe der quadrierten Abweichungen vom Durchschnitt).

Schätzfunktionen sind Statistiken, die zur Schätzung der Parameter herangezogen werden.

Unter Verwendung dieser Begriffe können wir nun auch genauer formulieren, was mit "optimalen Schätzregeln" gemeint ist: Es geht um die Frage, welche Statistiken für welche Parameter als optimale Schätzfunktion zu betrachten sind.

Die Theorie der Parameterschätzung beruht nun im wesentlichen auf dem zum Schluß des Abschnitts @D formulierten Prinzip, daß jede Stichprobenkennziffer eine Zufallsgröße ist, also eine Wahrscheinlichkeitsverteilung hat. Dieses Prinzip können wir mit den neuen Begriffen noch präziser formulieren: Es gilt nicht nur für die Kennziffern der "Beschreibenden Statistik", sondern für jede "Statistik", also für alles, was man aus den Daten einer Stichprobe berechnen kann, insbesondere also auch für Schätzfunktionen.

Um die verschiedenen Arten von Kennziffern in Formeln, Gleichungen usw. auseinanderhalten zu können, haben sich einige Konventionen eingebürgert:

- Parameter werden meist mit griechischen Buchstaben bezeichnet.
- Für Statistiken, die zur Beschreibung der Daten einer Stichprobe verwendet werden, benutzt man die in der Beschreibenden Statistik eingeführten Symbole.
- Für Schätzfunktionen verfährt man ähnlich wie in der Regressionsstatistik: Genauso, wie dort \hat{Y}_i für "geschätzter Wert von Y_i " geschrieben wird, verwendet man für Schätzfunktionen das Symbol für den zu schätzenden Parameter mit einem Dach darüber. Beispielsweise steht $\hat{\mu}_x$, $\hat{\sigma}_x$ bzw. $\hat{\sigma}_x^2$ für "geschätzter Wert von μ_x , σ_x bzw. σ_x^2 ".

Der Sinn dieser Unterscheidungen wird deutlich, wenn man für drei denkbare Gleichungen als Satz danebensreibt, was sie bedeuten:

Gleichung: Bedeutung:

$\mu_x = 101.3$ Das Populationsmittel der X -Werte ist 101.3.

$\bar{x} = 101.3$ Der Stichprobendurchschnitt der X -Werte ist 101.3.

$\hat{\mu}_x = 101.3$ Wir schätzen, daß das Populationsmittel 101.3 ist.

<<

Wie in Abschnitt @C.IV dargestellt, sind für die Mathematik "Stichproben aus Populationen" nur ein Spezialfall der mehrfachen Realisierung von unabhängigen Zufallsvariablen mit gleicher Wahrscheinlichkeitsverteilung, wobei diese Wahrscheinlichkeitsverteilung gleich der Populationsverteilung des Merkmals ist. Dementsprechend wird dann in der Mathematik die

Parameterschätzung auch ganz als Schätzung der Kennziffern von Wahrscheinlichkeitsverteilungen behandelt. Wenn es dann darum geht, Kennziffern einer Populationsverteilung zu schätzen, dann schätzt man eben die Kennziffern der in Abschnitt @C.III behandelten Wahrscheinlichkeitsverteilung, die beim Urnenziehen mit Zurücklegen entstehen würde. Für die Psychologie ist diese Sichtweise z.B. in der Diagnostik von Bedeutung. Wie in Abschnitt @C.IV dargestellt, kann man davon ausgehen, daß die Verhaltensdispositionen einzelner Menschen durch individuelle Wahrscheinlichkeitsverteilungen charakterisiert werden. Dann besteht die Aufgabe der Diagnostik darin, aufgrund von Daten (z.B. aus Tests) die Parameter dieser individuellen Wahrscheinlichkeitsverteilungen zu schätzen.

>>

II. Kriterien guter Schätzfunktionen

1. Erwartungstreue

In Abschnitt @D.I wurde angekündigt, daß der Stichprobendurchschnitt sich gut als Schätzfunktion für das Populationsmittel eignet, während dies für die Varianz nicht entsprechend gilt. In diesem Abschnitt @D.II.1 wird dies nun näher zu begründen sein.

Das Merkmal, auf das dabei ankommt, haben wir für den Durchschnitt bereits in der Gleichung

$$E(\bar{x}) = \mu_x$$

aus Abschnitt @C.V kennengelernt. In den Kontext der Parameterschätzung gestellt, bedeutet sie: Wenn wir die

$$\text{Schätzregel: } \hat{\mu}_x = \bar{x}$$

zugrundelegen, dann können wir uns zwar durch Stichprobenzufälle nach oben oder unten verschätzen, und deshalb ist unsere Schätzfunktion - genau wie der Durchschnitt - eine Zufallsgröße; aber für den Erwartungswert dieser Zufallsgröße gilt

$$E(\hat{\mu}_x) = \mu_x.$$

Der Erwartungswert der Schätzfunktion ist also genau der Wert, den wir schätzen wollen: Das Populationsmittel μ_x .

Die stillschweigende Annahme, daß etwas derartiges für alle Stichprobenkennziffern zutreffen muß, ist wohl der Hintergrund der am Anfang von Abschnitt @D.I erwähnten irrtümlichen Auffassung, daß man alle Stichprobenkennziffern als Schätzung für die entsprechenden Populationskennziffern verwenden kann. Aber für die Varianz gilt es nicht. Der Anfang des Gedankengangs ist zwar ähnlich wie beim Durchschnitt: Auch die Stichprobenvarianz

ist eine Zufallsgröße (wir können durch Zufall mehr oder weniger extreme Meßwerte in die Stichprobe bekommen), und deshalb hat auch die Stichprobenvarianz eine Wahrscheinlichkeitsverteilung; aber für den Erwartungswert dieser Wahrscheinlichkeitsverteilung läßt sich mathematisch die Gleichung

$$E(s_x^2) = \sigma_x^2 \cdot \frac{N - 1}{N}$$

beweisen. Würden wir also die Stichprobenvarianz als Schätzfunktion für die Populationsvarianz verwenden, dann wäre der Erwartungswert der Schätzfunktion nicht mehr - wie beim Durchschnitt - identisch mit dem zu schätzenden Parameter, sondern um den Faktor $(N - 1) / N$ kleiner als die zu schätzende Populationsvarianz σ_x^2 . Das bedeutet natürlich nicht, daß die Stichprobenvarianz nicht zufällig sogar größer als die Populationsvarianz werden könnte (nämlich dann, wenn wir zufällig viele extrem niedrige und viele extrem hohe Meßwerte in die Stichprobe bekommen); aber tendenziell (eben "im Mittel" oder "im Erwartungswert") wird die Populationsvarianz von der Stichprobenvarianz eher unterschätzt, und zwar um den Faktor $(N - 1) / N$.

<< Daß dies so ist, läßt sich zunächst plausibel machen. Wir legen ja der Varianz die Summe der quadrierten Abweichungen vom Stichprobendurchschnitt \bar{x} zugrunde. Diese ist aber aufgrund der Least-Squares-Eigenschaften des Durchschnitts (vgl. @Statistik I) in aller Regel etwas kleiner als die Summe der quadrierten Abweichungen vom tatsächlichen Populationsmittel μ_x (Ausnahme: Wenn der Stichprobendurchschnitt einmal zufällig genau identisch mit dem Populationsmittel ist; aber das ist nur mit geringer Wahrscheinlichkeit der Fall). Mit anderen Worten: Würden wir das genaue Populationsmittel μ_x kennen und die Varianz als arithmetisches Mittel der quadrierten Abweichungen von μ_x bestimmen, dann wäre der Erwartungswert der so berechneten Varianz gleich der Populationsvarianz σ_x^2 . Die Tatsache, daß wir stattdessen die quadrierten Abweichungen vom Stichprobendurchschnitt \bar{x} zugrundelegen, vermindert aber den Wert der Varianz, und deshalb wird die Populationsvarianz tendenziell unterschätzt. Ein genauerer Beweis folgt etwas später.

>> Da wir aber den Faktor kennen, um den die Populationsvarianz unterschätzt wird, können wir diese Eigenschaft kompensieren, indem wir die aus der Stichprobe berechnete Varianz s_x^2 mit einem entsprechenden Korrekturfaktor multiplizieren. Damit kommen wir zu folgender

$$\text{Schätzregel: } \hat{\sigma}_x^2 = s_x^2 \cdot \frac{N}{N - 1}$$

Diese Schätzregel können wir auch anders schreiben. Da die Stichprobenvarianz ja als $s_x^2 := QS_x / N$ definiert ist (wobei QS_x die Summe der quadrierten Abweichungen vom Stichprobendurchschnitt ist), können wir den Ausdruck QS_x / N anstelle von s_x^2 in obige Schätzregel einsetzen. Dann kürzt sich N weg, und wir erhalten die

$$\text{Schätzregel: } \hat{\sigma}_x^2 = \frac{QS_x}{N - 1}$$

Selbstverständlich ergibt sich nach beiden Schätzregeln derselbe Schätzwert $\hat{\sigma}_x^2$. Es bleibt natürlich dabei, daß dieser Schätzwert durch Stichproben-Zufälle größer oder kleiner ausfallen kann; aber für den Erwartungswert der nach den obigen Schätzregeln bestimmten Schätzfunktion $\hat{\sigma}_x^2$ gilt:

$$E(\hat{\sigma}_x^2) = \sigma_x^2.$$

Zusammengefaßt, erhalten wir also nach dieser Schätzregel eine ähnliche Situation, wie sie sich für den Durchschnitt ohne alle Korrekturen ergab: Der Erwartungswert der Schätzfunktion (oder genauer: der Erwartungswert der Wahrscheinlichkeitsverteilung der Schätzfunktion) ist gleich dem Wert des Parameters, der geschätzt werden soll. Eine Schätzfunktion mit dieser Eigenschaft nennt man auch *erwartungstreu* oder *unverzerrt* (engl.: unbiased).

Definition: Eine Schätzfunktion ist *erwartungstreu* oder *unverzerrt*, wenn der Erwartungswert ihrer Wahrscheinlichkeitsverteilung gleich dem Wert des zu schätzenden Parameters ist. Weicht dieser Erwartungswert dagegen vom Wert des zu schätzenden Parameters ab, dann ist die Schätzfunktion *verzerrt* (engl.: biased).

Mit dem zuletzt eingeführten Begriff können wir z.B. kurz formulieren, was gegen die Verwendung der Stichprobenvarianz als Schätzung für die Populationsvarianz spricht: Wir hätten eine verzerrte Schätzfunktion.

Hinweis: In einigen Lehrbüchern wird die Varianz von vorne herein so eingeführt, daß man die Summe der Abweichungsquadrate (also unser QS_x) durch $N - 1$ und nicht durch N dividiert. Das läuft aber leicht auf eine Verwechslung der beiden unterschiedlichen Ziele des statistischen Umgangs mit Stichprobendaten hinaus: Beschreiben, was in der Stichprobe los ist, und schätzen, was in der Population der Fall ist. Beides wird besser auseinandergehalten, wenn wir klar trennen: Zur *Beschreibung* der Varianz in der Stichprobe dividieren wir QS_x durch N ; zur *Schätzung* der Populationsvarianz dividieren wir QS_x dagegen durch $N - 1$.

Der Begriff der erwartungstreuen Schätzfunktion ist ein Beispiel für den Gewinn, den die Anwendung der Wahrscheinlichkeitstheorie auf Probleme der Inferenzstatistik bringt. Auf den ersten Blick könnte man versucht sein zu fragen: Wenn ich den Wert des Populationsparameters nicht kenne (sonst bräuchte ich ja keine Schätzung!), wie soll ich dann wissen, ob der Erwartungswert der Schätzfunktion gleich dem Wert des Parameters ist? Antwort: Um mathematisch zu beweisen, daß z.B. der Erwartungswert der Schätzfunktion $\hat{\sigma}_x^2$ aus unserer Schätzregel gleich der Populationsvarianz σ_x^2 ist, brauche ich keine Zahlen. Aber wenn ich das weiß, dann heißt das konkret: Ist die Populationsvarianz z.B. 225 oder 169, dann ist der Erwartungswert der Schätzfunktion ebenfalls 225 bzw. 169, und dasselbe gilt für jeden anderen

möglichen Wert der Populationsvarianz. Um festzustellen, daß dies eine wünschenswerte Eigenschaft ist, brauche ich die konkrete Zahl nicht zu kennen.

<< Weil an solchen Stellen ein allgemeiner mathematischer Beweis besonders wichtig ist, soll "für Spezialisten" gezeigt werden, wie man die Gleichung

$$E(QS_x) = (N - 1) \cdot \sigma_x^2$$

beweisen kann. (Daraus folgt dann alles weitere, was über die Erwartungswert von s_x^2 und von $\hat{\sigma}_x^2$ behauptet wurde.) Diese Gleichung ist aber nur die Zusammenstellung des ersten und letzten Glieds einer längeren Gleichungskette, die zunächst einmal hingeschrieben und dann Zeile für Zeile durchgegangen werden soll.

$$\begin{aligned} E(QS_x) &= E(\sum_i X_i^2 - N \cdot \bar{x}^2) \\ &= \sum_i E(X_i^2) - N \cdot E(\bar{x}^2) \\ &= N \cdot (\sigma_x^2 + \mu_x^2) - N \cdot (\sigma_x^2/N + \mu_x^2) \\ &= (N - 1) \cdot \sigma_x^2 \end{aligned}$$

Die erste Zeile beruht auf der aus @Statistik I bekannten Gleichung $QS_x = \sum_i X_i^2 - N \cdot \bar{x}^2$.

Für den Übergang zur zweiten Zeile ist an das in Abschnitt @B.III.1.g erläuterte Prinzip zu erinnern, daß sich die aus der beschreibenden Statistik bekannten Lehrsätze über Kennziffern sinngemäß übertragen lassen.

Der Übergang zur dritten Zeile ist das Kernstück des Ganzen. Um die Erwartungswerte $E(X_i^2)$ und $E(\bar{x}^2)$ weiter verarbeiten zu können, verwenden wir die in Abschnitt @B.3.1.e aus der Varianzformel hergeleitete Gleichung $E(X^2) = \sigma_x^2 + [E(X)]^2$, die für jede Zufallsgröße X gilt. Um diese Gleichung auf den Erwartungswert $E(X_i^2)$ anwenden zu können, erinnern wir uns an das Grundprinzip der Stichprobentheorie aus Abschnitt @C.III, nach welchem jeder Meßwert als eine Zufallsgröße betrachtet wird, deren Wahrscheinlichkeitsverteilung mit der Populationsverteilung des Merkmals identisch ist. Insbesondere ist also $E(X_i)$ (der Erwartungswert der Wahrscheinlichkeitsverteilung des i -ten Meßwerts) gleich dem Populationsmittel μ_x , und die Varianz dieser Wahrscheinlichkeitsverteilung ist gleich der Populationsvarianz σ_x^2 . Also können wir nach obiger Gleichung schreiben:

$$E(X_i^2) = \sigma_x^2 + \mu_x^2.$$

Kehren wir mit diesem Ergebnis zur zweiten Zeile der langen Gleichung zurück, dann können wir feststellen: Bei der Summe $\sum_i E(X_i^2)$ wird für jedes i (insgesamt also N mal) derselbe Wert $\sigma_x^2 + \mu_x^2$ aufaddiert. Daher können wir für diese Summe auch $N \cdot (\sigma_x^2 + \mu_x^2)$ schreiben, und damit ist der erste Teil des Übergangs von der zweiten zur dritten Zeile der langen Gleichung aufgeklärt.

Am Ende der zweiten Zeile taucht noch der Erwartungswert $E(\bar{x}^2)$ auf, und auf diesen Erwartungswert wenden wir wieder die obige allgemeine Gleichung für $E(X^2)$ an. Die Varianz der Wahrscheinlichkeitsverteilung des Stichprobendurchschnitts ist laut Abschnitt @C.V gleich σ_x^2/N (wobei σ_x^2 die Populationsverteilung des Merkmals ist), und aus demselben Abschnitt kennen wir die Gleichung $E(\bar{x}) = \mu_x$ (die wir ja auch im hiesigen Abschnitt zugrundegelegt

haben). Sinngemäß in die obige allgemeine Gleichung für $E(X^2)$ eingesetzt, ergibt das:

$$E(\bar{x}^2) = \sigma_x^2 / N + \mu_x^2.$$

Damit ist auch der letzte Teil des Übergangs von der zweiten zur dritten Zeile unserer langen Gleichung geklärt.

Schließlich ergibt sich die letzte Zeile aus der vorletzten durch einfache Klammer-Operationen usw..

>>

Abschließend soll das Konzept der Erwartungstreue noch einmal mit allgemein anwendbarer mathematischer Symbolik dargestellt werden. Dabei verwenden wir für den zu schätzenden Parameter das allgemeine Symbol π . Eine Schätzfunktion dieses Parameters nennen wir $\hat{\pi}$. Dann ist ϕ , der Fehler bei der Schätzung des Parameters π , gegeben durch die Definition⁷⁵

$$\phi := \hat{\pi} - \pi.$$

Beispiel: Wenn der zu schätzende Parameter das Populationsmittel μ_x ist, das den Wert $\mu_x = 62.3$ hat, und unser Stichprobendurchschnitt, den wir als Schätzfunktion für μ_x verwenden, 61.7 beträgt, dann gilt für den Fehler:

$$\phi = \hat{\mu} - \mu = 61.7 - 62.3 = -0.6.$$

In Worten: Wir haben das Populationsmittel um 0.6 Einheiten unterschätzt (wobei sich die Unterschätzung in dem negativen Vorzeichen von ϕ ausdrückt).

Mit diesen Symbolen können wir dann die Definition der Erwartungstreue folgendermaßen formulieren:

Eine Schätzfunktion $\hat{\pi}$ für einen Parameter π ist erwartungstreu oder unverzerrt, wenn die Gleichung

$$E(\hat{\pi}) = \pi$$

erfüllt ist, und diese ist gleichbedeutend mit

$$E(\phi) = 0.$$

⁷⁵In der Regressionsstatistik hatten wir den Schätzfehler als Differenz "wahrer Wert minus Schätzwert" (also $Y_i - \hat{Y}_i$) definiert. Nachträglich läßt sich ergänzen, warum das nützlich war: Dadurch kamen wir unmittelbarer zu der Zerlegung der Abweichung eines Meßwerts Y_i vom Durchschnitt \bar{y} nach der Gleichung

$$Y_i - \bar{y} = (Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{y}),$$

die dann zur Zerlegung der Varianz σ_y^2 in die Fehlervarianz und die determinierte Varianz führte. In der Theorie der Parameterschätzung wird dagegen davon ausgegangen, daß ein positiver Fehler vorliegt, wenn wir einen Parameter überschätzen, und negativer, wenn wir den Parameter unterschätzen. Das führt dann zu der Definition $\phi := \hat{\pi} - \pi$.

Da wir uns vor allem dafür interessieren, ob der Erwartungswert von ϕ null ist und wie groß die Varianz der Wahrscheinlichkeitsverteilung von ϕ ist, würde es keinen großen Unterschied machen, wenn wir - analog zur Regressionsstatistik - die Differenz $\pi - \hat{\pi}$ als Schätzfehler betrachten würden: An der Antwort zu den beiden genannten Fragen würde sich nichts ändern. In der Theorie der Parameterschätzung ist aber die andere Definition eher eingebürgert, und deshalb halten wir uns daran.

Anderenfalls (also bei $E(\hat{\pi}) \neq \pi$ bzw. $E(\phi) \neq 0$) ist die Schätzfunktion verzerrt.

2. Minimalvarianz

Die Erwartungstreue einer Schätzfunktion ist ein intuitiv einleuchtendes Kriterium; aber daß sie nicht ausreichend für eine gute Schätzfunktion ist, zeigt eine einfache Überlegung. Wenn wir die (offenbar unsinnige) Schätzregel zugrundelegen würden, einfach den Wert X_1 (also den ersten Meßwert unserer Stichprobe) als Schätzung für das Populationsmittel zugrundelegen, dann hätten wir immerhin eine erwartungstreue Schätzfunktion; denn (wie bei jedem Meßwert) ist auch der Erwartungswert von X_1 gleich dem Populationsmittel μ_x .

Daß diese Schätzregel unsinnig ist, liegt daran, daß dabei viel eher als bei der Verwendung des Stichprobendurchschnitts große Fehler nach oben oder nach unten auftreten können. Eine Kennziffer, die das ausdrückt, ist die Varianz der Wahrscheinlichkeitsverteilung der Schätzfunktion: Würden wir den Meßwert X_1 zugrundelegen, dann wäre diese Varianz gleich der Populationsvarianz σ_x^2 . Verwenden wir dagegen den Durchschnitt als Schätzfunktion, dann ist die Varianz der Wahrscheinlichkeitsverteilung dieser Schätzfunktion natürlich die (aus Abschnitt @C.V bekannte) Größe σ_x^2 / N , also erheblich kleiner.

Unter Verwendung der Symbole aus Abschnitt @D.II.1 können wir die Varianz der Wahrscheinlichkeitsverteilung von $\hat{\pi}$ als $\sigma_{\hat{\pi}}^2$ bezeichnen. Diese ist aber gleich der Varianz des Schätzfehlers ϕ ; es gilt also die Gleichung

$$\sigma_{\hat{\pi}}^2 = \sigma_{\phi}^2$$

Begründung: ϕ entsteht aus $\hat{\pi}$, indem wir die zwar unbekannte, aber doch feste Größe ("Konstante") π von $\hat{\pi}$ subtrahieren. Die Subtraktion einer Konstanten ändert aber nichts an der Varianz.

Diese Gleichung macht deutlich, warum die Varianz, deren Bedeutung am Beispiel erklärt wurde, auch als *Schätzfehlervarianz* bezeichnet wird. Ihre Quadratwurzel ist der *Standardschätzfehler*. Genauer spricht man vom Standardschätzfehler bei der Schätzung von π .

<<

Die Bezeichnung Schätzfehlervarianz kennen wir bereits aus der Regressions-Statistik. Dort konnten wir einfach das arithmetische Mittel der quadrierten Schätzfehler als Schätzfehlervarianz bezeichnen; denn da der Durchschnitt der Schätzfehler dort immer null war, konnten wir darauf verzichten, bei der Berechnung der Varianz der Schätzfehler das Quadrat dieses Durchschnitts vom arithmetischen Mittel der quadrierten Schätzfehler zu subtrahieren. Bei der Parameterschätzung können wir unter Rückgriff auf die Varianz-Formel $\sigma_x^2 = E(X^2) - [E(X)]^2$ zunächst schreiben:

$$\sigma_{\phi}^2 = E(\phi^2) - [E(\phi)]^2.$$

Auf die Subtraktion des quadrierten Erwartungswerts des Schätzfehlers können wir hier nur dann verzichten, wenn dieser Erwartungswert 0 ist, also bei erwartungstreuen Schätzfunktionen.

>>

Offenbar ist nun eine Schätzfunktion umso besser, je geringer diese Schätzfehlervarianz: Kommen nur geringe Fehler vor, so ist die Schätzfehlervarianz niedrig; kommen dagegen mit nennenswerter Wahrscheinlichkeit auch große Fehler vor, so ist die Schätzfehlervarianz, groß. Für eine gute Schätzfunktion ergibt sich daher zusätzlich zur Erwartungstreue die

Forderung nach *Minimalvarianz*: Habe ich die Wahl zwischen zwei Schätzregeln, bei denen die Schätzfehlervarianz der jeweiligen Schätzfunktion verschieden groß ist und die ansonsten gleich gut sind, so ist diejenige mit der geringeren Schätzfehlervarianz vorzuziehen.

Der Zusatz "und die ansonsten gleich gut sind" ist dabei wichtig; denn natürlich könnte man die Schätzfehlervarianz beliebig klein machen, indem man eine Schätzfunktion mit einer hinreichend nahe bei 0 liegenden Zahl multipliziert.⁷⁶ Dabei geht dann aber die Erwartungstreue verloren, falls sie vorher vorlag, und deshalb ist der Zusatz "und die ansonsten gleich gut sind" verletzt.

<<

Erwartungstreue und Minimalvarianz lassen sich auch zusammenfassen in der Formulierung, daß $E(\phi^2)$ (also der Erwartungswert des quadrierten Schätzfehlers) möglichst klein sein soll. Eine Schätzfunktion, die diese Forderung erfüllt, wird auch *effizient* genannt. Offenbar wird dabei das Prinzip der kleinsten Quadrate auch auf Probleme der Parameterschätzung übertragen.

>>

3. Konsistenz und weitere Kriterien

Es gibt noch einige weitere Kriterien für gute Schätzfunktionen. Am ehesten erwähnenswert ist die *Konsistenz*. Dazu gehört insbesondere, daß die Schätzfehlervarianz umso kleiner wird, je größer die Stichprobe wird, und für eine gegen unendlich strebende Stichprobengröße null wird. Beim Stichprobendurchschnitt als Schätzfunktion für das Populationsmittel ergibt sich dies z.B. aus der in Abschnitt @ eingeführten Formel $\sigma_{\bar{x}}^2 = \sigma_x^2 / N$: Da die Stichprobengröße N im Nenner steht, wird der Bruch bei einer gegen unendlich strebenden Stichprobengröße null.

<<

Genauer: Für jede noch so kleine positive Zahl ε muß die Wahrscheinlichkeit, daß $|\phi|$ (also der Absolutbetrag des Schätzfehlers) kleiner als ε wird, gegen 0 streben. In mathematischen Symbolen formuliert, muß für die Konsistenz die Gleichung

$$\lim_{N \rightarrow \infty} p(|\phi| < \varepsilon) = 1$$

für jede noch so kleine positive Zahl ε gelten.

⁷⁶Unmittelbar einzusehen ist, daß damit die Varianz $\sigma_{\hat{\pi}}^2$ klein wird; denn die Multiplikation einer Zufallsgröße (hier: der Schätzfunktion $\hat{\pi}$) mit einer Konstanten b führt ja zu einer Multiplikation der Varianz mit b^2 (vgl. Abschnitt @B.III.1.g). Da aber die Varianzen $\sigma_{\hat{\pi}}^2$ und σ_{ϕ}^2 gleich sind, führt die Multiplikation der Schätzfunktion mit einer hinreichend nahe bei 0 gelegenen Konstanten b auch zu einer beliebig kleinen Fehlervarianz.

Ein weiteres Prinzip, das *Maximum-Likelihood-Prinzip*, läßt sich folgendermaßen formulieren: Wähle als Schätzwert für einen Parameter diejenige Zahl, bei der die Wahrscheinlichkeit der beobachteten Daten maximal ist.

Beispiel: In einer Situation, in der die Voraussetzungen der Binomialverteilung erfüllt und n (die Zahl der "Versuche") bekannt ist, soll die Grundwahrscheinlichkeit als Parameter π geschätzt werden. Die "beobachteten Daten" bestehen in der Zahl X der Versuche, in denen das kritische Ereignis aufgetreten ist. Dann läßt sich, wie wir aus Abschnitt @ wissen, die Wahrscheinlichkeit dieser Daten angeben als

$$\binom{n}{X} \cdot \pi^X \cdot (1 - \pi)^{n-X}$$

Mit Hilfe der Differentialrechnung kann man nun zeigen, daß derjenige Wert des Parameters π , bei dem diese Wahrscheinlichkeit am größten ist, durch den Quotienten X/n gegeben wird. Also erfüllt die Schätzfunktion $\hat{\pi} = X/n$ das Maximum-Likelihood-Prinzip.

Nur angedeutet werden kann der Begriff der "erschöpfenden Schätzfunktion". Er besagt, daß die Schätzfunktion alle in den Daten enthaltene Information über den Parameter ausschöpft. Da diese Eigenschaft von einiger Bedeutung für die Testtheorie ist, wird der Begriff in einschlägigen Lehrbüchern (z.B. Rost, 1996) genauer behandelt.

>>

E "Signifikanztests" oder das "Testen von Hypothesen"

I. Das Problem

Nehmen wir an, daß wir aus der Population der deutschen Männer und aus der Population der deutschen Frauen jeweils eine Zufallsstichprobe gezogen hätten und daß in einem Intelligenztest die Frauen durchschnittlich 105 Punkte erzielt hätten, die Männer durchschnittlich 103 Punkte. Die beste Parameterschätzung wäre dann, daß das Populationsmittel der Frauen 105 Punkte beträgt, das Populationsmittel der Männer 103 Punkte. Diese Schätzung ist aber, wie unter D besprochen wurde, mit Zufallsfehlern behaftet, und es wäre denkbar, daß die Differenz nur durch solche Zufallsfehler entstanden ist, daß also in der Population Männer und Frauen gleich intelligent sind und daß wir rein zufällig mehr überdurchschnittliche Frauen und unterdurchschnittliche Männer in die Stichprobe bekamen.

Um zu überprüfen, ob eine solche Abweichung der tatsächlichen Daten von einer bestimmten Hypothese noch durch Zufall erklärt werden kann, oder ob sie als statistisch bedeutsam ("signifikant") anzusehen ist, verwenden wir Signifikanztests.

Wir nennen die als Gleichung formulierte Hypothese die *Nullhypothese* (H_0). Im obigen

Beispiel:

$$H_0: \mu_F = \mu_M$$

Die Gegenhypothese ("Alternativhypothese") wird auch H_1 genannt. Hier:

$$H_1: \mu_F \neq \mu_M$$

Manchmal gibt es auch mehrere Alternativhypothesen.

Beweisen können wir mit unserer Stichprobe keine der beiden Hypothesen, aber oft können wir aufzeigen, daß unsere Daten bei Gültigkeit der Nullhypothese sehr unwahrscheinlich sind. Das spräche dann für die Alternativhypothese. Dagegen können wir das Gegenteil nie aufzeigen. Beispiel: Selbst wenn in der Stichprobe die Durchschnitte von Frauen und Männern beide genau 104 Punkte betragen hätten, wäre es zwar möglich, daß gemäß H_0 gilt: $\mu_F = \mu_M = 104$; die Daten können aber mit fast gleicher Wahrscheinlichkeit auftreten, wenn

$$\mu_F = 103.999 \text{ und}$$

$$\mu_M = 104.001 \text{ wäre.}$$

Damit wäre aber die Alternativhypothese richtig. Es gibt also auch innerhalb der Alternativhypothese Möglichkeiten, unter denen die Daten nicht unwahrscheinlich sind.

<< Es gibt verschiedene Ansätze zur Beantwortung solcher Fragen. In der Psychologie wird in der Regel eine Art Mischform angewandt, der zwei solche Ansätze kombiniert: Der erste geht auf Fisher zurück, der zweite auf Neyman und Pearson. Zu Lebzeiten haben sich diese Autoren geradezu kriegsartig bekämpft. Daß die Psychologie sich nicht zwischen beiden entscheidet, sondern eine Mischung beider - ein "Hybrid"-Modell - verwendet, wird von einigen Autoren kritisiert (z.B. Gigerenzer). Die Wahl des Worts "Hybrid" (in der Genetik eine Bezeichnung für "Mischlinge") gibt allerdings Anlaß zu der Frage, warum denn Mischlinge automatisch minderwertig sind. Tatsächlich beruht ja Fortschritt in der Evolution nicht nur auf Mutationen, sondern oft auch auf der "Rekombination" von Erbanlagen, die sich in verschiedenen Umwelten als günstig erwiesen haben. Ähnlich gibt es gute Argumente für den Standpunkt, daß die Psychologie eine Umwelt ist, für die eine Kombination des "geistigen Erbguts" zweier Ansätze am günstigsten ist, die vor allem in Hinblick auf andere Umwelten entwickelt wurden: Der Fisher'sche für die agrarwissenschaftliche Züchtungsforschung, und der Neyman-Pearson'sche für wirtschaftliche Entscheidungen.

Der folgende Abschnitt II demonstriert den ursprünglichen Ansatz von Fisher, während Abschnitt III den stärker entscheidungstheoretischen Ansatz von Neyman und Pearson darstellt. Auf Vorteile des "Fisher-Neyman-Pearson-Hybrids" wird später eingegangen (S. @209), wenn an Beispielen dargestellt ist, wie es konkret aussieht.

>>

II. Drei Beispiele

1) Der Vorzeichentest

Als Beispiel betrachten wir ein Experiment: Bei 200 Vpn wurde die Schlaftiefe eine Stunde und fünf Stunden nach dem Einschlafen gemessen. Bei 70 Vpn war die Schlaftiefe nach fünf Stunden tiefer (dieses Ergebnis wollen wir auch "+" nennen), bei den übrigen 130 Vpn war das Gegenteil ("-") der Fall

Frage: Kann es sein, daß + und - "in der Population" gleich häufig auftreten, daß also gilt:

$$H_0 : p_+ = p_- = 0.5?$$

Die Alternativhypothese würde lauten:

$$H_1 : p_+ \neq p_-$$

oder

$$H_1 : p_+ \neq 0.5.$$

Wir fragen uns: *Wie groß ist die Wahrscheinlichkeit, daß eine so große Abweichung von der Nullhypothese, wie wir sie in unseren Daten auffanden, oder eine noch extremere Abweichung auftritt, wenn H_0 richtig ist?*

Wir nennen die Zahl der "+" Fälle einmal x . In unserem Fall ist also $x = 70$. Ist H_0 richtig, so erwarten wir ein x von etwa 100. Jedes $x \leq 70$ und jedes $x \geq 130$ weicht genau so stark wie unsere Daten oder noch stärker von H_0 ab. Wir stellen mit Hilfe der Binomialverteilung (die wir hier durch eine Normalverteilung approximieren können) fest, daß ein solches x bei Gültigkeit von H_0 äußerst unwahrscheinlich ist. Die Abweichung unserer Daten von H_0 ist also signifikant, so daß wir H_0 verwerfen können, und zwar mit einer äußerst geringen *Irrtumswahrscheinlichkeit*. (näheres S. @).

Durchführung der Berechnungen zum Schlafexperiment:

Die Wahrscheinlichkeit, einen x -Wert von höchstens 70 oder mindestens 130 zu erhalten, wenn die Nullhypothese stimmt, berechnen wir folgendermaßen:

Wenn die Nullhypothese stimmt, hat die Größe x eine Binomialverteilung mit $N=200$ Versuchen und einer Grundwahrscheinlichkeit von $p=0.5$. Erwartungswert und Varianz dieser Binomialverteilung können wir nach den allgemeinen Formeln für die Binomialverteilung bestimmen:

$$E(x) = N \cdot p = 200 \cdot 0.5 = 100$$

$$\sigma_x^2 = N \cdot p \cdot (1 - p) = 200 \cdot 0.5 \cdot 0.5 = 50$$

Diese Binomialverteilung können wir mit Hilfe einer Normalverteilung mit gleichem Mittelwert und gleicher Varianz approximieren. Dabei ist zu beachten: da die Normalverteilung kontinuierlich ist, die Binomialverteilung dagegen diskret, müssen wir davon ausgehen, daß dem Wert 70 in der Binomialverteilung das Intervall von 69.5 bis 70.5 entspricht, und dem Wert 130 in der Binomialverteilung entspricht das Intervall von 129.5 bis 130.5. Wir fragen also: Wie groß

ist die Wahrscheinlichkeit, in einer Normalverteilung mit Mittelwert 100 und Varianz 50 einen Wert von unter 70.5 oder über 129.5 zu bekommen?

Ergebnis der Berechnung (die als Übungsaufgabe selbst gerechnet werden kann): Die Wahrscheinlichkeit beträgt ungefähr 0.00003.

Also: Es ist äußerst unwahrscheinlich, eine so große Abweichung von der Nullhypothese, wie wir sie erhielten, oder eine noch größere durch Zufall zu erhalten, wenn in Wirklichkeit die Nullhypothese gilt.

2) Der *t*-Test für korrelierte Stichproben

Ein weiteres Beispiel für einen Signifikanztest: Wir wollen überprüfen, ob Koffein die Leistung in einem Test beeinflusst. Zu diesem Zweck testen wir 74 Vpn einmal nach der Verabreichung einer Coffeintablette (Testwert der *i*-ten Vp ist x_i) und einmal mit "Placebo"⁷⁷ (Testwert der *i*-ten Vp ist y_i). Hat Koffein keinen Einfluß auf die Testwerte, so erwarten wir zumindest im Populationsmittel gleiche Leistung, also:

$$H_0: \mu_x = \mu_y$$

oder

$$H_0: \mu_x - \mu_y = 0$$

Für jede Vp berechnen wir:

$$d_i = x_i - y_i$$

d_i drückt aus, wie weit der Testwert unter Koffein (x_i) über dem unter Placebo (y_i) liegt. Dann ist:

$$\mu_d = \mu_x - \mu_y$$

Wir können daher die Nullhypothese auch folgendermaßen formulieren:

$$H_0: \mu_d = 0$$

Die Alternativhypothese lautet:

$$H_1: \mu_d \neq 0.$$

Für unsere 74Vpn ergab sich:

$$\bar{x} = 105.6; \bar{y} = 105.3;$$

also ist

$$\bar{d} = 105.6 - 105.3 = 0.3$$

Unser tatsächliches \bar{d} weicht von der Nullhypothese ab. Frage: Kann das Zufall sein? Genauer:

⁷⁷"Placebo" ist eine Tablette, die genauso aussieht wie das Präparat, aber keine wirksamen Stoffe enthält. Die Vergleichswerte erheben wir nicht ohne Tablette, sondern unter "Placebo", um zu verhindern, daß eventuelle Unterschiede von x und y nicht auf das Koffein zurückgehen, sondern auf die psychischen Rückwirkungen auf das Bewußtsein, eine Tablette eingenommen zu haben.

Wie groß ist die Wahrscheinlichkeit, daß eine so große oder noch extremere Abweichung von der Nullhypothese auftritt, wenn H_0 richtig ist? Also: Wie groß ist die Wahrscheinlichkeit eines \bar{d} über 0.3 oder unter -0.3, wenn H_0 richtig ist?

Um diese Frage zu beantworten, werden wir eine neue Wahrscheinlichkeitsverteilung einführen müssen, die "t-Verteilung". Das Verständnis dieser Verteilung wird jedoch erleichtert, wenn wir die obige Frage zunächst für eine fiktive Situation beantworten, für die unsere bisherigen Kenntnisse ausreichen. Dazu soll *zunächst* angenommen werden, es sei bereits bekannt, daß die Populationsvarianz von d 4.00 ist und daß die Populationsverteilung der d-Werte eine Normalverteilung ist. Dann könnten wir (unter Rückgriff auf Ergebnisse aus Abschnitt @C: Stichprobentheorie) sagen: Gilt die Nullhypothese, dann ist die Wahrscheinlichkeitsverteilung des Stichprobendurchschnitts \bar{d} eine Normalverteilung mit Erwartungswert 0 und mit der Standardabweichung

$$\sigma_{\bar{d}} = \frac{\sigma_d}{\sqrt{N}} = \frac{2.0000}{\sqrt{74}} = 0.2325$$

Um die Wahrscheinlichkeit zu bestimmen, daß der Stichprobendurchschnitt \bar{d} bei Gültigkeit dieser Annahmen größer als 0.30 oder kleiner als -0.30 wird, wäre der Wert 0.30 in einen z-Wert umzurechnen, und die zweiseitige Überschreitungswahrscheinlichkeit dieses z-Werts könnten wir mit Hilfe der z-Tabelle bestimmen. Es ergäbe sich also:

$$z = \frac{\bar{d} - \mu_{\bar{d}_0}}{\sigma_{\bar{d}}} = \frac{0.30 - 0}{0.2325} = 1.29$$

Aus der z-Tabelle können wir für einen z-Wert von 1.29 eine kumulative Wahrscheinlichkeit von 0.90 entnehmen, so daß die "zweiseitige Überschreitungswahrscheinlichkeit" 0.20 ist.

Damit können wir die Berechnungen für die fiktive Situation (bekannte Populationsvarianz $\sigma^2_d=4.00$ und Normalverteilung der d-Werte) folgendermaßen zusammenfassen: Auch wenn die Nullhypothese $\mu_d=0$ zutrifft, beträgt die Wahrscheinlichkeit, einen Stichprobendurchschnitt von $\bar{d} \geq 0.30$ oder $\bar{d} \leq -0.30$ zu erzielen, etwa 20%. Nach einer Konvention, die erst später zu begründen sein wird, würde man diese Abweichung von der Nullhypothese noch nicht als signifikant betrachten.

Diese Berechnung war in dieser Form aber nur unter zwei vereinfachenden Annahmen gültig: Die Populationsvarianz sollte mit $\sigma^2_d=4.00$ bekannt, und die Populationsverteilung der d-Werte eine Normalverteilung sein. Nun könnten wir die Populationsvarianz ja aus den Daten schätzen. Dazu sei angenommen, daß sich in unserer Stichprobe eine Stichprobenvarianz der d-Werte von

$$s^2_d = 3.946$$

ergab. Daraus können wir die Populationsvarianz σ_d^2 nach einer aus Abschnitt @D bekannten Regel schätzen:

$$\hat{\sigma}_d^2 = s_d^2 \cdot \frac{N}{N-1} = 3.946 \cdot \frac{74}{73} = 4.000$$

Der geschätzte Wert ist also genau der Wert, der zuvor für die fiktive Situation als vorab bekannte Populationsvarianz angenommen wurde. Das berechtigt uns aber nicht, so zu tun, als ob die wahre Populationsvarianz tatsächlich 4.00 wäre. Es geht bei der Inferenzstatistik ja gerade darum, mögliche Stichprobenfehler unserer Schätzwerte einzukalkulieren. Für unseren Fall heißt das: Die Wahrscheinlichkeitsverteilung des z-Werts, den wir in unserem fiktiven Beispiel zugrundegelegt haben, ist nur dann eine Standard-Normalverteilung, wenn im Nenner der entsprechenden Formel die wahre Standardabweichung der Wahrscheinlichkeitsverteilung des Stichprobendurchschnitts \bar{d} steht. Ersetzen wir diese Standardabweichung durch eine Schätzung, dann haben wir zusätzlich zu den (im Zähler des Bruchs berücksichtigten) Zufallsfehlern des Stichprobendurchschnitts \bar{d} noch eine weitere Quelle von Zufallsfehlern: Auch die Stichproben-Varianz der d-Werte, aus der wir die Schätzung für den Wert im Nenner des z-Bruchs abgeleitet haben, ist mit Stichprobenfehlern behaftet. Um dies zu berücksichtigen, gibt es die "t-Verteilung", die ebenfalls tabelliert ist. Zunächst eine Faustregel für ihre Anwendung:

Faustregel: Ersetzen wir die Standardabweichung im Nenner eines standardnormalverteilten z-Bruchs durch eine Schätzung, die auf einer erwartungstreuen Varianzschätzung beruht, dann ist die Wahrscheinlichkeitsverteilung des so veränderten Bruchs meist⁷⁸ eine t-Verteilung.

Für unser Beispiel bedeutet das: Genau so, wie wir vorher in der fiktiven Situation aus der als bekannt vorausgesetzten Populationsvarianz σ_d^2 eine Standardabweichung σ_d und daraus eine Standardabweichung $\sigma_{\bar{d}}$ abgeleitet haben, können wir jetzt aus der geschätzten Populations-Varianz $\hat{\sigma}_d^2$ eine geschätzte Populations-Standardabweichung $\hat{\sigma}_d$ und eine geschätzte Standardabweichung $\hat{\sigma}_{\bar{d}}$ der Wahrscheinlichkeitsverteilung des Stichprobendurchschnitts \bar{d} berechnen:

$$\hat{\sigma}_d = \sqrt{\hat{\sigma}_d^2} = \sqrt{4.000} = 2.0000$$

und

⁷⁸"Meist" heißt konkret: In allen für uns in diesem Kurs relevanten Fällen.

$$\hat{\sigma}_{\bar{d}} = \frac{\hat{\sigma}_d}{\sqrt{N}} = \frac{2.0000}{\sqrt{74}} = 0.2325$$

Die Schätzwerte sind also dieselben, wie die zuvor in der fiktiven Situation mit bekannter Populationsvarianz zugrundegelegt wurden. Nun wenden wir die obige Faustregel an: Wenn wir den letzten dieser Schätzwerte im Nenner des z-Bruchs aus der fiktiven Situation einsetzen, dann hat der so veränderte Bruch eine t -Verteilung. Wir berechnen also

$$t = \frac{\bar{d} - \mu_{\bar{d}_0}}{\hat{\sigma}_{\bar{d}}} = \frac{0.30 - 0}{0.2325} = 1.29$$

Wir erhalten also denselben Wert, wie zuvor für z ; aber seine (zweiseitige) Überschreitungswahrscheinlichkeit ist nicht nach der Tabelle der Standard-Normalverteilung zu bestimmen, sondern nach einer Tabelle für die t -Verteilung.

Dazu muß man wissen: Es gibt unendlich viele t -Verteilungen, die sich voneinander durch die sogenannte Zahl der Freiheitsgrade (abgekürzt df für "degrees of freedom") unterscheiden. Diese Freiheitsgrade hängen von der Größe der Stichprobe ab, auf der die Varianzschätzung beruhte. Dazu gibt es zu allen Verfahren, bei denen t -Verteilungen zur Anwendung kommen, spezielle Regeln.

Im übrigen sind die t -Verteilungen symmetrisch um den Nullpunkt und "mit bloßem Auge" kaum von einer Standard-Normalverteilung zu unterscheiden. In Statistikbüchern sind sie meist so tabelliert, daß zu den wichtigsten Prozentwerten die t -Werte mit der entsprechenden einseitigen bzw. zweiseitigen Überschreitungswahrscheinlichkeit angegeben sind. Bei mehr als 120 Freiheitsgraden ist die t -Verteilung fast völlig gleich der Standard-Normalverteilung, so daß wir für mehr als 120 Freiheitsgrade die z -Tabelle benutzen können.

<< Mathematisch gilt: Ein standardnormalverteilter z -Wert ist der "Grenzfall" eines t -Werts für unendlich viele Freiheitsgrade. Als Gleichung formuliert: Bezeichnen wir mit $t(n, Q)$ den t -Wert, dessen einseitige Überschreitungswahrscheinlichkeit bei t -Verteilung mit n Freiheitsgraden Q ist, und mit $z(Q)$ den z -Wert, dessen einseitige Überschreitungswahrscheinlichkeit in der Standard-Normalverteilung Q ist, dann gilt für jede Wahrscheinlichkeit Q :

$$\lim_{n \rightarrow \infty} t(n, Q) = z(Q).$$

Entsprechendes gilt auch, wenn wir statt der einseitigen Überschreitungswahrscheinlichkeiten zweiseitige zugrundelegen. Daher sind in unserer t -Tabelle auch die z -Werte mit einseitiger Überschreitungswahrscheinlichkeit Q bzw. zweiseitiger Überschreitungswahrscheinlichkeit $2Q$ als t -Werte für unendlich viele Freiheitsgrade aufgeführt.

>>

Für die Freiheitsgrade gilt in unserem Fall die Regel:

$$df = N - 1$$

Für $N=74$ Vpn erhalten wir also $df = 74 - 1 = 73$. In unserer t -Tabelle ist die t -Verteilung für 73 Freiheitsgrade nicht enthalten. Wir können ihr aber entnehmen, daß bei 60 Freiheitsgraden die Wahrscheinlichkeit, einen t -Wert von über 1.296 oder unter -1.296 zu erhalten, 20% beträgt. Man sagt auch: Die *zweiseitige* Überschreitungswahrscheinlichkeit beträgt 20% (zweiseitig, weil der positive und der negative "Schwanz" der Verteilung einbezogen sind). Für 120 Freiheitsgrade lautet der entsprechende t -Wert 1.289. Der t -Wert für 73 Freiheitsgrade und 20% zweiseitige Überschreitungswahrscheinlichkeit läge dann zwischen 1.289 und 1.296, also ziemlich genau bei unserem Wert von 1.29.

Fassen wir zusammen: Wir haben die "Prüfgröße" t berechnet, die sich - anders als das x beim vorhergehenden Test - aus mehreren Stichprobenkennziffern bzw. Schätzfunktionen zusammensetzt. Der Wert 1.29, den diese Prüfgröße annimmt, weicht zwar etwas von dem unter H_0 zu erwartenden Wert ($t=0$) ab, doch können so große oder noch extremere Abweichungen auch bei Gültigkeit von H_0 noch mit etwa 20% Wahrscheinlichkeit auftauchen. Würden wir also aufgrund unserer Daten die Nullhypothese verwerfen und uns für H_1 entscheiden, so betrüge die Irrtumswahrscheinlichkeit 20%. Wir sollten also die Nullhypothese zwar nicht als bewiesen akzeptieren, aber doch weiterhin mit der Möglichkeit rechnen, daß sie richtig ist. Man sagt: Die Nullhypothese wird *beibehalten*.

<< Das Ergebnis stimmt also mit demjenigen überein, das wir für die fiktive Situation unter Zugrundelegung der Standard-Normalverteilung errechnet haben. Das liegt an unserer geringen Rechengenauigkeit. Bei Rechengenauigkeit mit zwei Stellen hinter dem Komma kann die t -Verteilung schon bei weniger als 120 Freiheitsgraden mit einer Normalverteilung approximiert werden. Umgekehrt könnten wir bei größerer Rechengenauigkeit feststellen, daß für die fiktive Ausgangssituation die (zweiseitige) Überschreitungswahrscheinlichkeit eines genauer berechneten z -Werts geringfügig kleiner ist, als die zweiseitige Überschreitungswahrscheinlichkeit desselben, als t -Wert mit 73 Freiheitsgraden interpretierten Werts. Das kann man indirekt auch unserer t -Tabelle entnehmen: Die z -Werte mit Überschreitungswahrscheinlichkeit Q bzw. $2Q$ sind als t -Werte mit unendlich vielen Freiheitsgraden angegeben. Sie sind durchweg etwas kleiner als t -Werte mit endlich vielen Freiheitsgraden, und das bedeutet auch: Die zweiseitige Überschreitungswahrscheinlichkeit derselben Zahl ist etwas kleiner, wenn wir sie als standardnormalverteilten z -Wert interpretieren, als wenn wir diese Zahl als t -Wert mit endlich vielen Freiheitsgraden verstehen.

>> Damit haben wir eine befriedigende Alternative für eine der beiden Annahmen unserer fiktiven Ausgangssituation: Wir brauchen die Populationsvarianz der d -Werte nicht zu kennen; wenn wir stattdessen unseren Berechnungen eine erwartungstreue Schätzung dieser Populationsvarianz zugrundelegen, dann brauchen wir lediglich an der Stelle, wo man bei bekannter Populationsvarianz die z -Tabelle zugrundelegt, zur t -Tabelle zu wechseln.

Für die Normalverteilungs-Annahme unserer fiktiven Ausgangssituation haben wir jedoch

noch keine Alternative, und das ist typisch für viele Signifikanztests: Die von ihnen verwendete "Prüfgröße" (in unserem Fall die Größe t) hat auch bei Gültigkeit der Nullhypothese die entsprechende Wahrscheinlichkeitsverteilung nur dann, wenn die Populationsverteilungen der untersuchten Variablen bestimmte "Zusatzvoraussetzungen" erfüllen. Tatsächlich weiß man natürlich in aller Regel nicht, ob sie erfüllt sind. Wie man mit diesem Problem umgeht, wird in Abschnitt @ zu klären sein. Einstweilen ist festzuhalten, daß unsere Prüfgröße t auch bei Gültigkeit der Nullhypothese nur dann exakt eine t -Verteilung hat, wenn die Populationsverteilung der d -Werte eine Normalverteilung ist.

3) Der U-Test

Wenn wir die Auswirkung von Coffein und Placebo auf Leistungen in einer Aufgabe bestimmen wollen, die sich nicht zweimal mit denselben Vpn durchführen läßt, können wir einen anderen Versuchsplan wählen: Wir teilen die Vpn per Zufall (z.B. im Losverfahren) in 2 Gruppen ein; die Vpn der einen Gruppe bekommen vor der Aufgabe eine Coffein-Tablette, die anderen ein Placebo. Man spricht dann von "unabhängigen Stichproben" (Näheres s. S.@ und @). Auch für solche Versuchspläne gibt es einen t -Test (der später behandelt wird, vgl. @); an dieser Stelle soll jedoch ein anderer Signifikanztest demonstriert werden, der aufgrund unserer Kenntnisse der Wahrscheinlichkeitstheorie zumindest in einer kleinen Stichprobe besser nachvollziehbar ist: Der U-Test von Mann und Whitney.

Wir nehmen an, daß 9 Vpn nach Zufall in 2 Gruppen eingeteilt werden; 4 Vpn lösen die Aufgabe unter Coffein und 5 unter Placebo. Gemessen wird jeweils die Lösungszeit.

Beim U-Test kommt es nun gar nicht auf die genauen Meßwerte der Vpn beider Gruppen an, sondern nur auf ihre Rangfolge: Aus welcher Stichprobe stammt der niedrigste Meßwert, der zweitniedrigste usw. Nehmen wir an, daß sich die folgende Abfolge von Mitgliedern der Coffein-Gruppe (C) und der Placebo-Gruppe (P) ergibt:

C C P C C P P P P

Es zeichnet sich eine Tendenz dahingehend ab, daß die Mitglieder der C-Gruppe eher am Anfang der Rangreihe liegen (niedrige Lösungszeiten) und die Mitglieder der P-Gruppe eher am Ende. Wie bei allen Signifikanztests stellt sich die Frage: Kann das noch Zufall sein? Oder genauer: Mit welcher Wahrscheinlichkeit können so große oder noch größere Abweichungen von der Zufallserwartung zufällig entstehen?

Um diese Frage mit dem U-Test zu beantworten, vergleichen wir jeden Wert der C-Gruppe mit jedem Wert der P-Gruppe und zählen aus, bei wie vielen dieser Vergleiche der Wert aus der C-Gruppe größer ist, als der aus der P-Gruppe. In unserem Fall wird also jede der 4 Personen der C-Gruppe mit jeder der 5 Personen verglichen; das sind insgesamt 20 Vergleiche. Von diesen 20 Vergleichen ergeben nur 2 einen höheren Wert in der C-Gruppe (nämlich die beiden Vergleiche

der an 4. und 5. Stelle stehenden Werte der C-Gruppe mit dem an 3. Stelle stehenden Wert der P-Gruppe). Diese Anzahl der Vergleiche, die "zugunsten der C-Gruppe ausgeht" (also einen höheren Wert in der C-Gruppe aufweisen⁷⁹), ergibt die "Prüfgröße" U. In unserem Fall ist also $U=2$.

<< Es kann natürlich auch vorkommen, daß bei einem solchen Vergleich 2 Meßwerte gleich groß sind. Ein solcher Fall würde in der Prüfgröße U mit einem halben Punkt berücksichtigt werden. So etwas führt aber zu Komplikationen bei den weiteren Berechnungen (Näheres s. u. @); hier soll nur der einfachere Fall behandelt werden, daß solche "Rangkopplungen" nicht vorkommen (was man im Falle von Lösungszeiten durch hinreichend genaue Zeitmessung praktisch immer erreichen kann).

>> Wenn es keinen "echten" Unterschied zwischen den beiden Durchführungsbedingungen (Coffein und Placebo) gibt und alles nur Zufall ist, dann ist zu erwarten, daß etwa gleich viele Vergleiche in der einen oder in der anderen Richtung ausgehen, in unserem Fall also jeweils etwa 10 Vergleiche in beiden Richtungen. Die im obigen Fall auch "mit bloßem Auge" erkennbare Tendenz, daß die Werte der C-Gruppe eher am Anfang liegen, drückt sich darin aus, daß deutlich weniger (nämlich 2) Vergleiche mit höherem C-Wert vorliegen, als per Zufall zu erwarten ist. Trotzdem stellt sich die Frage: Wie groß ist die Wahrscheinlichkeit, daß so große oder noch größere Abweichungen von der Zufallserwartung zufällig in unserer Stichprobe auftreten, wenn "in der Population" keine Unterschiede in der Verteilung der Lösungszeiten bestehen.

Mit Hilfe von Überlegungen der Kombinatorik (die für Interessenten weiter unten demonstriert werden) läßt sich zeigen, daß diese Wahrscheinlichkeit genau $8/126$, also etwa 6.35% beträgt. D.h.: Wenn man sich strikt an die verbreitete Faustregel hält und die Aussage "das kann kaum noch Zufall sein" für Fälle mit einer Zufallswahrscheinlichkeit von höchstens 5% reserviert, dann kann man eine solche Aussage in diesem Fall noch nicht treffen. Aber das ist natürlich kein Beleg dafür, daß es tatsächlich Zufall ist!

<< Abschließend noch "für Interessenten" die genaue Berechnung der Wahrscheinlichkeit. Sie erfolgt in 4 Schritten.

Im 1. Schritt fragen wir uns, wieviele Möglichkeiten es gibt, eine Rangfolge zu bilden, in der 4 mal der Buchstabe C und 5 mal der Buchstabe P auftritt. Die einfache Permutationsregel für Reihenfolgen können wir nicht anwenden, weil es sich (im Sinne der Fragestellung) nicht um 9 voneinander verschiedene Elemente handelt. Eine Lösung ergibt sich, wenn man die Frage umformuliert: Wieviele Möglichkeiten gibt es, aus den 9 Rangplätzen von 1 bis 9 eine Untergruppe von 4 Rangplätzen zu bilden, auf die man den Buchstaben C setzt? (Und auf die übrigen Rangplätze setzt man den Buchstaben P.) Aufgrund der Regel für Kombinationen (vgl.

⁷⁹Das Wort "zugunsten" ist dabei nicht wertend gemeint. Natürlich ist bei Problemlösungen eine niedrige Lösungszeit meist ein "günstigeres" Ergebnis als eine hohe. Wenn wir aber fragen, wieviele Vergleiche "zugunsten" der C-Gruppe ausgehen, dann ist damit gemeint, bei wie vielen Vergleichen der Meßwert aus der C-Gruppe höher ist.

S. @) ergibt sich, daß dies

$$\binom{9}{4} = \frac{9 \cdot 8 \cdot 7 \cdot 6}{1 \cdot 2 \cdot 3 \cdot 4} = 126$$

Möglichkeiten sind.

Im 2. Schritt vergegenwärtigen wir uns: Falls in der Population kein Unterschied in der Verteilung der Lösungszeiten zwischen den Bedingungen C und P besteht, sind diese 126 möglichen Reihenfolgen offenbar alle gleich wahrscheinlich.

Wenn man das genau überprüfen will, kann man sich z.B. die Wahrscheinlichkeit der Rangfolge CPCCPPCPP nach dem Modell des Urnenziehens ohne Zurücklegen ableiten. Als Übungsaufgabe? Ansatz: Wir haben in einer Urne 4 C-Kugeln und 5 P-Kugeln. Wenn die Populationsverteilungen der Lösungszeiten für beide Bedingungen gleich sind, dann ist die Wahrscheinlichkeit eines C an 1. Stelle $4/9$: die bedingte Wahrscheinlichkeit eine P an zweiter Stelle, gegeben ein C an 1. Stelle, ist $5/8$ usw. Wenn man diese Wahrscheinlichkeiten nach dem erweiterten Multiplikationssatz hintereinanderschreibt und in gleicher Weise die Wahrscheinlichkeit der Rangfolge PPCCPCPP hinschreibt, sieht man schnell, daß diese beiden bis auf Vertauschung von Faktoren gleich aufgebaut sind und demnach übereinstimmen müssen, und daß dasselbe auch für alle übrigen solchen Rangfolgen von 4 C's und 5 P's gelten muß.

Wenn aber alle 126 möglichen Reihenfolgen die gleiche Wahrscheinlichkeit haben, dann muß jede einzelne dieser Wahrscheinlichkeiten offenbar $1/126$ sein.

Im 3. Schritt überprüfen wir, wie viele von diesen 126 gleich wahrscheinlichen Reihenfolgen genausoweit oder noch weiter von der Zufallserwartung abweichen, wie die im Experiment erhaltene Reihenfolge CCPCCPPPP mit ihrem U-Wert von 2. In unserem überschaubaren Beispiel mit wenigen Vpn können wir das durch Aufzählen klären. Zunächst können wir feststellen, daß die 4 Reihenfolgen

CCCCPPPPP, CCCPCPPPP, CCCPPCPPP und CCPCPPPPP

alle einen U-Wert von 2 oder weniger haben und damit nach dem Kriterium der Prüfgröße U alle genausoweit oder noch weiter von der Zufallserwartung abweichen wie die im Experiment erhaltene. Wir müssen aber auch diejenigen Rangfolgen einbeziehen, die in umgekehrter Richtung von der Zufallserwartung abweichen (d.h. Tendenz zu hohen C-Werten). 4 solche Rangfolgen erhalten wir, indem wir die obigen "rückwärts lesen":

PPPPCCCC, PPPPCPPP, PPPCPPCCC und PPPCCPCC

Es läßt sich nun zeigen (mit systematischem Durchprobieren oder auch mit mathematischen Beweisen, die hier nicht vorgeführt werden) daß lediglich diese 8 Reihenfolgen genauso weit oder noch weiter von der Zufallserwartung abweichen, wie diejenige aus dem Experiment; alle anderen weichen weniger weit ab.

Im 4. Schritt kombinieren wir die Ergebnisse der vorangehenden Schritte: Wenn es 8 Reihenfolgen gibt, die genausoweit oder noch weiter von der Zufallserwartung abweichen wie die Daten unseres Experiments (Schritt 3) und wenn jede von ihnen eine Wahrscheinlichkeit von $1/126$ hat (Schritt 2), dann ergibt sich die Wahrscheinlichkeit, daß irgendeine von ihnen (also die 1. oder die 2. oder... oder die 8.) zufällig auftritt, nach dem einfachen Additionssatz als

$$1/126 + 1/126 + \dots + 1/126 = 8/126 = 0.06349.$$

>>

III. Das Entscheidungs-Modell der Signifikanzprüfung bei zweiseitigen Signifikanztests

Es gibt eine Modellvorstellung darüber, wie Entscheidungen über Hypothesen aufgrund von statistischen Daten verlaufen. Nach Vorbereitungen durch Fisher (@) wurde diese Modellvorstellung von Neyman und Pearson (@) entwickelt. In diesem Entscheidungsmodell spielen Überlegungen der in Abschnitt @II demonstrierten Art eine wesentliche Rolle, aber der Gesichtspunkt einer Entscheidung rückt noch stärker in den Vordergrund.

Grundlage ist eine Entscheidungstheorie, die vor allem für ökonomische Entscheidungen entwickelt wurde. Wenn die Kosten eventueller Fehlentscheidungen kalkulierbar sind, kann man nach dieser Entscheidungstheorie auch angeben, welche Risiken von Fehlentscheidungen noch akzeptiert werden können. Für Situationen dieser Art ist das Entscheidungsmodell von Neyman und Pearson geeignet. Dagegen kann man durchaus bezweifeln, daß dieses Modell auch der tatsächlichen Bedeutung von Signifikanztests in psychologischen Untersuchungen gerecht wird. Trotzdem werden in Veröffentlichungen die Ergebnisse psychologischer Untersuchungen meistens in einer Sprache berichtet und diskutiert, die aus dem Entscheidungsmodell von Neyman und Pearson stammt. Zumindestens in der Wahl der Sprache tut man gewissermaßen so, als ob man dieses Entscheidungsmodell zugrundelegen würde. Aus dieser Tatsache ergeben sich vor allem zwei Gründe, dieses Modell im folgenden "in reiner Form" darzustellen:

- Wenn man ein solches Modell kritisieren will, muß man es zunächst kennen.
- Um psychologische Untersuchungen mit Verständnis lesen und kritisieren zu können, muß man ebenfalls das Entscheidungsmodell kennen, aus dem die verwendeten Begriffe stammen.

Bei der folgenden Darstellung des entscheidungstheoretischen Modells wird zur Veranschaulichung auf die Beispiele aus Abschnitt II zurückgegriffen.

1) Entscheidungsregeln aufgrund kritischer Bereiche

Nachdem man eine *Nullhypothese* und eine (oder mehrere) *Alternativhypothese(n)* formuliert hat, entscheidet man sich, welche Prüfgröße man berechnen will (hier: x , t bzw. U). Das liegt vor allem an der zu untersuchenden Fragestellung; denn unterschiedliche Prüfgrößen prüfen unterschiedliche Hypothesen. Die Wahrscheinlichkeitsverteilung der Prüfgröße bei Gültigkeit von H_0 muß bekannt sein (Hier: Binomialverteilung, t -Verteilung, kombinatorische Überlegungen über die Wahrscheinlichkeit möglicher Werte der Prüfgröße U). Häufig ist allerdings diese Wahrscheinlichkeitsverteilung nur dann bekannt, wenn noch weitere *Zusatzvoraussetzungen* erfüllt sind, die selbst mit der Nullhypothese nichts zu tun haben (Beispiel: t -Verteilung gilt nur

bei Normalverteilung der d-Werte. Das kann genauso gut bei Gültigkeit von H_0 wie bei Gültigkeit von H_1 der Fall sein.).

Dann formulieren wir einen "kritischen Bereich" für die Prüfgröße. Er wird durch diejenigen Werte der Prüfgröße gebildet, bei denen die Abweichungen der Daten von der Nullhypothese als "signifikant" betrachtet werden. (Das wird in den folgenden Abschnitten näher erläutert.) Genauer wird der kritische Bereich mit der folgenden *Entscheidungsregel* verbunden:

- Fällt nach Erhebung der Daten die Prüfgröße in den kritischen Bereich, so verwerfen wir H_0 und entscheiden uns für H_1 .
- Fällt die Prüfgröße nicht in den kritischen Bereich, so rechnen wir weiterhin mit der Möglichkeit, daß H_0 richtig ist. Auch H_1 kann richtig sein. H_0 wird also nicht unter Verwerfung von H_1 akzeptiert, sondern lediglich als weitere Möglichkeit neben H_1 beibehalten.

Diese Formulierung des kritischen Bereichs muß *vor Erhebung der Daten* erfolgen!

2) Die Wahl des kritischen Bereichs

a) Beispiele

Wir hätten bei unseren Beispielen in II verschiedene kritische Bereiche festlegen können, z.B. zum Vorzeichentest:

1a) kritischer Bereich: $x \leq 90$ oder $x \geq 110$

Entscheidungsregel: Erzielen wir ein $x \leq 90$ oder ein $x \geq 110$, so verwerfen wir H_0 ; erzielen wir $90 < x < 110$, so behalten wir H_0 bei.

Wir hätten auch sagen können:

1b) kritischer Bereich: $x \leq 80$ oder $x \geq 120$

Entscheidungsregel entsprechend.

Zum t -Test hätten wir sagen können:

2a) kritischer Bereich: $t \leq -1.99$ oder $t \geq 1.99$

Entscheidungsregel: Erzielen wir ein $t \leq -1.99$ oder ein $t \geq 1.99$, so verwerfen wir H_0 , andernfalls wird H_0 beibehalten.

2b) kritischer Bereich: $t \leq -2.65$ oder $t \geq 2.65$

Entscheidungsregel entsprechend

Beim U-Test wären folgende Festlegungen denkbar:

3a) kritischer Bereich: $U \leq 2$ oder $U \geq 18$

Entscheidungsregel: Erzielen wir ein $U \leq 2$ oder ein $U \geq 18$, so verwerfen wir H_0 , andernfalls wird H_0 beibehalten.

3b) kritischer Bereich: $U \leq 1$ oder $U \geq 19$
Entscheidungsregel entsprechend

Bei den angenommenen Ergebnissen hätten wir die Nullhypothese nach jeder dieser Entscheidungsregeln beim Vorzeichentest verworfen, beim t -Test dagegen beibehalten. Beim U-Test hätten wir bei einem tatsächlichen $U=2$ die Nullhypothese nach Regel 3a verworfen, nach Regel 3b dagegen beibehalten. Bei der Aufstellung der Entscheidungsregel, also *vor* Erhebung der Daten, wäre dies alles aber unbekannt gewesen. Es hätte sich also in allen drei Fällen die Frage ergeben, ob man die "strengere" Regel (1b, 2b bzw. 3b) wählen soll oder die weniger strenge (1a, 2a bzw. 3a) oder noch eine andere. Allgemein stellt sich die Frage: Wie soll man den kritischen Bereich legen?

Bevor wir diese Frage in den folgenden Abschnitten beantworten, soll der dabei auftauchende, für die weiteren Überlegungen wichtige Begriff von mehr oder weniger "strengen" Entscheidungsregeln erläutert werden. Dieser Begriff geht davon aus, daß eine Entscheidungsregel umso strenger ist, je höhere Anforderungen gestellt werden, um die Nullhypothese zu verwerfen. Genauer gilt die

Definition: Eine Entscheidungsregel b ist dann strenger als eine Entscheidungsregel a, wenn die Nullhypothese in allen Fällen, in denen sie nach Regel b verworfen wird, auch nach Regel a verworfen wird und wenn es außerdem Fälle gibt, in denen die Nullhypothese nach Regel a verworfen würde, nach Regel b dagegen nicht.

Übungsaufgabe: Überprüfen Sie, daß nach dieser Definition in den obigen Beispielen die Regel b jeweils strenger ist als Regel a.

<< In der Sprache der Mengenlehre kann man den bei einer Entscheidungsregel zugrundegelegten kritischen Bereich auch als eine Menge denkbarer Werte der Prüfgröße betrachten und sagen:

Definition: Eine Entscheidungsregel b ist dann strenger als eine Entscheidungsregel a, wenn der bei Regel b zugrundegelegte kritische Bereich eine echte Untermenge des bei Regel a zugrundegelegten kritischen Bereichs ist.

Zu beachten ist, daß es auch Entscheidungsregeln a und b geben kann, von denen keine strenger als die andere ist. Für das Beispiel zum t -Test wäre auch eine Entscheidungsregel 2c denkbar, bei der die Nullhypothese nur bei $t \geq 1.67$ verworfen und bei allen Werten von $t < 1.67$ beibehalten wird, insbesondere also auch *allen* negativen Werten von t . (Situationen, in denen eine solche "einseitige Entscheidungsregel" zur Anwendung kommen kann, werden wir später kennenlernen.) Dann besteht der kritische Bereich nur aus den Werten mit $t \geq 1.67$. Vergleichen wir diese Entscheidungsregel mit der Entscheidungsregel 2a, dann stellen wir fest: Es ist weder 2c strenger als 2a noch 2a strenger als 2c.

Übungsaufgabe: Überprüfen Sie diese Behauptung, indem Sie für $t=1.80$ und für $t=-2.20$ jeweils die nach beiden Regeln zu treffenden Entscheidungen vergleichen.

Man kann auch sagen: Durch die Relation "Regel x ist höchstens so streng wie Regel y" wird in die Menge der zu einer Untersuchung denkbaren Entscheidungsregeln eine Ordnung gebracht, bei der aber nicht jedes Element dieser Menge mit jedem vergleichbar ist. Unter Rückgriff auf eine in Statistik I (@) eingeführte Terminologie können wir feststellen, daß diese Relation zwar transitiv

ist, aber nicht konnex. Die durch diese Relation in die Menge der möglichen Entscheidungsregeln gebrachte Ordnung ist also keine "schwache Ordnung". Unter Verwendung eines in der Mathematik gebräuchlichen Begriffs kann man jedoch von einer "partiellen Ordnung" sprechen.

>>

b) Fehlertypen und Risiken

In der Einleitung zu Abschnitt III @ wurde bereits angedeutet, daß es ein Hauptziel des entscheidungstheoretischen Modells der Hypothesenprüfung ist, *die Risiken von Fehlentscheidungen kalkulierbar zu machen*. Dabei sind mehrere mögliche Fehlertypen und Fehlerrisiken zu unterscheiden, die folgendermaßen definiert sind:

- Einen Fehler erster Art ("Fehler I") machen wir, wenn wir die Nullhypothese verwerfen, obschon sie richtig ist.
- Einen Fehler zweiter Art ("Fehler II") machen wir, wenn wir die Nullhypothese beibehalten, obschon sie falsch ist.
- Das Risiko I ist die Wahrscheinlichkeit, einen Fehler I zu machen. Diese Wahrscheinlichkeit nennt man auch α .
- Das Risiko II ist die Wahrscheinlichkeit, einen Fehler II zu machen. Man verwendet das Symbol β .

Die Methode der Signifikanztests läuft nun darauf hinaus, die Entscheidungsregel so zu wählen, daß diese Fehlerrisiken sich in akzeptablen Grenzen halten. Was dabei als "akzeptabel" gilt, wird noch zu erläutern sein. Aus Gründen, die ebenfalls in den folgenden Abschnitten näher dargestellt werden, wird vor allem das Risiko I beachtet. Dazu muß die Wahrscheinlichkeitsverteilung der Prüfgröße bei Gültigkeit der Nullhypothese angebar sein; dann ist auch das Risiko I berechenbar: Da die Nullhypothese verworfen wird, wenn die Prüfgröße in den kritischen Bereich fällt, brauchen wir nur anzugeben, wie groß bei Gültigkeit der Nullhypothese die Wahrscheinlichkeit ist, daß die Prüfgröße in den kritischen Bereich fällt.

Beispiel: Zur Berechnung des Risiko I bei Regel 1a können wir auf Ergebnisse von Abschnitt @ E II 1 zurückgreifen. Wie dort gezeigt wurde, hat die Prüfgröße x bei Gültigkeit der Nullhypothese eine Binomialverteilung mit $N=200$ und $p=0.50$. Die zu beantwortende Frage lautet also: Wie groß ist die Wahrscheinlichkeit, daß eine Größe x mit dieser Verteilung einen Wert von $x \leq 90$ oder $x \geq 110$ annimmt. Natürlich kann diese Binomialverteilung wieder durch eine Normalverteilung mit Erwartungswert 200 und Varianz 50 approximiert werden. Unter Anwendung der "Kontinuitätskorrektur" lautet unsere Frage also: Wie groß ist die Wahrscheinlichkeit, daß eine mit Erwartungswert 100 und Varianz 50 normalverteilte Zufallsgröße x einen Wert von $x \leq 90.5$ oder $x \geq 109.5$ annimmt? Wie sich leicht überprüfen läßt, beträgt diese Wahrscheinlichkeit rund 0.18. Also ist bei Regel 1a das Risiko I angebar als $\alpha \approx 0.18$.

Die Berechnung des Risiko II ist dagegen komplizierter; denn während beim Fehler I

vorausgesetzt wird, daß die Nullhypothese richtig und damit (für das Beispiel des Vorzeichentests) $p_+ = 0.50$ ist, wird beim Fehler II lediglich vorausgesetzt, daß die Nullhypothese falsch und damit $p_+ \neq 0.50$ ist. Daraus kann man natürlich noch nicht die Wahrscheinlichkeit berechnen, daß die Prüfgröße nicht in den kritischen Bereich fällt, wenn die Nullhypothese falsch ist. Wie wir in Abschnitt @ E III 2 e noch genauer sehen werden, hängt das Risiko II davon ab, wie weit die tatsächlichen Verhältnisse von der Nullhypothese abweichen.

c) Die Abhängigkeit der Risiken von der Wahl des kritischen Bereichs.

Die Höhe beider Risiken hängt offenbar von der Wahl des kritischen Bereichs ab. Verwerfe ich H_0 erst dann, wenn meine Prüfgröße sehr weit von dem der Nullhypothese entsprechenden Wert abweicht ("strenge Entscheidungsregel"), dann ist die Gefahr einer ungerechtfertigten Verwerfung der Nullhypothese, also das Risiko I, sehr gering. Andererseits steigt damit die Gefahr, daß ich die Nullhypothese zu Unrecht beibehalte, wenn sie falsch ist, also das Risiko II. Andererseits: Verwerfe ich H_0 allzu leichtfertig (weniger strenge Entscheidungsregel), dann sinkt zwar das Risiko II, aber das Risiko I steigt.

Beispiel: Wähle ich im Schlafexperiment die strengere Entscheidungsregel 1b, dann ist das Risiko I kleiner als bei Regel 1a. Andererseits ist aber die Wahrscheinlichkeit eines Fehlers II bei Regel 1b größer als bei Regel 1a.

Das läßt sich (mit einer kleinen, praktisch unbedeutenden Einschränkung) unmittelbar aus der Definition einer "strengeren" Entscheidungsregel (Abschnitt @) ableiten. Wenn eine Entscheidungsregel b strenger ist als Regel a, dann gilt:

- Die Nullhypothese wird nach Regel a in allen Fällen verworfen, in denen sie nach Regel b zu verwerfen ist, und außerdem in weiteren Fällen. Also ist die Wahrscheinlichkeit, die Nullhypothese zu verwerfen, obwohl sie richtig ist, bei Regel a größer.
- Die Nullhypothese wird nach Regel b in allen Fällen beibehalten, in denen sie nach Regel a beizubehalten ist, und außerdem in einigen weiteren. Also ist die Wahrscheinlichkeit, die Nullhypothese beizubehalten, obwohl sie falsch ist, bei Regel b größer.

<< Genauer können wir die möglichen Ergebnisse einer Untersuchung in drei Ereignisse A, B und C unterteilen:

- Ereignis A: Wir erhalten einen Wert der Prüfgröße, der bei beiden Entscheidungsregeln zum Verwerfen der Nullhypothese führt.
- Ereignis B: Wir erhalten einen Wert der Prüfgröße, der bei beiden Entscheidungsregeln zum Beibehalten der Nullhypothese führt.
- Ereignis C: Wir erhalten einen Wert der Prüfgröße, der bei Regel a zum Verwerfen und bei Regel b zum Beibehalten der Nullhypothese führt.

Dagegen folgt aus dem Begriff der "strengeren" Entscheidungsregel, daß es keine Fälle geben kann, in denen die Nullhypothese nach Regel a beizubehalten, nach Regel b dagegen zu verwerfen wäre.

Daraus folgt dann:

- Unter Regel a wird die Nullhypothese verworfen, wenn entweder Ereignis A oder Ereignis C auftritt, unter Regel b dagegen nur bei Ereignis A. Also ist das Risiko I (die Wahrscheinlichkeit, die Nullhypothese zu verwerfen, obwohl sie richtig ist) unter Regel a größer.
- Unter Regel b wird die Nullhypothese beibehalten, wenn entweder Ereignis B oder Ereignis C auftritt, unter Regel a dagegen nur bei Ereignis B. Also ist das Risiko II (die Wahrscheinlichkeit, die Nullhypothese beizubehalten, obwohl sie falsch ist) unter Regel b größer.

Die angekündigte Einschränkung bezieht sich auf Situationen, in denen die Wahrscheinlichkeit des Ereignisses C null ist. Theoretisch könnten wir etwa beim t -Test noch eine weitere Entscheidungsregel betrachten, bei der der kritische Bereich durch " $t > 1.99$ oder $t < -1.99$ " definiert ist. Formal ist diese Entscheidungsregel strenger als Regel a. Die Fälle, in denen die Nullhypothese nach dieser Regel beibehalten, nach Regel 1a dagegen verworfen würde, beschränken sich aber auf Situationen, in denen $t=1.99$ oder $t=-1.99$ genau zutrifft. Da die t -Verteilung eine kontinuierliche Verteilung ist, ist die Wahrscheinlichkeit dieses Ereignisses streng genommen null; aber eine kontinuierliche Verteilung der Prüfgröße t ergäbe sich in Wirklichkeit nur, wenn die Meßwerte kontinuierlich wären, und das ist in der Praxis nie der Fall. Eine kontinuierliche Meßwertskala ist eine mathematische Idealisierung; faktisch haben wir immer künstlich (oder natürlich) diskrete Skalen, und damit ist die tatsächliche Verteilung der Prüfgröße t auch nicht kontinuierlich.

>>

Wir stehen also in dem Dilemma, entweder ein hohes Risiko I oder ein hohes Risiko II in Kauf zu nehmen. Es gibt nun zwei Prinzipien, aus diesem Dilemma herauszukommen:

1. Ein Fehler I ist wesentlich schlimmer als ein Fehler II. Wir nehmen also lieber ein erhöhtes Risiko II in Kauf als ein zu hohes Risiko I (Prinzip des konservativen Testens; siehe unten: @d).
2. Haben wir uns dafür entschieden, das Risiko I niedrig zu halten, dann können wir durch eine ausreichend große Stichprobe doch auch das Risiko II beliebig niedrig machen (siehe unten: @e).

d) Näheres zum Prinzip des konservativen Testens

Bei einem Fehler I machen wir einen wirklichen Fehler: Wir behaupten, die Nullhypothese sei unrichtig, obschon sie richtig ist. Bei einem Fehler II dagegen behaupten wir nicht, die Alternativhypothese sei falsch, wir rechnen nur mit der Möglichkeit, daß H_0 richtig ist. Der Schaden beim Fehler I besteht darin, daß wir etwas Falsches behaupten; beim Fehler II haben wir lediglich ein bestimmtes Gesetz (H_1) nicht entdeckt, behaupten damit aber noch nicht, daß dieses Gesetz nicht zuträfe. Insofern ist ein Fehler I schlimmer als ein Fehler II.

Der Name "Prinzip des konservativen Testens" erklärt sich wie folgt: Vor dem Experiment

wissen wir nicht, ob H_0 oder H_1 richtig ist. Man könnte jetzt das Prinzip auch folgendermaßen formulieren: Lieber nach dem Experiment - genauso wie vorher - nicht wissen, ob H_0 oder H_1 richtig ist, als zu voreilig eine Entscheidung fällen, die dann falsch ist. Offenbar ein konservatives Prinzip! Wir wählen daher unseren kritischen Bereich so, daß das Risiko I niedrig ist. Meistens wählt man eine Risiko I ("eine Irrtumswahrscheinlichkeit") von 5%.

Beispiel: Bei unserem t -Test ermitteln wir aufgrund der t -Verteilungstabelle, welcher t -Wert bei 73 Freiheitsgraden eine zweiseitige Überschreitungswahrscheinlichkeit von 5% hat. Das ist der Wert 1.99. Wir kämen also genau zu dem oben unter 2a genannten kritischen Bereich. Wäre unser t in diesen kritischen Bereich gefallen, so würden wir auch sagen: "Wir verwerfen die Nullhypothese mit (weniger als) 5% Irrtumswahrscheinlichkeit", oder "Die Abweichung von der Nullhypothese ist auf dem 5%-Niveau signifikant". Überhaupt spricht man von signifikanten Abweichungen von der Nullhypothese nur, wenn die Irrtumswahrscheinlichkeit höchstens 5% beträgt.

In manchen Fällen fordern wir aber ein noch geringeres Risiko I, insbesondere dann, wenn H_1 besonders stark von allen bisherigen Erfahrungen abweicht. Meist wählt man dann ein "Signifikanzniveau" von 1% oder weniger. Hätten wir beispielsweise jemanden zu testen, der behauptet, ein Hellseher zu sein und jedem Menschen den Wochentag, an dem er geboren ist, sagen zu können, dann würde die Nullhypothese lauten: "Richtige Angabe des Geburtstages ist zufällig; auch bei unserem 'Hellseher' beträgt die Wahrscheinlichkeit eine richtige Angabe zu machen, wie bei jedem anderen Menschen $1/7$."

$$H_0: p = 1/7$$

$$H_1: p \neq 1/7$$

Diese Nullhypothese würden wir erst dann verwerfen, wenn unser Kandidat so häufig richtig "rät", daß seine Leistung oder eine noch extremere nur mit 1% Wahrscheinlichkeit auftreten würde, wenn es sich um ein reines Raten handelte. Ergebnisse, die auf dem 1%-Niveau signifikant sind, nennt man auch *sehr signifikant*.

Anderes Beispiel: Würde die Annahme, daß Koffein die Leistung in unserem Test beeinflußt, allen bisherigen Erfahrungen widersprechen, dann hätten wir einen anderen kritischen Bereich für unsere Prüfgröße t wählen müssen. Der t -Wert mit 1% zweiseitiger Überschreitungswahrscheinlichkeit beträgt bei 73 Freiheitsgraden laut Tabelle etwa 2.65. Mit diesem Wert kämen wir dann zu der Entscheidungsregel 2b.

Ein weiterer Grund, ein Signifikanzniveau von 1% zu wählen, liegt dann vor, wenn ein Fehler I neben der "Belastung" mit einem wissenschaftlichen Irrtum auch noch mit einer ethischen Belastung besonderer Art (oder mit einem besonders gravierenden Schaden anderer Art) verbunden ist. Beispiel: Wenn (medizinische oder psychotherapeutische) Behandlungsmethode A zwar wirksamer ist als eine andere Methode B, gleichzeitig aber mit Nebenwirkungen schwerer Art verbunden ist (z.B. "Elektroschock"), wird man unter anderem verlangen, daß die größere Wirksamkeit von A besonders abgesichert ist, also z.B. mit einer Irrtumswahrscheinlichkeit von höchstens 1% (oder sogar noch weniger) belastet ist. (Und auch dann wird man um eine

Abwägung zwischen größerer Wirksamkeit und Nebenwirkungen nicht umhinkommen! Ein verschärftes Signifikanzniveau ersetzt eine solche Abwägung nicht. Aber bevor man in eine solche Abwägung eintritt, sollte erst einmal die größere Wirksamkeit mit besonders niedriger Irrtumswahrscheinlichkeit feststehen.)

e) Das Risiko II und die Teststärke

Unter der Stärke eines Signifikanztests verstehen wir seine Fähigkeit, eine in der Population vorhandene Abweichung von der Nullhypothese auch tatsächlich als signifikant aufzuzeigen. Genauer: Die Teststärke ist die Wahrscheinlichkeit, daß eine in der Population vorhandene Abweichung von der Nullhypothese sich in Form eines signifikanten Ergebnisses niederschlägt. Die Wahrscheinlichkeit, daß kein signifikantes Ergebnis auftritt, ist gleich dem Risiko II, also gleich β . Folglich ist die

$$\text{Teststärke} = 1 - \beta$$

Es ist nun keineswegs so, daß jeder statistische Test immer die gleiche Stärke hat! Die Teststärke hängt vielmehr von drei Faktoren ab:

1. davon, wie weit sich die Populationsverhältnisse von der Nullhypothese entfernen,
2. von dem Risiko I, das wir mit unserer Entscheidungsregel in Kauf zu nehmen bereit sind, und weiteren Merkmalen der Entscheidungsregel (s.u. S.@), und
3. von der Stichprobengröße.

Beispiel: Nehmen wir an, daß in der Population der Schläfer die Wahrscheinlichkeit eines "+"-Ergebnisses 40% betrüge. Wir hätten weiterhin 200 Vpn getestet und uns für eine Entscheidungsregel 1a entschieden. Mit Hilfe der Binomialverteilung mit $N=200$ und $p=0.4$ können wir berechnen, daß wir dann mit einer Wahrscheinlichkeit von 0.065 ein x über 90 und unter 110 erhalten hätten. (Übungsaufgabe: Überprüfen Sie die Angabe und einige der folgenden.) In diesem Fall hätten wir einen Fehler II gemacht. β beträgt also 0.065. Die Teststärke wäre die Wahrscheinlichkeit, ein signifikantes Ergebnis, also ein x von höchstens 90 und mindestens 110 zu erzielen:

$$\text{Teststärke} = 1 - \beta = 1 - 0.065 = 0.935$$

Wir erhalten also mit 93.5% Wahrscheinlichkeit ein signifikantes Ergebnis.

Hätten wir dagegen Entscheidungsregel 1b verwendet, so betrüge β (hier die Wahrscheinlichkeit eines Ergebnisses über 80 und unter 120) 0.471, die Teststärke wäre also 0.529.

Wäre dagegen die Populationswahrscheinlichkeit eines "+"-Ergebnisses 0.45, so würden wir die Binomialverteilung mit $N=200$ und $p=0.45$ zugrunde legen. Wir hätten dann bei Entscheidungsregel 1a ein Risiko II von $\beta = 0.911$ und die Teststärke 0.089. Das letzte Ergebnis ist besonders auffallend: Obschon H_0 laut Voraussetzung falsch ist, ist es viel wahrscheinlicher,

daß wir die Nullhypothese beibehalten müssen, als daß wir sie verwerfen können. Warum?

Bei einer Binomialverteilung mit $N=200$ und $p=0.45$ beträgt der Erwartungswert 90. Dieser Erwartungswert liegt noch eindeutig in dem Bereich über 80 und unter 120, also in dem Bereich der Prüfgröße x , der zur Beibehaltung von H_0 führt! Indem wir also die strenge Regel 1b angenommen haben, haben wir ein hohes Risiko II und eine geringe Teststärke in Kauf genommen.

Ein letztes *Beispiel*: Nehmen wir an, wir hätten eine kleinere Stichprobe von $N=44$ gehabt und uns für den kritischen Bereich $x \leq 17$ oder $x \geq 27$ entschieden.

Entscheidungsregel: Bei einem x über 17 und unter 27 behalten wir H_0 bei; andernfalls verwerfen wir H_0 . Diese Entscheidungsregel entspricht insofern der Regel 1a, als bei beiden das Risiko I etwa 0.18 beträgt, und bei Gültigkeit von H_0 (also $p_+ = 0.5$) liegen die Grenzen des kritischen Bereichs etwas mehr als 1.3 Standardabweichungen unterhalb und oberhalb des Erwartungswerts der Prüfgröße x . (Überzeugen Sie sich davon!) Beide Entscheidungsregeln würden also zu einem viel zu hohen Risiko I führen.

Bei einem p_+ von 0.40 in der Population hätten wir ein Risiko II von $\beta = 0.509$, also eine Teststärke von 0.491. Wäre dagegen die Populationswahrscheinlichkeit eines "+"-Schläfers $p=0.45$, so wäre das Risiko II gleich 0.736, die Teststärke 0.264.

In der folgenden Tabelle sind die Teststärken unserer sechs Beispiele nochmals zusammengefaßt. Die im letzten Fall angewandte Entscheidungsregel ist dabei wegen ihrer Ähnlichkeit zu Regel 1a als Regel 1a' bezeichnet.

Tabelle 9: Die Abhängigkeit der Teststärke von der Entscheidungsregel, vom Ausmaß der Populationsabweichung von H_0 und von der Stichprobengröße.

	$p_+ = 0.40$	$p_+ = 0.45$
$N=200$, Regel 1b	0.529	0.089
$N=200$, Regel 1a	0.935	0.531
$N=44$, Regel 1a'	0.491	0.264

Im Überblick zusammengefaßt, zeichnen sich folgende drei Gesetzmäßigkeiten ab, die auch bei anderen Signifikanztests gelten:

1. Je stärker die Population von der Nullhypothese abweicht, um so stärker ist der Test (also um so größer die Wahrscheinlichkeit eines signifikanten Ergebnisses und um so geringer das Risiko II).

Vergleichen Sie dazu die beiden Spalten miteinander. In der ersten Spalte ist mit einem p_+ von 0.40 die Abweichung von der Nullhypothese ($p_+=.50$) größer als in der zweiten Spalte, in der wir ein p_+ von 0.45 annehmen.

Es bedarf kaum einer Erklärung, warum das immer so ist.

2. Je strenger die Entscheidungsregel, je geringer das Risiko I, das wir in Kauf zu nehmen bereit sind, um so schwächer ist der Test.

Vergleichen Sie dazu die beiden ersten Zeilen der Tabelle 9. In der ersten verwenden wir die strenge Regel 1b, die nur ein geringes Risiko I mit sich bringt, in der zweiten die weniger strenge Regel 1a mit einem an sich viel zu hohen Risiko I. Zur allgemeinen Begründung dieser Regeln vgl. oben Abschnitt e.

3. Je größer die Stichprobe, um so stärker wird der Test.

Vergleichen Sie die zweite und die dritte Zeile der Tabelle 9! Allgemeine Begründung: Je größer die Stichprobe, um so größer die Wahrscheinlichkeit, daß ihre Kennziffern die Populationsverhältnisse gut widerspiegeln (Wurzel-N-Gesetz!⁸⁰). Also können wir auch um so eher annehmen, daß evtl. Abweichungen der Stichprobe von der Nullhypothese keine Zufallsprodukte sind.

f) Konsequenzen für die Praxis

Aus unseren theoretischen Überlegungen ergeben sich folgende Regeln für die Praxis:

Man entscheidet sich nach der Formulierung der Hypothesen und Wahl einer Prüfgröße zunächst gemäß den besprochenen Kriterien für ein bestimmtes Signifikanzniveau. Dann bestimmt man die Stichprobengröße und fragt sich dabei: Wie weit werden die Populationsverhältnisse wohl von der Nullhypothese abweichen? Wie groß muß dann meine Stichprobe sein, damit der Test bei einer bestimmten Abweichung der Populationsverhältnisse von der Nullhypothese noch eine akzeptable Stärke hat? Oft fragt man sich auch umgekehrt: Wie groß müßte die Abweichung der Populationsverhältnisse von H_0 sein, damit der Test bei einer bestimmten Stichprobengröße noch eine akzeptable Stärke hat? Vermutet man, daß die Populationsabweichung schwächer ist, als bei der zunächst in Aussicht genommenen Stichprobengröße erforderlich wäre, so wählt man eine größere Stichprobe und umgekehrt.

Bei vielen Signifikanztests ist allerdings die Berechnung der Teststärke schwieriger als in unserem

⁸⁰Explizit haben wir das Wurzel-N-Gesetz nur für $\sigma_{\bar{x}}$ - die Standardabweichung der Wahrscheinlichkeitsverteilung des Stichprobendurchschnitts kennengelernt. Aus der Formel

$$\sigma_{\bar{x}} = \sigma_x / \sqrt{N}$$

geht hervor, daß er um so kleiner ist, je größer die Stichprobengröße N . Damit ist die Schätzung des Populationsmittels μ_x durch Stichprobendurchschnitt \bar{x} um so präziser, je größer die Stichprobe. Ähnliches gilt auch für viele andere Kennziffern, die wir als Schätzfunktionen verwenden.

Beispiel. Es hat sich daher die Gewohnheit eingebürgert, die Stichprobengröße "über den Daumen zu peilen". (vgl. Cohen @@ oder Faul & Erdfelder (1992) für ein Computerprogramm zur Berechnung der Teststärke).

Aus unseren Überlegungen zur Teststärke ergibt sich aber auch eine Möglichkeit, das Grundprinzip abzumildern, nach welchem ein nicht signifikantes Ergebnis es nicht erlaubt, die Nullhypothese zu akzeptieren. Haben wir eine große Stichprobe und trotz einer nicht zu strengen Entscheidungsregel ein nicht signifikantes Ergebnis, dann können wir immerhin sagen: Wenn eine erhebliche Abweichung von H_0 vorläge, dann hätte bei einer großen Stichprobe die Teststärke für ein signifikantes Ergebnis ausreichen müssen. Nun ist bei vielen psychologischen Fragestellungen eine minimale Abweichung von H_0 psychologisch fast genau so zu bewerten wie ein Zutreffen der H_0 . Wenn es z.B. um die Untersuchung des Erfolgs einer Therapie geht, dann ist ein minimaler Erfolg psychologisch so gut wie kein Erfolg. In solchen Situationen ist eine Grundvoraussetzung des Signifikanztest-Modells problematisch: Daß jede noch so geringfügige Abweichung von H_0 bereits als Erfüllung von H_1 gewertet wird. In solchen Situationen kann man bei großen Stichproben und nicht signifikantem Ergebnis die Schlußfolgerung ziehen, daß eine Abweichung von H_0 - sofern es überhaupt eine gibt - höchstens so geringfügig ist, daß sie unter dem psychologischen Erkenntnisinteresse faktisch gleichbedeutend mit einem Zutreffen von H_0 ist.

Damit ist die allgemeine Theorie der Signifikanztests noch nicht ganz erschöpft. Weitere Fragen werden dort behandelt, wo sie sich im Zusammenhang ergeben.

IV Einzelne Signifikanztests

Es sind in der Statistik sehr viele Techniken zur Prüfung von Hypothesen entstanden. Sie unterscheiden sich insbesondere nach den folgenden Kriterien:

- a) Welche Fragestellung soll beantwortet werden? D.h.: Wie lautet die Nullhypothese, wie die Alternativhypothese?
- b) Welche Zusatzvoraussetzungen macht der einzelne Signifikanztest?
- c) Wie groß ist die Teststärke?
- d) Wie groß ist der Rechenaufwand?

Alle vier Kriterien sollten wir berücksichtigen, wenn wir uns für einen bestimmten Signifikanztest entscheiden. Wir wollen hier nur einige behandeln.

1) Signifikanztests zur Prüfung von Hypothesen über einzelne Parameter

a) Der t-Test für einen einzelnen Mittelwert

Überprüft, ob ein Populationsmittel einen bestimmten Wert (μ_0) hat.

$$H_0: \mu_x = \mu_0 \quad H_1: \mu_x \neq \mu_0$$

Zusatzvoraussetzung: Normalverteilung der x-Werte in der Population.

Prüfgröße:

$$t = \frac{\bar{x} - \mu_0}{\hat{\sigma}_x} = \frac{\bar{x} - \mu_0}{\sqrt{\frac{QS_x}{N \cdot (N - 1)}}$$

Dabei ist QS_x die Summe der quadrierten Abweichungen vom Durchschnitt.

Bei Gültigkeit von H_0 hat diese Prüfgröße eine t -Verteilung mit $N-1$ Freiheitsgraden.

Wichtigstes Anwendungsgebiet: Als t -Test für Paardifferenzen oder t -Test für zwei korrelierte Stichproben. (s.o. II.2)

Beispiel: Es ist bekannt, daß in der Gesamtpopulation der durchschnittliche IQ 100 ist!

Frage: Ist der Durchschnitts-IQ in der Population der Sextaner ebenfalls 100?

$$H_0: \mu_x = 100$$

$$H_1: \mu_x \neq 100$$

Beide Hypothesen beziehen sich nicht auf die (deutsche) Gesamtpopulation, sondern auf die Population der Sextaner.

Wir nehmen uns vor, 41 Sextaner zu testen. Den jeweiligen IQ-Wert nennen wir x . Das Signifikanzniveau legen wir auf 5% fest. Wir wollen unsere Entscheidung nach der Prüfgröße t richten.

Kritischer Bereich: $t \leq - 2.02$ oder $t \geq 2.02$

(Diesen Wert entnehmen wir der t -Tabelle für 40 Freiheitsgrade).

Wir erheben Daten und erhalten einen Durchschnitt von $\bar{x} = 106.3$ und eine Quadratsumme QS_x von 5 763. Die Stichprobenverteilung der X-Werte zeigt keine deutlichen Abweichungen von einer Normalverteilung.

Wir setzen also in die Formel ein:

$$t = \frac{106.3 - 100}{\sqrt{\frac{5763}{41 \cdot 40}}} \approx 3.36$$

Dieser t -Wert liegt im kritischen Bereich, also verwerfen wir H_0 und entscheiden uns für H_1 : Der Durchschnitts-IQ der Sextanerpopulation ist nicht gleich dem (als Durchschnitts-IQ der Gesamtpopulation bekannten) Wert 100.

Die Entscheidung ist nach der bisher verwendeten Regel mit einer Irrtumswahrscheinlichkeit von $\alpha = 5\%$ behaftet. Aus unseren t -Tabellen können wir aber entnehmen, daß wir zu der gleichen Entscheidung gekommen wären, wenn wir den kritischen Bereich für das 0.2%-Niveau zugrundegelegt hätten: " $t \leq -3.31$ oder $t \geq 3.31$ "; in kürzerer Form: $|t| \geq 3.31$. Unser t -Wert ist also auch auf dem 0.2%-Niveau signifikant; die Irrtumswahrscheinlichkeit beim Verwerfen von H_0 beträgt nur 0.2%! Es ist üblich, neben dem zunächst gewählten Signifikanzniveau auch noch das niedrigste Signifikanzniveau mitzuteilen, mit dem H_0 verworfen werden kann: Je niedriger die Irrtumswahrscheinlichkeit, um so gesicherter der Befund!

Zur Schreibweise: Das Risiko I, das unserer ursprünglichen Entscheidungsregel zugrundelag, wird als α bezeichnet. Für die Überschreitungswahrscheinlichkeit des tatsächlich erzielten Werts der Prüfgröße wird in der Literatur oft der Buchstabe p verwendet. Aus der Tabelle können wir für einen t -Wert von 3.31 eine zweiseitige Überschreitungswahrscheinlichkeit von 0.2% entnehmen. Also ist die zweiseitige Überschreitungswahrscheinlichkeit unseres tatsächlich erzielten t -Werts von 3.36 etwas kleiner als 0.2%. Bei der Darstellung von Untersuchungsergebnissen in der Fachliteratur wird dann ein Ergebnis wie das oben festgestellte häufig in der folgenden Kurzform dargestellt:

$$\bar{x} = 106.3, t = 3.36, p < 0.002 \text{ (zweiseitig)}.$$

Der Zusatz "zweiseitig" weist ausdrücklich darauf hin, daß die Angabe $p < 0.002$ sich auf die zweiseitige Überschreitungswahrscheinlichkeit des vorangehenden t -Werts bezieht.

<<

Wenn man den genauen Unterschied zwischen den Ansätzen von Fisher und von Neyman und Pearson kurz beschreiben will, dann ist es besonders wichtig, zwischen den Wahrscheinlichkeiten α und p zu unterscheiden. Daher noch eine andere Formulierung:

- α ist die Überschreitungswahrscheinlichkeit der in der Entscheidungsregel festgelegten Grenzen des kritischen Bereichs.
- p ist die (bisher immer zweiseitige) Überschreitungswahrscheinlichkeit des tatsächlich gefundenen Werts der Prüfgröße.

Dann läßt sich der Unterschied der Ansätze etwa folgendermaßen darstellen:

- Geht man (wie Neyman und Pearson) davon aus, daß man *vor* der Datenerhebung eine Entscheidungsregel zu formulieren hat, dann kommt es auf α an, und es ist nach dieser Auffassung geradezu Mogelei, wenn man im nachhinein so tut, als wäre man von einer

strengeren Entscheidungsregel ausgegangen, indem man ein niedrigeres p angibt.

- Heute sieht man empirische Forschung dagegen eher als einen Prozeß in einer Wissenschaftler-Gemeinschaft, zu dem jede Einzeluntersuchung beiträgt. Dazu paßt es dann schlecht, daß der Einzelforscher für die Wissenschaftler-Gemeinschaft entscheiden soll, welches α für die gegebene Fragestellung zugrunde zu legen ist. Passender ist es unter diesem Verständnis, wenn (entsprechend dem Ansatz von Fisher) die Überschreitungswahrscheinlichkeit p des empirischen Werts der Prüfgröße mitgeteilt wird, so daß jeder selbst entscheiden kann, ob dieses p klein genug ist, um die Nullhypothese zu verwerfen..

Auch wenn der Ansatz von Fisher dieser Sichtweise eher entgegenkommt, ist andererseits gerade für die Psychologie die Wichtigkeit des auf dem Neyman-Pearson-Ansatz beruhenden Konzepts der Teststärke zu betonen. Beide Vorteile zu kombinieren, könnte ein Motiv sein, der Kritik am "Fisher-Neyman-Pearson-Hybrid" (vgl. die Einleitung zu Abschnitt @E) als "selbstbewußter Mischling" entgegenzutreten.

>>

Exkurs: Einseitige Entscheidungsregeln

Was hätten wir getan, wenn wir ein \bar{x} von 93.7 erhalten hätten? Nach unserer bisherigen Entscheidungsregel hätten wir H_0 verworfen: Wir würden ein t von -3.36 erhalten, und das fällt in den kritischen Bereich. In der Realität würden wir diesem Ergebnis aber nicht ganz trauen: Sollte es wirklich so sein, daß die Population der Sextaner weniger intelligent ist als die Gesamtpopulation? Da liegt die Annahme, daß selbst diese extreme Abweichung durch Zufall (oder einen Fehler bei der Testdurchführung usw.) entstanden ist, noch näher! Kurz: Irgend etwas "ist faul".

Eigentlich war unsere Entscheidungsregel also nicht richtig formuliert. Sie hätte eher lauten sollen: Wenn $t \geq 2.02$ ist, verwerfen wir H_0 und akzeptieren H_1 , wenn $t \leq -2.02$ ist, ist irgend etwas faul; wenn t zwischen beiden Werten liegt, behalten wir H_0 bei.

Solche Entscheidungsregeln sind natürlich nur dann gerechtfertigt, wenn wir *vor* dem Experiment wissen, daß eine der möglichen Abweichungen von der Nullhypothese nicht in Frage kommt.

In unserem Beispiel: Wir haben H_1 folgendermaßen formuliert:

$$H_1: \mu_x \neq 100$$

Wir können auch schreiben:

$$H_1: \mu_x < 100 \text{ oder } \mu_x > 100$$

Die Alternative $\mu_x < 100$ kommt aus Gründen, die uns schon vor dem Aufstellen der Entscheidungsregel bekannt sind nicht in Frage. Folglich bleibt übrig:

$$H_1: \mu_x > 100.$$

Wir könnten jetzt aber auch die Entscheidungsregel folgendermaßen formulieren:

Erhalten wir ein $t \geq 1.68$, so verwerfen wir H_0 ;

erhalten wir ein $t \leq -1.68$, so ist irgend etwas faul;

erhalten wir ein t zwischen -1.68 und $+1.68$. so behalten wir H_0 bei.

Die Wahrscheinlichkeit, einen Fehler I zu machen, ist hier gleich der Wahrscheinlichkeit eines t -Wertes von wenigstens 1.68 , und diese Wahrscheinlichkeit ist 5% . Wir hätten also eine einseitige Entscheidungsregel mit 5% Irrtumswahrscheinlichkeit. Bei dieser Entscheidungsregel ist aber die Wahrscheinlichkeit, daß ein signifikantes Ergebnis entsteht, wenn H_1 richtig ist, größer als bei der zuerst gewählten: Immer wenn wir ein t zwischen 1.68 und 2.02 erhalten, würden wir nach der alten Regel H_0 beibehalten, nach der neuen Regel aber H_0 verwerfen. Wir würden allerdings bei einem $t \leq -2.02$ H_0 jetzt nicht mehr verwerfen, aber die Wahrscheinlichkeit eines solchen t -Wertes ist natürlich sehr niedrig, wenn H_1 richtig ist.

Zusammengefaßt:

Durch die Wahl einer einseitigen Entscheidungsregel können wir die Wahrscheinlichkeit eines signifikanten Ergebnisses bei Gültigkeit von H_1 , also die Teststärke, erhöhen, ohne eine höhere Irrtumswahrscheinlichkeit in Kauf zu nehmen.

Dies ist ein weiterer, wichtiger Aspekt des bereits in III.2.c und e behandelten Grundsatzes, daß die Teststärke auch von der Entscheidungsregel abhängt. Im Unterschied zu den dort besprochenen Fällen nehmen wir hier aber mit einer einseitigen Entscheidungsregel keine Erhöhung des Risiko I in Kauf und senken doch das Risiko II.

Übrigens würden wir in unserem Experiment auch nach der folgenden Entscheidungsregel H_0 verwerfen:

Wenn $t \geq 3.31$, verwerfen wir H_0 ;

wenn $t \leq -3.31$, ist etwas faul;

ansonsten wird H_0 beibehalten.

Hier beträgt das Risiko I nur noch 0.1% . Ziehen wir also die Tatsache in Betracht, daß nur die eine Hälfte der Alternativhypothese in Frage kam, und testen wir aus diesem Grund einseitig, so beträgt unsere Irrtumswahrscheinlichkeit beim Verwerfen von H_0 nur noch 0.1% .

Die "Kurzform" der Ergebnismitteilung unter Verwendung des Buchstaben p würde also lauten:

$$\bar{x} = 106.3, t = 3.36, p < 0.001 \text{ (einseitig).}$$

Ende des Exkurses!

b) Signifikanzprüfungen für einzelne Korrelationskoeffizienten

Auch Korrelationskoeffizienten sind mit einem Stichprobenfehler behaftet.

Beispiel: Wir wollen überprüfen, ob die Zeit, die eine Vpn zur Beantwortung einer Frage braucht, mit ihrer Intelligenz zusammenhängt. Wir testen 102 Vpn. $x = \text{Zeit}$; $y = \text{IQ}$.

Wir erhalten ein r_{xy} von -0.23. Es wäre theoretisch denkbar, daß die Populationskorrelation $\rho_{xy} = 0$ ist und wir durch Zufall mehr intelligente und schnelle sowie dumme und langsame Vpn in unsere Stichprobe bekamen.

Frage: Läßt sich diese Hypothese bei unserem r_{xy} von -0.23 noch halten? Zur Beantwortung solcher Fragen gibt es verschiedene Verfahren. Wir wollen insbesondere drei besprechen: einen t -Test, Tabellen und die Fisher'sche z' -Transformation.

ba) Der t -Test für die Signifikanz einer Korrelation

Überprüft, ob die Populationskorrelation null ist.

$$H_0: \rho_{xy} = 0$$

Zusatzvoraussetzungen:

a) Alle bedingten x -Verteilungen sind Normalverteilungen

b) Alle bedingten x -Verteilungen haben die gleiche Varianz ("Homoskedastizität")

Es genügt auch, wenn die zwei Voraussetzungen für die bedingte y -Verteilung gelten.

Prüfgröße:

$$t = r_{xy} \cdot \sqrt{\frac{N - 2}{1 - r_{xy}^2}}$$

Bei Gültigkeit von H_0 hat diese Prüfgröße eine t -Verteilung mit $N-2$ Freiheitsgraden. In unserem Beispiel hätten wir vor der Erhebung der Daten formulieren müssen:

$$H_0: \rho_{xy} = 0$$

$$H_1: \rho_{xy} \neq 0$$

Wir wählen eine zweiseitige Alternativhypothese, da es ja auch denkbar ist, daß intelligente Vpn mehr zu unserer Frage zu sagen haben und daher länger für die Beantwortung brauchen, ρ_{xy} also positiv ist.

Es bestehen keinerlei Bedenken gegen ein Signifikanzniveau von 5%.

Entscheidungsregel: Liegt t im kritischen Bereich des 5% Niveaus für zweiseitige Fragestellungen, verwerfen wir H_0 .

(Allgemein: Man kann seine Entscheidungsregel auch so formulieren, daß man den kritischen Bereich nicht vor Erhebung der Daten genau ausrechnet, sondern nur festlegt, welche Prüfgröße

man berechnet, ob man einseitig oder zweiseitig testet und welches Signifikanzniveau man verlangt. Dann ist ja der kritische Bereich festgelegt.)

Wir wollen einmal annehmen, daß die Zusatzvoraussetzungen erfüllt sind und setzen in die Formel ein:

$$t = -0.23 \cdot \sqrt{\frac{102 - 2}{1 - 0.23^2}} \approx -2.36$$

Wir haben $102 - 2 = 100$ Freiheitsgrade. Falls 100 Freiheitsgrade in unserer Tabelle enthalten sind, sehen wir: Der kritische Wert von t für 100 Freiheitsgrade bei 5% zweiseitiger Überschreitungswahrscheinlichkeit betrüge 1.98. Also liegt t im kritischen Bereich.

Ist die t -Verteilung für 100 Freiheitsgrade nicht in unserer Tabelle enthalten, könnten wir sie folgendermaßen ablesen: Der entsprechende t -Wert bei 40 Freiheitsgraden wäre 2.02. Je größer die Zahl der Freiheitsgrade, um so kleiner der Betrag des kritischen t -Wertes. Unser t liegt also bereits im kritischen Bereich für 40 Freiheitsgrade und damit erst recht für 100 Freiheitsgrade.

Wir verwerfen die Nullhypothese also mit einer Irrtumswahrscheinlichkeit von weniger als 5% (aus einigen Tabellen läßt sich entnehmen, daß die Irrtumswahrscheinlichkeit ziemlich genau 2% beträgt), wir akzeptieren die Alternativhypothese: Die Abweichung unseres r_{xy} vom Wert 0 kann kaum durch Zufall entstanden sein. Die Abweichung ist signifikant. Man sagt auch einfach: Die Korrelation ist signifikant.

bb) Signifikanzprüfung über Tabellen

In einigen Büchern findet man Tabellen, in denen die niedrigsten signifikant von null verschiedenen Korrelationskoeffizienten zusammengestellt sind (@ Guilford Tabelle D; Hofstätter-Wendt Tafel D). Aus diesen Tabellen hätten wir entnehmen können, daß bei 100 Freiheitsgraden ein r_{xy} von mindestens 0.195 erforderlich ist, um auf dem 5% Niveau (zweiseitig) signifikant zu sein. Entsprechend ist auch ein r_{xy} von höchstens -0.195 signifikant; also auch unseres.

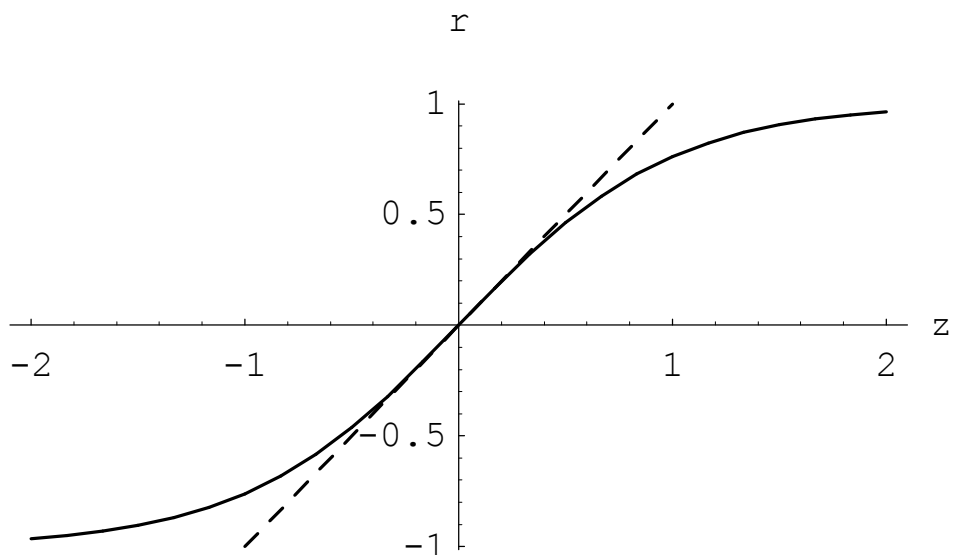
Allerdings ist dies eigentlich kein eigenständiger Signifikanztest; denn die Tabellen geben lediglich an, bei welchen Korrelationskoeffizienten die Prüfgröße t gerade den kritischen Wert annimmt.

bc) Signifikanzprüfung über die Fisher'sche z' -Transformation

Die Schwierigkeit bei der Signifikanzprüfung von Korrelationen beruht darauf, daß die Wahrscheinlichkeitsverteilung der Stichprobenkennziffer r_{xy} ziemlich kompliziert ist. Beim Stichprobendurchschnitt \bar{x} konnten wir sagen: Ist die Stichprobe groß oder ist die Populationsverteilung von x eine Normalverteilung, dann ist auch die

Wahrscheinlichkeitsverteilung von \bar{x} eine Normalverteilung. Das vereinfachte alle Signifikanztests sehr.

Für die Korrelation konnte der Statistiker R.A. Fisher nachweisen, daß zwar die Wahrscheinlichkeitsverteilung von r_{xy} keine Normalverteilung ist, daß aber unter bestimmten Bedingungen eine Normalverteilung entsteht, wenn wir den Korrelationskoeffizienten einer geeigneten Transformation unterwerfen. Die transformierten Werte werden üblicherweise als z' bezeichnet. Für diese Transformation gibt es im Anhang @ eine Tabelle., und die folgende Abbildung stellt diese Transformation graphisch dar.



Darstellung der Transformation von r_{xy} (y-Achse) in z' -Werte (x-Achse). Gestrichelt: $x=y$.

Zu beachten ist:

- Zur besseren Darstellung sind die Werte von r_{xy} auf der y-Achse und die zugehörigen Werte von z' auf der x-Achse abzulesen.
- Die durchgezogene Linie stellt die Transformation dar, und die gestrichelte Linie entspricht der Gleichung $x=y$.
- Für eine Darstellung der gesamten Transformation wäre eine unendliche x-Achse erforderlich, denn für $r_{xy} \rightarrow 1$ geht $z' \rightarrow +\infty$, und für $r_{xy} \rightarrow -1$ geht $z' \rightarrow -\infty$.

Aus der Abbildung geht (ebenso wie aus der Tabelle) hervor, daß für Korrelationen von -0.30 bis +0.30 die Werte von r_{xy} und z' nahezu übereinstimmen, und erst wenn der Betrag der Korrelation größer als 0.50 wird, ergeben sich nennenswerte Unterschiede. Für die Benutzung der Tabelle ist außerdem zu beachten, daß dort nur der Betrag von r_{xy} ohne Vorzeichen aufgeführt ist. (Vor jeder Zeile stehen die ersten beiden Stellen hinter dem Komma, und über jeder Spalte die dritte; der entsprechende Tabellenwert ist dann z' .) Das Vorzeichen von z' ist gleich dem Vorzeichen von r_{xy} .

<< Die Gleichung für die Transformation ist:

$$z' = \frac{1}{2} \cdot \ln\left(\frac{1 + r_{xy}}{1 - r_{xy}}\right)$$

Dabei bedeutet \ln "natürlicher Logarithmus".

>>

Beispiel: Wir wollen feststellen, ob ein Schulleistungstest (x) geeignet ist, die Schulleistung von Erstkläßlern vorherzusagen. Als Maß der Schulleistung gilt die Gesamtnote im Abschlußzeugnis der ersten Klasse (y). Es besteht der Verdacht, daß die Vorhersage-Validität dieses Tests (die Populationskorrelation ρ_{xy}) nur etwa 0.30 beträgt. In diesem Fall würde es sich nicht lohnen, den Test einzuführen.

Wir wollen also eine Untersuchung durchführen, bei der wir an einer Zufallsstichprobe von $N=103$ Schulneulingen den Test aufnehmen und die Testergebnisse nach einem Jahr mit der Gesamtnote⁸¹ korrelieren.

Entscheidungsregel: Ist (unter Zugrundelegung eines α von 5%) die Korrelation r_{xy} signifikant größer als 0.30, dann wird der Test eingeführt. Ist r_{xy} nicht signifikant von 0.30 verschieden, dann wird der Test nicht eingeführt. Ist r_{xy} signifikant kleiner als 0.30, wird der Test ebenfalls nicht eingeführt. Das ist eine einseitige Entscheidungsregel!

Allgemein: Wenn eine der beiden Abweichungen von H_0 genau dieselben Konsequenzen hat, wie ein Zutreffen von H_0 , formulieren wir eine einseitige Entscheidungsregel.

$$H_0: \rho_{xy} = 0.30$$

$$H_1: \rho_{xy} > 0.30$$

Da ein $\rho_{xy} < 0.30$ für uns die gleiche praktische Konsequenz hätte wie ein $\rho_{xy} = 0.30$, könnte man auch formulieren:

$$H_0: \rho_{xy} \leq 0.30$$

In vielen Büchern wird bei einseitiger Fragestellung die Nullhypothese immer in dieser Form formuliert. Das ist natürlich dann nicht sinnvoll, wenn wir - wie bisher - unsere einseitige Entscheidungsregel damit begründen, daß die eine Hälfte der Alternativhypothese aufgrund unserer Vorkenntnisse nicht zutreffen kann.

Ein für die Prüfung unserer H_0 geeigneter Signifikanztest läßt sich allgemein folgendermaßen darstellen: Er überprüft die Hypothese, daß ρ_{xy} einen bestimmten Wert ρ_0 hat:

$$H_0: \rho_{xy} = \rho_0$$

⁸¹Wenn in dem entsprechenden Schulsystem auf den Abschlußzeugnissen der 1. Klasse keine "Gesamtnoten" vergeben werden, ist eine entsprechende Beurteilung durch die Lehrer zu verwenden.

Zusatzvoraussetzung:

Bivariate Normalverteilung von x und y in der Population; $N \geq 50$.

Prüfgröße:

$$z = (z'_{xy} - \zeta'_0) \cdot \sqrt{N-3}$$

mit

z'_{xy} = der z' -Wert, der zu unserem empirischen r_{xy} gehört.

ζ'_0 = der z' -Wert, der zu ρ_0 gehört.

N = Stichprobengröße.

Bei Gültigkeit von H_0 hat diese Prüfgröße z approximativ eine Standardnormalverteilung.

Auf unser Beispiel angewandt: Bisher wurde die Entscheidungsregel so formuliert, als ob r_{xy} die Prüfgröße wäre. Sie läßt sich aber leicht in eine Entscheidungsregel unter Verwendung der beschriebenen Prüfgröße z übersetzen. Aus der Standardnormalverteilungstabelle entnehmen wir für eine einseitige Überschreitungswahrscheinlichkeit von 5% einen z -Wert von 1.64. Dann können wir die folgende Entscheidungsregel aufstellen: Erhalten wir für die Prüfgröße z einen Wert von mindestens 1.64, dann wird der Test eingeführt. Anderenfalls wird der Test nicht eingeführt. Das mit dieser Entscheidungsregel verbundene Risiko I ist 5%; denn wie oben angegeben, hat die Prüfgröße z bei Gültigkeit von H_0 näherungsweise eine Standardnormalverteilung, und das bedeutet auch: Bei Gültigkeit von H_0 beträgt die Wahrscheinlichkeit, für unsere Prüfgröße einen Wert von $z \geq 1.64$ zu erhalten, etwa 5%.

<< Nähere Begründung: Aus der Formel für z sieht man leicht, daß z genau dann größer als 0 ist, wenn z'_{xy} (also der unserem empirischen r_{xy} entsprechende z' -Wert) größer ist als ζ'_0 (der zu ρ_0 gehörende z' -Wert). Da die z' -Transformation aber streng monoton ist, ist dies genau dann der Fall, wenn r_{xy} größer ist als ρ_0 . Entsprechend ist auch r_{xy} genau dann signifikant größer als ρ_0 , wenn der Wert der Prüfgröße z signifikant größer als 0 ist.⁸²

>>

Nachdem die Entscheidungsregel aufgestellt ist, können wir Daten erheben. Dabei erhalten wir zwischen den Punktwerten aus dem Schulreife-test und der Gesamtnote eine Korrelation von $r_{xy}=0.47$. Um die Prüfgröße z zu berechnen, entnehmen wir die Werte von z'_{xy} und ζ'_0 der Tabelle für die z' -Transformation: Zu einem $r_{xy}=0.47$ gehört ein $z'_{xy}=0.5101$, und für $\rho_0=0.30$ ergibt sich $\zeta'_0=0.3095$. Setzen wir diese Werte in die Formel für die Prüfgröße z ein, dann erhalten wir:

$$z = (0.5101 - 0.3095) \cdot \sqrt{103 - 3} = 2.006$$

Damit ist die Nullhypothese zu verwerfen, und zwar mit einer Irrtumswahrscheinlichkeit von nur

⁸²Bei zweiseitiger Entscheidungsregel würde gelten: r_{xy} unterscheidet sich genau dann signifikant von ρ_0 , wenn der Wert der Prüfgröße z sich signifikant von 0 unterscheidet.

2.3% (s. z-Tabelle). Der Test wird also eingeführt.

<< Die genauere Begründung dieses Tests beruht auf der folgenden, von Fisher nachgewiesenen Tatsache: Ist die gemeinsame Verteilung zweier Zufallsgrößen x und y eine bivariate Normalverteilung und ist die z' -transformierte Korrelation dieser Zufallsgrößen gleich ζ' , dann hat in einer Zufallsstichprobe der Größe N die z' -transformierte Korrelation der Meßwerte (also unser z'_{xy}) approximativ eine Normalverteilung mit Erwartungswert ζ' und Varianz $1/(N-3)$. Dabei ist diese Approximation umso genauer, je größer die Stichprobe.

Für uns heißt das: Stimmt die Nullhypothese (ist also die z' -transformierte Korrelation in der Population gleich ζ'_0), dann hat z'_{xy} approximativ eine Normalverteilung mit Erwartungswert ζ'_0 und Varianz $1/(N-3)$. Nun kann man die obige Formel für die Prüfgröße z folgendermaßen umschreiben:

$$z = \sqrt{N-3} \cdot z'_{xy} - \zeta_0 \cdot \sqrt{N-3}$$

Das kann man auch so sehen, daß zwischen z'_{xy} und unserer Prüfgröße z eine lineare Beziehung

$$z = b \cdot z'_{xy} + a$$

besteht mit

$$b = \sqrt{N-3}$$

und

$$a = - \zeta_0 \cdot \sqrt{N-3}$$

Unter Anwendung von uns bekannten Sätzen über die Auswirkungen von Lineartransformationen kommen wir (unter der Annahme der Gültigkeit von H_0 und der Zusatzvoraussetzungen) zu folgenden Ergebnissen:

- Die Wahrscheinlichkeitsverteilung von z ist ebenfalls (approximativ) eine Normalverteilung (vgl. @)
- Der Erwartungswert dieser Normalverteilung ist $b \cdot \zeta'_0 + a = 0$.
- Die Varianz dieser Normalverteilung ist $(1/(N-3)) \cdot b^2 = 1$.

Zusammengefaßt: Bei Gültigkeit der Nullhypothese und der Zusatzvoraussetzungen hat z in der Tat approximativ eine Normalverteilung mit Erwartungswert 0 und Varianz 1, also eine Standard-Normalverteilung.

>>

2) Signifikanztests zum Vergleich zweier Populationsparameter anhand von unabhängigen Stichproben

Zwei Stichproben heißen unabhängig, wenn die Meßwerte in der einen Stichprobe stochastisch unabhängig von den Meßwerten der anderen Stichprobe sind.

Wichtigste Beispiele für unabhängige Stichproben:

- Es werden mehrere Zufallsstichproben aus derselben Population gezogen.
- Es werden Zufallsstichproben aus verschiedenen Subpopulationen (z.B. Männer und

Frauen) gezogen.

- Die Personen einer Zufallsstichprobe werden zufällig in Teilstichproben eingeteilt ("Randomisierung")

Gegenbeispiele:

- Korrelierte Stichproben (s. @ IV.3)
- Abhängigkeit durch die Versuchsbedingungen: Wenn man die Lernfähigkeit von männlichen und weiblichen Ratten in einem Labyrinth vergleichen will, so muß man nach jedem Lauf das Labyrinth reinigen, um Geruchsspuren zu tilgen. Sonst würden beispielsweise die Männchen die Geruchsspur der Weibchen verfolgen; die Läufe wären also nicht stochastisch unabhängig voneinander.
- Würde man beim Vergleich der Intelligenz von Männern und Frauen Kommunikation zulassen, so wären die Meßwerte der Männer und Frauen nicht stochastisch unabhängig voneinander.

a) Der F -Test für den Vergleich zweier Varianzen

Überprüft, ob zwei Populationsvarianzen σ_A^2 und σ_B^2 gleich sind.

$$H_0: \sigma_A^2 = \sigma_B^2$$

Zusatzvoraussetzungen:

1. Normalverteilung der Meßwerte
2. unabhängige Stichproben

In diesem Test kommen zwei Prüfgrößen zur Anwendung:

$$F_1 = \frac{\hat{\sigma}_A^2}{\hat{\sigma}_B^2} \quad F_2 = \frac{\hat{\sigma}_B^2}{\hat{\sigma}_A^2}$$

Es werden jeweils die erwartungstreuen Varianzschätzungen eingesetzt. Bei Gültigkeit von H_0 haben F_1 und F_2 beide eine " F -Verteilung". Bei der F -Verteilung (die übrigens nach dem bereits aus der allgemeinen Signifikanztest-Theorie bekannten Autor Fisher benannt ist) unterscheiden wir zwei Freiheitsgradzahlen: Die Freiheitsgrade des Zählers und die des Nenners.

Exkurs: Der Begriff der Freiheitsgrade

Eine Varianzschätzung beruht meistens auf einer Quadratsumme (QS). Diese ist die Summe der quadratischen Abweichungen vom Durchschnitt:

$$QS = \sum_i a_i^2$$

mit $a_i = x_i - \bar{x}$.

Wie im Kurs I gezeigt wurde (vgl. S. @), gilt für die a_i :

$$\sum_{i=1}^N a_i = 0$$

Haben wir beispielsweise eine Stichprobe der Größe $N = 4$ und wissen: $a_1 = 3$, $a_2 = 4$, $a_3 = -9$, so ist damit auch a_4 festgelegt: Es ist der Wert, der notwendig ist, damit sich alle a_i zu 0 aufsummieren, also $a_4 = 2$.

Allgemein: In einer Stichprobe der Größe N sind nur $N-1$ a_i -Werte "frei beweglich". Sind sie einmal gegeben, so ist auch der letzte fest. Das drückt man aus, indem man sagt, die Quadratsumme habe $N-1$ Freiheitsgrade. Die erwartungstreue Schätzung einer Populationsvarianz ergibt sich, indem man die Quadratsumme durch die Zahl ihrer Freiheitsgrade dividiert:

$$\hat{\sigma}_x^2 = \frac{QS_x}{df.}$$

Es gibt aber auch Fälle, in denen die Zahl der Freiheitsgrade etwas anders bestimmt wird, allerdings nach einem ähnlichen Grundgedanken. Wenn wir wüßten, daß die Varianz des Merkmals x in der Population der Frauen (A) und in der Population der Männer (B) gleich ist, so könnten wir zur Schätzung dieser Varianz aus beiden Populationen je eine Stichprobe der Größe N_A (Zahl der Frauen) und N_B (Zahl der Männer) ziehen. Selbst wenn die Durchschnitte in beiden Populationen verschieden sind, könnten wir die beiden Quadratsummen addieren:

$$QS_x = QS_A + QS_B = \sum_{i=1}^{N_A} a_i^2 + \sum_{i=1}^{N_B} b_i^2$$

Dabei ist a_i die Abweichung der i -ten Frau vom Stichprobendurchschnitt der Frauen, b_i die Abweichung des i -ten Mannes vom Stichprobendurchschnitt der Männer. Aus ähnlichen Überlegungen wie oben ergibt sich, daß N_A-1 a_i -Werte und N_B-1 b_i -Werte "frei beweglich" sind. Die Freiheitsgrade von QS_x sind also $N_A-1 + N_B-1 = N_A + N_B - 2$.

Als erwartungstreue Varianzschätzung ergibt sich dann:

$$\hat{\sigma}_x^2 = \frac{QS_x}{df.} = \frac{QS_A + QS_B}{N_A + N_B - 2}$$

Aufgrund der Ergebnisse dieses Exkurses läßt sich nun auch eine vorläufige Feststellung über die Zahl der Freiheitsgrade von t -Werten aus Abschnitt @ präzisieren. Dort wurde in einer "Faustregel" zunächst festgestellt, daß t -verteilte Größen (meist) dadurch entstehen, daß man im Nenner eines üblichen z -Bruchs die Standardabweichung durch eine Schätzung ersetzt, die auf einer erwartungstreuen Varianzschätzung beruht. Jetzt können wir hinzufügen: Die Zahl der Freiheitsgrade dieser Varianzschätzung ergibt auch die Freiheitsgrade der t -Verteilung.

<< Präziser läßt sich das folgendermaßen darstellen:

Lehrsatz:

Jede t -verteilte Größe läßt sich darstellen als Quotient einer normalverteilten Größe mit Erwartungswert 0 und einer Schätzung ihrer Standardabweichung; die Zahl der Freiheitsgrade von t ist gleich der Zahl der Freiheitsgrade der Quadratsumme, auf der die Varianz - und damit die Standardabweichungsschätzung beruht.

>>

Für die F -Verteilung, die gleich eingeführt wird, gilt:

Jede F -verteilte Größe läßt sich darstellen als Quotient zweier Schätzungen numerisch gleicher Varianzen. Die Freiheitsgrade von Zähler und Nenner ergeben sich aus der Zahl der Freiheitsgrade in den entsprechenden Quadratsummen.⁸³

Ende des *Exkurses*; zurück nach @ IV.2.a.

Beispiel für den F-Test:

Wir wollen die Hypothese überprüfen, ob die Varianz der Punktwerte in einem Intelligenztest bei den Frauen (A) und den Männern (B) gleich ist.

$$H_0: \sigma_A^2 = \sigma_B^2$$

Wegen des Fehlens von Gegengründen entscheiden wir uns für das 5%-Niveau und eine zweiseitige Alternativhypothese:

$$H_1: \sigma_A^2 \neq \sigma_B^2$$

Wir wählen $N_A = N_B = 31$. Beide Varianzschätzungen werden also 30 Freiheitsgrade haben. Für 30 Freiheitsgrade im Zähler und 30 Freiheitsgrade im Nenner finden wir in der Tabelle der F -

⁸³Die beiden Sätze sind nicht umkehrbar: Nicht jeder Quotient aus einer mit Erwartungswert 0 normalverteilten Größe und einer Schätzung ihrer Standardabweichung ist t -verteilt, und nicht jeder Quotient zweier erwartungstreuer Schätzungen numerisch gleicher Varianzen ist F -verteilt. Weitere Voraussetzungen sind erforderlich, wenn man einen Signifikanztest mathematisch stringent herleiten will. Für ein intuitives Nachvollziehen der von Mathematikern entwickelten Tests reichen aber die obigen Sätze.

Verteilung einen kritischen Wert von 1.84. Dieser hat eine kumulative Wahrscheinlichkeit von 95%; die *einseitige* Überschreitungswahrscheinlichkeit dieses Wertes 1.84 beträgt also 5%.

Versuchen wir eine Entscheidungsregel:

Erste Entscheidungsregel:

Wir berechnen die Prüfgröße

$$F_1 = \frac{\hat{\sigma}_A^2}{\hat{\sigma}_B^2}$$

Erhalten wir ein F_1 von mehr als 1.84, so verwerfen wir H_0 . Bei Gültigkeit von H_0 hat diese Prüfgröße eine F -Verteilung; die Wahrscheinlichkeit, bei Gültigkeit von H_0 durch Zufall ein signifikantes Ergebnis zu erzielen, (also Risiko I) betrüge demnach 5%. Aber: Sollte σ_B^2 größer als σ_A^2 sein, so gilt die Nullhypothese nicht; das würde sich aber nicht in einem großen F_1 , sondern in einem großen F_2 niederschlagen. Die erste Entscheidungsregel wäre also nur dann sinnvoll, wenn wir die folgende einseitige Alternativhypothese gehabt hätten:

$$H_1 : \sigma_A^2 > \sigma_B^2$$

Unsere bisherigen Überlegungen legen eine *zweite Entscheidungsregel* nahe: Wenn eine der beiden Prüfgrößen F_1 oder F_2 größer als 1.84 ist, verwerfen wir H_0 . Anders ausgedrückt: Wir dividieren die größere durch die kleinere Varianzschätzung. Ist der Quotient F größer als 1.84, so verwerfen wir H_0 .

Wie groß ist hier das Risiko I?

$$\begin{aligned} p(F > 1.84) &= p(F_1 > 1.84 \vee F_2 > 1.84) \\ &= p(F_1 > 1.84) + p(F_2 > 1.84) \\ &= 0.05 + 0.05 \\ &= 0.10 \end{aligned}$$

Es gilt ja offenbar:

$$F_1 = 1/F_2$$

Daher kann nur eine von beiden Größen größer als 1 sein; die Ereignisse $F_1 > 1.84$ und $F_2 > 1.84$ schließen also einander aus, so daß wir oben den einfachen Additionssatz verwenden konnten. Da F_1 und F_2 beide eine F -Verteilung mit je 30 Freiheitsgraden im Zähler und Nenner haben, ist die Wahrscheinlichkeit der beiden Ereignisse $F_1 > 1.84$ und $F_2 > 1.84$ je 0.05, solange H_0 zutrifft.

Fassen wir zusammen: Die Überschreitungswahrscheinlichkeit des Wertes 1.84 beträgt 10%, wenn wir den F -Wert nach der zweiten Entscheidungsregel berechnen.

Allgemein: Bei zweiseitiger Alternativhypothese berechnen wir bei diesem Test:

$$F = \frac{\text{größere Varianzschätzung}}{\text{kleinere Varianzschätzung}}$$

Die Zahl der Freiheitsgrade von Zähler und Nenner ergibt sich, indem man die zugehörige Stichprobengröße um 1 vermindert.

Wichtig: Die Irrtumswahrscheinlichkeit beim Verwerfen von H_0 ist doppelt so groß wie die in der Tabelle berücksichtigte einseitige Überschreitungswahrscheinlichkeit des F -Wertes!!

Bei einseitiger Alternativhypothese dividieren wir die laut H_1 größere Varianz durch die laut H_1 kleinere Varianz. Hier können wir die Überschreitungswahrscheinlichkeit gemäß der Tabelle ablesen.

In unserem Beispiel hätten wir, da wir eine zweiseitige H_1 haben, den F -Wert mit 2.5% einseitiger Überschreitungswahrscheinlichkeit in die zweite Entscheidungsregel einbauen müssen. Diese ist bei je 30 Freiheitsgraden im Zähler und Nenner 2.07 (Hays, 1973, S. 278!).

Wir erheben Daten

$$s_A^2 = 102.6; \quad s_B^2 = 103.4$$

Daraus können wir ermitteln

$$QS_A = 102.6 \cdot 31 = 3180.6; \quad \hat{\sigma}_A^2 = 3180.6/30 = 106.02$$

$$QS_B = 103.4 \cdot 31 = 3205.4; \quad \hat{\sigma}_B^2 = 3205.4/30 = 106.85$$

$$F = 106.85/106.02 = 1.008$$

Unser F -Wert ist also sogar viel zu gering, um die Nullhypothese mit 10% Irrtumswahrscheinlichkeit zu verwerfen. Wir behalten H_0 bei. Der Unterschied der beiden Stichprobenvarianzen *kann* durch Zufall entstanden sein.

b) Der t -Test für den Vergleich zweier Populationsmittel bei homogener Varianz und unabhängigen Stichproben

Überprüft, ob zwei Populationsmittel sich um einen bestimmten Betrag unterscheiden:

$$H_0: \mu_A - \mu_B = \Delta_0$$

wichtigster Fall: $\Delta_0 = 0$, also

$$H_0: \mu_A = \mu_B$$

Zusatzvoraussetzungen:

1. Die beiden Populationsverteilungen, deren Mittel verglichen werden, sind Normalverteilungen.
2. Sie haben die gleiche ("homogene") Varianz.

3. Unabhängige Stichproben.

Prüfgröße:

$$t = \frac{\bar{x}_A - \bar{x}_B - \Delta_0}{\sqrt{\frac{QS_A + QS_B}{N_A + N_B - 2} \cdot \frac{N_A + N_B}{N_A \cdot N_B}}}$$

Mit:

- \bar{x}_A, \bar{x}_B = Stichprobendurchschnitte der beiden Stichproben
- QS_A, QS_B = Abweichungs-Quadratsummen in beiden Stichproben
- N_A, N_B = Stichprobengrößen der beiden Stichproben
- Δ_0 = Differenzbetrag der beiden Populationsmittel laut H_0 (meistens 0)

Bei Gültigkeit von H_0 hat diese Prüfgröße eine t -Verteilung.

Freiheitsgrade:

$$df = N_A + N_B - 2$$

<< Den ersten Bruch unter dem Wurzelzeichen kennen wir aus dem Exkurs über Freiheitsgrade: Er ist die Schätzung der - laut Zusatzvoraussetzung in beiden Populationen gleichen - Populationsvarianz.

Knobelaufgabe: Der Zähler des Gesamtbruchs hat eine Wahrscheinlichkeitsverteilung: Durch Zufälle kann er größer oder kleiner sein. Beweisen Sie, daß diese Wahrscheinlichkeitsverteilung eine Normalverteilung ist, daß der gesamte Ausdruck unter dem Wurzelzeichen die Schätzung der Varianz dieser Normalverteilung ist, und daß der Erwartungswert dieser Normalverteilung bei Gültigkeit von H_0 null ist.

>>

Beispiel: Anhand der in IV.2.a erhobenen Daten wollen wir überprüfen, ob Männer (B) und Frauen (A) in dem dort verwendeten Intelligenztest im Durchschnitt gleich Gutes leisten.

$$H_0: \mu_A = \mu_B$$

also: $\Delta_0 = 0.$

Wir entscheiden uns für eine zweiseitige Alternativhypothese, da frühere Experimente gezeigt haben, daß in manchen Intelligenzbereichen die Frauen überlegen sind, in anderen die Männer.

$$H_1: \mu_A \neq \mu_B$$

Wir wählen ein α von 5%. Den t -Test können wir verwenden, wenn die zwei Zusatzvoraussetzungen erfüllt sind.

Exkurs: Das Problem der Zusatzvoraussetzungen bei Signifikanztests

a) Der traditionelle Weg: "Vortests"

Häufig wurde - und wird auch heute noch - vorgeschlagen, die Zusatzvoraussetzungen von Signifikanztests durch Vorschaltung anderer Signifikanztests zu überprüfen. Die Zusatzvoraussetzung der Varianzhomogenität würde man mit dem in IV.2.a besprochenen F -Test überprüfen; Normalverteilungsannahmen überprüft man mit sog. Tests für die Güte der Anpassung.

In IV.2.a fanden wir keine signifikante Abweichung von der Hypothese, daß die beiden Populationsvarianzen gleich sind. Diese Hypothese ist damit aber noch nicht bewiesen!

Allgemein: In den "Vortests" tauchen die Zusatzvoraussetzungen, die für den Haupttest überprüft werden sollen, meistens als Nullhypothesen auf. Eine nicht signifikante Abweichung von H_0 ist aber noch kein Beweis von H_0 . (Nullhypothesen lassen sich nie beweisen, sondern nur widerlegen!) Es kann sein, daß die H_1 richtig ist - daß also die Zusatzvoraussetzung, die wir überprüfen, nicht erfüllt ist, daß aber die Teststärke des Vortests zu gering war, um ein signifikantes Ergebnis zu liefern. Z.B.: Die Stichprobe war zu klein. Viele Zusatzvoraussetzungen sind aber um so wichtiger, je kleiner die Stichprobe ist. Also: Ausgerechnet da, wo eine Überprüfung der Zusatzvoraussetzungen wichtig wäre - eben bei kleinen Stichproben - versagen die Vortests, weil es wegen geringer Teststärke leicht vorkommt, daß eine in der Population vorhandene Abweichung von den Zusatzannahmen sich nicht als signifikant nachweisen läßt.

b) Andere Verfahren

Die Überprüfung von Zusatzvoraussetzungen anhand von Vortests ist aus den genannten Gründen nicht befriedigend. Heute verwendet man andere Methoden.

ba) Skalentransformation

Die Normalverteilungsannahme wird für die Stichprobe erfüllt, indem man die Maßskala so transformiert, daß eine Normalverteilung entsteht. Die Signifikanztests führt man dann mit den transformierten Testscore durch.

Kritik:

1. Wenn die Stichprobenverteilung über die transformierte Skala eine Normalverteilung ist, braucht das noch nicht für die Populationsverteilung gelten. Allerdings wird die Populationsverteilung nicht mehr stark von einer Normalverteilung abweichen.
2. Die Skalentransformation, die hier erforderlich wird, ist nicht linear. Damit ändert sich die Aussage der Hypothesen: Denn nur bei linearer Transformation würde aus

$$\mu_{X'A} = \mu_{X'B} \text{ bzw. } \mu_{X'A} \neq \mu_{X'B}$$

folgen, daß

$$\mu_{xA} = \mu_{xB} \text{ bzw. } \mu_{xA} \neq \mu_{xB}$$

Anders formuliert: Wenn man die Nullhypothese für die transformierten Meßwerte (x') verwerfen kann, kann man sie noch nicht für die Ausgangswerte (x) verwerfen.

Ein wichtiger Artikel zum Problem der Skalentransformation: Lienert (1962).

bb) Versuchsplanerische Maßnahmen

Es läßt sich beweisen, daß die Zusatzvoraussetzung der Varianzhomogenität sehr erheblich an Gewicht verliert, wenn die Stichproben gleich groß sind. Wenn wir gleich große Stichproben A und B zugrunde legen, können wir auf die Annahmen der Varianzhomogenität weitgehend verzichten. Wir verlagern also das Problem der Varianzhomogenität in das Stadium der Versuchsplanung. Ähnlich läßt sich die Normalverteilungsannahme durch versuchsplanerische Maßnahmen entschärfen: Je größer die Stichprobe, um so unbedeutender wird die Normalverteilungsannahme.

Faustregel: Nur bei starken Abweichungen von der Normalverteilung braucht man eine Stichprobengröße über 100, um die Normalverteilungsannahme zu entschärfen.

bc) Verwendung von Signifikanztests mit weniger Zusatzvoraussetzungen

Schon seit längerer Zeit gibt es einen t -Test, der auch für heterogene Varianzen anwendbar ist. Insbesondere sind hier aber die "verteilungsfreien" oder "parameterfreien" Verfahren zu erwähnen, die in Abschnitt @ behandelt werden. Nachteil aller dieser Verfahren: Die Teststärke ist geringer als bei den "klassischen".

bd) Abschätzung des Fehlers, der bei einer Verletzung der Zusatzannahmen auftritt

<<

Wir unterscheiden das angenommene ("theoretische", "nominelle") Risiko I und das tatsächliche Risiko I. Das angenommene Risiko I ist das Risiko I, das wir zugrunde legen, wenn wir unter Annahme der Zusatzvoraussetzungen aus unseren Tabellen einen kritischen Wert der Prüfgröße entnehmen. Wenn nun die Zusatzvoraussetzungen unseres Tests verletzt sind, kann das Risiko I - also die Wahrscheinlichkeit eines signifikanten Ergebnisses trotz Gültigkeit von H_0 - anders sein ("tatsächliches Risiko I"), als wir bei der Aufstellung der Entscheidungsregel zugrunde legten.

Ist das tatsächliche Risiko I geringer als das angenommene (konservativer Fehler oder konservative Verletzung der Voraussetzungen unseres Signifikanztests), so ist dies nach dem Prinzip des konservativen Testens weit weniger schlimm als das Gegenteil (radikaler Fehler oder radikale Verletzung der Zusatzvoraussetzungen).

Wir überschätzen die Irrtumswahrscheinlichkeit beim konservativen Fehler; beim radikalen Fehler unterschätzen wir sie, was offenbar schlimmer ist. (Beispiel: s.u.: χ^2 -Test).

>>

Ende des *Exkurses*. Zurück nach IV.2.b.

Die Zusatzannahme der Varianzhomogenität ist in unserem Fall unwichtig, da $N_A = N_B$. Wir wollen annehmen, daß die Stichprobenverteilung der Meßwerte bereits so gut an eine Normalverteilung angeglichen ist, daß eine Skalentransformation nicht lohnen würde.

Als Stichprobendurchschnitte erhalten wir:

$$\bar{x}_A = 103.62; \quad \bar{x}_B = 98.41$$

Wir setzen ein:

$$t = \frac{103.62 - 98.41 - 0}{\sqrt{\frac{3180.06 + 3205.4}{31 + 31 - 2} \cdot \frac{31 + 31}{31 \cdot 31}}} = 1.988$$

$$df = 31 + 31 - 2 = 60$$

Der kritische t -Wert für 60 Freiheitsgrade und 5% zweiseitige Überschreitungswahrscheinlichkeit wäre 2.000. Unser Ergebnis ist also nicht signifikant, allerdings fast signifikant. In einem solchen Fall würden wir die Nullhypothese "nur noch mit einem halben Auge" beibehalten. Die Irrtumswahrscheinlichkeit beim Verwerfen von H_0 wäre wenig mehr als 5%. Die Wahl der Zahl 5% als Signifikanzgrenze ist ja nur eine relativ willkürliche Konvention. Hätten wir einen größeren t -Wert erhalten und H_1 eindeutig akzeptiert, so könnten wir sagen: Die Frauen erzielen in unserem Test durchschnittlich mehr Punkte als die Männer.

Abschließend noch exemplarisch ein Hinweis auf Fragen des Skalenniveaus: Aus der Aussage "Frauen erzielen in unserem Test durchschnittlich mehr Punkte als Männer" folgt nur bei Messung auf Intervallskalenniveau die Schlußfolgerung "Frauen sind in der von diesem Test gemessenen Intelligenzfunktion intelligenter als Männer". Manchmal wird sogar das Intervallskalenniveau unter den Zusatzvoraussetzungen des t -Tests aufgeführt. Aus Gründen, die in Statistik I bereits angedeutet wurden (@), kann aber die Frage, ob sich die Durchschnitte von Meßwerten unterscheiden, auch in Situationen auftreten, in denen es gar nicht um die entsprechenden Durchschnitte des gemessenen Merkmals geht. Das läßt sich folgendermaßen verallgemeinern:

Allein für die Überprüfung der Frage, ob eine oder mehrere Populationsverteilungen von Meßwerten von einer Nullhypothese abweichen, sind keinerlei Anforderungen an das Skalenniveau erforderlich. Das Skalenniveau wird erst bei Rückschlüssen auf die Verteilung(en) des gemessenen Merkmals wichtig.

c) Der z -Test für den Vergleich zweier Produktmomentkorrelationen

Überprüft, ob zwei Produkt-Moment-Korrelationen in der Population gleich sind:

$$H_0: \rho_1 = \rho_2.$$

ρ_1 und ρ_2 können die Korrelationen der gleichen Variablen in zwei verschiedenen Populationen sein (z.B.: Korrelieren die Tests x und y bei Frauen und Männern gleich hoch?); es können aber auch die Korrelationen verschiedener Variablen in der gleichen Population (s.u.: Beispiel) oder in verschiedenen Populationen sein.

Zusatzvoraussetzungen:

1. Die empirischen Korrelationen werden an unabhängigen Stichproben erhoben. Die Stichprobengröße soll mindestens 50 sein. (Anmerkung: kleinere Stichproben zulässig, wenn beide gleich groß).
2. Die zweidimensionalen Populationsverteilungen, deren Korrelationen verglichen werden, sind bivariate Normalverteilungen.

Prüfgröße:

$$z = \frac{z'_1 - z'_2}{\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}}$$

mit:

z'_1, z'_2 = die z' -Werte, die zu den zwei empirischen Korrelationen gehören.

N_1, N_2 = Stichprobengrößen

Bei Gültigkeit von H_0 hat die Prüfgröße eine Normalverteilung. (Knobelaufgabe: Beweisen Sie diesen Satz!)

Beispiel: Es ist zu untersuchen, ob zwei Schulreifetests die Schulleistung gleich gut vorhersagen.

Versuchsplan 1: Man erhebt beide Schulreifetests an den gleichen Kindern und korreliert die Ergebnisse mit den Gesamt-Schulnoten nach einem Jahr.

Fehler: Es sind nicht zwei unabhängige Stichproben.

Versuchsplan 2: Der erste Test (x) wird mit 303 Kindern durchgeführt, der zweite (y) mit 153 anderen Kindern; nach einem Jahr korreliert man mit den Gesamtschulnoten (z). Man achtet darauf, daß die schulischen Voraussetzungen und die Kriterien bei einer Notenvergabe für beide Gruppen gleich sind.

Bei so großen Stichproben ist die Normalverteilungsannahme schon weitgehend entschärft.

$\alpha = 5\%$, zweiseitig.

Ergebnisse:

$$r_{xz} = 0.584 \quad r_{yz} = 0.640$$

zugehörige z' -Werte:

$$z'_1 = 0.6685 \quad z'_2 = 0.7582$$

$$z = \frac{0.6685 - 0.7582}{\sqrt{\frac{1}{303 - 3} + \frac{1}{153 - 3}}} = -0.897$$

Die zweiseitige Überschreitungswahrscheinlichkeit dieses z-Wertes ist etwa 37%. Wir behalten H_0 bei: Der Unterschied der empirischen Korrelationen *kann* zufällig sein; es *kann* sein, daß beide Tests gleich gut sind. Allerdings sind die obigen Korrelationen die besten Parameterschätzungen: Es ist wahrscheinlicher, daß Test y besser ist, doch ist die Hypothese, daß beide gleich gut sind, nicht mit einer hinreichend niedrigen Irrtumswahrscheinlichkeit zu widerlegen.

<<

Mit diesem Signifikanztest kann man auch eine allgemeinere H_0 überprüfen

$$\zeta'_1 - \zeta'_2 = \Delta_0$$

Die Größe Δ_0 wäre dann im Zähler der Prüfgrößenformel zu subtrahieren.

>>

3) Signifikanztests zum Vergleich zweier Populationsparameter anhand von korrelierten Stichproben

Zwei Stichproben heißen korreliert, wenn (wie bei einer Korrelation) Meßwertpaare vorliegen, von denen je ein Meßwert zu jeder der beiden Stichprobe gehört.

Wichtigste Beispiele für korrelierte Stichproben:

- Die zwei Messungen x und y , für deren Populationsvergleich der Test durchgeführt wird, werden an denselben Vpn erhoben.
- Die Meßwerte, die ein Paar bilden, werden an Personen erhoben, die einander ähnlich sind.

Bei der ersten Form der Gewinnung korrelierter Stichproben spricht man von "Meßwiederholung", und bei der zweiten von "parallellisierten Paaren" (engl.: 'matched pairs'). Für diese zweite Form bildet man Paare von Vpn, die sich in irgendwelchen bekannten, für die untersuchte Variable relevanten Merkmalen ähnlich sind. Möglicherweise erhebt man dafür sogar "Vortests". Im Experiment wird dann bei jedem Paar zufällig entschieden, welche Person in welche Stichprobe kommt.⁸⁴

⁸⁴Wenn man vermutet, daß die in einem Experiment untersuchte abhängige Variable stark von der Tageszeit oder auch vom Wochentag abhängt ("Montags- und Wochenend-Flaute"), kann man auch Paare von einander ähnlichen Versuchsterminen bilden und dann für jedes Termin-Paar zufällig entscheiden, an welchem Termin welche Versuchsbedingung realisiert wird. Durch diese Zufallsentscheidung werden dann implizit auch die Paare von Personen, die sich für zwei ähnliche Termine entschieden haben, zufällig auf die Gruppen verteilt.

Wichtig: Abgesehen von der durch die Paarbildung bedingten Abhängigkeit müssen die Stichproben voneinander unabhängig sein.

Vorteil der Versuchspläne mit abhängigen Stichproben: Größere Teststärke, solange die Meßwerte, die ein Paar bilden, positiv korrelieren.⁸⁵

a) Der t-Test für Paardifferenzen

Auch t-Test für korrelierte (oder abhängige) Stichproben genannt. Es handelt sich um den in II. 2 ausführlich behandelten Test.

$$H_0: \mu_x - \mu_y = \Delta_0$$

Häufigster Fall: $\Delta_0 = 0$

Wir bilden die Differenzen: $d_i = x_i - y_i$

und testen die Hypothese: $H_0: \mu_d = \Delta_0$

anhand des t-Tests für einen einzelnen Mittelwert.

Es läßt sich zeigen, daß:

$$\begin{aligned} \bar{d} &= \bar{x} - \bar{y} \\ s_d^2 &= s_x^2 - 2 r_{xy} \cdot s_x \cdot s_y + s_y^2 \\ &= s_x^2 - 2 s_{xy} + s_y^2. \end{aligned}$$

Diese Formeln verwenden wir dann, wenn wir x, y, s_x^2, s_y^2 und r_{xy} schon ausgerechnet haben.

b) Die Überprüfung der Varianzgleichheit

<<

Dieser vom Verfasser des vorliegenden Skripts entwickelte Test gehört nicht zum "Standardrepertoire".

Die zu untersuchende Nullhypothese ist

$$H_0: \sigma_x^2 = \sigma_y^2.$$

Für den Signifikanztest definiert man Größen u und v durch die Gleichungen $u := x + y$ und $v := x - y$. Dann läßt sich zeigen,⁸⁶ daß die Aussage dieser Nullhypothese gleichbedeutend ist

⁸⁵Die Korrelation ist dabei so zu verstehen, als ob die Meßwerte, die ein Paar bilden, von derselben Person stammen.

⁸⁶Vgl. die Formel

$$\text{Cov}(V + W, X + Y) = \text{Cov}(V, X) + \text{Cov}(V, Y) + \text{Cov}(W, X) + \text{Cov}(W, Y)$$

aus dem Abschnitt "Die Kovarianz von Lineartransformationen und von Summen" von Statistik I. Ersetzt man in dieser Formel V und X durch x sowie W durch y und Y durch -y, so erhält man

$$\begin{aligned} \text{Cov}(x + y, x - y) &= \text{Cov}(x, x) + \text{Cov}(x, -y) + \text{Cov}(y, x) + \text{Cov}(y, -y) \\ &= \sigma_x^2 - \sigma_{xy} + \sigma_{xy} - \sigma_y^2. \end{aligned}$$

Für das zweite Gleichheitszeichen ist zu beachten:

- Die Kovarianz einer Variablen mit sich selbst ist ihre Varianz.

mit:

$$H_0: \rho_{uv} = 0.$$

Die so umgeformte H_0 kann mit dem entsprechenden t -Test geprüft werden.

>>

Nachtrag zu IV.2 und IV.3

Bei den Tests, bei denen zwei Populationsparameter verglichen werden, sind in der Praxis zwei Fälle zu unterscheiden: Entweder handelt es sich um zwei verschiedene Populationen oder um verschiedene Maße, deren Verteilungen in der gleichen Population (hinsichtlich bestimmter Parameter, z.B. hinsichtlich des Mittels) verglichen werden.

Wichtig: Das hat nur am Rande etwas mit der Unterscheidung von unabhängigen und korrelierten Stichproben zu tun, wie wir uns leicht an vier Beispielen veranschaulichen können:

1.) Vergleich zweier Populationen

a) mit unabhängigen Stichproben

Wir vergleichen die Intelligenz von Männern und Frauen mit dem in IV.2 behandelten Versuchsplan.

b) mit korrelierten Stichproben

Wir wollen untersuchen, ob Erstgeborene mit jüngeren Geschwistern extravertierter sind als ihre zweitgeborenen Geschwister. Hierzu testen wir 50 Geschwisterpaare mit einem Extraversionstest - jeweils ein Erstgeborenes bildet mit seinem nächstjüngeren Geschwister ein Beobachtungspaar. Offenbar korrelierte Stichproben; verglichen wird die Extraversion der Population der Erstgeborenen mit der Extraversion der Zweitgeborenen.

In beiden Fällen würde man die Nullhypothese formulieren:

$$\mu_{xA} = \mu_{xB}$$

(Also: μ_x ist in beiden Populationen gleich.)

2.) Vergleich der Verteilung zweier Maße in der gleichen Population

Wir wollen die Reaktionszeiten nach Alkoholgenuß mit den Reaktionszeiten in nüchternem Zustand vergleichen. Die Nullhypothese soll lauten: In der Population der deutschen Erwachsenen ist die durchschnittliche Reaktionszeit unter Alkohol (x) gleich der durchschnittlichen Reaktionszeit ohne Alkohol (y).

$$H_0: \mu_x = \mu_y$$

Hier werden nicht, wie bei den vorhergehenden Beispielen, die Populationsmittel zweier Populationen (Männer und Frauen bzw. Erst- und Zweitgeborene), sondern die Mittelwerte der

-
- Die Beziehung von y und $-y$ kann als "lineare Umpolung" betrachtet werden, bei der sich das Vorzeichen der Kovarianz ändert, während der Betrag erhalten bleibt.

Entsprechend den Definitionen $u := x + y$ und $v := x - y$ läßt sich das Ergebnis auch als

$$\sigma_{uv} = \sigma_x^2 - \sigma_y^2$$

schreiben. Die Gleichung $\sigma_{uv} = 0$ (und damit $\rho_{uv} = 0$) gilt also genau dann, wenn die Varianzen σ_x^2 und σ_y^2 gleich sind.

Verteilungen zweier Maße (x und y) in der gleichen Population (deutsche Erwachsene) verglichen.

Auch hier können wir einen Versuchsplan mit unabhängigen oder mit korrelierten Stichproben ansetzen:

a) mit unabhängigen Stichproben

Wir ziehen zwei Stichproben; die eine wird unter Alkohol, die andere nüchtern getestet.

b) mit korrelierten Stichproben

Wir erheben beide Maße an denselben Vpn.

Oder: Wir ziehen zunächst eine große Stichprobe und bilden aufgrund eines Vortests Vpn-Paare mit etwa gleicher Reaktionsfähigkeit.

Wie wir an den vier Beispielen sehen, sind sowohl Vergleiche zweier Populationen als auch Vergleiche zweier Maße in der gleichen Population möglich, sowohl mit korrelierten als auch mit unabhängigen Stichproben.

4) Verteilungsfreie oder parameterfreie Signifikanztests

a) Allgemeines

Die bisher besprochenen "klassischen Signifikanztests" haben zweierlei gemeinsam:

1. Es werden immer Hypothesen über ganz bestimmte, genau beschriebene Parameter getestet.
2. Die Tests machen Zusatzvoraussetzungen über die Populationsverteilungen (insbesondere Normalverteilung).

Die jetzt zu besprechenden Tests unterscheiden sich in beiden Punkten von klassischen.

1. Die Hypothesen sind unspezifischer. Die Alternativhypothese lautet beispielsweise: Es bestehen Unterschiede in der zentralen Tendenz.

Nachteil: Wenn wir diese Alternativhypothese akzeptieren, wissen wir noch nicht, ob sich nun die Mediane, die Erwartungswerte oder andere Kennziffern der zentralen Tendenz in den beiden Populationen unterscheiden. Aus den genannten Gründen heißen diese Tests "parameterfrei".

Bei den meisten Tests für Unterschiede in der zentralen Tendenz treten allerdings vor allem die Mediane in den Vordergrund.

Vorteil: Aussagen über die Mediane der Meßwerte lassen sich auch dann in Aussagen über die gemessene Größe übersetzen, wenn nur Rangskalenniveau vorliegt. (Gegenteil bei den klassischen Tests: vgl. IV.2.b, Schlußbemerkungen.)

2. Weniger (meistens keine) Zusatzvoraussetzungen über Populationsverteilungen (daher "verteilungsfrei").

Der Vorteil liegt auf der Hand. Er wird allerdings mit einem Verlust der Teststärke erkaufte. Man sagt auch: Die verteilungsfreien Verfahren haben eine verminderte Effizienz.

Nun läßt sich ein solcher Teststärkeverlust durch Stichprobenvergrößerung wieder ausgleichen. Wenn wir beispielsweise den t-Test für Paardifferenzen mit 94 Meßwertpaaren durchführen, so erzielen wir eine größere Teststärke, als wenn wir einen entsprechenden verteilungsfreien Test - den Wilcoxon-Test für Paardifferenzen - anwenden. Um diese Teststärke mit dem Wilcoxon-Test zu erzielen, braucht man 100 Vpn. Man sagt auch: Die Effizienz des Wilcoxon-Tests beträgt 94% der Effizienz des t-Tests für Paardifferenzen.

Allgemein *definiert* man: Das Effizienzverhältnis zweier Tests ist das umgekehrte Verhältnis der zur Erzielung gleicher Teststärke erforderlichen Stichprobengrößen:

$$E_1 : E_2 = N_2 : N_1.$$

Gibt man die Effizienz eines verteilungsfreien Tests einfach als "soundsoviel %" an, so bedeutet das "soundsoviel % der Effizienz des entsprechenden klassischen Tests".

Im obigen Beispiel:

$$E_W : E_t = 94 : 100$$

$$E_W = 94\%$$

Dabei bedeutet

E_W = Effizienz des Wilcoxon Tests;

E_t = Effizienz des t-Tests für Paardifferenzen.

Wir sehen: Die Effizienz eines Tests ist keine absolute Größe, sondern eine Verhältnisgröße.

Wir wollen hier nur drei der vielen verteilungsfreien Signifikanztests besprechen: Den Vorzeichentest, den Mann-Whitney U-Test und den χ^2 -Test. Für die übrigen wird auf die Literatur verwiesen, z.B. Siegel (1976, eine leicht verständliche Darstellung), Bortz, Lienert und Boehnke (1990, ein umfassendes Handbuch) sowie Büning & Trenkler (1978, vor allem für mathematische Ableitungen brauchbar).

b) Der Vorzeichentest

Diesen Test haben wir unter II.2 näher betrachtet (auf einen Spezialfall angewendet).

Zweck des Tests: Vergleich der zentralen Tendenz mit korrelierten Stichproben.

H_0 : Es besteht kein Unterschied in der zentralen Tendenz.

Man überprüft bei jedem Meßwertpaar, ob der zweite Meßwert größer oder kleiner ist als der erste. Ist der zweite Meßwert größer, so gibt man diesem Paar das Zeichen +; ist der zweite Meßwert kleiner, so gibt man dem Paar das Zeichen -. Sind beide Meßwerte gleich, so gibt man dem Paar das Zeichen 0.

Man zählt aus:

f_+ = Zahl der Paare mit Zeichen +.

f_- = Zahl der Paare mit Zeichen -.

f_0 = Zahl der Paare mit Zeichen 0.

Man berechnet: $N = f_+ + f_-$.

Die Meßwerte mit gleichen Meßwerten läßt man hier also bei der Bestimmung von N (und bei

allen weiteren Berechnungen) ausfallen. Für den Signifikanztest gilt die Größe f_+ als binomialverteilt mit $N = f_+ + f_-$ und $p = 0.5$, wenn H_0 richtig ist.

Beispiel: Wir wollen überprüfen, ob eine bestimmte Methode der Psychotherapie Einfluß auf die vegetative Reaktionsbereitschaft der Patienten hat. Als Maß der vegetativen Reaktionsbereitschaft verwenden wir die Psychogalvanische Reaktion (PGR). Diese besteht in einem Absinken des Hautwiderstandes unmittelbar nach einem Schrecken.

Vor der Therapie werde die Patienten einem lauten Knall ausgesetzt; die PGR wird gemessen. Nach der Therapie wird das Experiment wiederholt.

Da es sich um einen Vergleich der zentralen Tendenz in korrelierten Stichproben handelt, entscheiden wir uns für den Vorzeichentest. Der t -Test ist nicht angebracht, da die Meßwerte erfahrungsgemäß nicht normalverteilt sind.

H_0 : In der Population der Patienten würde es keinen Unterschied in der zentralen Tendenz der PGR vor und nach der Therapie geben. Eventuelle Stichprobeneffekte sind zufällig.

H_1 : Es bestehen Unterschiede in der zentralen Tendenz der PGR vor und nach der Therapie.

$\alpha = 5\%$; zweiseitig.

Wir führen das Experiment mit 53 Patienten durch. Bei 17 Patienten ist die PGR nach der Therapie stärker, bei 32 Patienten ist sie nach der Therapie schwächer; bei den restlichen Patienten (4) ist die PGR vor und nach der Therapie gleich stark. Also

$$f_+ = 17$$

$$f_- = 32$$

$$f_0 = 4$$

$$N = 17 + 32 = 49$$

Ist H_0 richtig, so gilt f_+ als binomialverteilt mit $N = 49$ und $p = 0,5$. Der Erwartungswert dieser Binomialverteilung wäre $49 \cdot 0,5 = 24,5$; die Standardabweichung wäre $\sigma = 3,5$.

Normalverteilungsapproximation der Binomialverteilung:

$$z = (17,5 - 24,5)/3,5 = -2,00$$

Wir können die Hypothese, daß vor und nach der Psychotherapie keine Unterschiede in der zentralen Tendenz der PGR bestehen, mit 4,55% Irrtumswahrscheinlichkeit gerade noch verwerfen: Die größere Häufigkeit eines Absinkens der PGR dürfte nicht zufällig sein.

c) Der Mann-Whitney U-Test

Die Arbeitsweise dieses Tests wurde bereits an einem überschaubaren Beispiel demonstriert (vgl. S. @). Hier nun noch einmal eine allgemeinere Darstellung.

Der U-Test von Mann und Whitney ist ein parameterfreier "Bruder" des t -Test für unabhängige Stichproben; d.h.: Er überprüft Unterschiede in der zentralen Tendenz bei unabhängigen Stichproben. Die Prüfgröße U ergibt sich, indem wir jeden Meßwert der einen

Stichprobe mit jedem Meßwert der anderen vergleichen und auszählen, wie viele dieser Vergleiche zugunsten der ersten Stichprobe ausgehen (wobei es vom Ergebnis her gleich ist, welche Stichprobe wir als "erste" bezeichnen).

ca) Definition und Berechnung der Prüfgröße U

Wie häufig in der Statistik, ist auch bei diesem Signifikanztest die zum Verständnis geeignete Darstellung für Rechenzwecke weniger geeignet. Dies gilt auch für die Prüfgröße U: Wenn man alle Meßwerte in Rangplätze transformiert und für eine der beiden Teilstichproben die Summe dieser Rangplätze (T) bildet, dann läßt sich die Prüfgröße U - die Zahl der Vergleiche, die zugunsten dieser Stichprobe ausfällt - nach der folgenden Formel berechnen:

$$U = T - N_1 \cdot (N_1 + 1) / 2$$

Dabei ist N_1 der Umfang der Stichprobe, deren Rangsumme zugrundegelegt wurde.

Beispiel: In der auf S. @ behandelten Rangreihe CCPCPPPP stehen die Meßwerte der Gruppe C an den Rangplätzen 1, 2, 4 und 5; die Summe dieser Rangplätze ist $T=12$. Da die Gruppe C 4 Vpn umfaßt, berechnen wir

$$U = 12 - 4 \cdot (4 + 1) / 2 = 2,$$

und das ist genau der U-Wert, den wir auch unmittelbar aufgrund der Definition von U ermittelt haben.

Der bei diesem Verfahren ersparte Rechenaufwand ist normalerweise erheblich größer, als in dem Beispiel, in dem für die direkte Ermittlung von U nur 20 Vergleiche erforderlich waren: Wenn z.B. die beiden Teilstichproben je 40 Vpn haben, wären für die Berechnung von U aufgrund von Einzelvergleichen bereits $40 \cdot 40 = 1600$ solche Vergleiche erforderlich.

cb) Die Signifikanzprüfung von U

Die Signifikanzprüfung von U läßt sich erheblich vereinfachen: Bei Gültigkeit von H_0 und "hinreichend großen" (s.u.) Stichproben ist die Wahrscheinlichkeitsverteilung der Prüfgröße näherungsweise identisch mit einer Normalverteilung mit folgenden Kennwerten:

$$E(U) = N_1 \cdot N_2 / 2$$

$$\text{Varianz}(U) = N_1 \cdot N_2 \cdot (N_1 + N_2 + 1) / 12$$

Dabei sind N_1 und N_2 die Größen der beiden Teilstichproben.

Beispiel: Wir haben das Coffein-Placebo-Experiment wiederholt mit $N_1=54$ Vpn in der C-Gruppe und $N_2=55$ Vpn in der P-Gruppe. Für C ergab sich eine Rangsumme von $T=2549$, also ein U von

$$U = 2549 - 54 \cdot (54+1)/2 = 1064$$

Da die Stichproben hinreichend groß sind (s.u.), ist U bei Gültigkeit von H_0 näherungsweise normalverteilt mit

$$E(U) = 54 \cdot 55 / 2 = 1485$$

$$\text{Varianz}(U) = 54 \cdot 55 \cdot (54+55+1) / 12 = 27225$$

Die Standardabweichung ist die Quadratwurzel dieser Varianz, d.i. 165.

Der tatsächliche Wert von U (1064) liegt erheblich unter dem bei Gültigkeit von H_0 anzusetzenden Erwartungswert von 1485. Kann das noch Zufall sein? D.h.: Mit welcher Wahrscheinlichkeit nimmt eine mit Erwartungswert 1485 und Standardabweichung 165 normalverteilte Größe einen Wert von 1064 oder weniger an? (Und diese Wahrscheinlichkeit ist erst die einseitige Überschreitungswahrscheinlichkeit; diese müssen wir bei zweiseitiger Fragestellung verdoppeln!). Wie bei Normalverteilungen üblich, berechnen wir diese Überschreitungswahrscheinlichkeit, indem wir die normalverteilte Größe (U) in einen z -Wert umrechnen und dann in der z -Tabelle nachschlagen. Also:

$$z = (1064 - 1485) / 165 = - 2.55$$

Die einseitige Überschreitungswahrscheinlichkeit dieses z -Werts können wir aus der z -Tabelle als 0.005 bestimmen; die zweiseitige ist also 0.01.

Ergebnis: Die Wahrscheinlichkeit, daß so große oder noch größere Abweichungen von der Zufallserwartung bei Gültigkeit von H_0 zufällig auftreten, beträgt 0.01; die vorgefundene Abweichung von der Zufallserwartung ist also sehr signifikant.

<< Wenn man es ganz genau nimmt, müßte man bei der Berechnung eines z -Werts berücksichtigen, daß die Prüfgröße U nur ganzzahlige Werte annehmen kann. Ähnlich wie bei der Normalverteilungsapproximation der Binomialverteilung (vgl. S. 153) müßte man eine Kontinuitätskorrektur vornehmen und sagen: Der U -Wert von 1064 steht für das Intervall 1063.5 bis 1064.5; daher müßten wir genau genommen fragen: Wie groß ist die Wahrscheinlichkeit eines U -Werts von weniger als 1064.5. Faktisch macht das aber nur einen minimalen Unterschied: Mit dieser Korrektur ergibt sich ein z -Wert von $(1064.5-1485)/165=-2.5485$, ohne die Korrektur ein z -Wert von -2.5515 . Da die Effekte einer solchen Korrektur derart geringfügig sind, ist sie beim

U-Test nicht üblich, obwohl sie eigentlich erforderlich wäre.⁸⁷

>>

Das hier vorgestellte Verfahren der Signifikanzprüfung von U über eine Normalverteilungsapproximation ist nur bei hinreichend großen Stichproben angebracht; bei kleineren Stichproben kann man Überschreitungswahrscheinlichkeiten bzw. kritische U-Werte aus Tabellen entnehmen, die sich z.B. in den auf S. @ aufgeführten Büchern finden.

<<

Diese Tabellen beruhen auf kombinatorischen Überlegungen, wie sie im Beispiel S. @ demonstriert wurden. In dem Buch von Büning und Trenkler ist Tabelle L anzuwenden; dort ist allerdings nicht die Prüfgröße U zugrundegelegt, sondern die Rangsumme (unser T), für die dort das Symbol W_N verwendet wird (und deshalb heißt der Test dann nicht mehr U-Test, sondern W_N -Test).

>>

Die Frage, was eine hinreichend große Stichprobe ist, hängt natürlich davon ab, mit welcher Genauigkeit der Approximation man sich zufriedengibt. Für die Praxis kann man von folgendem Verfahren ausgehen: Wenn beide Stichproben mindestens je 25 Vpn umfassen, kann man problemlos die Normalverteilungsapproximation verwenden. Anderenfalls geht man davon aus, daß die Stichprobengrößen, für die eine solche Approximation nicht angebracht ist, in den Tabellen erfaßt sind. D.h.: Wenn unsere Stichprobengrößen in der Tabelle nicht enthalten sind, verwenden wir die Normalverteilungsapproximation.

cc) Die Behandlung von Rangkopplungen

Wir sind bisher davon ausgegangen, daß die Meßwerte sich eindeutig in eine Rangreihe bringen lassen. Dies ist nicht mehr der Fall, wenn es schon in den Ausgangswerten - also vor der Rangreihenbildung - für mehrere Vpn den gleichen Meßwert gibt. In solchen Fällen spricht man von einer Rangkopplung (engl. tie) oder auch von Bindungen mehrerer Meßwerte. Im Zusammenhang mit diesen Rangkopplungen müssen sowohl für die Bestimmung der Prüfgröße als auch für die Signifikanzprüfung Zusatzregeln herangezogen werden. Die Bildung der Prüfgröße U erfolgt dann in der Weise, daß bei jedem Vergleich zweier gleicher Meßwerte eine Art "Punkteteilung" stattfindet: Für diesen Vergleich gibt es für beide Stichproben je einen halben Punkt. Beispiel: In 2 Gruppen A und B fallen die folgenden Meßwerte an:

Gruppe A: 3 5 8 9

Gruppe B: 7 8 10 12

⁸⁷Begründung für den Verzicht auf die Kontinuitätskorrektur: Selbst bei einer Größe von 25 der beiden Stichproben, die im allgemeinen für die Normalverteilungsapproximation gefordert wird, ergibt sich für U eine Varianz von 2656.25 und damit eine Standardabweichung von 51.5. (Wollen Sie diese Angaben aufgrund der Formel für die Varianz von U überprüfen?) Daher bewirkt eine Veränderung um 0.5 im Zähler des z-Bruchs selbst bei dieser Stichprobengröße nur eine Änderung des z-Werts von 0.01. Bei größeren Stichproben ist die Veränderung noch geringfügiger, wie am Beispiel gezeigt wurde.

Wir können leicht überprüfen, daß von den 16 Vergleichen 3 eindeutig zugunsten von A ausgehen (8 und 9 verglichen mit 7) und 12 Vergleiche eindeutig zugunsten von B (Vergleiche von 3 und 5 aus Gruppe A mit sämtlichen Werten aus Gruppe B sowie die Vergleiche von 8 und 9 aus A mit 10 und 12 aus B). Für den Vergleich der beiden Meßwerte der Höhe 8 gibt es für beide Gruppen einen halben Punkt; insgesamt wird also so gerechnet, daß 3.5 Vergleiche zugunsten von A und 12.5 zugunsten von B ausgegangen sind.

Bei der Berechnung von U über die Rangsumme T kommen wir zum gleichen Ergebnis, wenn wir für einen Block von gleichen Ausgangswerten das arithmetische Mittel der auf diese Ausgangswerte entfallenden Rangplätze vergeben. Im obigen Beispiel: Auf die beiden Meßwerte der Höhe 8 würden die Rangplätze 4 und 5 entfallen; also vergeben wir für beide den Rangplatz 4.5. Damit ergeben sich für Gruppe A die Rangplätze 1, 2, 4.5 und 6, also eine Rangsumme von $T=13.5$, aus der wir nach der bereits eingeführten Formel einen Wert von

$$U = T - N_1 \cdot (N_1 + 1) / 2 = 13.5 - 4 \cdot (4 + 1) / 2 = 3.5$$

berechnen, der mit dem durch Auszählen von Vergleichen ermittelten Wert übereinstimmt. - Dasselbe wird auch bei mehr als 2 übereinstimmenden Meßwerten gemacht: Wenn z.B. in einem anderen Datensatz auf die Rangplätze 12-14 drei gleich hohe Ausgangswerte entfallen, dann wird für alle drei der Rangplatz 13 vergeben und die auf dieser Grundlage ermittelte Rangsumme in einen U-Wert umgerechnet.

Die auf diese Weise ermittelte Prüfgröße U hat aber nicht genau dieselbe Wahrscheinlichkeitsverteilung wie bei Stichproben ohne Rangkopplungen. Vielmehr sind Korrekturformeln für die Berechnungen anzuwenden, die aber nur selten zu nennenswerten Veränderungen des Ergebnisses führen und deshalb allenfalls in Situationen praktisch bedeutsam werden, in denen das Ergebnis knapp die Signifikanz verfehlt. (Und auch da kann man sich fragen, welchen Unterschied es wirklich macht, ob die Irrtumswahrscheinlichkeit nun 5.1% oder 4.9% ist.)

<< Da die Behandlung von Rangkopplungen im U-Test typisch für viele parameterfreie Signifikanztests ist, wird sie im folgenden "für Interessenten" dargestellt. Falls wir mit der Normalverteilungsapproximation arbeiten, können wir in der Formel für die Varianz ein Korrekturglied einfügen, das nach folgender Formel berechnet wird:

$$C = \sum_g f_g \cdot (g^3 - g)$$

Dabei steht g für "Kopplungsgrad", und das ist die Zahl der Meßwerte, die bei einer Kopplung auf den gleichen Rangplatz kommen: Bei 2 ranggleichen Meßwerten ist $g=2$ usw.. f_g ist die Häufigkeit, mit der Rangkopplungen des Kopplungsgrads g auftreten. (Dabei sind - abweichend von der Angabe in vielen Lehrbüchern - auch Rangkopplungen einzubeziehen, bei denen alle gekoppelten Meßwerte der gleichen Stichprobe entstammen: Das ist mathematisch zulässig und erhöht die Teststärke.) Summiert wird über alle g, d.h. über alle tatsächlich auftretenden Kopplungsgrade.

Beispiel: In 2 Stichproben der Größen $N_1=54$ und $N_2=55$ ergeben sich insgesamt 5 Zweier-

kopplungen und 2 Viererkopplungen. Dann ist

$$C = 5 \cdot (2^3 - 2) + 2 \cdot (4^3 - 4) = 150$$

Das auf diese Weise berechnete Korrekturglied für Rangkopplungen wird in die folgende Formel für die Varianz von U eingesetzt:

$$\text{Varianz}(U) = N_1 \cdot N_2 \cdot (N+1 - C/(N \cdot (N-1))) / 12$$

Dabei ist N die Größe der Gesamtstichprobe, also $N = N_1 + N_2$. In unserem Beispiel ergibt sich mit $N = 54 + 55 = 109$:

$$\text{Varianz}(U) = 54 \cdot 55 \cdot (110 - 150/(109 \cdot 108)) / 12 = 27221.85$$

Die Wurzel aus dieser Varianz (also die Standardabweichung) beträgt 164.99 und unterscheidet sich nur minimal von dem Wert von 165, den wir oben für die gleichen Stichprobengrößen ohne Rangkopplungen ermittelt hatten.

Dieses Verfahren zur Behandlung von Rangkopplungen ist aber nur anwendbar, wenn eine Normalverteilungsapproximation zulässig ist. Bei kleinen Stichproben müßte man theoretisch eine kombinatorische Berechnung der Überschreitungswahrscheinlichkeit vornehmen; denn die Tabellen für den U-Test gelten streng genommen nur für Daten ohne Rangkopplungen. Diese Berechnung ist aber ziemlich kompliziert.

Bei zwei Gruppen A und B mit $N_A = N_B = 15$, bei denen sich 3 Zweierkopplungen und eine Dreierkopplung ergeben, müßte man sich fragen: Wie viele Möglichkeiten gibt es, eine Rangreihe mit 3 Zweierkopplungen und einer Dreierkopplung zu bilden, so daß die Buchstaben A und B je 15 mal auftreten? Wieviele davon haben ein mindestens genauso weit von der Zufallserwartung abweichendes U wie unsere Daten?

Es läßt sich gut rechtfertigen, in solchen Fällen die Prüfgröße U aufgrund der mittleren Rangplätze zu berechnen (s.o.) und diesen U-Wert anhand der Tabellen so zu prüfen, als ob keine Rangkopplungen vorlägen. Die unter Berücksichtigung von Rangkopplungen kombinatorisch ermittelte Irrtumswahrscheinlichkeit ist nämlich immer geringfügig niedriger, als die Irrtumswahrscheinlichkeit desselben U-Werts ohne Rangkopplungen, die den Tabellen zugrundeliegt; wir machen also allenfalls einen geringfügigen konservativen Fehler, wenn wir in der Tabelle so nachschlagen, als wenn keine Kopplungen vorlägen. Nur in ganz wenigen Fällen würde ein U-Wert, der lt. Tabelle die Signifikanz knapp verfehlt, durch Berücksichtigung der Rangkopplungen signifikant werden. Wo so etwas überhaupt möglich ist, wäre es wenig sinnvoll, wenn man die Entscheidung über eine Hypothese tatsächlich davon abhängig macht, ob die Irrtumswahrscheinlichkeit 4.9% oder 5.1% beträgt - auch wenn das rein formal im Signifikanztestmodell so vorgesehen ist.

Abzuraten ist dagegen von einem anderen Verfahren der Auflösung von Rangkopplungen, das manchmal vorgeschlagen wird: Man entscheidet durch Zufall (z. B. durch Münzwurf oder mit einem Würfel), in welcher Reihenfolge die gleich hohen Meßwerte in die Rangreihe zur Ermittlung der Rangsumme aufgenommen werden, und führt dann die Berechnungen so aus, wie sie in Fällen ohne Rangkopplungen vorgesehen sind. Dieses Verfahren ist zwar insofern mathematisch korrekt, als dabei tatsächlich die Wahrscheinlichkeit eines zufällig signifikanten Ergebnisses bei Gültigkeit von H_0 mit dem zugrundegelegten Signifikanzniveau übereinstimmt.

Andererseits verschenkt man damit unnötig Teststärke (denn die ist bei dieser Art von "Randomisierung" meistens geringer, als wenn man U aufgrund des mittleren Rangplatzes berechnet und dann so überprüft, als ob keine Rangkopplungen vorlägen). Vor allem aber macht es wenig Sinn, in den Grenzfällen, in denen das aufgrund der mittleren Rangplätze ermittelte U nur knapp die Signifikanz verfehlt, die Entscheidung über eine Hypothese aufgrund von Münzwurf zu treffen.

>>

cd) Teststärke und Zusatzvoraussetzungen beim U-Test

Der U-Test hat im Vergleich zum t-Test für unabhängige Stichproben bei großen Stichproben eine Effizienz von über 90%. D.h. also: Seine Teststärke ist so groß wie die des t-Tests mit ca 10% weniger Vpn. Bei kleinen Stichproben liegt sie teilweise noch darüber.

<< Genauer: Wenn die Stichprobengröße gegen unendlich strebt, strebt die Effizienz gegen $3/\pi=0.955$ ($\pi=3.14\dots$). Man sagt auch: Die asymptotische relative Effizienz zum t-Test beträgt 0.955. Dieser Wert gilt dann, wenn die Normalverteilungsvoraussetzung des t-Tests erfüllt ist. Bei Daten, die in der Population nicht normalverteilt sind, ist die Teststärke z.T. sogar erheblich größer als beim t-Test.

>>

Dem "Preis" einer meist geringfügig niedrigeren Teststärke steht als Gewinn gegenüber, daß die Zusatzvoraussetzungen des t-Tests (Normalverteilung, Varianzhomogenität und Intervallskalenniveau) entfallen. Ordinalskalenniveau genügt.

<< In einigen Büchern findet man eine Angabe, die auf eine weitergehende Zusatzvoraussetzung des U-Tests hinauslaufen würde. Nach diesen Büchern behauptet die H_0 des U-Tests, daß die den beiden Gruppen zugrundeliegenden Populationsverteilungen identisch sind, und H_1 , daß die beiden Verteilungen sich nur durch eine rechts-links-Verschiebung unterscheiden. Das würde auf die Zusatzvoraussetzung hinauslaufen: Man muß vorher wissen, daß die beiden Populationsverteilungen bis auf solche rechts-links-Verschiebungen identisch sind. Tatsächlich wird die Wahrscheinlichkeitsverteilung der Prüfgröße bei Gültigkeit von H_0 meistens unter der Annahme gleicher Populationsverteilungen hergeleitet. Dieselben mathematischen Ableitungen sind aber auch aufgrund einer wesentlich weniger anspruchsvollen Annahme möglich: Wenn wir mit p_A und p_B die Wahrscheinlichkeit bezeichnen, daß beim zufälligen Herausgreifen je eines Werts aus beiden Populationsverteilungen der A-Wert bzw. der B-Wert größer ausfällt, dann genügt für die Herleitung der Wahrscheinlichkeitsverteilung von U bei Gültigkeit von H_0 die Annahme $H_0: p_A=p_B$. Das ist also die präzise Formulierung der Nullhypothese des U-Tests. Streng genommen ist das nicht völlig identisch mit einer Gleichheit der Mediane beider Verteilungen. Es lassen sich mathematisch Fälle konstruieren, in denen trotz Mediangleichheit der Populationsverteilungen die obige H_0 nicht erfüllt ist und umgekehrt. Für praktische Anwendungszwecke sind solche Feinheiten aber ohne Bedeutung.

>>

d) Der χ^2 -Test nach Pearson

Der χ^2 -Test ist vielseitig anwendbar. Ein Beispiel haben wir bereits ganz zu Beginn von Statistik I kennengelernt. In einem Therapie-Experiment ergaben sich für eine Experimentalgruppe und eine Kontrollgruppe die folgenden Häufigkeiten von Meßwerten in der "oberen Hälfte" und der "unteren Hälfte" einer Gesamtverteilung von Meßwerten der Selbstexploration (SE):

	"untere Hälfte"	"obere Hälfte"	Summe
Experimentalgruppe:	8	17	25
Kontrollgruppe:	16	9	25
Summe:	24	26	50

Unter der Nullhypothese, daß die Verteilung auf "obere" und "untere Hälfte" in beiden Gruppen gleich sind, wären die folgenden Häufigkeiten zu erwarten:

	"untere Hälfte"	"obere Hälfte"	Summe
Experimentalgruppe:	12	13	25
Kontrollgruppe:	12	13	25
Summe:	24	26	50

Gewisse Abweichungen der beobachteten Häufigkeiten von den erwarteten Häufigkeiten können zufällig entstehen; aber mit Hilfe des χ^2 -Tests stellten wir dann fest, daß die Wahrscheinlichkeit, so große oder noch größere Abweichungen wie in unseren Daten zufällig zu erhalten, wenn in Wirklichkeit die Nullhypothese zutrifft, kleiner als 5% ist.

Außer diesem "Mediantest" gibt es noch viele verschiedene Einsatzmöglichkeiten des χ^2 -Tests. Ihre Gemeinsamkeit läßt sich folgendermaßen beschreiben: Die Daten bestehen aus absoluten Häufigkeiten, und der Test überprüft, ob diese "beobachteten Häufigkeiten" signifikant von den Häufigkeiten abweichen, die aufgrund einer Nullhypothese "idealerweise" zu erwarten wären.

Die verschiedenen Einsatzmöglichkeiten dieses Tests, von denen einige besonders wichtige in den nachfolgenden Abschnitten da bis db behandelt werden, unterscheiden sich durch Typen von Nullhypothesen, aus denen die erwarteten Häufigkeiten abgeleitet werden. In allen Fällen gibt es für die Prüfgröße χ^2 , nach der dieser Test benannt ist, zwei Varianten:

χ^2_c (mit Kontinuitätskorrektur)

χ^2_u (ohne Kontinuitätskorrektur).

Wann diese Varianten zum Einsatz kommen, wird später @ im Zusammenhang mit den Zusatzvoraussetzungen behandelt. Die Formeln sind:

$$\chi_u^2 = \sum_k \frac{(f_{ok} - f_{ek})^2}{f_{ek}} = \sum_k \frac{f_{ok}^2}{f_{ek}} - N$$

und

$$\chi_c^2 = \sum_k \frac{(|f_{ok} - f_{ek}| - 0.5)^2}{f_{ek}}$$

Dabei bedeutet:

f_{ok} : Die in Klasse k beobachtete Häufigkeit (**f**requency **o**bserved).

f_{ek} : Die in Klasse k gemäß H_0 erwartete Häufigkeit (**f**requency **e**xpected). Sie berechnet sich nach der Formel $f_{ek} = p_k \cdot N$, wobei p_k die Wahrscheinlichkeit dieser Klasse gemäß H_0 ist.

Beide Prüfgrößen haben approximativ eine χ^2 -Verteilung; auch bei χ^2 -Verteilungen haben wir Freiheitsgrade zu beachten. Sie berechnen sich jedoch bei den verschiedenen Verwendungen des χ^2 -Tests verschieden.

Zusatzvoraussetzungen: s.u.

da) Als χ^2 -Test für die Güte der Anpassung ohne zusätzliche Parameterschätzung

Die Nullhypothese behauptet eine ganz bestimmte Wahrscheinlichkeitsverteilung der Beobachtungen auf die einzelnen Klassen.

Beispiel:

Wir wollen überprüfen, ob ein Würfel ungefälscht ist.

H_0 : Die Wahrscheinlichkeit jeder Augenzahl ist 1/6.

Da der Würfel nach allen Regeln der Kunst hergestellt wurde, würde es allen bisherigen Erfahrungen widersprechen, wenn verfälschte Ergebnisse auftauchen würden.

$$\alpha = 1\%$$

Wie aus beiden Formeln hervorgeht, führt eine Abweichung von f_o und f_e immer zu großen χ^2 -Werten. *Der Test ist seinem Wesen nach einseitig:* Nur bei großem χ^2 verwerfen wir H_0 . Wir würfeln 600 mal. Wenn H_0 stimmt, erwarten wir, daß jede Augenzahl 100 mal auftritt. f_e ist also für alle Klassen 100.

Wir erzielen die einzelnen Augenzahlen mit den folgenden Häufigkeiten (f_o):

Augenzahl	f_o	f_e
1	106	100
2	91	100
3	92	100
4	103	100
5	96	100
6	112	100
	600	

Tabelle 10

Welche der beiden Prüfgrößen berechnen wir? χ^2_u oder χ^2_c ?

Regel:

Die Verteilung beider Prüfgrößen ist nur approximativ eine χ^2 -Verteilung. χ^2_c ist etwas konservativ; χ^2_u ist bei kleinen Stichproben radikal. Also bei kleinen Stichproben χ^2_c ; bei großen χ^2_u . Eine Stichprobe gilt hier als klein, wenn mindestens ein f_e -Wert kleiner als 10 ist. Je größer die Zahl der Freiheitsgrade (s.u.), umso großzügiger kann man mit dieser Regel umgehen.

In unserem Fall können wir ohne Bedenken χ^2_u berechnen.

$$\begin{aligned}\chi^2_u &= \frac{(106 - 100)^2}{100} + \frac{(91 - 100)^2}{100} + \frac{(92 - 100)^2}{100} + \frac{(103 - 100)^2}{100} \\ &+ \frac{(96 - 100)^2}{100} + \frac{(112 - 100)^2}{100} = 3.50\end{aligned}$$

Um diesen Wert mit den in den χ^2 -Tabellen angegebenen kritischen Werten von χ^2 vergleichen zu können, müssen wir die Zahl der Freiheitsgrade von χ^2 kennen. In diesem Fall einfach: Wenn die ersten 5 f_o -Werte gegeben sind, ist der 6. festgelegt, da die Summe aller f_o -Werte N ergeben muß. Also 5 Freiheitsgrade.

Allgemein: Beim χ^2 -Test für die Güte der Anpassung ohne vorgeschaltete Parameterschätzung ist $df = \text{Klassenzahl} - 1$.

Für 5 Freiheitsgrade finden wir in der χ^2 -Tabelle für 1% Überschreitungswahrscheinlichkeit einen kritischen χ^2 -Wert von 15.09. Unser χ^2_u ist also viel zu klein, um H_0 zu verwerfen. Die Irrtumswahrscheinlichkeit betrüge mehr als 50%. Die Hypothese, daß der Würfel ungefälscht ist, wird also beibehalten.

Wichtige Zusatzvoraussetzungen in allen χ^2 -Tests nach Pearson:

1. Die Einordnung aller Beobachtungen in die Klassen ist stochastisch unabhängig.
2. Die f_e -Werte dürfen nicht zu klein sein.
Faustregel: Kein f_e sollte kleiner als 5 sein. Je größer die Zahl der Freiheitsgrade, um so eher kann man hier großzügig sein.
3. Die Klassen müssen einander ausschließen.

db) Der χ^2 -Test für die Güte der Anpassung mit vorgeschalteter Parameterschätzung

Dieser Test untersucht, ob eine Populationsverteilung (Wahrscheinlichkeitsverteilung), aus der eine Stichprobe gezogen wurde, zu einer bestimmten Familie gehört ("eine bestimmte Form hat"). Der Wert bestimmter Parameter ist dabei für die zu prüfende H_0 irrelevant. Es kommt nur auf die Familie an.

Beispiele:

Unsere Vpn haben in einem Geschicklichkeitstest die folgende Verteilung von "Treffern" erzielt (erste Spalte: Trefferzahl; zweite Spalte: empirische Häufigkeit).

x	f_o	f_e	Zsf. f_e
56- 60	2	2.26	2.26 + 5.92 = 8.18
61- 65	7	5.92	
66- 70	10	14.44	
71- 75	26	27.66	
76- 80	47	39.38	
81- 85	46	41.42	
86- 90	34	34.20	
91- 95	10	20.56	
96-100	11	9.72	
101-105	5	3.34	3.34 + 1.10 = 4.44
106-110	2	1.10	
N = 200			

Tabelle 11

Die dritte und vierte Spalte lassen wir vorläufig außer acht.

Die Häufigkeitsverteilung hat eine leichte positive Schiefe.

Frage: Kann es trotzdem sein, daß das ein Zufall ist und daß die Populationsverteilung eine Normalverteilung ist?

H_0 : Die Populationsverteilung der Treffer ist eine Normalverteilung.

H_1 : Die Populationsverteilung ist keine Normalverteilung.

$\alpha = 5\%$

Die Nullhypothese spezifiziert noch nicht genau die Verteilung; sie läßt μ und σ offen. Diese müssen wir zuerst aus den Daten schätzen. Der Stichprobendurchschnitt \bar{x} und die

Quadratsumme QS_x lassen sich wie üblich aus der Häufigkeitsverteilung berechnen, und dann ergeben sich die folgenden Schätzungen für μ und σ :

$$\hat{\mu} = \bar{x} = 81.75$$

$$\hat{\sigma} = \sqrt{\frac{QS_x}{N-1}} = 9.334$$

Jetzt können wir die f_e -Werte für die einzelnen Klassen berechnen, die in der dritten Spalte stehen.

Beispiel: Die Klasse "76 - 80" hat die Intervallgrenzen 75.5 und 80.5. In einer Normalverteilung mit Erwartungswert 81.75 und Standardabweichung 9.334 entspricht diesen Werten z-Werte von -0.67 bzw. -0.13. Die Wahrscheinlichkeit eines z-Wertes zwischen diesen Grenzen ist⁸⁸

$$\Phi(-0.13) - \Phi(-0.67) = 0.4483 - 0.2514 = 0.1969.$$

Damit ist also auch die Wahrscheinlichkeit eines Intervalls 75.5 - 80.5 in unserer Verteilung gleich 0.1969. Die erwartete Häufigkeit (f_e) beträgt dann $0.1969 \cdot N = 0.1969 \cdot 200 = 39.38$. In der gleichen Weise wurde die gesamte dritte Spalte berechnet. Der erste Wert (2.26) ist dabei die erwartete Häufigkeit des Gesamtbereichs unterhalb von 60.5. Entsprechend ist der letzte Wert (1.10) der Erwartungswert der Häufigkeit des gesamten Bereichs oberhalb von 105.5.

Jetzt haben wir die f_o - und die f_e -Werte, aber wir können noch keinen χ^2 -Test rechnen, da einige Klassen ein zu geringes f_e haben. Wir fassen die untersten und die obersten Klassen zusammen. Die f_e -Werte der zusammengefaßten Klassen stehen in der vierten Spalte. Damit erhalten wir eine neue Tabelle:

⁸⁸Da später mit $N = 200$ multipliziert wird, müssen die Wahrscheinlichkeiten aus der Tabelle der Standardnormalverteilung mit vier Kommastellen abgelesen werden, damit die f_e -Werte eine Genauigkeit von 2 Kommastellen haben.

x	f _o	f _e
≤ 65	9	8.18
66 - 70	10	14.44
71 - 75	26	27.66
76 - 80	47	39.38
81 - 85	46	41.42
86 - 90	34	34.20
91 - 95	10	20.56
96 - 100	11	9.72
≥ 101	7	4.44
	N=200	

Tabelle 12

Noch immer ist ein f_e -Wert kleiner als 5; aber da unser χ^2 in diesem Fall - wie noch zu zeigen ist - immerhin 6 Freiheitsgrade hat, können wir uns eine so geringfügige Verletzung leisten.

Wir berechnen χ^2_c , da einige f_e -Werte deutlich kleiner als 10 sind:

$$\chi^2_c = \frac{(|9 - 8.18| - 0.5)^2}{8.18} + \frac{(|10 - 14.44| - 0.5)^2}{14.44} + \dots + \frac{(|7 - 4.44| - 0.5)^2}{4.44} \approx 8.769$$

Die Zahl der Freiheitsgrade berechnet sich bei, χ^2 -Test mit vorgeschalteter Parameterschätzung nach der Formel:

$$df = \text{Klassenzahl} - 1 - \text{Zahl der aus den Daten geschätzten Parameter}$$

Wir haben zwar mit 11 Klassen angefangen, führen aber den χ^2 -Test nur noch mit 9 Klassen durch. Wir haben zwei Parameter μ und σ aus den Daten geschätzt. Also: $df = 9 - 1 - 2 = 6$.

Für 6 Freiheitsgrade und 5% Überschreitungswahrscheinlichkeit finden wir in der χ^2 -Tabelle einen kritischen Wert von 12.59. Wir behalten H_0 bei. Die Irrtumswahrscheinlichkeit betrüge sogar mehr als 10%, wenn wir die H_0 verwerfen würden, da unser χ^2 auch kleiner als 10.64 ist.

Hinweis: Der Test gewinnt häufig an Stärke, wenn man noch mehr Klassen zusammenfaßt. Es ist aber unzulässig, solange Klassenzusammenfassungen auszuprobieren, bis man zu einem signifikanten χ^2 kommt. Die Wahrscheinlichkeit, bei irgendeiner dieser Zusammenfassungen ein signifikantes χ^2 zu bekommen, auch wenn H_0 richtig ist, (also das Risiko I), wäre viel zu hoch.

dc) Der χ^2 -Test für stochastische Unabhängigkeit in zweidimensionalen Verteilungen

Beispiel: Wir wollen untersuchen, ob eine stochastische Abhängigkeit zwischen dem Bildungsstand und dem Krankheitsbild von Patienten besteht.

Wir erheben Daten und erhalten:

Krankheit	Bildungsstand					Zeilen- summe
	I	II	III	IV	V	
Neurose	19	42	96	41	34	232
Endogene Psychose	23	101	42	21	31	218
Exogene Psychose	48	73	29	16	18	184
Andere Erkrankungen	34	40	45	38	9	166
Spaltensumme	124	256	212	116	92	$N=800$

Tabelle 13

Beim Bildungsstand bedeutet:

I: keine abgeschlossene Volksschule

II: Volksschule

III: abgeschlossener Besuch einer Schule, die weiter führt als die Volksschule, jedoch nicht zum Abitur führt (Mittelschule etc.)

IV: Abitur ohne abgeschlossenes Studium

V: Akademiker

Patienten, die gleichzeitig mehrere Krankheiten haben, sind in die Krankengruppe eingeordnet worden, die nach dem Urteil des behandelnden Arztes beim jeweiligen Patienten am stärksten ausgeprägt ist. (Erforderlich, damit sich die Klassen gegenseitig ausschließen).

Tabelle 13 besagt dann: Wir haben 19 Neurotiker aus der Bildungsgruppe I, 42 endogene Psychotiker in der Bildungsgruppe III usw.; insgesamt also $19+42+96+41+34 = 232$ Neurotiker; $19+23+48+34 = 124$ Angehörige der Bildungsgruppe I, insgesamt eine Stichprobe von $N = 800$ Patienten. (Da es sich um eine zweidimensionale Häufigkeitsverteilung handelt, ist N die Summe der Zeilensummen oder die Summe der Spaltensummen. Beides berechnet man als "Probe" auf Rechenfehler.

Unsere Tabelle 13 legt nahe, daß ein Zusammenhang zwischen Bildungsstand und Krankheit des Patienten besteht. In der Bildungsgruppe I finden sich vor allem "exogene Psychosen" und "andere Erkrankungen"; in Bildungsgruppe II überwiegen eindeutig die Psychosen, insbesondere die endogenen, in den Gruppen III-V finden sich überwiegend Neurosen; außerdem sind in Gruppe IV die "anderen Erkrankungen", in Gruppe V die endogenen Psychosen stark ausgeprägt.

Es wäre aber theoretisch denkbar, daß dies Zufall ist und daß in der Population aller Patienten Bildungsstand und Krankheiten voneinander stochastisch unabhängig sind. Dies wäre

die Nullhypothese.

Allgemein: H_0 : Zwei Merkmale sind stochastisch unabhängig.

Wir wählen $\alpha = 5\%$.

Welche Häufigkeiten würden wir dann in den einzelnen Feldern erwarten? Wir würden beispielsweise erwarten, daß sich die 124 Vpn aus Bildungsgruppe I genau im gleichen Verhältnis auf die Krankheitsgruppen verteilen wie die Gesamtheit der Patienten;

also: 232 : 218 : 184 : 166.

Das wäre erfüllt, wenn wir die folgenden Häufigkeiten erzielen würden:

35.96; 33.79; 28.52; 25.73.

Allgemein: Wenn die beiden Variablen (hier Krankheit und Bildungsstand) voneinander unabhängig wären, dann müßten die bedingten Wahrscheinlichkeitsverteilungen gleich den entsprechenden Randverteilungen sein.

Entsprechend müßten sich also die Häufigkeiten in den einzelnen Spalten genau so zueinander verhalten (das gleiche Verhältnis zueinander haben) wie die Zeilensummen, und die Häufigkeiten in den Zeilen das gleiche Verhältnis zueinander wie die Spaltensummen. Z.B. müßten sich die 166 Fälle mit "anderen Erkrankungen" im Verhältnis 124 : 256 : 212 : 116 : 92 auf die fünf Bildungsgruppen verteilen.

Wie berechnet man die entsprechenden erwarteten Häufigkeiten?

Man könnte aufgrund der mathematischen Gesetze der Proportion zu einer Lösung kommen. Diese Lösung wäre genau dieselbe wie die, die wir auf dem folgenden Weg erzielen.

Zunächst muß ja für das Feld ganz oben links gelten:

$$f_e = p(\text{Neu}, I) \cdot N$$

$$\text{denn } (f_{ek} = p_k \cdot N)$$

dabei bedeutet $p(\text{Neu}, I)$ "Wahrscheinlichkeit, daß ein Patient, der in die Stichprobe kommt, Neurotiker ist und zur Bildungsgruppe I gehört". Wenn H_0 richtig ist (wenn also Krankheit und Bildungsstand voneinander stochastisch unabhängig sind), dann muß nach dem einfachen Multiplikationssatz gelten:

$$p(\text{Neu}, I) = p(\text{Neu}) \cdot p(I).$$

$p(\text{Neu})$ (also die Wahrscheinlichkeit, daß eine Vp, die in unsere Stichprobe kommt, Neurotiker ist), ist gleich dem Anteil der Neurotiker an der Gesamtpopulation aller Patienten, aus der unsere Stichprobe stammt. Dieser Anteil wäre ein Populationsparameter, den wir aus unserer Stichprobe schätzen können. Als Schätzfunktion würden wir den Stichprobenanteil verwenden:

$$\hat{p}(\text{Neu}) = 232 / 800$$

Entsprechend würden wir die Wahrscheinlichkeit, daß ein Patient zur Bildungsgruppe I gehört, schätzen:

$$\hat{p}(I) = 124 / 800$$

Setzen wir diese zwei Werte einmal in die Gleichung für $p(\text{Neu}, I)$ ein, dann erhalten wir:

$$\hat{p}(\text{Neu}, I) = 232/800 \cdot 124/800$$

Diesen Ausdruck setzen wir in die Gleichung für f_e ein und erhalten:

$$f_e = (232/800) \cdot (124/800) \cdot 800 = (232 \cdot 124) / 800 = 35.96$$

Allgemein: $f_e = (\text{Zeilensumme} \cdot \text{Spaltensumme}) / N$

Gemeint ist die Spaltensumme *der* Spalte und die Zeilensumme *der* Zeile, der das Feld angehört, dessen erwartete Häufigkeit (f_e) wir berechnen wollen.

Weiteres Beispiel: In dem Feld "Exogene Psychosen der Bildungsgruppe III" würden wir erhalten:

$$f_e = (184 \cdot 212) / 800 = 48.76$$

Mit dieser Methode erhalten wir die folgende Tabelle 14 erwarteter Häufigkeiten (f_e).

Krankheit	Bildungsstand					Zeilensumme
	I	II	III	IV	V	
Neurose	35.96	74.24	61.48	33.64	26.68	232.00
Endogene Psychose	33.79	69.76	57.77	31.61	25.07	218.00
Exogene Psychose	28.52	58.88	48.76	26.68	21.16	184.00
Andere Erkrankungen	25.73	53.12	43.99	24.07	19.09	166.00
Spaltensumme	124.00	256.00	212.00	116.00	92.00	N=800

Tabelle 14

Rechenprobe: Die Spalten- und Zeilensummen müssen bis auf Rundungsfehler dieselben sein, wie bei den beobachteten Häufigkeiten!

Jetzt haben wir also für alle 20 Klassen unserer zweidimensionalen Verteilung die f_o -Werte (1. Tabelle) und die f_e -Werte (2. Tabelle). Da kein f_e -Wert unter 10 liegt, können wir χ^2_u berechnen. Wir verwenden einmal die zweite Formel für χ^2_u :

$$\begin{aligned} \chi^2_u &= \frac{19^2}{35.96} + \frac{42^2}{74.24} + \dots + \frac{34^2}{26.68} \\ &+ \frac{23^2}{33.79} + \dots + \frac{9^2}{19.09} - 800 = 120.468 \end{aligned}$$

Wie viele Freiheitsgrade hat dieser χ^2 -Wert?

Man kann vorgehen wie in cb). Wir haben einige Parameter (die Populationsanteile der einzelnen Krankheitsgruppen und der Bildungsgruppen) geschätzt, und daraus die laut H_0 gültigen f_e -Werte berechnet. Auf den ersten Blick würden wir sagen, wir haben 9 Populationsparameter geschätzt (nämlich die den Spaltensummen mal den Zeilensummen entsprechenden Randwahrscheinlichkeiten).

Wir können aber auch anders argumentieren (und es läßt sich beweisen, daß diese Überlegung besser ist). Wenn wir die Populationswahrscheinlichkeit der ersten drei Krankengruppen geschätzt haben, dann ist damit die vierte festgelegt, denn alle vier zusammen müssen ja 1.00 ergeben. Entsprechend bei den Zeilen: Wenn wir die Populationsanteile der ersten vier Bildungsgruppen geschätzt haben, ist derjenige der fünften festgelegt.

Allgemein: Haben wir k Zeilen und m Spalten, so müssen wir nur $k - 1 + m - 1$ Parameter schätzen.

Da das Gesamtfeld rechteckig ist, hat unsere zweidimensionale Verteilung $k \times m$ Klassen (Felder; man spricht daher auch vom $k \times m$ -Felder χ^2 -Test).

Nach der alten Regel ergibt sich:

$$\begin{aligned} df &= \text{Zahl der Klassen} - 1 - \text{Zahl der geschätzten Parameter} \\ &= k \times m - 1 - (k - 1 + m - 1) \\ &= k \times m - k - m + 1 \end{aligned}$$

oder:

$\begin{aligned} df &= (k - 1) \cdot (m - 1) \\ &= (\text{Zeilenzahl} - 1) \cdot (\text{Spaltenzahl} - 1) \end{aligned}$
--

In unserem Fall:

$$df = (4 - 1) \cdot (5 - 1) = 12$$

Aus der χ^2 -Tabelle entnehmen wir: Für 12 Freiheitsgrade ist auf dem 5%-Niveau ein χ^2 von mindestens 21.0261 erforderlich. Das von uns erzielte χ^2 von 120.468 liegt bei weitem darüber; wir verwerfen die Hypothese, daß Krankheitsbild und Bildungsstand voneinander stochastisch unabhängig sind mit einer Irrtumswahrscheinlichkeit von erheblich weniger als 0.1%. (D.h.: Die Wahrscheinlichkeit, daß sich bei Gültigkeit von H_0 in den Daten ein mindestens so enger Zusammenhang zwischen Bildungsstand und Krankheitsbild finden würde, wie in unseren Daten, wäre erheblich weniger als 0.1%.)

Zusammenfassung:

Beim " $k \times m$ "-Felder χ^2 -Test berechnet man zunächst die Zeilen- und Spaltensummen sowie die Stichprobengröße.

Dann berechnet man die f_e -Werte für jedes Feld nach der Formel:

$$f_e = \frac{\text{Zeilensumme} \cdot \text{Spaltensumme}}{N}$$

Jetzt berechnet man χ^2_c bzw. χ^2_u . Ist die Zahl der Spalten und/oder der Zeilen gleich 2, so gibt es vereinfachte Verfahren zur Berechnung von χ^2 . (Vgl. Bortz, Lienert & Boehnke, 1990, Gl. 5.24ff.). Bei diesen Verfahren braucht man f_e nicht zu berechnen. Allerdings muß man vorab prüfen, ob nicht ein f_e kleiner als 5 oder kleiner als 10 ist. Man berechnet:

$$\text{kleinstes } f_e = \frac{\text{kleinste Zeilensumme} \cdot \text{kleinste Spaltensumme}}{N}$$

Ist das kleinste f_e mindestens 5, kann man den χ^2 -Test verwenden; ist das kleinste f_e mindestens 10, braucht man keine Kontinuitätskorrektur.

Es gilt immer:

$$df = (\text{Zeilenzahl} - 1) \cdot (\text{Spaltenzahl} - 1).$$

Ein weiteres Beispiel:

Wir wollen untersuchen, ob Schüler und Schülerinnen, die vor dem Abitur stehen, die folgende Frage gleich häufig mit "ja" beantworten: "Spielt die wirtschaftliche Seite eines Berufes bei ihrer bevorstehenden Berufswahl eine nennenswerte Rolle?"

Versuchsplan: Wir stellen diese Frage 50 Schülern und Schülerinnen vor dem Abitur; wir verwenden den χ^2 -Test und setzen $\alpha = 5\%$.

H_0 : Schüler und Schülerinnen der untersuchten Population beantworten die Frage gleich häufig mit ja. (D.h.: Geschlecht und Antwort auf die Frage sind stochastisch unabhängig voneinander.)

H_1 : H_0 ist unrichtig.

Wir erhalten folgende Ergebnisse:

	Schüler	Schülerinnen	Zeilensumme
ja	42	38	80
nein	8	12	20
Spaltensumme	50	50	$N=100$

Tabelle 15

Auch Tabelle 15 legt einen Zusammenhang zwischen Geschlecht und Antwort auf die gestellte Frage nahe. Bei beiden Geschlechtern ist zwar die Antwort "ja" vorherrschend, aber bei den Schülern noch stärker als bei den Schülerinnen. Ist dieser Unterschied signifikant?

Wir untersuchen zunächst, ob man den χ^2 -Test verwenden kann:

$$\text{kleinstes } f_e = (\text{kleinste Zeilensumme} \cdot \text{kleinste Spaltensumme}) / N = (20 \cdot 50) / 100 = 10$$

Das kleinste f_e ist 10, wir können also sogar ohne Kontinuitätskorrektur rechnen. Es ergibt

sich ein χ^2_u von⁸⁹

$$\chi^2_u = \frac{(42 - 40)^2}{40} + \frac{(38 - 40)^2}{40} + \frac{(8 - 10)^2}{10} + \frac{(12 - 10)^2}{10} = 1.0$$

mit

$$df = (2-1) \cdot (2-1) = 1.$$

Damit ist χ^2 nicht signifikant. Wir behalten H_0 bei.

Das Beispiel ist noch aus zwei weiteren Gründen interessant:

1. Wir haben nicht aus der Population der Schülerinnen und Schüler eine Stichprobe gezogen, sondern willkürlich festgelegt: Wir testen 50 Schüler und Schülerinnen. Wir haben also gleichsam aus den zwei Populationen ("Schüler" und "Schülerinnen") je eine Stichprobe gezogen. Das ist erlaubt.

Allgemein: Wir können den $k \times m$ -Felder χ^2 -Test auch durchführen, um die gesamte Verteilung eines Merkmals in mehreren Populationen zu vergleichen, indem wir aus jeder der Populationen eine Stichprobe von willkürlicher Größe ziehen. (Am günstigsten für die Teststärke: gleich große Stichproben!)

- << 2. Wir hätten schon vor der Datenerhebung gute Gründe gehabt, die Alternativhypothese folgendermaßen zu formulieren:

Schüler beantworten die Frage häufiger mit "ja" als Schülerinnen. In gewisser Weise ist das eine einseitige Alternativhypothese. Andererseits ist der χ^2 -Test bereits seinem Wesen nach einseitig. Wir behelfen uns, indem wir die folgende Entscheidungsregel formulieren:

Ist χ^2 größer als der kritische Wert auf dem 10%-Niveau (d.i. 2.71) und beantworten außerdem die Schüler häufiger die Frage mit ja als die Schülerinnen, so verwerfen wir H_0 .

Wie groß ist das Risiko I bei dieser Entscheidungsregel?

Wir wollen zur Berechnung dieses Risiko I zwei Symbole definieren: Der Buchstabe M für "Männer" soll das Ereignis bezeichnen, daß Schüler häufiger mit ja antworten als Schülerinnen. Das Gegenteil wollen wir mit dem Buchstaben F für "Frauen" bezeichnen. Das Risiko I ist dann die bei Gültigkeit von H_0 bestehende Wahrscheinlichkeit, daß χ^2 größer als 2.71 wird und das Ergebnis M auftritt, d.h.: $p(\chi^2 > 2.71; M)$. Ist H_0 richtig, so gilt aber offenbar:

$$p(\chi^2 > 2.71; F) = p(\chi^2 > 2.71; M) \quad (\text{Gl. 1})$$

Andererseits gilt aber auch:

$$p(\chi^2 > 2.71) = p[(\chi^2 > 2.71; M) \vee (\chi^2 > 2.71; F)] \quad (\text{Gl. 2})$$

In Worten: Die Wahrscheinlichkeit, daß χ^2 größer als 2.71 ist, ist gleich der Wahrscheinlichkeit, daß *entweder* χ^2 größer als 2.71 ist und das Ergebnis M auftritt *oder* daß χ^2 größer als 2.71 und das Ergebnis F auftritt.

Beide Möglichkeiten schließen sich aber aus; folglich gilt nach dem einfachen Additionssatz die Gleichung

$$p[(\chi^2 > 2.71; M) \vee (\chi^2 > 2.71; F)]$$

⁸⁹Wir könnten auch mit der auf S. @ erwähnten Spezialformel für 2 x 2-Feldertafel aus Bortz, Lienert & Boehnke, 1990, Gl. 5.24 ff. rechnen.

$$p(\chi^2 > 2.71; M) + p(\chi^2 > 2.71; F) \quad (\text{Gl. 3})$$
$$2 \cdot p(\chi^2 > 2.71; M)$$

Das letzte Gleichheitszeichen gilt wegen unserer ersten Gleichung. Kombinieren wir die Gleichungen 2 und 3, so erhalten wir:

$$p(\chi^2 > 2.71) = 2 \cdot p(\chi^2 > 2.71; M)$$

Andererseits ist $p(\chi^2 > 2.71)$ aber laut χ^2 -Tabelle gleich 10%. Damit ist die gesuchte Wahrscheinlichkeit $p(\chi^2 > 2.71; M)$, die gleich dem Risiko I ist, nur 5%.

Zusammenfassend können wir sagen: Obschon der χ^2 -Test nach Pearson von Natur aus einseitig ist, da nur große χ^2 -Werte signifikant sind, können wir noch einmal die Irrtumswahrscheinlichkeit halbieren: Wir nehmen uns in der Entscheidungsregel vor, die Nullhypothese nur dann zu verwerfen, wenn eine der beiden Möglichkeiten einer Abweichung von H_0 in unseren Daten auftritt. Wir haben in gewisser Weise eine "halbseitige" Entscheidungsregel verwandt. Diese Möglichkeit besteht in dieser Form nur bei 2 x 2-Feldern!

>>

dd) Schlußbemerkungen zu den χ^2 -Tests nach Pearson

1. Wie wir in @dc gesehen haben, lassen sich χ^2 -Tests auch bei Nominalskalen durchführen.
2. In der hier beschriebenen Form kann man den χ^2 -Test nur mit absoluten Häufigkeiten rechnen, nicht jedoch mit prozentualen oder relativen Häufigkeiten. (Dazu wären andere Formeln erforderlich!)
3. Es sei nochmals auf die zum Schluß von @da genannten Zusatzvoraussetzungen hingewiesen. Ist ein f_e kleiner als 5, so gibt es zwei Möglichkeiten: a) Klassenzusammenfassungen (aber nicht die Methode herausuchen, die das größte χ^2 ergibt! Vor Erhebung der Daten sollte eindeutig festgelegt werden, nach welchem Prinzip man Klassen zusammenfaßt!) b) Übergang zu anderen Signifikanztests (Fisher-Yates exakter Wahrscheinlichkeitstest; Multinomialtest; vgl. Maxwell, 1961).
4. Die Regel dafür, welche f_e -Werte nötig sind, um χ^2_c bzw. χ^2_u verwenden zu können, sind weitgehend Ermessenssache. Fast jedes Buch hat andere Regeln.

<<

5. Die χ^2 -Tabellen erfassen häufig nur die χ^2 -Verteilungen bis zu 30 Freiheitsgraden. Oberhalb von 30 Freiheitsgraden kann man die kritischen Werte von χ^2 nach der folgenden Formel bestimmen:

$$\chi^2(v, Q) = \frac{1}{2} \cdot \left(z_Q + \sqrt{2v-1} \right)^2$$

mit

$\chi^2(v, Q)$:= der χ^2 -Wert mit v Freiheitsgraden, dessen einseitige Überschreitungswahrscheinlichkeit nach oben gleich Q ist.

z_Q := der z -Wert, dessen einseitige Überschreitungswahrscheinlichkeit nach oben gleich Q ist.

v := Zahl der Freiheitsgrade.

Beispiel: Wir haben 150 Freiheitsgrade und suchen den χ^2 -Wert mit einer einseitigen Überschreitungswahrscheinlichkeit von 5%. Der entsprechende z-Wert ($z_{0,05}$) wäre 1.6449. Damit ergibt sich

$$\chi^2(150, 5\%) = \frac{1}{2} \cdot (1.6449 + \sqrt{2 \cdot 150 - 1})^2 = 179.2958$$

6. Es gibt noch viele weitere Raffinessen beim χ^2 -Test: Vereinfachte Formeln; weitere Anwendungen; vgl. dazu z.B. Maxwell (1961).
7. Eine besondere Anwendungstechnik für die χ^2 -Tests findet sich in der "Konfigurations-Frequenz-Analyse" (KFA): es wird untersucht, ob die Auftretenshäufigkeit bestimmter Merkmalskonfigurationen signifikant größer oder kleiner ist, als per Zufall zu erwarten ist (vgl. Krauth & Lienert, 1973 sowie Lautsch & von Weber, 1995).

>>

Übersicht über Signifikanztests

Signifikanztests für einzelne Stichproben

I Eindimensionale Verteilung

- 1) Hypothesen über die zentrale Tendenz
 - t-Test für einen einzelnen Mittelwert
 - (Vorzeichentest)
 - (Wilcoxon-Test)
 - (Fisher-Pitman-Test für Paardifferenzen)
 - (Welsh-Test)
- 2) Hypothesen über die Dispersion
 - χ^2 -Test
- 3) Hypothesen über die gesamte Verteilung oder ihre Form (z.B. Normalität)
 - χ^2 -Test für die Güte der Anpassung
 - Kolmogoroff-Smirnoff-Test für die Güte der Anpassung
 - kombinatorischer Test für die Güte der Anpassung
- 4) siehe S. @

II Mehrdimensionale Verteilungen

- 1) Hypothesen über Korrelationen
 - t-Test für eine einzelne Produkt-Moment-Korrelation
 - Signifikanztest über Fisher'sche z'-Transformation
 - Signifikanztest der parameterfreien Korrelationsmaße
 - Eckentest
 - F-Test für multiple Korrelation
- 2) Linearitätsannahme
 - Varianzanalyse
- 3) stochastische Unabhängigkeit bzw. Nullhypothesen über Spezialformen stochastischer Abhängigkeit
 - k x m-Felder χ^2 -test
 - Fisher-Yates exakter Wahrscheinlichkeitstest
 - CSM-Test
 - Wilson-Test
 - Symmetrietest von Bowker

Hinweis zu II:

Ist eines der beiden Merkmale in Klassen gegeben, so ist häufig eine Umformulierung des Problems möglich und behilflich, indem man den ganzen Datensatz als Versuchsplan mit mehreren Stichproben auffaßt. Beispiel 1: Fragestellung: besteht eine punktbiseriale Korrelation

zwischen Geschlecht und Intelligenz? Umformuliert: Besteht zwischen Männern und Frauen ein Unterschied in der zentralen Tendenz der Intelligenz? Beispiel 2: Fragestellung: Besteht eine stochastische Abhängigkeit zwischen sozialer Schicht und Merkmal x? Umformuliert: Ist die Verteilung des Merkmals x in den sozialen Schichten gleich? Vergleich mehrerer Stichproben hinsichtlich ihrer Gesamtverteilung

Signifikanztests für den Vergleich mehrerer unabhängiger Stichproben

I Vergleich in einem Einzelmerkmal

Man beachte den Hinweis zu I.3

1) Vergleich hinsichtlich der zentralen Tendenz

t-Test für den Vergleich zweier Mittelwerte

Varianzanalyse

U-Test von Mann-Whitney

x-Test von Van Der Werden

Mediantest

Fisher-Pitman-Test für unabhängige Stichproben

Lagetest von Rosenbaum

Rang-Varianzanalyse nach Kruskal und Wallis (H-Test)

Rang-Varianzanalyse nach Jonckheere

Multiple Vergleiche

2) Vergleich hinsichtlich der Dispersion (Variabilität)

F-Test

Bartlett χ^2 -Test

Moses-Test

Kamat-Test

Dispersionstest von Rosenbaum

3) Vergleich hinsichtlich der Gesamtverteilung (Omnibus-Test)

Hinweis: Falls feststeht, daß sich die Populationsverteilungen höchstens hinsichtlich einer Kennziffer unterscheiden, können die Omnibustests eingesetzt werden, um diesen Unterschied zu überprüfen. Beispiel: Falls feststeht, daß sich zwei Populationsverteilungen höchstens hinsichtlich der zentralen Tendenz unterscheiden, kann der Kolmogoroff-Smirnoff-Test eingesetzt werden, um die Nullhypothese zu testen, daß keine Unterschiede in der zentralen Tendenz bestehen.

k x m-Felder χ^2 -Test

Kolmogoroff-Smirnoff-Test für zwei unabhängige Stichproben

Iterationstest

II Vergleich der Korrelation

z-Test über Fisher'sche z'-Transformation

χ^2 -Test über Fisher'sche Transformation

Wilson-Test

III Vergleich mehrerer unabhängiger Stichproben aufgrund von mehreren Merkmalen

Diskriminanzanalyse

generalisierte Varianzanalyse ("multivariate Varianzanalyse")

"Konfigurations-Analyse" ("Pattern-Analysis")

"Klassifikation"

Signifikanztests für den Vergleich korrelierter Stichproben

Vorfrage: Sollen wirklich

- (a) zwei Populationverteilungen hinsichtlich einzelner Parameter oder hinsichtlich der Gesamtverteilung untersucht werden?

Oder handelt es sich um

- b) ein Problem zu einer mehrdimensionalen Verteilung?

Beispiel zu a): Fragestellung: Sind Erstgeborene extravertierter als ihre zweitgeborenen Geschwister? D.h.: Bestehen Unterschiede in der zentralen Tendenz der Populationsverteilungen? Vergleich zweier korrelierter Stichproben hinsichtlich ihrer zentralen Tendenz.

Beispiel zu b): Fragestellung: Besteht ein Zusammenhang zwischen der Extraversion von Erstgeborenen und der Extraversion ihrer zweitgeborenen Geschwister? Korrelieren sie? D.h. $H_0: \rho_{xy} = 0$ mit $x =$ Extraversion des Erstgeborenen, $y =$ Extraversion des Zweitgeborenen, s. @ II, 1

Hinweis: Wo beide Fragestellungen gleichzeitig zur Diskussion stehen, handelt es sich manchmal um ein Problem der "intraklassischen Korrelation".

- 1) Vergleich hinsichtlich der zentralen Tendenz

t-Test für Paardifferenzen

Varianzanalyse

multiple Vergleiche

Fisher Pitman-Test für Paardifferenzen

Wilcoxon-Test

Vorzeichentest

Rang-Varianzanalyse nach Friedman

(χ^2 -Test von McNemar)

(Q-Test von Cochran)

- 2) Vergleich hinsichtlich der Dispersion

Man überprüfe: $\rho_{xy} = 0$, mit $u = x + y$; $v = x - y$ (vgl. S. @).

Parameterfreie Dispersionstests (vgl. Bortz, Lienert & Boehnke, 1990, S. 292).

- 3) Vergleich hinsichtlich bestimmter Klassenhäufigkeiten

χ^2 -Test von McNemar)
Q-Test von Cochran

Fortsetzung von: D. Parameterschätzung

Wir haben bisher vor allem drei Eigenschaften kennengelernt, die eine gute Schätzfunktion erfüllen soll: Sie soll erwartungstreu (unverzerrt) und konsistent sein und die Forderung nach Minimalvarianz erfüllen. Es gibt noch weitere Eigenschaften einer guten Schätzfunktion, die wir hier nicht weiter erörtern wollen: Sie soll "medianstreu", "erschöpfend" und eine "Maximum-Likelihood-Schätzung" sein. Es wäre nützlich (aber für die Klausur nicht erforderlich), wenn Sie sich die Bedeutung dieser Begriffe anhand der Literatur erarbeiten würden.

III Konfidenzintervalle

1) Konfidenzintervalle für das Populationsmittel

Wir wollen annehmen, wir hätten einen Intelligenztest an 121 Vpn durchgeführt und die folgenden Werte erhalten:

$$\bar{x} = 64.34$$
$$QS_x = 14665,49$$

Nach unseren Überlegungen zur Parameterschätzung würden wir das Populationsmittel μ wie folgt schätzen:

$$\hat{\mu}_x = \bar{x} = 64.34$$

Das ist eine Punktschätzung; wir haben einen bestimmten *Punkt* der x-Skala als Schätzwert angegeben! Wir wissen aber, daß das nur eine Schätzung ist: Das wahre Populationsmittel kann darunter oder darüber liegen, und es liegt nahe zu fragen, um wieviel Punkte der x-Skala der Stichprobendurchschnitt vom wahren Populationsmittel abweichen kann. Daraus können wir dann ein sog. *Konfidenzintervall* für das Populationsmittel μ_x berechnen.

Aus der Stichprobentheorie wissen wir, daß der Stichprobendurchschnitt mit einem Standardfehler behaftet ist, den man folgendermaßen schätzen kann:

$$\hat{\sigma}_x = \sqrt{\frac{QS_x}{N \cdot (N - 1)}} = \sqrt{\frac{14665,49}{121 \cdot (121 - 1)}} = 1.005$$

Wenn dies der genaue Standardfehler wäre, könnten wir folgendermaßen fortfahren: Da die Wahrscheinlichkeitsverteilung des Stichprobendurchschnitts bei $N = 121$ zumindest approximativ als Normalverteilung angenommen werden kann⁹⁰, weicht der Stichprobendurchschnitt \bar{x} mit einer Wahrscheinlichkeit von 95% um höchstens $1.96 \cdot 1.005$ Punkte auf der x -Skala vom Populationsmittel μ_x ab (wobei 1.96 der z -Wert mit einer zweiseitigen Überschreitungswahrscheinlichkeit von 5% ist). Da der Wert 1.005 aber nur eine Schätzung des Standardschätzfehlers ist, nehmen wir anstelle des z -Werts den entsprechenden t -Wert mit 120 Freiheitsgraden und einer zweiseitigen Überschreitungswahrscheinlichkeit von 5%. Dieser t -Wert ist 1.98. Dann können wir sagen: Der Stichprobendurchschnitt \bar{x} weicht mit einer Wahrscheinlichkeit von 95% um höchstens $1.98 \cdot 1.005 = 1.99$ Punkte vom Populationsmittel μ_x ab. Daher können wir das Populationsmittel im Bereich 64.34 ± 1.99 , also im Intervall von 62.35 bis 66.33 vermuten, und damit haben wir das gesuchte Konfidenzintervall.

Die bei dieser Überlegung zugrundegelegte Wahrscheinlichkeit von 95% (oder 0.95) nennt man das *Verlässlichkeitsniveau* des Konfidenzintervalls. Die genaue Bedeutung dieser Wahrscheinlichkeit wird noch zu klären sein. Im Moment soll nur ergänzt werden, daß wir in gleicher Weise z.B. auch ein Konfidenzintervall für ein Verlässlichkeitsniveau von 99% aufstellen könnten. Das wird vereinfacht, wenn wir den Gedankengang, der zu diesem Intervall führte, in der folgenden Regel verallgemeinern:

Allgemeine Regel:

Grenzen des Konfidenzintervalls für μ_x :

$$\bar{x} \pm t_{\text{tab}} \cdot \hat{\sigma}_{\bar{x}}$$

Dabei ist t_{tab} der aus der t -Tabelle entnommene t -Wert, dessen zweiseitige Überschreitungswahrscheinlichkeit bei $N-1$ Freiheitsgraden gleich $1 - V$ ist. V ist das Verlässlichkeitsniveau (hier: 0.95). Drückt man V in Prozent aus, dann muß die zweiseitige Überschreitungswahrscheinlichkeit des t_{tab} -Wertes $100\% - V$ sein.

Für ein Verlässlichkeitsniveau von 99% müßten wir also den t -Wert mit einer zweiseitigen Überschreitungswahrscheinlichkeit von 1% bei 120 Freiheitsgraden zugrundelegen. Dieser t -Wert ist 2.62, und daher liegen die Grenzen des Konfidenzintervalls für das 99%-Verlässlichkeitsniveau bei $64.34 + 2.62 \cdot 1.005 = 66.98$ und $64.34 - 2.62 \cdot 1.005 = 61.70$.

<< Es bleibt noch die genaue Bedeutung des Verlässlichkeitsniveaus V zu klären. Auf den ersten Blick könnte man versucht sein, unser letztes Ergebnis in folgendem Satz zusammenzufassen: "Bei unseren Daten liegt das Populationsmittel μ_x mit einer Wahrscheinlichkeit von 99% im

⁹⁰Genauer: Es soll angenommen werden, daß in der Population keine extremen Abweichungen von der Normalverteilung vorliegen. Dann reicht $N = 121$ aus, um aus dem zentralen Grenzwertsatz eine approximative Normalverteilung des Stichprobendurchschnitts \bar{x} abzuleiten.

Intervall von 61.70 bis 66.98." Das wäre aber eine "Wahrscheinlichkeit von Populationsverhältnissen, gegeben bestimmte Daten", obwohl die von uns behandelte Inferenzstatistik immer nur "Wahrscheinlichkeiten von Daten, gegeben bestimmte Populationsverhältnisse" liefert. Richtig ist dagegen eine andere Formulierung: "Die Wahrscheinlichkeit, daß wir Daten erhalten, bei denen das nach der obigen allgemeinen Regel berechnete Konfidenzintervall für ein Verlässlichkeitsniveau V den wahren Wert μ_x enthält, beträgt V ."

Man ist nun leicht versucht, folgendermaßen weiterzudenken: Wenn wir also das Konfidenzintervall nach einer solchen Regel bestimmt haben, dann beträgt die Wahrscheinlichkeit, daß der wahre Wert μ_x sich in diesem Intervall befindet, V . Dieser Schluß ist aber unzulässig. Da es sich hier aber um ein diffiziles Problem handelt, das nur wenig Konsequenzen für praktische Anwendungen hat, beschränken wir uns mit einem Hinweis auf die Behandlung bei Hays (1973, S. 376).

>>

Das Konfidenzintervall mit einem Verlässlichkeitsniveau von 95% hat noch eine weitere interessante Eigenschaft, aus der sich auch eine zusätzliche Bedeutung des Verlässlichkeitsniveaus ableiten läßt. Man kann sich fragen, von welchen denkbaren Werten des Populationsmittels unser Stichprobendurchschnitt \bar{x} auf dem 5%-Niveau signifikant abweicht und von welchen nicht. Der entsprechende Signifikanztest wäre der t -Test für einen einzelnen Mittelwert. Diese Frage können wir folgendermaßen beantworten: Bei einem μ_0 außerhalb unseres Konfidenzintervalls würde unser \bar{x} signifikant von μ_0 abweichen, während die Abweichung von einem μ_0 innerhalb des Konfidenzintervalls nicht signifikant wäre. Das kann man folgendermaßen verallgemeinern:

Regel: Das Konfidenzintervall für das Populationsmittel μ_x mit dem Verlässlichkeitsniveau V ist die Gesamtheit aller denkbaren Werte des Populationsmittels, von denen der tatsächlich erzielte Stichprobendurchschnitt \bar{x} auf dem Signifikanzniveau $1-V$ (bzw. $100\%-V$) nicht signifikant abweicht.

<< Wenn man ganz exakt sein will, muß man die Endpunkte des Intervalls genauer betrachten. Diese Endpunkte werden üblicherweise noch als zum Konfidenzintervall gehörend betrachtet. Andererseits wird aber der kritische Bereich für einen t -Test so definiert, daß der Tabellenwert t_{tab} schon zum kritischen Bereich gehört. Das bedeutet aber, daß die Abweichung des Stichprobendurchschnitt \bar{x} von den Endpunkten des Konfidenzintervalls bereits auf dem Signifikanzniveau $1-V$ signifikant ist. Genau würde die obige Regel also nur stimmen, wenn man entweder den Tabellenwert t_{tab} zum kritischen Bereich zählen und das Konfidenzintervall als "offenes Intervall" betrachten würde⁹¹ oder aber den Tabellenwert schon nicht mehr zum kritischen Bereich zählen und das Konfidenzintervall als "geschlossenes Intervall" ansehen würde.

>>

⁹¹Ein "offenes Intervall" zwischen zwei Zahlen a und b enthält die Zahlen a und b nicht. Dagegen gehören beide Zahlen zum "geschlossenen Intervall" von a bis b .

2) Konfidenzintervalle für die Produkt-Moment-Korrelation ρ_{xy}

Da die Produkt-Moment-Korrelation eine ungünstige Wahrscheinlichkeitsverteilung hat, berechnen wir zunächst Konfidenzintervalle für ζ' (den z' -Wert, der zu ρ_{xy} gehört) und transformieren die Grenzen dieser Konfidenzintervalle zurück in ρ -Werte.

Regel: Grenzen des Konfidenzintervalls für ζ' :

$$z' \pm z_{\text{tab}} \cdot \sigma_{z'}$$

mit

z_{tab} = der z -Wert, dessen zweiseitige Überschreitungswahrscheinlichkeit $1 - V$ (bzw. $100\% - V$) ist.

und

$$\sigma_{z'} = \frac{1}{\sqrt{N - 3}}$$

Beispiel: Wir haben in einer Stichprobe von 403 Vpn zwischen x und y eine Produkt-Moment-Korrelation von 0.450 gefunden. Dazu gehört ein z' -Wert von 0.4847. Wir wollen das Konfidenzintervall auf einem Verlässlichkeitsniveau von 99% haben. Dann ist.

$$z_{\text{tab}} = 2.58$$

$$\sigma_{z'} = \frac{1}{\sqrt{403 - 3}} = 0.05$$

Grenzen des Konfidenzintervalls für ζ' :

$$0.4847 + 2.58 \cdot 0.05 = 0.6137$$

$$0.4847 - 2.58 \cdot 0.05 = 0.3557$$

Daraus erhält man dann ein Konfidenzintervall für die Populationskorrelation ρ_{rx} , indem man die Tabelle für die z' -Tabelle rückwärts liest. Zu $z'=0.3557$ bzw. $z'=0.6137$ gehören r -Werte von 0.341 und 0.547, und das sind die Grenzen des 99% Konfidenzintervalls für ρ_{xy} .

3) Konfidenzintervalle bei Schätzungen aufgrund von Regressionsgleichungen

Probleme der Schätzung ergeben sich auch bei der Verwendung von Regressionsgleichungen.

Beispiel: Wir wollen annehmen, in dem oben besprochenen Experiment sei x ein Test für die Fähigkeit einer bestimmten Fabrikarbeit und y die Zahl der erfolgreich bearbeiteten Werkstücke an einem Arbeitstag gewesen.

Weiterhin haben wir folgende Werte erhalten:

$$r_{xy} = 0.45; \quad \bar{x} = 24.30; \quad \bar{y} = 44.80; \quad \hat{\sigma}_x = 10.23; \quad \hat{\sigma}_y = 15.42$$

Jetzt soll ein neuer Arbeiter eingestellt werden. Wir wollen vorhersagen, wie viele Arbeitsstücke dieser Arbeiter an einem Tag richtig bearbeiten wird. Hierzu testen wir ihn mit dem Eignungstest. Der Arbeiter erzielt einen Punktwert von $x = 34$. Wir können nun die folgende Regressionsgleichung verwenden:

$$\hat{y} = 0.67x + 28.52$$

(Berechnung dieser Regressionsgleichung siehe Kurs I!)

In unserem Fall also:

$$\hat{y} = 0.67 \cdot 34 + 28.52 = 51.30$$

Wir schätzen also, daß der Arbeiter 51.30 Werkstücke erfolgreich bearbeiten wird. Das ist eine Punktschätzung. Wir können aber auch ein Konfidenzintervall aufstellen:

Grenzen des Konfidenzintervalls für die Schätzung von y aus x :

$$\hat{y} \pm z_{tab} \cdot \hat{\sigma}_{y.x}$$

mit

$$\hat{\sigma}_{y.x} = \text{geschätzter Standardfehler} = \hat{\sigma}_y \cdot \sqrt{1 - r_{xy}^2}$$

(Vgl. Kurs I)

Wir wollen in unserem Fall ein Konfidenzintervall aufstellen, innerhalb dessen die "Erfolgsquote" y unseres Arbeiters mit 95% Wahrscheinlichkeit liegt.

Wir berechnen also:

$$z_{tab} = 1.96$$

$$\hat{\sigma}_{y.x} = 15.42 \cdot \sqrt{1 - 0.45^2} = 13.77$$

Grenzen des Konfidenzintervalls für y

$$51.30 + 1.96 \cdot 13.77 = 78.29$$

$$51.30 - 1.96 \cdot 13.77 = 24.31$$

Wir können also mit einer Verlässlichkeit von 95% annehmen, daß unsere Arbeiter an einem Tag zwischen 24 und 78 Stück erfolgreich bearbeiten wird. Wir sehen: Der Test läßt nur eine sehr ungenaue Vorhersage zu. (Das liegt daran, daß der Schätzungseffekt bei einem r_{xy} von 0.45 nur 11% beträgt.)

Knobelaufgabe: Wieso spielt hier der Schätzungseffekt eine Rolle?)

Wichtig: Diese Berechnung eines Konfidenzintervalls setzt voraus, daß die Schätzfehler normalverteilt sind und daß bei allen Werten von x mit etwa gleichem Schätzfehler zu rechnen ist. Es gibt aber auch Fälle, in denen z.B. im unteren Bereich der x -Skala kleinere (oder größere)

Schätzfehler bei der Schätzung von x zu erwarten sind als im oberen Bereich. Beispiel: Ein gewisses Mindestmaß an Intelligenz ist notwendige Voraussetzung für Kreativität, aber auch sehr intelligente Menschen können unkreativ sein. Konsequenz: Bei niedriger Intelligenz können wir mit geringfügigem Schätzfehler sagen, daß auch eine niedrige Kreativität zu erwarten ist, aber die Vorhersage einer höheren Kreativität bei hoher Intelligenz ist mit einem erheblich größeren Schätzfehler behaftet.

4) Konfidenzintervalle und Signifikanztests

Die Konfidenzintervalle für das Populationsmittel μ_x und für die Produkt-Moment-Korrelation ρ_{xy} haben offenkundige Parallelen zu den entsprechenden Signifikanztests für einen einzelnen Mittelwert bzw. für eine einzelne Produkt-Moment-Korrelation. Für solche Konfidenzintervalle gibt es eine Interpretation, die sich an den Signifikanztest anlehnt und die für den Fall des Populationsmittels bereits eingeführt wurde: Das Konfidenzintervall mit Verlässlichkeitsniveau V ist die Gesamtheit aller denkbaren Werte des Populationsparameters, von denen die entsprechende Stichproben-Statistik auf dem Signifikanzniveau $1 - V$ (bzw. $100\% - V$) nicht signifikant abweicht. Ähnlich kann man auch für viele andere Populationsparameter Konfidenzintervalle aufstellen und interpretieren.

<<

Diese Möglichkeit besteht insbesondere auch für "abgeleitete Populationsparameter" wie z.B. die Differenz zweier Populationsmittel. Der t-Test für unabhängige Stichproben konnte ja nicht nur zur Überprüfung der Nullhypothese $\mu_A = \mu_B$ angewandt werden, sondern auch für $H_0 : \mu_A - \mu_B = \Delta_0$, wobei Δ_0 ein von der Nullhypothese behaupteter Differenzbetrag ist. Dementsprechend kann man auch für die Populationmittel-Differenz $\mu_A - \mu_B$ ein Konfidenzintervall aufstellen:

Grenzen des Konfidenzintervalls für die Populationsmittel-Differenz $\mu_A - \mu_B$:

$$\bar{x}_A - \bar{x}_B \pm t_{\text{tab}} \cdot \hat{\sigma}_{\text{diff}}$$

Dabei ist t_{tab} der Tabellenwert für eine zweiseitige Überschreitungswahrscheinlichkeit $1 - V$ bei $N_A + N_B - 2$ Freiheitsgraden, und $\hat{\sigma}_{\text{diff}}$ ist der aus den Daten geschätzte Wert für den Standardschätzfehler der Mittelwerts-Differenz $\bar{x}_A - \bar{x}_B$. Dieser Standardschätzfehler steht im Nenner der Formel für die Prüfgröße t beim t-Test für unabhängige Stichproben; es ist also

$$\hat{\sigma}_{\text{diff}} = \sqrt{\frac{QS_A + QS_B}{N_A + N_B - 2} \cdot \frac{N_A + N_B}{N_A \cdot N_B}}$$

Beispiel: Wenn wir in zwei unabhängigen Stichproben der Größe $N_A = N_B = 61$ Durchschnitte von $\bar{x}_A = 66.79$ und $\bar{x}_B = 62.81$ sowie Quadratsummen von $QS_A = 2895.72$ und $QS_B = 2913.38$ erhalten, dann ergibt sich aus obiger Formel ein Schätzwert von $\hat{\sigma}_{\text{diff}} = 1.26$ für den Standardschätzfehler, mit der die Mittelwertsdifferenz $\bar{x}_A - \bar{x}_B$ behaftet ist. Da der Tabellenwert t_{tab} mit zweiseitiger Überschreitungswahrscheinlichkeit 5% für 120 Freiheitsgrade 1.98 beträgt,

erhalten wir für ein Verlässlichkeits-Niveau von 95% als Grenzen des Konfidenzintervalls für die Populationsmittel-Differenz $\mu_A - \mu_B$ die Werte $66.79 - 62.81 + 1.98 \cdot 1.26 = 6.47$ und $66.79 - 62.81 - 1.98 \cdot 1.26 = 1.49$. Das heißt also: Die tatsächlich gefundene Mittelwertsdifferenz von $\bar{x}_A - \bar{x}_B = 3.98$ weicht nicht signifikant von allen denkbaren Differenzen im Bereich von 1.49 bis 6.47 ab.

Was bedeutet das im Vergleich zum Ergebnis eines t-Tests für unabhängige Stichproben mit der Nullhypothese $\mu_A = \mu_B$ und $\alpha = 0.05$ (zweiseitig)? Zunächst ist festzustellen, daß die von dieser Nullhypothese behauptete Mittelwerts-Differenz $\mu_A - \mu_B = 0$ außerhalb des Konfidenzintervalls liegt, und das bedeutet, daß die Abweichung von H_0 signifikant ist. Genauer: Beim t-Test erhalten wir einen Wert der Prüfgröße von $t = 3.16$, und das entspricht $p = 0.002$ (s. t-Tabelle). Die Nullhypothese ließe sich also auch noch auf einem wesentlich strengeren Signifikanzniveau verwerfen. Aber die damit zu akzeptierende H_1 besagt lediglich, daß die wahre Populationsmittel-Differenz nicht gleich 0 ist. Erinnern wir uns: Bei hinreichend großer Stichprobe können auch minimale Abweichungen von H_0 mit nennenswerter Teststärke zu einem signifikanten Ergebnis führen! Dagegen ist das Konfidenzintervall informativer: Mittelwerts-Differenzen unterhalb von 1.49 liegen außerhalb des Konfidenzintervalls, und damit erhalten wir (sofern wir mit einem Verlässlichkeitsniveau von 95% zufrieden sind) auch einen Beitrag zu einer Antwort auf die Frage, ob eine nennenswerte Mittelwerts-Differenz vorliegt. Natürlich hängt es von der jeweiligen Fragestellung und von der verwendeten Skala ab, ob eine Mittelwerts-Differenz von 1.49 schon als nennenswert zu betrachten wäre; aber der Beitrag des Konfidenzintervalls (die Untergrenze von 1.49) ist für die Psychologie meist wichtiger als die zusätzliche Information des p-Werts von 0.002, die besagt, daß die Nullhypothese sich auch noch auf einem wesentlich strengeren Signifikanzniveau verwerfen ließe.

Aufgrund solcher Überlegungen wird z.Zt. (2001) in der methodischen Fachliteratur ein Streit um "Signifikanztests oder Konfidenzintervalle?" ausgetragen, der fast an einen Glaubenskrieg erinnert.⁹² Der Verfasser dieser Skripts hält das Argument für wichtig, daß das Konfidenzintervall hinsichtlich der Frage des Ausmaßes einer evtl. Abweichung von H_0 informativer ist. Andererseits soll eine Statistik-Einführung vor allem mit den Methoden bekanntmachen, die faktisch in der psychologischen Fachliteratur verwendet werden, und das sind (einstweilen) vor allem Signifikanztests. Zudem ist niemand gehindert, inferenzstatistische Ergebnisse sowohl in der Sprache der Signifikanztests als auch in Form von Konfidenzintervallen mitzuteilen, was einstweilen wohl die beste Lösung in Fällen ist, wo es beides gibt. Bedauerlicherweise bieten gängige Statistik-Software-Pakete allerdings weniger Möglichkeiten zur Berechnung von Konfidenzintervallen, als es derzeit möglich wäre.

>>

Die Parallelen zwischen Konfidenzintervallen und Signifikanztests gehen noch weiter: Die Aufstellung von Konfidenzintervallen ist ebenfalls an Zusatzvoraussetzungen geknüpft. Für das Konfidenzintervall von μ_x muß X in der Population normalverteilt sein; für die beiden übrigen Konfidenzintervalle muß die gemeinsame Populationsverteilung von x und y eine bivariate Normalverteilung sein. Für das Konfidenzintervall von ρ_{xy} soll die Stichprobengröße mindestens 50 betragen; für das Konfidenzintervall der Schätzung von y aus x muß die Korrelation r_{xy} mit

⁹²Überblick z.B. bei Brandstätter, J. (1999). Confidence intervals as an alternative to significance testing. *Methods of Psychological Research Online*, 4(2), 33-46. Internet: <http://www.mpr-online.de>

deren Hilfe die Regressionsgrade bestimmt wurde, aus einer Stichprobe von mindestens 100 Vpn bestimmt worden sein.

Für diese Zusatzannahmen gilt das gleiche, wie für die Zusatzannahmen von Signifikanztests: Normalverteilungsannahmen verlieren bei großen Stichproben an Gewicht; aber wenn man ein Konfidenzintervall für die Schätzung von y aus x berechnet, handelt es sich um einen Einzelfall, also $N = 1$ (auch wenn die Regressionsgerade aus einer größeren Stichprobe ermittelt wurde), und damit ist hier die Zusatzveraussetzung der bivariaten Normalverteilung ziemlich kritisch!

<< Es ist noch anzumerken, daß es auch einseitig begrenzte Konfidenzintervalle gibt. In unserem Beispiel des Produkt-Moment-Korrelations-Konfidenzintervalls hätten wir auch sagen können: Die Wahrscheinlichkeit eines z -Wertes über -1.64 ist 95%. Also ist ζ' mit einer Verlässlichkeit von 95% größer als $0.4847 - 1.64 \cdot 0.05 = 0.4027$. Diesem z' -Wert entspricht eine Produkt-Moment-Korrelation von etwa 0.382. wir haben damit nur eine untere Grenze für ρ_{xy} aufgestellt. Es kann für den Praktiker wichtig sein, nur eine solche untere Grenze eines Validitätskoeffizienten zu wissen; denn die untere Grenze des Konfidenzintervalls für den Validitätskoeffizienten entscheidet darüber, ob mit der nötigen Sicherheit angenommen werden kann, daß der Test für die Praxis valide genug ist.
Für solche und andere Probleme im Zusammenhang mit Konfidenzintervallen wird insbesondere auf die Behandlung der Konfidenzintervalle bei Hays (1973). verwiesen.

>>

F Varianzanalyse

I Die Einweg-Varianzanalyse

1) Ein Experiment (nach Hays, 1973)

Ein Forscher möchte das "Anspruchsniveau" seiner Vpn messen. Hierzu gibt es in der Psychologie häufig verwendete Techniken: Man läßt eine Vp schätzen, wieviele Punkte sie bei einer Aufgabe erzielen wird. Vpn mit hohem Anspruchsniveau erwarten von sich hohe Leistungen und umgekehrt.

Unser Forscher möchte nun untersuchen, ob es etwas an dem gemessenen Anspruchsniveau ausmacht, wenn er den Vpn vorher eine Information darüber gibt, ob ihre Leistung in einem ersten Versuch überdurchschnittlich, durchschnittlich oder unterdurchschnittlich ist.

Versuchsplan:

60 Vpn werden nach Zufall in drei Gruppen (I, II und III) eingeteilt. Alle führen eine Geschicklichkeitsaufgabe einmal durch. Unabhängig von ihrer tatsächlichen Leistung wird ihnen gesagt, sie hätten 31 Punkte erzielt. Was ihnen außerdem gesagt wird, ist verschieden:

Den Vpn der Gruppe I wird gesagt: "Sie haben 31 Punkte erzielt. *Das ist eine*

überdurchschnittliche Leistung. Wieviele Punkte glauben Sie beim nächsten Versuch zu erzielen?"

Gruppe II: "Sie haben 31 Punkte erzielt. *Das ist eine durchschnittliche Leistung*. Wie viele Punkte glauben Sie beim nächsten Versuch zu erzielen?"

Gruppe III: "Sie haben 31 Punkte erzielt. *Das ist eine unterdurchschnittliche Leistung*. Wie viele Punkte glauben Sie beim nächsten Versuch zu erzielen?"

Der Meßwert X_i von V_p i ist dann die Punktzahl, die sie als Ziel für den nächsten Versuch angibt.

Es ergeben sich folgende Werte:

Tab. V1: Meßwerte der Versuchspersonen aus 3 Gruppen
aus einem Experiment zum Anspruchsniveau.

Gruppe I ("überdurchschnittlich")	Gruppe II ("durchschnittlich")	Gruppe III ("unterdurchschnittlich")
52	28	27
48	35	24
43	34	33
50	32	31
43	34	24
44	27	30
46	31	31
46	27	26
43	29	30
49	25	24
38	43	33
42	34	35
42	33	28
35	42	36
33	41	28
38	37	36
39	37	30
34	40	29
33	36	32
34	35	27
<hr/>		
832	680	594 (Gruppensummen)

Aus den Gruppensummen (unter dem Strich) können wir für jede Gruppe einen Gruppendurchschnitt berechnen:

$$\bar{x}_I = 832/20 = 41.6$$

$$\bar{x}_{II} = 680/20 = 34.0$$

$$\bar{x}_{III} = 594/20 = 29.7$$

Für den Gesamtdurchschnitt aller 60 Meßwerte erhalten wir die Gesamtsumme der Meßwerte durch Addition der Gruppensummen. Das ergibt dann:

$$\bar{x} = (832+680+594)/60 = 35.1$$

Wir sehen: Die Vpn der Gruppe I haben sich im Durchschnitt mehr zugetraut als die Vpn der Gruppe II, und diese wieder mehr als die Vpn der Gruppe III.

Es geht nun um die Frage, ob diese Unterschiede signifikant sind.

2) Das Radikalitäts-Problem bei mehrfacher Anwendung des t -Tests

Mit unseren bisherigen Methoden könnten wir lediglich dreimal den t -Test für unabhängige Stichproben durchführen: $t_{I,II}$ wäre der t -Wert aus dem Vergleich von Gruppe I und II; entsprechend wären die Größen $t_{I,III}$ und $t_{II,III}$ definiert als die t -Werte aus dem Vergleich von I und III bzw. II und III. Die Hypothese, daß die "Vorinformation" keinen Einfluß auf die x -Werte hat, wäre unsere Nullhypothese:

$$H_0: \mu_{xI} = \mu_{xII} = \mu_{xIII} (= \mu_x)$$

In Worten: Es gibt ein Populationsmittel von x (μ_x), das für alle drei Versuchssituationen I, II und III dasselbe ist. Oder: Das Populationsmittel von x ist in allen Situationen gleich.

$$H_1: \text{Eine der Gleichungen aus } H_0 \text{ ist unrichtig.}$$

Wir würden H_1 akzeptieren, wenn mindestens einer der drei t -Tests ein signifikantes t ergäbe. Aber: Nehmen wir an, wir führen die t -Tests auf dem 5%-Niveau durch. Das Risiko I (hier also die Wahrscheinlichkeit, daß irgendeiner der drei t -Werte zufällig signifikant wird, obschon H_0 richtig ist) wäre größer als 5%! In Formeln:

$$\begin{aligned} \text{Risiko I} &= p(t_{I,II} \text{ sign.} \vee t_{I,III} \text{ sign.} \vee t_{II,III} \text{ sign.} \mid H_0 \text{ richtig}) \\ &> p(t_{I,II} \text{ sign.} \mid H_0 \text{ richtig}) \end{aligned}$$

Die letztere Wahrscheinlichkeit ist aber 5%, wenn wir die t -Tests auf dem 5%-Niveau durchführen.

Ergebnis der Überlegungen: Würden wir die Nullhypothese durch mehrfache Anwendung des t -Tests überprüfen, so wäre das Risiko I erhöht! Wir würden eine evtl. richtige H_0 zu leichtfertig verwerfen. Unsere Signifikanzprüfung würde radikal werden. Wir können daher die Signifikanzprüfung nicht durch mehrfache Anwendung des t -Test durchführen. Der geeignete Signifikanztest wäre die Einweg-Varianzanalyse. Sie überprüft - allgemein gesprochen - die Nullhypothese, daß mehrere Populationsmittel gleich sind.

<<

Grundsätzlich gäbe es noch eine andere Möglichkeit, die Überhöhung des Risiko I bei mehrfacher Anwendung des t -Tests zu vermeiden: Man legt den einzelnen t -Tests ein strengeres α zugrunde und wählt dieses so, daß die Wahrscheinlichkeit, trotz Gültigkeit von H_0 mindestens

einen signifikanten t-Wert zu erhalten, gerade die erwünschte Größe hat (z.B. 5%). Man spricht dann auch von einer " α -Adjustierung". Aber das wäre mit einem Verlust an Teststärke verbunden, und außerdem hat die Varianzanalyse noch weitere Vorteile, die allerdings erst bei der später zu behandelnden "Mehrweg-Varianzanalyse" voll zum Tragen kommen.

>>

3) Zusatzvoraussetzungen

Unabhängige Stichproben.

Die Populationsverteilungen, deren Mittel verglichen werden, müssen Normalverteilungen sein (Normalitätsannahme) und gleiche Varianzen haben (Homogenitätsannahme).

Die Normalitätsannahme verliert - wie beim t -Test - erheblich an Gewicht, wenn die Stichproben groß sind.

Die Homogenitätsannahme verliert an Gewicht, wenn die Stichproben etwa gleich groß sind, behält allerdings dann einiges Gewicht, wenn die Zahl der Stichproben groß ist.

Die laut Zusatzannahme in allen Populationen gleiche Varianz wollen wir einfach σ^2 nennen (griechischer Buchstabe, da Populationsvarianz!). In unserem Fall haben wir drei gleich große unabhängige Stichproben; also ist die erste Zusatzannahme erfüllt und wir können die Homogenitätsannahme vernachlässigen. Die Stichproben sind zwar zu klein, um schon aus o.g. Grund die Normalitätsannahme zu vernachlässigen; die drei Stichprobenverteilungen sind aber schon so gut an die Normalverteilung angepaßt, daß sich eine Skalentransformation nicht lohnen würde.

In vielen Büchern findet man außerdem die Zusatzvoraussetzung, daß auf Intervallskalenniveau gemessen wird. Besser würde man wie beim t -Test sagen (vgl. die Schlußbemerkung zu E, IV, 2b): Die mit der Varianzanalyse erzielten Aussagen über die Populationsmittel des Meßwertes x lassen sich nur dann auf die Populationsmittel der von x gemessenen Größe übertragen, wenn Intervallskalenniveau vorliegt.

4) Die Aufspaltung der Varianz in Varianzquellen.

Wenn wir die in den 3 Gruppen unseres Experiments erzielten Meßwerte ansehen, dann können wir 2 Quellen von Unterschieden feststellen:

- Einerseits gibt es Unterschiede innerhalb der Gruppen, die sich in den Abweichungen jedes einzelnen Meßwerts vom zugehörigen Gruppendurchschnitt ausdrücken.
- Andererseits gibt es auch Unterschiede zwischen den Gruppen, die sich in den Abweichungen der Gruppendurchschnitte vom Gesamtdurchschnitt ausdrücken.

Nun können ja - wie wir es für 2 Gruppen vom t -Test für unabhängige Stichproben wissen - Unterschiede zwischen Gruppen auch durch Zufall entstehen. "Durch Zufall" heißt dann:

Dadurch, daß diejenigen Einflußgrößen, die für die Unterschiede innerhalb der Gruppen verantwortlich sind, sich bei der Bildung der Gruppen nicht völlig gleichmäßig auf die Gruppen verteilt haben. In unserem Experiment dürften beispielsweise Unterschiede innerhalb der Gruppen in erheblichem Umfang durch das Anspruchsniveau (als Persönlichkeitsvariable) bestimmt sein. Wenn nun in die eine Gruppe "zufällig" mehr Vpn mit hohem Anspruchsniveau gekommen sind als in die andere, dann wäre insoweit ein Unterschied zwischen den Gruppen zufällig aufgrund der Einflußgrößen entstanden, die für die Unterschiede innerhalb der Gruppen verantwortlich sind.

Die Varianzanalyse überprüft nun, ob die Unterschiede zwischen den Gruppen so groß sind, daß sie in einem solchen Sinne nicht mehr (bzw. kaum noch) durch Zufall entstanden sein können.

a) Die Zerlegung der Abweichungen vom Gesamtdurchschnitt.

Um die verschiedenen Quellen der Unterschiedlichkeit von Meßwerten quantitativ miteinander vergleichen zu können, wird - wie der Name "Varianzanalyse" nahelegt - auf die Varianz als Maß der Unterschiedlichkeit von Meßwerten zurückgegriffen.

Eine zentrale Rolle bei der Definition der Varianz spielte die Abweichung (Differenz) der einzelnen Meßwerte vom Stichprobendurchschnitt \bar{x} : Die Varianz ist das arithmetische Mittel der Quadrate dieser Abweichungen. Für praktische Rechenzwecke ist es dann günstiger, andere Formeln zu verwenden, aber um zu verstehen, was Varianz ist, ist es zunächst einmal sinnvoller, von den Abweichungswerten auszugehen. Ähnliches gilt auch bei der Varianzanalyse: Für das Verständnis ihres Grundansatzes ist es wohl am sinnvollsten, von den Abweichungswerten auszugehen. Für Rechenzwecke gibt es dann praktischere Formeln, die in einem späteren Abschnitt behandelt werden.

Im Grunde beruht die ganze Varianzanalyse darauf, die Abweichung eines Meßwerts vom Gesamtdurchschnitt zu zerlegen in die Abweichung des Meßwerts vom zugehörigen Gruppendurchschnitt und die Abweichung dieses Gruppendurchschnitts vom Gesamtdurchschnitt. Am Beispiel unseres ersten Meßwerts sieht das so aus:

$$\begin{array}{rclcl} 52 - 35.1 & = & (52 - 41.6) & + & (41.6 - 35.1) \\ \text{Abweichung eines} & & \text{Abweichung des Meßwerts} & & \text{Abweichung des} \\ \text{Meßwerts vom} & = & \text{vom zugehörigen} & + & \text{Gruppendurchschnitts vom} \\ \text{Gesamtdurchschnitt} & & \text{Gruppendurchschnitt} & & \text{Gesamtdurchschnitt} \end{array}$$

Eine solche Zerlegung der Abweichung eines Meßwerts vom Gesamtdurchschnitt kann man für jeden einzelnen Meßwert durchführen. Weil sie von zentraler Bedeutung für die ganze Varianzanalyse ist, ist sie in der folgenden Tabelle V2 jeweils für den ersten, zweiten und letzten Meßwert jeder Gruppe durchgeführt.

Tab. V2: Zerlegung der Abweichung des Meßwerts vom Gesamtdurchschnitt in die Abweichung des Meßwerts vom zugehörigen Gruppendurchschnitt und die Abweichung des Gruppendurchschnitts vom Gesamtdurchschnitt (Daten nach Tab. V1).

$$(52-35.1) = (52-41.6) + (41.6-35.1)$$

$$(48-35.1) = (48-41.6) + (41.6-35.1)$$

.
 . usw. Gruppe I

$$(34-35.1) = (34-41.6) + (41.6-35.1)$$

$$(28-35.1) = (28-34.0) + (34.0-35.1)$$

$$(35-35.1) = (35-34.0) + (34.0-35.1)$$

.
 . usw. Gruppe II

$$(35-35.1) = (35-34.0) + (34.0-35.1)$$

$$(27-35.1) = (27-29.7) + (29.7-35.1)$$

$$(24-35.1) = (24-29.7) + (29.7-35.1)$$

.
 . usw. Gruppe III

$$(27-35.1) = (27-29.7) + (29.7-35.1)$$

Daß das alles nicht nur in diesem Beispiel stimmt, sondern ganz allgemein, ist unmittelbar einleuchtend, wenn man die Klammern auflöst. Wenn man es als allgemeine Formel schreiben will, dann könnte es so aussehen:

$$(X_i - \bar{x}) = (X_i - \bar{x}_{g(i)}) + (\bar{x}_{g(i)} - \bar{x})$$

Dabei ist g(i) die Gruppe, zu der der i-te Meßwert gehört, und entsprechend ist $\bar{x}_{g(i)}$ der zugehörige Gruppendurchschnitt.

b) Die Zerlegung der Abweichungs-Quadratsumme.

Da die Varianz auf den Quadraten der bisher betrachteten Abweichungen beruht, wollen wir uns ansehen, was aus unserer Zerlegungsgleichung aus dem vorigen Abschnitt wird, wenn wir

beide Seiten der Gleichung quadrieren. Für den ersten Meßwert ergibt sich z.B.:

$$(52-35.1)^2 = ((52-41.6) + (41.6-35.1))^2 = (52-41.6)^2 + 2(52-41.6)(41.6-35.1) + (41.6-35.1)^2$$

Diese Gleichung ergibt sich durch Quadrieren der Zerlegungsgleichung des vorigen Abschnitts. Das letzte Gleichheitszeichen beruht auf der bekannten Formel

$$(a+b)^2 = a^2 + 2ab + b^2$$

wobei in diese Formel für a und b die Klammerausdrücke (52-41.6) und (41.6-35.1) eingesetzt sind.

Wir sehen: Wenn man die Abweichungswerte quadriert, dann gibt es keine so einfache Zerlegung mehr, wie bei den unquadrierten Abweichungswerten. Wir werden aber gleich sehen, daß sich wieder eine ähnlich einfache Zerlegung ergibt, wenn wir Quadratsummen bilden. Dazu ist in der folgenden Tabelle V3 eine entsprechende Rechnung für alle Zeilen aus Tabelle V2 durchgeführt, und damit das, worauf es nachher ankommt, leichter zu erkennen ist, sind die einzelnen Ausdrücke in etwas anderer Reihenfolge angeordnet, als in der obigen Gleichung.

Tab. V3: Zerlegung der Abweichungsquadrate.

$(52-35.1)^2$	$=$	$(52-41.6)^2 + (41.6-35.1)^2 + 2(41.6-35.1)(52-41.6)$	
$(48-35.1)^2$	$=$	$(48-41.6)^2 + (41.6-35.1)^2 + 2(41.6-35.1)(48-41.6)$	
.			
. usw.		Gruppe I	
$(34-35.1)^2$	$=$	$(34-41.6)^2 + (41.6-35.1)^2 + 2(41.6-35.1)(34-41.6)$	
$(28-35.1)^2$	$=$	$(28-34.0)^2 + (34.0-35.1)^2 + 2(34.0-35.1)(28-34.0)$	
$(35-35.1)^2$	$=$	$(35-34.0)^2 + (34.0-35.1)^2 + 2(34.0-35.1)(35-34.0)$	
.			
. usw.		Gruppe II	
$(35-35.1)^2$	$=$	$(35-34.0)^2 + (34.0-35.1)^2 + 2(34.0-35.1)(35-34.0)$	
$(27-35.1)^2$	$=$	$(27-29.7)^2 + (29.7-35.1)^2 + 2(29.7-35.1)(27-29.7)$	
$(24-35.1)^2$	$=$	$(24-29.7)^2 + (29.7-35.1)^2 + 2(29.7-35.1)(24-29.7)$	
.			
. usw.		Gruppe III	
$(27-35.1)^2$	$=$	$(27-29.7)^2 + (29.7-35.1)^2 + 2(29.7-35.1)(27-29.7)$	
2895.4	$=$	1443.0 + 1452.4	<u>(Summe über alle Vpn!)</u>
QS_{tot}	$=$	QS_{in} + QS_{zw}	

Der "Knüller" dieser Tabelle liegt in den letzten zwei Zeilen. In der vorletzten Zeile sind jeweils die einzelnen quadrierten Klammerausdrücke, die darüber stehen, für jede Vp ausgerechnet und über alle Vpn summiert (alle aus Tab. V1, auch die, die in Tab. V3 nicht enthalten sind!). Wer will, kann's nachrechnen. Worauf es ankommt: Die Gleichung geht unter dem Strich auch dann auf, wenn man die "doppelten Produkte" am Schluß der vorhergehenden Zeilen unter den Tisch fallen läßt. Warum das so ist, werden wir gleich sehen. Erst mal ist als Ergebnis etwas festzuhalten, was allgemein in der Einweg-Varianzanalyse gilt: So, wie sich jede einzelne Abweichung additiv in zwei Bestandteile zerlegen läßt, so auch die Summe der Abweichungsquadrate:

Die Summe der quadrierten Abweichungen aller Meßwerte vom Gesamtdurchschnitt läßt sich in 2 Quadratsummen zerlegen:

- die Summe der quadrierten Abweichungen der Meßwerte vom jeweiligen Gruppendurchschnitt und...
- die Summe der quadrierten Abweichungen der Gruppendurchschnitte vom

Gesamtdurchschnitt, (wobei auch diese quadrierte Abweichung für jede V_p neu aufzusummieren ist - bzw., was auf dasselbe hinausläuft, für jede Gruppe so oft, wie es Meßwerte in der Gruppe gibt.)

Alle drei Quadratsummen spielen eine Rolle in der Varianzanalyse und haben deshalb einen besonderen Namen:

- Die Summe der quadrierten Abweichungen aller Meßwerte vom Gesamtdurchschnitt heißt Gesamt-Quadratsumme oder "totale Quadratsumme" (Abkürzung: QS_{tot}).
- Die Abweichungen der Meßwerte vom jeweiligen Gruppendurchschnitt spiegeln die Unterschiede der Meßwerte innerhalb der Gruppen wider; die Summe der entsprechenden Abweichungsquadrate heißt deshalb "Quadratsumme innerhalb der Gruppen" oder auch einfach "Quadratsumme innerhalb" (Abk.: QS_{in}).
- Die Abweichungen der Gruppendurchschnitte vom Gesamtdurchschnitt drücken die Unterschiede zwischen den Gruppen aus. Die Summe der entsprechenden Abweichungsquadrate heißt entsprechend "Quadratsumme zwischen den Gruppen" oder "Quadratsumme zwischen", Abk. QS_{zw} .

Mit diesen Definitionen läßt sich die additive Zerlegung, um die es hier geht, dann auch so schreiben wie in der letzten Zeile von Tab. V3:

$$QS_{\text{tot}} = QS_{\text{in}} + QS_{\text{zw}}$$

<< Bleibt nachzutragen, warum das so ist, bzw. genauer: Warum bei der Bildung der Quadratsummen die doppelten Produkte (also die Glieder hinter dem letzten Pluszeichen jeder Zeile) wegfällen. Man kann leicht zeigen, daß sich sogar in jeder einzelnen Gruppe diese doppelten Produkte zu 0 addieren (und damit natürlich auch in der Gesamtgruppe). Sehen wir uns z.B. in Tab. V3 die doppelten Produkte der Gruppe I an. Sie beginnen alle mit $2(41.6-35.1)$. Diesen Ausdruck können wir "ausklammern", wenn wir die Summe der doppelten Produkte bilden. Wir können also für die Summe der doppelten Produkte auch schreiben:

$$2(41.6-35.1) \cdot [(52-41.6)+(48-41.6)+\dots+(34-41.6)].$$

Die Differenzen $(52-41.6)$, $(48-41.6)$ usw., die dabei in der eckigen Klammer stehen, sind die Abweichungen der Meßwerte der Gruppe I vom Gruppendurchschnitt dieser Gruppe - und diese Abweichungen müssen sich innerhalb die Gruppe zu 0 addieren. Genau diese Summe steht aber in den eckigen Klammern, und damit ist dann auch die gesamte Summe der doppelten Produkte in dieser Gruppe 0. Und das ist offensichtlich nicht nur in dieser Gruppe so, sondern auch in jeder anderen.

>> Zu ergänzen ist noch, daß man die verschiedenen Gesichtspunkte, unter denen wir bisher Abweichungswerte berechnet haben, auch als "Varianzquellen" bezeichnet. In der Einwegvarianzanalyse, um die es hier geht, gibt es also die Varianzquellen "innerhalb der Gruppen", "zwischen den Gruppen" und "Gesamt".

c) Vereinfachte Berechnung der Quadratsummen.

Jeden Abweichungswert zu berechnen und zu zerlegen, um dann zu quadrieren und die Quadratsummen zu bilden, bedeutet eine Menge Rechenaufwand. Wie bereits angekündigt, gibt es einfachere Formeln für die Berechnung der Quadratsummen - ganz ähnlich wie bei der einfachen Varianzberechnung. In der folgenden Darstellung sind die allgemein gültigen Formeln und Definitionen durch Fettdruck gekennzeichnet.

Bei der einfachen Varianzberechnung galt die Gleichung:

$$QS = RQ - C$$

für die Summe der Abweichungsquadrate (QS). Dabei ist RQ, die "rohe Quadratsumme", die Summe der quadrierten Rohwerte, und das Korrekturglied C ist definiert als

$$C = (\sum_i X_i)^2/N$$

<< Bekanntlich gibt es auch eine andere Definition dieses Korrekturglieds:

$$C = N \cdot \bar{x}^2$$

Für die einfache Varianzberechnung reicht diese Formel völlig aus, da man leicht abschätzen kann, welche Rechengenauigkeit für den Durchschnitt erforderlich ist, damit Rundungsfehler sich nicht "fortpflanzen". In der Varianzanalyse kann das schwieriger werden, da mehrere Korrekturglieder berechnet werden. In der Varianzanalyse empfiehlt es sich daher, die Korrekturglieder in der Weise zu berechnen, daß man - entsprechend der ersten Formel - die quadrierte Meßwertsumme durch die Zahl der Meßwerte dividiert. In unserem Beispiel macht das zwar keinen Unterschied, weil die Zahlen so "gebastelt" sind, daß mit einer Stelle hinter dem Komma (bei den Durchschnitten) keine Rundungsfehler mehr auftreten, aber damit das Rechenschema auch für nicht gebastelte Situationen übernommen werden kann, wird auch hier die sicherere Methode angewandt.

Und dazu gleich noch ein Rechenrick: Hat man einen Taschenrechner, der "aussteigt", weil das Quadrat der Meßwertsumme zu groß ist, dann dividiert man die Meßwertsumme durch die Stichprobengröße und multipliziert dann nochmal mit der Meßwertsumme.

>>

Die Quadratsumme QS, um die es bei der einfachen Varianzberechnung geht, ist offenbar das, was hier die totale Quadratsumme ist (wir betrachten einfach alle Gruppen zusammen als eine Stichprobe, und dann ist QS_{tot} genau das, was bei der einfachen Varianzberechnung QS ist.)

In unserem Fall wäre also

$$RQ = 52^2 + 48^2 + \dots + 34^2 + 28^2 + 35^2 + \dots + 35^2 + 27^2 + 24^2 + \dots + 27^2 = 76816$$

Für die Berechnung des Korrekturglieds C müssen wir als N natürlich die Gesamtzahl aller

Meßwerte in allen Gruppen einsetzen; in unserem Fall gilt also (mit einer Meßwertsumme von $832+680+594 = 2106$, vgl. S.@):

$$C = 2106^2 : 60 = 73920.6$$

und somit

$$\mathbf{QS_{tot}} = \mathbf{RQ} - \mathbf{C} = 76816 - 73920.6 = 2895.4$$

Für die beiden anderen Quadratsummen brauchen wir die Korrekturglieder, die sich ergeben, wenn wir jede Gruppe als Stichprobe für sich betrachten, also

$$C_I = 832^2 : 20 = 34611.2$$

$$C_{II} = 680^2 : 20 = 23120.0$$

$$C_{III} = 594^2 : 20 = 17641.8$$

Es muß also für C_g - das Korrekturglied der Gruppe g - die zugehörige Meßwertsumme quadriert und dieses Quadrat durch die entsprechende Gruppengröße N_g dividiert werden. Bezeichnen wir die Meßwertsumme der Gruppe g als S_g , dann können wir es als Formel schreiben:

$$C_g = (S_g)^2 / N_g$$

Die Summe dieser Korrekturglieder der Gruppen bezeichnen wir als SCG:

$$\mathbf{SCG} = \sum_g C_g = 34611.2 + 23120.0 + 17641.8 = 75373,0$$

Dann gelten für die verschiedenen Quadratsummen die folgenden Formeln:

$$\mathbf{QS_{in}} = \mathbf{RQ} - \mathbf{SCG} = 76816 - 75373.0 = 1443.0$$

$$\mathbf{QS_{zw}} = \mathbf{SCG} - \mathbf{C} = 75373.0 - 73920.6 = 1452.4$$

und schließlich (s.o.):

$$\mathbf{QS_{tot}} = \mathbf{RQ} - \mathbf{C} = 76816 - 73920.6 = 2895.4$$

d) Die Freiheitsgrade der verschiedenen Varianzquellen.

Bereits im Zusammenhang mit dem t-Test für unabhängige Stichproben (und anderen Anwendungen der t-Verteilung) haben wir das Konzept der Freiheitsgradzahl kennengelernt, die einer Quadratsumme zugeordnet ist. Dieses Konzept spielt auch in der Varianzanalyse eine wichtige Rolle.

In dem Exkurs zum Begriff der Freiheitsgrade (S. @) ergab sich: Die Quadratsumme beruht jeweils auf (quadrierten) Abweichungswerten; die Freiheitsgradzahl ist dann die Zahl der (nicht quadrierten) Abweichungswerte, die "frei beweglich" sind, bevor alle weiteren Abweichungswerte aufgrund von mathematischen Begrenzungen ("Restriktionen") aus den frei beweglich vorgegebenen berechnet werden können. Eine solche Restriktion kann z.B. dadurch entstehen, daß die Summe der Abweichungen vom Durchschnitt einer Stichprobe 0 sein muß.

In der Varianzanalyse ist jeder Quadratsumme eine Freiheitsgradzahl zugeordnet, die nach diesen Prinzipien bestimmt wird. Sie werden - ebenso wie die Quadratsummen, zu denen sie jeweils gehören - durch einen Index voneinander unterschieden. Wir haben also df_{tot} , df_{in} und df_{zw} .

Die Abkürzung df steht wieder für "degrees of freedom". Ohne Punkte, weil es sonst (mit Index) ein zu langes Symbol ergäbe.

Am einfachsten zu bestimmen ist die Freiheitsgradzahl für die totale Quadratsumme. Sie beruht auf den Abweichungen der Meßwerte vom Gesamtdurchschnitt, und da diese zusammen 0 ergeben müssen ("Restriktion"), ist

$$df_{tot} = N - 1.$$

Dabei ist N die Gesamtzahl der Meßwerte in allen Gruppen zusammen. In unserem Beispiel also

$$df_{tot} = 60 - 1 = 59$$

Die Freiheitsgrade der "Quadratsumme innerhalb" werden bestimmt wie beim t-Test für unabhängige Stichproben. Die Quadratsumme beruht auf den Abweichungen der Meßwerte vom jeweiligen Gruppendurchschnitt. Diese müssen in jeder einzelnen Gruppe zusammen 0 ergeben; daher sind - in unserem Beispielexperiment - in jeder Gruppe nur 19 Abweichungen vom Gruppendurchschnitt frei beweglich; der 20. ist jeweils durch die vorhergehenden festgelegt. Insgesamt sind also 57 Abweichungen frei beweglich. Allgemein: Haben wir insgesamt (in allen Gruppen zusammen) N Meßwerte (und damit auch N Abweichungen vom jeweiligen Gruppendurchschnitt) und G Gruppen, dann ist in jeder Gruppe der letzte Abweichungswert vom Gruppendurchschnitt durch die vorhergehenden festgelegt. Frei beweglich sind also

$$df_{in} = N - G = 60 - 3$$

Abweichungen vom Gruppendurchschnitt.

Der Quadratsumme zwischen den Gruppen liegen die (quadrierten) Abweichungen der Gruppendurchschnitte vom Gesamtdurchschnitt zugrunde. Diese werden zwar für jede V_p neu gebildet und aufsummiert, aber innerhalb jeder Gruppe ist dieselbe Abweichung zu verwenden. Auch das können wir als eine Restriktion auffassen, die die Zahl der frei beweglichen Abweichungen begrenzt: Es kann überhaupt nur G verschiedene Abweichungen der Gruppendurchschnitte vom Gesamtdurchschnitt geben. Dann gibt es noch eine weitere Restriktion: Offensichtlich muß bei gleich großen Gruppen die Summe der Abweichungen der Gruppendurchschnitte vom Gesamtdurchschnitt ebenfalls 0 sein.

<< Bei unterschiedlicher Gruppengröße gilt eine andere Restriktion, die im nächsten Abschnitt behandelt wird, die aber hinsichtlich der Freiheitsgrade auf das gleiche Ergebnis hinausläuft. Und wem das oben für gleich große Gruppen Behauptete nicht offensichtlich ist, der kann es sich aus dieser allgemeineren Restriktion für beliebige Gruppengrößen leicht selbst ableiten.

>> Das bedeutet: Sind $G-1$ Abweichungen der Gruppendurchschnitte vom Gesamtdurchschnitt festgelegt, dann ist damit durch die Restriktion auch die letzte festgelegt. Die Freiheitsgradzahl der auf diesen Abweichungen beruhenden Quadratsumme zwischen den Gruppen ergibt sich demnach als

$$df_{zw} = G - 1 = 3 - 1 = 2.$$

Wenn wir nun die Formeln für die verschiedenen Freiheitsgradzahlen (und die Rechenergebnisse aus unserem Beispiel) vergleichen, dann können wir feststellen, daß für die Freiheitsgrade die gleiche additive Zerlegung gilt wie für die Quadratsummen: Es ist

$$df_{tot} = df_{in} + df_{zw}.$$

Das gilt allgemein in der Varianzanalyse: Wenn die Quadratsumme einer Varianzquelle additiv zerlegt wird in Quadratsummen für mehrere einzelne Varianzquellen, dann gilt eine entsprechende additive Zerlegung auch für die zugehörigen Freiheitsgrade.

e) Gesamtdurchschnitt und Freiheitsgradzahl zwischen den Gruppen bei unterschiedlicher Gruppengröße.

Wenn wir verschiedene Gruppen unterschiedlicher Größe haben und die jeweiligen Gruppendurchschnitte kennen, dann können wir daraus leicht den Gesamtdurchschnitt berechnen.

Beispiel: Wir haben 4 Gruppen, deren Größe und Durchschnitt in der 2. und 3. Spalte der folgenden Tabelle angegeben ist:

Gruppe	Größe	Durchschnitt	Meßwertsumme	$\bar{x}_g - \bar{x}$	$N_g \cdot (\bar{x}_g - \bar{x})$
I	25	43.6	1090	4.0	100.0
II	34	36.0	1224	-3.6	-122.4
III	36	40.5	1458	0.9	32.4
IV	50	39.4	1970	-0.2	-10.0

Auch wenn die in der vierten Spalte der Tabelle angegebenen Meßwertsummen nicht bekannt sind, können wir sie aus Größe und Durchschnitt der Einzelgruppen leicht rekonstruieren, indem wir beide miteinander multiplizieren - denn der Durchschnitt ist ja entstanden, indem die Meßwertsumme durch die Gruppengröße dividiert wurde. Beispiel Gruppe I: $25 \cdot 43.6 = 1090$. Die Größe und die Meßwertsumme der Gesamtgruppe, die sich durch Zusammenfassung der Einzelgruppen ergibt, erhalten wir offenbar, indem wir Größen und Meßwertsummen der Einzelgruppen addieren, und damit ergibt sich der Gesamtdurchschnitt \bar{x} als

$$\bar{x} = \frac{1090+1224+1458+1970}{25+34+36+50} = \frac{5742}{145} = 39.60$$

(Die beiden letzten Spalten der Tabelle werden später behandelt.)

Verallgemeinert: Wir addieren die (durch Multiplikation von Gruppengröße und Gruppendurchschnitt rekonstruierten) Meßwertsummen der Einzelgruppen und dividieren durch die Summe der Gruppengrößen. Als Formel:

$$\bar{x} = \frac{\text{Gesamt-Meßwertsumme}}{\text{Gesamt-Gruppengröße}} = \frac{\sum_g N_g \cdot \bar{x}_g}{\sum_g N_g}$$

Dabei sind N_g und \bar{x}_g Größe und Durchschnitt der Gruppe g .

Das Ergebnis läßt sich auch folgendermaßen formulieren: Der Gesamtdurchschnitt ist ein gewichteter Durchschnitt der Gruppendurchschnitte, wobei die Gruppengrößen die Gewichtszahlen sind.

Aus dieser Gleichung ergibt sich aber mit einfachen algebraischen Umformungen eine andere, in der die für die Quadratsumme und die Freiheitsgrade zwischen den Gruppen bedeutsamen Abweichungen der Gruppendurchschnitte vom Gesamtdurchschnitt auftauchen (wer sich überzeugen will, kann's selber herleiten):

$$\sum_g N_g \cdot (\bar{x}_g - \bar{x}) = 0$$

oder ausgeschrieben:

$$N_I \cdot (\bar{x}_I - \bar{x}) + N_{II} \cdot (\bar{x}_{II} - \bar{x}) + \dots = 0$$

Für das obige Beispiel ist das in den letzten zwei Spalten der Tabelle vorgerechnet.

Die in unserer letzten Gleichung in Klammern stehenden Ausdrücke sind aber die Abweichungen der Gruppendurchschnitte vom Gesamtdurchschnitt, und damit stellt diese Gleichung eine Restriktion für diese Abweichungen dar, die auch bei ungleichen Gruppengrößen gilt. Sie läßt sich folgendermaßen formulieren: Die mit den Gruppengrößen multiplizierten Abweichungen der Gruppendurchschnitte vom Gesamtdurchschnitt müssen zusammen 0 ergeben.

Und wegen dieser Restriktion gilt auch bei ungleicher Gruppengröße für die Freiheitsgrade zwischen den Gruppen die Gleichung $df_{zw} = G-1$.

5) Logik und Durchführung der Einwegvarianzanalyse.

Wir haben im vorigen Abschnitt gesehen, daß sich sowohl die Summe der quadrierten Abweichungen vom Gesamtdurchschnitt als auch die mit dieser Quadratsumme verbundenen N-1 Freiheitsgrade additiv auf 2 Varianzquellen aufteilen lassen. Wenn man den Zusammenhang von Quadratsummen, Freiheitsgraden und Varianzschätzungen kennt, dann schreibt man eine solche Situation geradezu danach, die verschiedenen Quadratsummen durch die zugehörigen Freiheitsgrade zu dividieren. Das wird in der Varianzanalyse auch gemacht. Das Ergebnis ist ein "Mittelquadrat" (abgekürzt MQ, wobei jeweils der Index "zw", "in" oder "tot" angefügt wird - je nach dem, welche Quadratsumme mit zugehöriger Freiheitsgradzahl zugrundegelegt wird). Also:

$$MQ_{in} = QS_{in} : df_{in} = 1443.0 : 57 = 25.316$$

$$MQ_{zw} = QS_{zw} : df_{zw} = 1452.4 : 2 = 726.2$$

$$MQ_{tot} = QS_{tot} : df_{tot} = 2895.4 : 59 = 49.075$$

Von Interesse in der Einwegvarianzanalyse sind die ersten beiden Mittelquadrate. Der Signifikanztest für die Nullhypothese der Varianzanalyse besteht darin, daß mit einem einseitigen F-Test geprüft wird, ob MQ_{zw} signifikant größer ist als MQ_{in} :

$$F = MQ_{zw} : MQ_{in} = 726.2 : 25.316 = 28.69$$

Dieser F-Wert ist hochsignifikant (näheres dazu s.u.); wir verwerfen also die Nullhypothese: Die Unterschiede (Varianz) zwischen den Gruppen sind zu groß, als daß sie noch mit nennenswerter Wahrscheinlichkeit zufällig entstehen könnten.

<< Begründung für diesen Schluß: Über die beiden Mittelquadrate, die in den F-Test eingehen, läßt sich mathematisch folgendes beweisen (vgl. Abschnitt @FI8):

- MQ_{in} - das Mittelquadrat innerhalb der Gruppen - ist eine erwartungstreue Schätzung der wahren Varianz innerhalb der Gruppen, die lt. Homogenitätsannahme (s.o.: Zusatzveraussetzungen) für alle Gruppen gleich ist.

- In MQ_{zw} - das Mittelquadrat zwischen den Gruppen - geht neben der wahren Varianz zwischen den Gruppen auch die wahre Varianz innerhalb der Gruppen ein, wie man leicht an einem Extremfall sehen kann: Selbst wenn es keine wahre Varianz zwischen den Gruppen gibt, wenn also die den Gruppennennern entsprechenden Populationsmittel alle gleich ist und somit die Nullhypothese der Varianzanalyse zutrifft, werden in aller Regel zufällige Unterschiede zwischen den Gruppennennern auftreten.

Genauer läßt sich beweisen, daß in diesem Fall - also bei Gültigkeit der Nullhypothese - die zu erwartenden zufälligen Unterschiede zwischen den Gruppen gerade so groß sind, daß MQ_{zw} ebenfalls - genauso wie MQ_{in} - eine erwartungstreue Schätzung der wahren Varianz innerhalb der Gruppen ist.

Gibt es dagegen eine echte, nicht nur auf Zufall zurückzuführende Varianz zwischen den Gruppen, dann ist MQ_{zw} entsprechend größer, und zwar um so mehr, je größer die Gruppen. Genauer: Der Erwartungswert von MQ_{zw} enthält außer der wahren Varianz innerhalb der Gruppen noch ein von der Gruppengröße abhängiges Vielfaches der wahren Varianz zwischen den Gruppen.

Das bedeutet dann aber: Gilt die Nullhypothese, dann werden bei dem oben definierten F-Wert zwei erwartungstreue Schätzungen derselben Varianz dividiert - und damit liegt es nahe, daß die Wahrscheinlichkeitsverteilung dieser Prüfgröße bei Gültigkeit der Nullhypothese eine F-Verteilung ist. Gilt dagegen die Nullhypothese nicht (gibt es also echte Varianz zwischen den Gruppen), dann nimmt die Prüfgröße größere Werte an, als bei Gültigkeit von H_0 .

Der F-Test in der Varianzanalyse ist also seinem Wesen nach einseitig (Genau wie der χ^2 -Test nach Pearson). Und genau so wie dort (vgl. S. @) gibt es in bestimmten Fällen die Möglichkeit, eine Art "halbseitigen" F-Test durchzuführen, wenn man sich vor der Datenerhebung dafür entscheidet, nur in bestimmten Fällen der Überschreitung des kritischen F-Wertes die H_0 zu verwerfen.

Hier haben wir einmal einen Fall eines einseitigen Signifikanztests, bei dem aus logischen Gründen nur eine der beiden möglichen Abweichungen von der Nullhypothese in Frage kommt: Die von MQ_{zw} geschätzte Varianz enthält einmal die (auch von MQ_{in} erfaßte) wahre Varianz innerhalb der Gruppen und darüber hinaus noch ggf. ein von der Gruppengröße abhängiges Vielfaches der wahren Varianz zwischen den Gruppen. Und da letztere wie alle Varianzen nicht negativ sein kann, kann die von MQ_{zw} geschätzte Varianz aus logischen Gründen nicht kleiner, sondern höchstens größer als die von MQ_{in} geschätzte sein.

Das schließt allerdings nicht aus, daß im Einzelfall MQ_{zw} deutlich kleiner ist als MQ_{in} - aber selbst in diesem Fall würden wir uns nicht zu der Aussage verleiten lassen, daß dann eben die wahre Varianz zwischen den Gruppen negativ ist - denn das gibt keinen Sinn.

>>

Beim F-Test haben wir immer 2 Freiheitsgradzahlen - eine für den Zähler und eine für den Nenner. In unserem Fall ist das natürlich df_{zw} für den Zähler und df_{in} für den Nenner.

In unserem Experiment hat der F-Wert also 2 Freiheitsgrade im Zähler und 57 im Nenner. Und der erzielte F-Wert von 28.69 übersteigt bei weitem alle in unseren Tabellen für diese Freiheitsgradzahlen entnehmbaren Werte.

Da 57 Freiheitsgrade im Nenner in unserer F-Tabelle nicht aufgeführt sind, können wir folgendermaßen die Tabelle benutzen: Wir stellen fest, daß jeweils innerhalb einer Spalte der Tabelle die kritischen F-Werte mit zunehmender Freiheitsgradzahl für den Nenner immer kleiner werden. Der kritische F-Wert für 2 gegen 57 Freiheitsgrade (d.h. 2 Freiheitsgrade im Zähler und 57 im Nenner) muß also kleiner sein, als der für 2 gegen 40 Freiheitsgrade, den wir in der Tabelle haben. Da der aus unseren Daten berechnete F-Wert aber viel größer ist als der Tabellenwert für 2 gegen 40 Freiheitsgrade, ist er erst recht größer als der (in unserer Tabelle nicht enthaltene) Tabellenwert für 2 gegen 57 Freiheitsgrade.

Ergebnis: Das Mittelquadrat zwischen den Gruppen ist signifikant größer als das Mittelquadrat innerhalb. D.h., daß die Unterschiede zwischen den Gruppen kaum zufällig sein dürften. Wir verwerfen also die Nullhypothese, daß es keine wahren Unterschiede (Varianz) zwischen den Gruppen gibt.

Allgemein kann man die Funktion des F-Tests in der Einweg-Varianzanalyse folgendermaßen zusammenfassen:

Der F-Test in der Einweg-Varianzanalyse überprüft, ob die Unterschiede zwischen den Gruppen größer sind, als es aufgrund der Unterschiede innerhalb der Gruppen bei Gültigkeit der Nullhypothese auch zufällig zu erwarten ist.

Es ist üblich, die Ergebnisse einer Varianzanalyse in einer Tabelle der folgenden Art zusammenzufassen:

Quelle	QS	df	MQ	F
Zwischen den Gruppen	1452.4	2	726.2	28.69**
innerhalb der Gruppen	1443.0	57	25.316	
Gesamt	2895.4	59		

** : $p < .01$

Zum Standardaufbau einer solchen Tabelle gehören folgende Regeln, die hier gleich so allgemein formuliert sind, daß sie auch bei anderen Arten der Varianzanalyse angewandt werden können:

1. Die Reihenfolge der Spalten (Quelle - QS - df - MQ - F) ist fest.

2. Für die Reihenfolge der Zeilen gilt, daß Varianzquellen, deren Quadratsummen und Freiheitsgrade diejenigen anderer Varianzquellen zusammenfassen, unter den zusammengefaßten Varianzquellen stehen, und zwar meistens mit einem Querstrich getrennt. (Hier: "Gesamt" unter "zwischen..." und "innerhalb...").

Außerdem werden 2 Varianzquellen, die zusammen in einen F-Bruch eingehen, so angeordnet, daß die Varianzquelle, die in den Zähler eingeht, oberhalb der Varianzquelle steht, die in den Nenner geht. (Hier: "zwischen" oberhalb von "innerhalb".) Allerdings können bei Varianzanalysen mit mehr Varianzquellen (die wir noch kennenlernen) weitere Varianzquellen zwischen den beiden stehen.

3. Es werden nur diejenigen Mittelquadrate in der Tabelle aufgeführt, die in einen F-Wert eingehen

4. F-Werte stehen in der Zeile der Varianzquelle, die in den Zähler des F-Bruchs eingeht (hier: "zwischen...").

5. Bei signifikanten F-Werten wird mit einem Zeichen (hier: **) das Signifikanzniveau gekennzeichnet, wobei das Zeichen in einer Fußnote zu erläutern ist.

Unser signifikantes Ergebnis besagt: Wir können die Nullhypothese, daß die den Gruppendurchschnitten entsprechenden Populationsmittel alle gleich sind, mit sehr niedriger Irrtumswahrscheinlichkeit verwerfen. Dabei müssen wir aber eines beachten: Diese Nullhypothese zu verwerfen bedeutet lediglich, daß sie nicht in vollem Umfang zutrifft, daß also nicht alle Populationsmittel gleich sind. Es kann aber durchaus sein, daß zwei der 3 Populationsmittel gleich sind und nur das dritte sich von den beiden anderen unterscheidet. Wenn man bei einer signifikanten Varianz zwischen den Gruppen genau wissen will, welche Populationsmittel sich unterscheiden und welche möglicherweise nicht, dann muß man weitere Signifikanztests durchführen - die sog. post-hoc-Tests. Sie werden in einschlägigen Lehrbüchern behandelt und hier nicht.

6) "Effekte" und "Fehler"

Die Abweichungen der Meßwerte vom zugehörigen Gruppendurchschnitt und die Abweichungen der Gruppendurchschnitte vom Gesamtdurchschnitt, die wir bisher betrachtet haben, sind aufgrund der jeweiligen Stichprobendurchschnitte definiert. Eng damit verwandt sind 2 Begriffe, die ebenfalls von Bedeutung sind: "Fehler" und "Effekte". Sie unterscheiden sich von den bisher betrachteten Abweichungen dadurch, daß sie aufgrund der Populationsdurchschnitte und nicht aufgrund der Stichproben definiert sind.

a) Der Begriff des Effekts

In unserem Beispiel haben wir H_0 verworfen:

Die Populationsmittel dürften sich mit an Sicherheit grenzender Wahrscheinlichkeit voneinander unterscheiden. Wir könnten auch sagen: Die "Vorinformation" über den angeblichen Leistungsstand hat eine Auswirkung (einen Effekt) auf die Meßwerte.

Wir wollen das Ausmaß dieses Effektes nun quantifizieren. Wir betrachten dazu die Populationsmittel μ_I , μ_{II} , μ_{III} und μ . μ_I wäre das Populationsmittel, das entstehen würde, wenn wir alle Mitglieder der Population unter Bedingung I testen würden; entsprechend wären μ_{II} und μ_{III} definiert.

μ sei das arithmetische Mittel dieser drei Populationsmittel:

$$\mu = (\mu_I + \mu_{II} + \mu_{III}) / 3.$$

<< Sind die Gruppen nicht (wie in unserem Fall) gleich groß, so gelten die folgenden Überlegungen nur dann, wenn man μ als das nach den Stichprobengrößen gewichtete Mittel der Einzelmittel definiert:

$$\mu := \frac{\sum_g N_g \cdot \mu_g}{\sum_g N_g}$$

Man könnte auch sagen: Es wird davon ausgegangen, daß die Gruppengrößen "repräsentativ" sind.

>>

Offenbar können wir diese Populationsmittel folgendermaßen schätzen:

$$\hat{\mu}_I = \bar{x}_I = 41.6$$

$$\hat{\mu}_{II} = \bar{x}_{II} = 34.0$$

$$\hat{\mu}_{III} = \bar{x}_{III} = 29.7$$

$$\hat{\mu} = (\hat{\mu}_I + \hat{\mu}_{II} + \hat{\mu}_{III}) / 3 = (41.6 + 34.0 + 29.7) / 3 = 35.1$$

$$\text{Oder auch: } \hat{\mu} = \bar{x} = 35.1$$

Wir wollen für einen Augenblick außer acht lassen, daß es sich bei diesen Werten nur um Schätzungen handelt, und annehmen, daß diese Werte die tatsächlichen Populationsmittel seien. Es sei also in der Population:

$$\mu_I = 41.6; \mu_{II} = 34.0; \mu_{III} = 29.7; \mu = 35.1$$

Interpretieren wir zunächst den letzten Wert: $\mu = 35.1$

Abgesehen von allen Effekten der verschiedenen "Vorinformationen" vermuten die Vpn im Durchschnitt, im zweiten Versuch etwa den gleichen Wert bzw. eine geringfügige Verbesserung zu erzielen. (Es war ihnen ja gesagt worden, sie hätten im ersten Versuch 31 Punkte erzielt.) Welchen Effekt hat nun die Vorinformation I? Sie schraubt die Erwartung in die Höhe: Man hofft

sich zu verbessern, wenn einem gesagt wurde, der angeblich erreichte Punktwert von 31 sei ein überdurchschnittlicher Punktwert. Das Ausmaß dieses Effekts der Vorinformation I wollen wir a_I nennen. Wir berechnen es, indem wir μ von μ_I subtrahieren:

$$a_I = \mu_I - \mu = 41.6 - 35.1 = 6.5$$

In Worten: Gegenüber dem Gesamtmittel μ besteht unter Bedingung I eine Hebung des Anspruchsniveaus um 6.5 Einheiten auf der x-Skala. Oder auch: Die Vorinformation I führt zu einer Hebung des Anspruchsniveaus um 6.5 Punkte.

Anmerkung: Bei diesem speziellen Experiment könnte man auch versucht sein zu sagen: Wir subtrahieren nicht das Gesamtmittel μ , sondern den Wert 31 (also die angebliche Punktzahl im Erstversuch) von μ_I . Wir hätten dann gleichsam einen anderen Bezugspunkt auf der x-Skala für die Beurteilung des einzelnen Effekts gewählt. Grundsätzlich könnte man verschiedene Punkte der x-Skala als Bezugspunkte wählen. Daß wir hier noch einen speziellen Bezugspunkt zur Auswahl haben (eben den Wert 31), ist offenbar eine Besonderheit dieses Experiments. Da es aber um allgemeine Definitionen für die Varianzanalyse geht, wird ein Bezugspunkt gewählt, der immer verfügbar ist, und das ist das Gesamtmittel μ .

Wie würde sich der Effekt der "Vorinformation II" berechnen?

Er würde sich berechnen nach der Formel:

$$a_{II} = \mu_{II} - \mu = 34.0 - 35.1 = -1.1$$

Formulieren Sie das Ergebnis in Worten!

In Worten: Gegenüber dem Gesamtmittel μ besteht unter Bedingung II eine Senkung des Anspruchsniveaus um 1.1 x-Einheiten.

Berechnen Sie auch den Effekt der Vorinformation III!

$$a_{III} = \mu_{III} - \mu = 29.7 - 35.1 = -5.4$$

In Worten?

Gegenüber dem Gesamtmittel μ besteht unter Bedingung III eine Senkung des Anspruchsniveaus um 5.4 Einheiten.

Könnten Sie eine allgemeine Formel für a_g , den Effekt der g-ten Bedingung geben?

Allgemeine Formel für a_g , also den Effekt der g-ten Bedingung:

$$a_g = \mu_g - \mu$$

In Worten: Ein Effekt ist definiert als die Abweichung des zu einer Gruppe gehörenden Populationsmittels μ_g vom Gesamtmittel μ .

Addieren wir einmal die Effekte:

$$a_I + a_{II} + a_{III} = 6.5 + (-1.1) + (-5.4) = 0$$

Das ist kein Zufall. Es gilt immer, daß sich die Effekte zu 0 addieren (sich gegenseitig aufheben), solange die Stichproben gleich groß sind:

$$\sum_g a_g = 0$$

Sind die Stichproben nicht gleich groß, so gilt (wie wir noch sehen werden)

$$\sum_g N_g \cdot a_g = 0$$

Wir wollen uns jetzt aber daran erinnern, daß die von uns eingesetzten Zahlenwerte nur Schätzungen der Populationsmittel sind. Die Gültigkeit der Überlegungen wird dadurch zwar nicht angetastet, solange es sich um die in Buchstaben geschriebenen Formeln handelt. Mit den Zahlen würden wir z.B. für Bedingung II besser schreiben:

$$\hat{a}_{II} = \hat{\mu}_{II} - \hat{\mu} = 34.0 - 35.1 = -1.1$$

In Worten: Wir *schätzen*, daß der Effekt der Bedingung II in einer Senkung des Anspruchsniveaus um 1.1 Einheiten besteht.

Allgemein lautet die Formel für unsere Schätzung des g-ten Effekts:

$$\hat{a}_g = \hat{\mu}_g - \hat{\mu} = \bar{x}_g - \bar{x}$$

Hinweis: Es gibt auch hier die Möglichkeit, neben der Punktschätzung eines Effekts ein Konfidenzintervall für diesen Effekt aufzustellen.

b) Organismische und experimentelle Effekte

Wir wollen noch ein anderes Beispiel betrachten:

Wir wollen annehmen, wir hätten aus den fünf Bildungsschichten, wie sie auf S.@ definiert sind, je eine Zufallsstichprobe gezogen. Die Größe der Stichprobe sei proportional dem Prozentsatz dieser Schichten in der Population. Wir haben die Intelligenz aller Vpn mit einem Test gemessen und die folgenden Stichprobendurchschnitte der Punktzahlen erhalten:

$$\bar{x}_I = 53.8;$$

$$\bar{x}_{II} = 64.7;$$

$$\bar{x}_{III} = 70.3;$$

$$\bar{x}_{IV} = 77.9;$$

$$\bar{x}_V = 83.4$$

$$\bar{x} = 68.5$$

Wir wollen annehmen, daß sich in der Varianzanalyse diese Unterschiede als signifikant herausgestellt haben.

Schätzen Sie die Effekte!

Wir schätzen die Effekte:

$$\hat{a}_I = 53.8 - 68.5 = -14.7$$

$$\hat{a}_{II} = 64.7 - 68.5 = -3.8$$

$$\hat{a}_{III} = 70.3 - 68.5 = 1.8$$

$$\hat{a}_{IV} = 77.9 - 68.5 = 9.4$$

$$\hat{a}_V = 83.4 - 68.5 = 14.9$$

Da die Stichproben nicht gleich groß sind, addieren sich die Effekte nicht zu 0. (Es würde aber gelten: $N_I \cdot \hat{a}_I + N_{II} \cdot \hat{a}_{II} + \dots + N_V \cdot \hat{a}_V = 0$).

Versuchen Sie, Effekt I in Worten zu formulieren!

Wenn wir die geschätzten Effekte hier in Worten formulieren wollen, müssen wir beachten:

Wir wissen nicht, ob die Verminderung der Intelligenz in der ersten Bildungsschicht eine Folge der Zugehörigkeit zu dieser Bildungsschicht ist oder ob die Vpn umgekehrt die Volksschule nicht absolviert haben, weil sie zu unintelligent sind. Im vorigen Experiment konnten wir eine eindeutige Kausalaussage machen: Die "Vorinformation I" führt zu einer Hebung des Anspruchsniveaus. Hier dagegen können wir keine solche Aussage über die Kausalrichtung machen. Wir können nur feststellen:

Gegenüber dem Gesamtmittel μ besteht in der Bildungsgruppe I eine Verminderung der Intelligenz um 14.7 Punkte.

Woher kommt es, daß wir in einem Fall eine Kausalaussage machen können, im anderen nicht?

Bei dem Experiment zur Intelligenz verschiedener Bildungsschichten brachte jede Vp die Zugehörigkeit zu der einen oder anderen Gruppe mit ins Experiment. Wir konnten die Vpn nicht beliebig den einzelnen Gruppen zuweisen. Da ist es dann möglich, daß die gemessene Größe (die Intelligenz) mit dazu beiträgt, in welche Gruppe eine Vp kommt. Im Experiment zum Anspruchsniveau haben wir dagegen willkürlich (durch Zufall) entschieden, in welche Gruppe eine Vp kommt, d.h. welche "Vorinformation" sie erhält. Es ist daher sichergestellt, daß nicht das verschiedene Anspruchsniveau der Vpn zur Gruppenzugehörigkeit führt.

Verallgemeinern wir den Gedankengang!

Wir *definieren* zunächst: Hängt die Zugehörigkeit zur einen oder anderen Gruppe von Eigenschaften ab, die die Vpn mit ihrem Organismus in das Experiment mitbringen, so nennen wir die Variable, die die verschiedenen Gruppen definiert, auch eine organismische Variable. Wird dagegen die Zugehörigkeit eines Meßwertes zur einen oder anderen Meßwertgruppe ganz vom Experimentator manipuliert, so unterscheiden sich die Gruppen durch eine "experimentell manipulierte" oder ganz einfach durch eine "experimentelle" Variable.

Welche der beiden bisher besprochenen Variablen ist organismisch, welche experimentell manipuliert?

Im ersten Experiment wurden die Vpn den verschiedenen Vorinformationen, durch die sich die Gruppen unterscheiden, vom Versuchsleiter durch Zufall zugewiesen. Die Variable "Vorinformation" ist also eine experimentelle manipulierte Variable. Im zweiten Experiment hing die Zugehörigkeit zu einer der fünf Bildungsgruppen dagegen von einer Eigenschaft ab, die die Vpn mit ins Experiment brachten: Von der schulischen Vergangenheit der Vpn. Daher ist die Variable "Bildungsschicht" eine organismische Variable.

Welchen Zusammenhang hat die Unterscheidung von experimentell manipulierten und organismischen Variablen mit der Frage der Kausalrichtung?

Bei experimentell manipulierten Variablen können wir die Kausalrichtung von Effekten eindeutig festlegen: Signifikante Unterschiede der Populationsmittel von x können nur auf die verschiedenen experimentellen Bedingungen zurückgehen, unter denen die Meßwerte erhoben

wurden, da ein Einfluß der Meßwerte auf die Gruppenzugehörigkeit nicht in Frage kommt. Die Gruppenzugehörigkeit hängt ja vom Zufall ab. Im Beispiel: Es ist nicht möglich, daß das Anspruchsniveau in irgendeiner Weise einen Einfluß darauf hat, welche Vorinformation die Vpn erhielten. Die Vpn wurden den einzelnen Vorinformationen ganz nach Zufall zugewiesen. (Wie man so etwas macht, wird im Experimental-Praktikum besprochen.) Bei experimentell manipulierten Gruppenunterschieden kann man also eindeutig eine unabhängige Variable (im Beispiel die Vorinformation) und eine abhängige Variable (im Beispiel das Anspruchsniveau) unterscheiden. Bei organismisch bedingten Gruppenunterschieden kann es dagegen vorkommen, daß die Gruppenzugehörigkeit teilweise oder ganz von den x-Werten abhängt, deren Populationsverteilungen verglichen werden. Beispiel: Es kann also sein, daß die Zugehörigkeit zu einer Bildungsschicht ganz oder teilweise von der Intelligenz abhängt. Daneben gibt es aber auch die Möglichkeit, daß beide sich gegenseitig beeinflussen oder von einer dritten Variablen (etwa von der sozialen Schicht des Elternhauses) abhängen.

Woran erinnert das?

An das Problem des Kausalzusammenhanges bei Korrelationen. Genau so wie dort ist es auch hier eine Frage der Sachkenntnis und nicht der Statistik, ob man sich für eine der möglichen Kausalinterpretationen entscheidet. Man kann jedenfalls hier nicht einfach sagen, daß die Variable, durch die die Gruppenzugehörigkeit definiert ist, die unabhängige ist, und daß eventuelle Meßwerte, in denen sich die Gruppen signifikant unterscheiden, eine von der Gruppenzugehörigkeit abhängige Variable darstellen. Die Verwendung der Begriffe "abhängig" und "unabhängig", die sich in der Literatur häufig auch bei organismischen findet, ist allenfalls in einem ähnlich abstrakten Sinn statthaft, wie wir ihn im folgenden für den "Effekt"-Begriff entwickeln.

Unsere Überlegungen gingen davon aus, daß wir bei der Verbalisierung von "Effekten" einer organismischen Variablen keine Kausalaussage machen konnten. Es läge nun nahe, hier das Wort "Effekt" ganz zu vermeiden, da es ja schon eine Kausalaussage beinhaltet. Die Verwendung des Wortes Effekt hat sich hier aber so weit eingebürgert, daß es nicht sinnvoll wäre, es zu umgehen. Wir *definieren* den Begriff vielmehr so, daß sich keine Schwierigkeiten ergeben:

Ein Effekt ist in der Varianzanalyse die Differenz eines (wahren) Gruppenschnitts und des (wahren) Gesamtdurchschnitts

$$a_g = \mu_g - \mu$$

Der Begriff des Effekts ist also in der Varianzanalyse so definiert, daß er keine Kausalaussage impliziert, sondern lediglich das Ausmaß der Abweichung eines Gruppenschnitts vom Gesamtdurchschnitt wiedergibt.

c) Differentielle Effekte

Von differentiellen Effekten spricht man dann, wenn ein Effekt nicht bei allen Vpn gleich ist. Es

ist in unserem Beispiel denkbar, daß die eine Vp erheblich stärker auf die Information reagiert, ihre Leistung sei überdurchschnittlich, als eine andere Vp. Es ist sogar denkbar, daß die gleiche "Vorinformation" bei der einen Vp zu einer Steigerung, bei der anderen zu einer Senkung des Anspruchsniveaus führt. In diesem Fall wäre der in Abschnitt a bestimmte "Effekt der Vorinformation I" das Populationsmittel des für alle Vpn verschiedenen Effekts.

Allgemein: Liegen differentielle Effekte vor, so sind die nach den Regeln aus Abschnitt a bestimmten Effekte die Populationsmittel dieser Effekte.

Daraus können wir zwei Schlußfolgerungen ziehen:

1. Findet sich irgendwo ein bestimmter signifikanter Effekt, so kann es trotzdem möglich sein, daß bei einigen Vpn kein Effekt, bei anderen sogar der gegenteilige Effekt auftritt.
2. Selbst wenn in der Population der mittlere Effekt gleich null ist, kann das dadurch zustande kommen, daß sich positive und negative differentielle Effekte gegenseitig aufheben.

d) Der Begriff des Fehlers

Für die Varianz innerhalb der Gruppen findet sich in der Literatur häufig der Begriff "Fehlervarianz". Das ist aus der Geschichte der Varianzanalyse zu erklären. Die Varianzanalyse wurde nicht für die Psychologie, sondern für die Agrarwissenschaft entwickelt. Aber auch in der Psychologie standen zunächst experimentelle Methoden im Vordergrund, aus denen die Verwendung des Begriffs "Fehlervarianz" verständlich wird. Wir wollen beispielsweise annehmen, daß ein Forscher die Reaktionszeit einer einzigen Vp in drei verschiedenen Situationen (I, II, III) je 10 mal gemessen hat und daß sich die folgenden Durchschnitte der Reaktionszeit x (in Millisekunden) ergeben:

$$\bar{x}_I = 80 \quad \bar{x}_{II} = 83 \quad \bar{x}_{III} = 85$$

Es ist durchaus sinnvoll, in diesen Durchschnitten eine Schätzung für die wahre Reaktionszeit der Vp in den verschiedenen Situationen zu sehen. Es handelt sich dabei um drei Gruppen von Meßwerten. Innerhalb dieser Gruppen von Meßwerten findet sich aber allenfalls eine gewisse Variation, eine "Varianz innerhalb". Wenn die Vp beispielsweise im ersten Versuch der Meßwertgruppe I eine Reaktionszeit von 77 Millisekunden hatte, so können wir den Differenzbetrag zum wahren Wert als Meßfehler auffassen, der entweder durch Unzulänglichkeiten der Apparatur oder durch mangelnde Stabilität der Vp zustande gekommen ist. Falls der Wert 80 die genaue "Wahre Reaktionszeit" der Vp ist, würden wir den Meßfehler berechnen als

$$e_1 = x_1 - \mu_1 = 77 - 80 = -3$$

In Worten: Der Meßfehler in der ersten Beobachtung ist die Differenz des Meßwerts und des wahren Mittels der Meßwertgruppe I, zu der er gehört. Der Meßfehler besteht in einer Abweichung von 3 Millisekunden nach unten (letzteres wird durch das negative Vorzeichen

ausgedrückt).

Allgemein würde man sagen:

$$e_i = x_i - \mu_{g(i)},$$

wobei $g(i)$ die Gruppe ist, zu der die Person i gehört. Der Meßfehler bei der i -ten Beobachtung ist also die Differenz des i -ten Meßwerts vom zugehörigen Gruppenmittel.

Diesen Begriff des Fehlers sollte man aber nicht unbesehen auf andere Experimente übertragen.

In dem Experiment, an dem wir die Varianzanalyse geübt haben, liegt beispielsweise der erste Meßwert (52) um 10.4 Einheiten über dem zugehörigen Gruppenmittel. Das *kann* an einem "Meßfehler" liegen. Es kann aber auch daran liegen, daß ein differentieller Effekt vorliegt oder auch daran, daß diese Vp ohnehin ein hohes Anspruchsniveau hat und auch unter jeder der anderen zwei Bedingungen ein überdurchschnittliches x genannt hätte.

Allgemein: Die Varianz innerhalb der Gruppen beruht nicht nur auf eigentlichen Meßfehlern, sondern auch auf differentiellen Effekten und auf überdauernden Eigenschaften der Vpn.

Allerdings gibt es eine andere Rechtfertigung für die Verwendung des Begriffs "Fehlervarianz" für die "Varianz innerhalb": Mit Hilfe der "Varianz innerhalb" schätzen wir ab, mit welchem Stichproben-Fehler die Schätzung der einzelnen Populationsmittel μ_I , μ_{II} und μ_{III} behaftet sind, also ob die Unterschiede der Stichprobendurchschnitte aufgrund von Stichprobenfehlern entstanden sein könnten. Solange wir also mit dem Begriff "Fehlervarianz" nur diese Funktion der "Varianz innerhalb" meinen, ist die Verwendung des Begriffs Fehlervarianz gerechtfertigt.

Wir können der Schwierigkeit bei der Verwendung des Fehlerbegriffs aber noch auf eine andere Weise begegnen. Genau so, wie wir bei den organismischen Effekten den Begriff Effekt mathematisch abstrakt definieren als Abweichung des Gruppendurchschnitts vom Gesamtdurchschnitt und damit alle kausalen Interpretationen aus dem statistischen Begriff des Effekts ausklammerten, so können wir auch den "Fehler" abstrakt definieren:

Definition:

$$e_i := x_i - \mu_{g(i)},$$

wobei $g(i)$ die Gruppe ist, zu der VP i gehört. In Worten: Der "Fehler" bei der i -ten Beobachtung ist die Abweichung des i -ten Meßwerts (x_i) vom zugehörigen Populationsmittel $\mu_{g(i)}$.

Diesen Fehler können wir schätzen als

$$\hat{e}_i = x_i - \hat{\mu}_{g(i)} = x_i - \bar{x}_{g(i)}$$

In diesem Fehler können differentielle Effekte usw. definitionsgemäß enthalten sein. Anders

formuliert: Wenn wir die Abweichung eines Meßwerts vom Gruppenmittel als "Fehler" bezeichnen, dann ist damit noch nicht behauptet, daß es sich um Meßfehler im engeren Sinn des Wortes handelt!

e) Allgemeines Modell der Einweg-Varianzanalyse

Definitionsgemäß gelten die folgenden Gleichungen:

$$\mu_{g(i)} - \mu = a_{g(i)}$$

$$x_i - \mu_{g(i)} = e_i$$

Addieren wir beide Gleichungen, so erhalten wir:

$$x_i - \mu = a_{g(i)} + e_i$$

In Worten: Die Abweichungen eines Meßwerts vom Populationsmittel besteht aus zwei Komponenten: Dem Gruppeneffekt und dem Fehler. Das sind aber auch die zwei Komponenten, in die wir die Varianz von x aufgelöst haben.

Addieren wir auf beiden Seiten μ , so erhalten wir

$$x_i = \mu + a_{g(i)} + e_i$$

In Worten: Jeder Meßwert x_i ergibt sich nach dem Modell der Einweg-Varianzanalyse durch die Addition des zugehörigen Gruppeneffekts $a_{g(i)}$ und eines Meßfehlers e_i zum Gesamtmittel μ . Diese Gleichung ist keine Zusatzannahme der Varianzanalyse; sie ergibt sich vielmehr aus der Definition von Effekten und Fehlern.

Wir können die Nullhypothese der Varianzanalyse auch anders formulieren: Wenn alle μ_g gleich dem Populationsmittel μ sind, dann sind alle Effekte gleich null. Also:

$$H_0: a_g = 0 \text{ für alle } g$$

$$H_1: a_g \neq 0 \text{ für irgendein } g.$$

Wichtig: Wenn wir H_0 verwerfen, so müssen wir doch damit rechnen, daß der eine oder andere Effekt gleich null ist! Es wird beim Verwerfen von H_0 nur behauptet, daß nicht alle Effekte gleich null sind.

Genau so in der früheren Formulierung:

Verwerfen wir die Nullhypothese

$$H_0: \mu_I = \mu_{II} = \mu_{III} = \dots$$

dann ist doch die Möglichkeit gegeben, daß zwei Populationsmittel gleich sind. Verwerfen wir H_0 , so behaupten wir nur, daß sie nicht *alle* gleich sind.

7) Varianzanalytische Parameterschätzung als Teil der Korrelationsstatistik

Bei der Produkt-Moment-Korrelation konnten wir formulieren:

Der Determinationskoeffizient (r^2_{xy}) drückt aus, welcher Teil der Varianz von x durch y determiniert ist - genau gilt das nur bei linearer Regression. Ist die Regression nicht linear, so drückt der Koeffizient *Eta-Quadrat* (η^2) aus, welcher Varianzanteil der einen Variablen durch die

andere determiniert ist.

Dabei enthält das Wort "determiniert" bekanntlich keinerlei Hypothesen über die Kausalrichtung!

In einem varianzanalytischen Experiment können wir ähnliche Aussagen machen. Wir können beispielsweise untersuchen, welcher Teil der Varianz der Intelligenz durch die Zugehörigkeit zu einer der Bildungsschichten determiniert ist. Das hängt offenbar davon ab, wie groß die Varianz zwischen den Bildungsschichten im Verhältnis zur Totalvarianz ist. Es ist allerdings aus verschiedenen Gründen nicht möglich, einfach das "Mittelquadrat zwischen den Gruppen" durch ein " MQ_{total} " zu dividieren, um eine derartige Aussage machen zu können.

Methoden der Berechnung entsprechender Koeffizienten werden beispielsweise bei Hays (1973) besprochen.

8. Anhang: Herleitung der Erwartungswerte von Quadratsummen und Mittelquadraten in der Einweg-Varianzanalyse.

<<

In Abschnitt @FI5 (Logik und Durchführung der Einwegvarianzanalyse) wurden einige Aussagen über die Erwartungswerte von MQ_{zw} und MQ_{in} gemacht, die in diesem Absatz näher unter die Lupe genommen werden sollen. Das erfolgt in zwei Schritten. Zunächst werden die Sätze, mit denen in Abschnitt @FI5 die Eigenschaften, auf die es ankommt, formuliert wurden, in die Sprache von Formeln übersetzt, und dann werden diese Formeln bewiesen.

Es geht um folgende Eigenschaften:

- Unabhängig davon, ob die H_0 zutrifft oder nicht, ist MQ_{in} - das Mittelquadrat innerhalb der Gruppen - eine erwartungstreue Schätzfunktion für die "wahre Varianz innerhalb der Gruppen", die lt. Zusatzvoraussetzung (Varianzhomogenität) für alle Gruppen gleich ist.
- Für MQ_{zw} - das Mittelquadrat zwischen den Gruppen, wurden zwei Behauptungen aufgestellt: Bei Gültigkeit der H_0 ist MQ_{zw} ebenfalls eine erwartungstreue Schätzfunktion der wahren Varianz innerhalb der Gruppen. Trifft dagegen H_0 nicht zu, dann erwarten wir größere Werte von MQ_{zw} .

Um die Bedeutung dieser Sätze genauer zu verstehen, ist ein Rückgriff auf die allgemeine Theorie der Parameterschätzung hilfreich. Einige dort (in Abschnitt D) dargestellte Begriffe und Überlegungen lassen sich folgendermaßen auf die Mittelquadrate MQ_{in} und MQ_{zw} anwenden:

- Jede aus den Daten berechnete Größe ist eine Statistik; also sind auch die Mittelquadrate MQ_{in} und MQ_{zw} Statistiken.
- Wie jede Statistik können auch die Mittelquadrate MQ_{in} und MQ_{zw} als Zufallsgrößen betrachtet werden, da sie durch Stichprobenzufälle größer oder kleiner ausfallen können.
- Als Zufallsgrößen haben die Mittelquadrate eine Wahrscheinlichkeitsverteilung. In den folgenden Überlegungen geht es vor allem um die Erwartungswerte dieser Wahrscheinlichkeitsverteilungen.
- Eine Statistik ist eine erwartungstreue Schätzfunktion für einen Parameter, wenn der Erwartungswert der Wahrscheinlichkeitsverteilung der Statistik gleich dem Wert des Parameters ist. Wenn es also darum geht, ob bzw. unter welchen Bedingungen die Mittelquadrate MQ_{in} und MQ_{zw} erwartungstreue Schätzfunktionen der "wahren Varianz innerhalb der Gruppen" ist, dann ist diese Varianz der Parameter. Mit der Bezeichnung σ^2

für diese Varianz lassen sich die obigen Behauptungen über die Mittelquadrate MQ_{in} und MQ_{zw} also folgendermaßen formulieren:

- Unabhängig davon, ob H_0 zutrifft oder nicht, ist $E(MQ_{in}) = \sigma^2$.
- Trifft H_0 zu, dann ist auch $E(MQ_{zw}) = \sigma^2$.
- Trifft H_0 dagegen nicht zu, dann ist $E(MQ_{zw}) > \sigma^2$.

Der Beweis dieser Eigenschaften wird vereinfacht, wenn wir zunächst die Quadratsummen QS_{in} und QS_{zw} betrachten. (Auch diese Quadratsummen sind Größen, die sich aus den Daten berechnen lassen, also Statistiken, die eine Wahrscheinlichkeitsverteilung haben!) Für die Erwartungswerte dieser Quadratsummen werden später die folgenden Gleichungen bewiesen, die sowohl bei Gültigkeit als auch bei Ungültigkeit von H_0 zutreffen:

$$E(QS_{zw}) = (G-1) \cdot \sigma^2 + \sum_g N_g \cdot a_g^2$$

$$E(QS_{in}) = (N-G) \cdot \sigma^2$$

Dabei ist...

σ^2 = die (als identisch vorausgesetzte) Populationsvarianz innerhalb der Gruppen

N = Gesamtzahl aller Meßwerte (V_{pn}) in allen Gruppen zusammen.

G = Zahl der Gruppen

N_g = Zahl der Meßwerte in Gruppe g

a_g = Wahrer Effekt der Zugehörigkeit zu Gruppe g

Bevor die obigen Gleichungen für die Erwartungswerte der Quadratsummen hergeleitet werden, soll zunächst gezeigt werden, was sich daraus für die Erwartungswerte der Mittelquadrate ergibt, um die es ja eigentlich geht. Aus der Gleichung für $E(QS_{zw})$ folgt⁹³

$$E(MQ_{zw}) = E(QS_{zw}/df_{zw}) = E(QS_{zw})/(G-1) = \sigma^2 + \sum_g N_g \cdot a_g^2 / (G-1),$$

und aus der Gleichung für $E(QS_{in})$ ergibt sich

$$E(MQ_{in}) = E(QS_{in}/df_{in}) = E(QS_{in})/(N-G) = \sigma^2$$

Die letzte dieser Gleichungen ist bereits die behauptete Eigenschaft von MQ_{in} als erwartungstreue Schätzfunktion der Varianz σ^2 . Die von der Gültigkeit von H_0 abhängigen Behauptungen über MQ_{zw} ergeben sich folgendermaßen aus der Gleichung für den Erwartungswert $E(MQ_{zw})$:

- Trifft H_0 zu, trifft also für alle wahren Gruppeneffekte die Gleichung $a_g = 0$ zu, dann ist auch $\sum_g N_g \cdot a_g^2 / (G-1) = 0$ und damit $E(MQ_{zw}) = \sigma^2$.
- Trifft dagegen H_0 nicht zu, ist also für einige (oder alle) Gruppeneffekte $a_g \neq 0$, dann ist für diese Gruppeneffekte $a_g^2 > 0$, und damit ergibt sich $\sum_g N_g \cdot a_g^2 / (G-1) > 0$. Dann folgt aber $E(MQ_{zw}) > \sigma^2$ aus der Gleichung für $E(MQ_{zw})$.

⁹³Für das zweite Gleichheitszeichen wird der im folgenden dargestellte Hilfssatz H2 sowie die Gleichung $df_{zw} = G-1$ verwendet.

Nachdem die Konsequenzen aufgezeigt sind, die sich aus den Gleichungen für die Erwartungswerte $E(QS_{zw})$ und $E(QS_{in})$ ergeben, brauchen nur noch diese Gleichungen hergeleitet zu werden. Dafür ist es günstig, auf die "Korrekturglieder-Formeln" für die Quadratsummen aus Abschnitt @FI4c zurückzugreifen (also $QS_{zw}=SCG-C$ und $QS_{in}=RQ-SCG$). Dazu ist zu beachten: Ebenso wie die Mittelquadrate und die Quadratsummen sind auch die rohe Quadratsumme RQ und die diversen Korrekturglieder Statistiken und damit Zufallsgrößen. Die Erwartungswerte der Quadratsummen sind leicht zu bestimmen, wenn wir zunächst die Erwartungswerte der diversen Korrekturglieder und von RQ bestimmen, und dabei können wir folgende "Hilfssätze" über Erwartungswerte und Varianzen von Zufallsvariablen verwenden. (Diese Hilfssätze sind unmittelbare Konsequenzen der Eigenschaften von Erwartungswerten und Varianzen, die bei der Einführung von Wahrscheinlichkeitsverteilungen behandelt wurden.⁹⁴)

H1: Für jede Zufallsgröße X gilt: $E(X^2)=\sigma^2_X+(E(X))^2$.

(Dieser Satz ist eine unmittelbare Konsequenz der Formel $\sigma^2_X=E(X^2)-(E(X))^2$.)

H2: Ist X eine Zufallsgröße und b eine Konstante, dann ist $E(b \cdot X)=b \cdot E(X)$.

H3: Der Erwartungswert einer Summe von Zufallsgrößen ist gleich der Summe der Erwartungswerte der einzelnen Zufallsgrößen.

Entsprechend gilt für Differenzen: $E(X-Y) = E(1 \cdot X + (-1) \cdot Y) = E(X) - E(Y)$.

H4: Ist S eine Summe von unkorrelierten Zufallsgrößen, dann ist die Varianz von S die Summe der Varianzen der einzelnen Zufallsgrößen.

H5: Ist S eine Summe von n unkorrelierten Zufallsgrößen X_1, X_2, \dots, X_n mit Erwartungswerten $\mu_1, \mu_2, \dots, \mu_n$ und Varianzen $\sigma^2_1, \sigma^2_2, \dots, \sigma^2_n$, dann ist

$$E(S^2) = \sigma^2_S + (E(S))^2 = \sum_{i=1..n} \sigma^2_i + (\sum_{i=1..n} \mu_i)^2.$$

(Dieser Hilfssatz ist eine Anwendung von H1 auf die Zufallsgröße S, für deren Erwartungswert und Varianz auf H3 und H4 zurückgegriffen wird.)

Nun also zu den Erwartungswerten von RQ, C und C_g in der Einweg-Varianzanalyse. Nach den Voraussetzungen der Einweg-Varianzanalyse (in Verbindung mit Grundprinzipien der Stichprobentheorie) sind die Meßwerte X_i unabhängige (und damit unkorrelierte) Zufallsgrößen, deren Erwartungswert das zum jeweiligen Meßwert gehörende "wahre Gruppenmittel" μ_g ist. Die Varianz jedes X_i ist die (lt. Zusatzvoraussetzung für alle Gruppen gleiche) Populationsvarianz σ^2 . Dann ergibt sich aus H1 und H3 für den Erwartungswert von RQ:

$$E(RQ) = N \cdot \sigma^2 + \sum_g N_g \cdot \mu_g^2.$$

Beachte dazu: Für jeden zur Gruppe g gehörenden Meßwert X_i läßt sich der Erwartungswert von X_i^2 nach H1 angeben als $E(X_i^2) = \sigma^2 + \mu_g^2$. Nun ist RQ die Summe dieser X_i^2 für alle Meßwerte der Gesamtstichprobe. Summieren wir (gemäß H3) die oben angegebenen Erwartungswerte von X_i^2 über alle i, dann kommt in dieser Summe N mal die Varianz σ^2 vor, und jedes μ_g^2 kommt N_g mal vor.

⁹⁴Zu H3 und H4 vgl. die vorletzte Seite des Statistik-I-Skripts: Das arithmetische Mittel von $X_i + Y_i$ ist $\bar{x} + \bar{y}$, und die Varianz dieser Summe ist $s^2_x + 2s_{xy} + s^2_y$. Bei unkorrelierten Größen ist $s_{xy} = 0$. Diese Regeln lassen sich auf Summen von mehr als zwei Größen ausdehnen (da $x+y+z = (x+y)+z$), und derartige Regeln über das Verhalten von Kennziffern bei Lineartransformation, Linearkombinationen etc. gelten auch für die entsprechenden Kennziffern von Wahrscheinlichkeitsverteilungen.

Für das Gesamt-Korrekturglied C legen wir die Formel $C=S^2/N$ zugrunde (wobei S die Summe aller Meßwerte aller Gruppen ist). Dann ergibt sich für den Erwartungswert von C:

$$E(C) = E(S^2/N) = (1/N) \cdot E(S^2) = (1/N) \cdot (N \cdot \sigma^2 + (\sum_g N_g \cdot \mu_g)^2) = \sigma^2 + N \cdot \mu^2.$$

Für das zweite Gleichheitszeichen verwenden wir H2 (mit $b=1/N$), für das dritte H5, und für das letzte die (als Definition zu verstehende) Beziehung $\mu := (\sum_g N_g \cdot \mu_g) / N$.

Ganz entsprechend gilt für $C_g = S_g^2 / N_g$ (also für das Korrekturglied der Gruppe g, wobei S_g die Summe der Meßwerte in Gruppe g ist):

$$E(C_g) = E(S_g^2 / N_g) = (1/N_g) \cdot E(S_g^2) = (1/N_g) \cdot (N_g \cdot \sigma^2 + (N_g \cdot \mu_g)^2) = \sigma^2 + N_g \cdot \mu_g^2.$$

Für SCG - die Summe der Korrekturglieder der Gruppen - gilt dann nach H3:

$$E(SCG) = E(\sum_g C_g) = \sum_g E(C_g) = G \cdot \sigma^2 + \sum_g N_g \cdot \mu_g^2.$$

Wie angekündigt, lassen sich nun (unter Verwendung der Differenzenregel in Hilfssatz H3) die Erwartungswerte der Quadratsummen leicht aus den Erwartungswerten der Korrekturglieder etc. herleiten:

$$E(QS_{in}) = E(RQ - SCG) = E(RQ) - E(SCG) = (N - G) \cdot \sigma^2.$$

$$E(QS_{zw}) = E(SCG - C) = E(SCG) - E(C) = (G - 1) \cdot \sigma^2 + \sum_g N_g \cdot \mu_g^2 - N \cdot \mu^2 = (G - 1) \cdot \sigma^2 + \sum_g N_g \cdot a_g^2.$$

Zum letzten Gleichheitszeichen: Aus der Definition $a_g := \mu_g - \mu$ ergibt sich

$$\begin{aligned} \sum_g N_g \cdot a_g^2 &= \sum_g N_g \cdot (\mu_g - \mu)^2 = \sum_g N_g \cdot (\mu_g^2 - 2\mu_g \mu + \mu^2) = \sum_g N_g \cdot \mu_g^2 - 2\mu \cdot \sum_g N_g \cdot \mu_g + \mu^2 \cdot \sum_g N_g \\ &= \sum_g N_g \cdot \mu_g^2 - N \cdot \mu^2. \end{aligned}$$

Beachte zum letzten Schritt: $\sum_g N_g \cdot \mu_g = N \cdot \mu$, und $\sum_g N_g = N$.

>>

II. Die Mehrweg-Varianzanalyse

1) Ein Experiment zur Zweiweg-Varianzanalyse

Es soll untersucht werden, welchen Einfluß die Arbeitsplatzgestaltung auf die Arbeit von Männern und Frauen hat. Eine bestimmte Arbeit wird unter drei Bedingungen durchgeführt:

Bedingung:

"nüchtern": Es handelt sich um saubere, aber nüchterne Arbeitsplätze.

"hübsch": Durch geschmackvolle Raumgestaltung (Vorhänge, Tapeten etc.) soll für

ein angenehmes Arbeitsklima gesorgt werden.

"altmodisch": Traditionelle Fabrikatmosphäre.

Unter diesen Bedingungen arbeiten Männer und Frauen getrennt. Gemessen wird als Erfolgsquote die "Zahl der erfolgreich bearbeiteten Stücke (x)". Es ergeben sich die in Tabelle 16 zusammengestellten Meßwerte (x).

Frauen			Männer		
Arbeitsbedingung			Arbeitsbedingung		
nüchtern	hübsch	altmodisch	nüchtern	hübsch	altmodisch
142	179	120	173	140	119
149	172	131	164	150	127
136	169	139	180	153	126
146	177	136	172	161	139
141	168	128	174	138	133
144	159	121	161	131	113
152	167	125	183	142	125
140	164	135	168	146	134
147	157	117	179	144	128
143	164	127	184	151	124
149	172	125	179	153	118
148	168	132	162	141	125
156	168		165	139	127
151	161		184	138	126
146	175		179	154	130
159			156	147	122
161			164	145	
154			173	150	
			168	153	
			184	145	
			169		
			176		
			166		
			165		
2664	2520	1536	4128	2920	2016

Tabelle 16: Ausgangsdaten (Unter dem Strich: Gruppensummen.)

2) "Faktoren" und "Stufen"

a) Begriffe

Wir haben insgesamt sechs Gruppen von Vpn. Es besteht aber ein wesentlicher Unterschied gegenüber der Einweg-Varianzanalyse: Die Gruppen unterscheiden sich unter zwei Aspekten: Nach der Arbeitsbedingung und nach dem Geschlecht.

Allgemein nennt man solche Aspekte, nach denen sich Gruppen von Meßwerten unterscheiden, auch Faktoren. In unserem Fall hätten wir also die zwei Faktoren "Arbeitsbedingung" und "Geschlecht". Die verschiedenen Ausprägungsstufen der einzelnen Faktoren nennt man auch "Stufen" des Faktors (auch "Niveaus", engl. "levels"). Der Faktor "Geschlecht" hat die Stufen "männlich" und "weiblich".

Welche Stufen hat der Faktor Arbeitsbedingung?

Die Stufen des Faktors "Arbeitsbedingung" sind "hübsch", "nüchtern" und "altmodisch".

Wieviele Faktoren und welche Stufen hatten wir in unserem Experiment zur Einweg-Varianzanalyse?

Wir hatten einen Faktor ("Vorinformation") mit den drei Stufen "überdurchschnittlich", "durchschnittlich" und "unterdurchschnittlich".

Eine Einweg-Varianzanalyse ist dann angebracht, wenn sich die Gruppen nur in einem Faktor unterscheiden. In unserem neuen Experiment wollen wir dagegen die Auswirkungen der beiden Faktoren getrennt betrachten.

Wir analysieren in zwei Richtungen. Daher spricht man auch von einer Zweiweg-Varianzanalyse.

In Tabelle @17 sind die Stichprobengrößen der verschiedenen Gruppen in unserem Experiment zusammengestellt.

	Frauen	Männer
nüchtern	$N_{11} = 18$	$N_{12} = 24$
hübsch	$N_{21} = 15$	$N_{22} = 20$
altmodisch	$N_{31} = 12$	$N_{32} = 16$

Tabelle 17: Stichprobengrößen

b) Die Unterscheidung von Zeilen und Spalten

Die ganze Struktur eines solchen Experiments wird durchsichtiger, wenn wir in Tabellen den einen Faktor mit den Zeilen (Reihen), den anderen Faktor mit den Spalten der Tabelle in Zusammenhang bringen. Den Stufen des einen Faktors (hier: Arbeitsbedingungen) entsprechen die verschiedenen Zeilen der Tabelle. Man nennt diesen Faktor dann den "Zeilenfaktor". Entsprechend nennt man den Faktor, dessen Stufen die Spalten der Tabelle entsprechen, den "Spaltenfaktor" (hier: Geschlecht). Natürlich ist es willkürlich, welchen Faktor man zum Zeilenfaktor und welchen man zum Spaltenfaktor macht. Man sollte sich aber einmal festlegen und in allen weiteren Tabellen eine entsprechende Unterteilung beibehalten.

c) Die Verwendung der Indices in der Zweiweg-Varianzanalyse

Wir sehen in Tabelle 17, daß die verschiedenen "Zellen" (das sind die Kreuzungsfelder der verschiedenen Zeilen und Spalten) nicht einfach durchnummeriert sind. Die verschiedenen N tragen Doppelindices. So ist N_{11} (lies: "N eins eins") die Stichprobengröße in der Zelle, die die Kreuzung der 1. Zeile und der 1. Spalte ist, also in der Gruppe der Frauen, die unter Arbeitsbedingung "nüchtern" auftreten. Der erste Index bezeichnet die Zeile, der zweite die Spalte der Tabelle 17. Diese Festlegung ist erforderlich, um N_{12} und N_{21} voneinander unterscheiden zu können.

Nehmen wir die verschiedenen Zellen einer Zeile zusammen und berechnen eine Gesamtstichprobengröße für diese Zeile, so setzen wir anstelle des Spaltenindex (also des zweiten) einen Punkt. $N_{1.}$ wäre also das Gesamt- N der ersten Zeile. Entsprechend bedeutet es eine Zusammenfassung der verschiedenen Zellen einer Spalte, wenn wir den Zeilenindex (also den ersten) durch einen Punkt ersetzen. So wäre $N_{.1}$ das Gesamt- N der ersten Spalte.

Was würde $N_{2.}$ bedeuten?

Gesamt- N in der 2. Zeile, also $N_{21} + N_{22}$

Was würde $N_{.2}$ bedeuten?

Gesamt- N in der 2. Spalte, also $N_{12} + N_{22} + N_{32}$

Was würde $N_{..}$ bedeuten?

Gesamte Stichprobengröße nach Zusammenfassung der Zellen aus allen Zeilen und Spalten, wofür wir auch einfach N schreiben.

Entsprechend verfahren wir später auch beim Durchschnitt (vgl. Tabelle 18): \bar{x}_{11} ist der Durchschnitt der Zelle "11", also der Durchschnitt aus allen Frauen, die unter der Bedingung "nüchtern" arbeiten. $\bar{x}_{1.}$ ist der Durchschnitt der Leistung in Zeile 1, $\bar{x}_{.1}$ der Durchschnitt der Leistung in Spalte 1 und $\bar{x}_{..}$ der Gesamtdurchschnitt, für den wir auch einfach \bar{x} schreiben.

3) Effekte in der Zweiweg-Varianzanalyse

a) Die Durchschnitte

Aus den Daten unseres Experiments ergeben sich die folgenden Durchschnitte:

	Frauen	Männer	Gesamt
nüchtern	$\bar{x}_{11} = 148$	$\bar{x}_{12} = 172$	$\bar{x}_{1.} = 161.71$
hübsch	$\bar{x}_{21} = 168$	$\bar{x}_{22} = 146$	$\bar{x}_{2.} = 155.43$
altmodisch	$\bar{x}_{31} = 128$	$\bar{x}_{32} = 126$	$\bar{x}_{3.} = 126.86$
Gesamt	$\bar{x}_{.1} = 149.33$	$\bar{x}_{.2} = 151.07$	$\bar{x} = 150.32$

Tabelle 18

Die Frauen, die unter Arbeitsbedingung "nüchtern" getestet wurden, erzielten also durchschnittlich 148 Punkte, usw.

Faßt man die Männer aus allen drei Arbeitsbedingungen zusammen zu einer Stichprobe, so ergibt sich hier ein Durchschnitt von 151.07 Punkten. Faßt man umgekehrt die Männer und Frauen unter der Bedingung "hübsch" zusammen, so ergibt sich ein Durchschnitt von 155.43 Punkten.

Warum ist dieser Durchschnitt nicht das Mittel von 168 und 146 (das wäre ja 157)?

Weil der Durchschnitt einer zusammengefaßten Gruppe normalerweise nur dann das Mittel der Einzeldurchschnitte ist, wenn die Einzelgruppen gleich groß sind.

Faßt man schließlich alle sechs Gruppen zusammen, so ergibt sich ein Durchschnitt von 150.32.

b) Effekte des Zeilenfaktors

Wir können nun - genau wie bei der Einweg-Varianzanalyse - Effekte schätzen. Dort war ja jeder Effekt definiert als die Abweichung eines zu einer bestimmten Stufe gehörenden Populationsmittels μ_g vom Populationsmittel μ :

$$a_g = \mu_g - \mu$$

Wir schätzen die Effekte, indem wir die in einer Untersuchung erzielten Durchschnitte als Schätzungen für die entsprechenden Populationsmittel einsetzen:

$$\hat{a}_g = \bar{x}_g - \bar{x}$$

Ganz entsprechend verfahren wir hier.

Wie würde man z.B. a_h , also den Effekt der Arbeitsbedingung "hübsch", schätzen und verbal beschreiben?

$$\hat{a}_h = \bar{x}_h - \bar{x} = 155.43 - 150.32 = +5.11$$

In Worten: Wir schätzen, daß der Effekt der Arbeitsbedingung "hübsch" in einer Hebung der Erfolgsquote um durchschnittlich 5.11 Einheiten besteht. (Das Wort durchschnittlich fügen wir ein, um eventuellen differentiellen Effekten Rechnung zu tragen.)

Schätzen und beschreiben Sie in gleicher Weise den Effekt der beiden anderen Arbeitsbedingungen!

$$\hat{a}_n = \bar{x}_n - \bar{x} = 161.71 - 150.32 = 11.39$$

Wir schätzen, daß der Effekt der Arbeitsbedingung "nüchtern" in einer Hebung der

Erfolgsquote um durchschnittlich 11.39 Einheiten besteht.

$$\hat{a}_a = \bar{x}_a - \bar{x} = 126.86 - 150.32 = -23.46$$

Wir schätzen, daß der Effekt der Arbeitsbedingung "altmodisch" in einer Senkung (negatives Vorzeichen) der Erfolgsquote um durchschnittlich 23.46 Einheiten besteht.

Diese Effekte des Zeilenfaktors nennt man auch einfach die Zeileneffekte. Eine allgemeine Formel für die Schätzung eines Zeileneffekts wäre:

$$\hat{a}_r = \text{Zeilenmittel} - \text{Gesamtmittel} = \bar{x}_r - \bar{x}$$

Dabei wird (in diesem Skript) als "allgemeiner Index" für die Zeile (Reihe) der Buchstabe r verwendet. Dies geschieht (obschon das Wort Zeile in solchem Zusammenhang geläufiger ist), damit wir später den Index z frei haben, um Zellen zu bezeichnen.

Die obigen Werte sind natürlich nur Schätzungen. Den wahren Effekt würde man erhalten, wenn man statt der Stichprobendurchschnitte die entsprechenden Populationsdurchschnitte einsetzen würde:

$$a_r = \mu_r - \mu$$

c) Effekte des Spaltenfaktors

Um die Effekte des Spaltenfaktors (die Spalteneffekte) deutlich von denen des Zeilenfaktors unterscheiden zu können, verwenden wir hier den Buchstaben b.

Wie würden Sie den Effekt des Geschlechts "weiblich" schätzen und in Worte fassen?

$$\hat{b}_w = \bar{x}_w - \bar{x} = 149.33 - 150.32 = -0.99$$

Der Effekt des Geschlechts weiblich besteht schätzungsweise in einer Senkung der Erfolgsquote um durchschnittlich 0.99 Punkte.

Dasselbe für männlich?

$$\hat{b}_m = \bar{x}_m - \bar{x} = 151.07 - 150.32 = +0.75$$

Den Effekt des Geschlechts männlich schätzen wir als Hebung um durchschnittlich 0.75 Punkte.

Könnten Sie eine allgemeine Formel für den Effekt der Spalte s und seine Schätzung angeben?

$$b_s = \mu_{.s} - \mu$$

$$\hat{b}_s = \bar{x}_{.s} - \bar{x}$$

d) Wechselwirkung

In Abschnitt c haben wir nur sehr schwache Effekte des Faktors "Geschlecht" festgestellt. Wir berücksichtigten dabei aber nur Gesamtdurchschnitte von Männern und Frauen. Betrachten wir aber die Tabelle der Durchschnitte (Tabelle 18), dann sehen wir eine weitere Auswirkung des

Geschlechts, die sich in den Gesamtdurchschnitten nicht niederschlägt: Die Auswirkungen des Faktors "Geschlecht" sind auf den verschiedenen Stufen des Faktors "Arbeitsbedingung" verschieden:

Unter nüchternen Arbeitsbedingungen sind die Männer besser, unter hübschen Arbeitsbedingungen die Frauen, während unter altmodischen Bedingungen beide fast gleich gut sind. Man spricht hier von einer Wechselwirkung.

Allgemein spricht man in der Varianzanalyse von einer Wechselwirkung, wenn die Effekte des einen Faktors auf den verschiedenen Stufen des anderen Faktors nicht gleich sind.

Wir könnten den gleichen Sachverhalt in unserem Experiment auch andersherum betrachten: Auf den verschiedenen Stufen des Faktors "Geschlecht" ist die Auswirkung des Faktors "Arbeitsbedingung" verschieden:

Bei den Männern ist die Bedingung "nüchtern" die günstigste, bei den Frauen die Bedingung "hübsch".

Allgemein: Jede Wechselwirkung läßt sich aus der Perspektive beider Faktoren betrachten.

Nehmen wir einmal an, wir hätten im Faktor "Arbeitsbedingung" nur die zwei Stufen "hübsch" und "altmodisch" gehabt und die gleichen Durchschnitte erzielt.

Wäre dann noch eine Wechselwirkung vorhanden?

Zunächst würde man versucht sein, die Existenz einer Wechselwirkung zu leugnen: Unter beiden Arbeitsbedingungen sind die Frauen besser als die Männer. Aber das Ausmaß dieses Unterschieds ist verschieden groß. Die Effekte des Faktors "Geschlecht" haben also auf beiden Stufen des Faktors "Arbeitsbedingung" zwar das gleiche Vorzeichen (+ für Frauen; - für Männer), aber einen verschiedenen Betrag. Es würde also auch hier eine Wechselwirkung vorliegen. - Wie schon erwähnt, kann man jede Wechselwirkung aus der Perspektive beider Faktoren betrachten.

Wie sähe die zweite Perspektive hier aus?

Bei beiden Geschlechtern ist zwar die Arbeitsbedingung "altmodisch" ungünstiger als die Arbeitsbedingung "hübsch", aber das Ausmaß dieses Effekts ist verschieden groß.

Allgemein: Von einer Wechselwirkung spricht man auch dann, wenn sich die Effekte des einen Faktors auf den verschiedenen Stufen des anderen Faktors nicht dem Vorzeichen, sondern nur dem Betrag nach unterscheiden.

Wir wollen zu unserem Versuchsplan mit den drei Stufen des Faktors "Arbeitsbedingung" zurückkehren und uns fragen, wie die Tabelle aussähe, wenn es keine Wechselwirkung gäbe. Dann müßte es *einen* Spalteneffekt b_w und *einen* Spalteneffekt b_m geben, die man auf allen Stufen

des Faktors "Arbeitsbedingung" zu dem jeweiligen Zeilendurchschnitt addiert, um zum Zellendurchschnitt zu kommen. Die Zeilendurchschnitte ergeben sich aber, indem man den jeweiligen Zeileneffekt zum Gesamtdurchschnitt \bar{x} addiert. (Überzeugen Sie sich davon anhand der Definitionsformel für \hat{a}_r !)

Zusammengefaßt ergäbe sich:

$$\mu_z = \mu + a_r + b_s$$

mit

μ_z = Populationsmittel, das zur Zelle z gehört

In Worten: Wenn es in der Population keine Wechselwirkung gäbe, würde sich jedes Zellenmittel ergeben, indem man den zugehörigen Zeileneffekt und den zugehörigen Spalteneffekt zum Gesamtmittel addiert.

Diese Feststellung läßt sich nicht beweisen. Sie ist als Definition zu verstehen: Eine Wechselwirkung liegt definitionsgemäß dann vor, wenn es mindestens eine Zelle gibt, bei der die obige Formel für eine Situation ohne Wechselwirkung nicht gilt.

Diese Formel, die für das Populationsmittel μ_z geschrieben ist und die in der Population gültigen Effekte a_r und b_s zugrundelegt, könnten wir analog auf die Verhältnisse in der Stichprobe übertragen:

Wenn es in unseren Daten keine Wechselwirkung gäbe, müßte gelten:

$$\bar{x}_z = \bar{x} + \hat{a}_r + \hat{b}_s$$

Davon ausgehend können wir nun einen numerischen Wert für den Wechselwirkungseffekt angeben, der in einer bestimmten Zelle, also durch die Kombination einer bestimmten Stufe des Zeilenfaktors und einer bestimmten Stufe des Spaltenfaktors wirksam wird.

Wir fragen uns einfach:

Um wieviel liegt der Durchschnitt jeder einzelnen Zelle oberhalb bzw. unterhalb des Wertes, der sich bei Additivität der zugehörigen Spalten- und Zeileneffekte ergeben würde? Hierzu bilden wir die entsprechende Differenz:

$$\hat{c}_z = \bar{x}_z - (\bar{x} + \hat{a}_r + \hat{b}_s)$$

Der nach dieser Formel ermittelte Wert \hat{c}_z ist der geschätzte Wechselwirkungseffekt der zugehörigen Zelle.

Beispiel aus unserem Experiment:

$$\begin{aligned}\hat{c}_{11} &= \bar{x}_{11} - (\bar{x} + \hat{a}_1 + \hat{b}_1) \\ &= 148 - [150.32 + 11.39 + (-0.99)] \\ &= 148 - 160.72 \\ &= -12.72\end{aligned}$$

In Worten: Die Frauen sind unter der Arbeitsbedingung "nüchtern" um 12.72 Punkte schlechter, als es ohne Wechselwirkung zu erwarten wäre.

Übungsaufgabe: Schätzen und verbalisieren Sie die Effekte der Wechselwirkung für weitere Zellen in ähnlicher Weise.

Im Unterschied zu diesen Wechselwirkungseffekten, die sich aus der Kombination einer

bestimmten Zeile mit einer bestimmten Spalte ergeben, nennt man die in den vorhergehenden Abschnitten behandelten Effekte der Zeilen und Spalten auch Haupteffekte.

Was mit einer Wechselwirkung gemeint ist, läßt sich auch gut an der Darstellung der Mittelwerte in Abbildung 6 sehen:

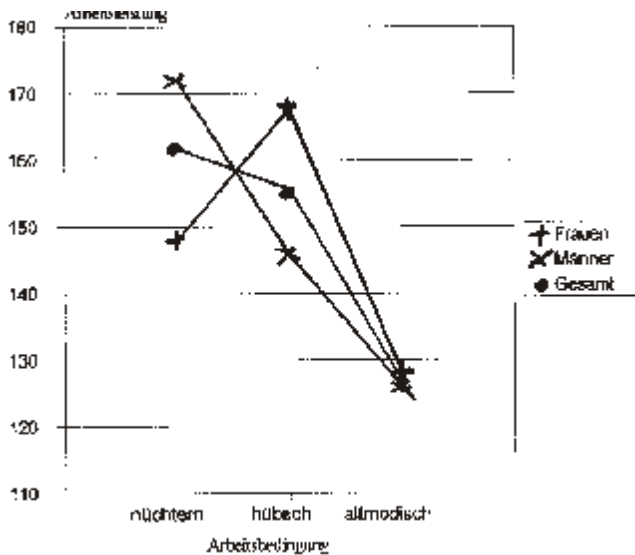


Abb. 6: Wechselwirkungsdiagramm

Die Wechselwirkung besteht - wie wir gesehen haben - darin, daß der Effekt des Geschlechts auf den verschiedenen Stufen des Faktors "Arbeitsbedingung" unterschiedlich ist. In unserer Darstellung zeigt sich das darin, daß die Kurven nicht parallel verlaufen. Wenn nämlich der Effekt des Geschlechts unter allen Arbeitsbedingungen gleich wäre, dann müßte die Kurve der Männer überall um den gleichen Betrag oberhalb der "Gesamt"-Kurve verlaufen. Beide müßten also parallel zur "Gesamt"-Kurve und damit auch parallel zueinander verlaufen.

Allgemein: Was bei einer Wechselwirkung genau vorliegt, kann man sich verdeutlichen, indem man eine solche graphische Darstellung anlegt und feststellt, in welcher Weise die Kurven von einem parallelen Verlauf abweichen.

Wir hatten vorhin festgestellt, daß eine Wechselwirkung auch dann noch vorläge, wenn wir nur die Arbeitsbedingungen "hübsch" und "altmodisch" gehabt hätten und die gleichen Durchschnitte erhalten hätten. Die Besonderheit bestand darin, daß hier zwar unter beiden Arbeitsbedingungen die Frauen besser und andererseits bei beiden Geschlechtern die Leistung unter Arbeitsbedingung "hübsch" besser als unter "altmodisch" ist; aber der Betrag dieser Effekte ist jeweils verschieden. Auch diese Art von Wechselwirkung zeigt sich in der graphischen

Darstellung: Von "hübsch" nach "altmodisch" fallen zwar alle Kurven ab, verlaufen aber nicht parallel.

Es ist wichtig drei verschiedene Arten von Effekten zu unterscheiden:

1. Den Effekt, der insgesamt mit einer Stufe eines Faktors verbunden ist (sogenannter *Haupteffekt*).

Beispiele: Der Effekt der Arbeitsbedingung "nüchtern" (S. @).

Der Effekts des Geschlechts "weiblich" (S. @).

Ein solcher Effekt gibt an, um welchen Betrag der zu dieser Stufe gehörende Durchschnitt nach Zusammenfassung aller Stufen des anderen Faktors vom Gesamtdurchschnitt abweicht.

2. Der Effekt, der mit einer Stufe eines Faktors auf einer bestimmten Stufe eines anderen Faktors verbunden ist.

Beispiele: Der Effekt der Arbeitsbedingung "nüchtern" bei den Männern beträgt 19.93 Punkte. Auf der Stufe "männlich" liegt nämlich der Durchschnitt unter Arbeitsbedingung "nüchtern" (172) um 19.93 Punkte über dem Gesamtdurchschnitt dieser Stufe (151.07).

Der Effekt des Geschlechts "weiblich" unter der Arbeitsbedingung "hübsch" beträgt 12.57. Auf der Stufe "hübsch" liegt nämlich der Durchschnitt der Frauen (168) um 12.57 Punkte über dem Gesamtdurchschnitt dieser Stufe (155.43).

Effekte dieser Art haben in der Varianzanalyse keinen besonderen Namen. Wir haben sie hier nur eingeführt, weil sich an der Unterschiedlichkeit solcher Effekte am ehesten aufzeigen läßt, was eine Wechselwirkung ist. Solche Effekte selbst sind aber noch keine Wechselwirkungseffekte. Diese Wechselwirkungseffekte sind folgendermaßen definiert:

3. Effekte, die in der Abweichung eines Zellendurchschnitts von demjenigen Wert bestehen, der aufgrund der Haupteffekte (d.h. aufgrund der Zeilen- und Spalten-Effekte) allein zu erwarten wäre (*Wechselwirkungseffekt einer Zelle*).

Beispiel (vgl. Tabelle 18): Für die Zelle "Frauen" unter Arbeitsbedingung "nüchtern" würden wir allein aufgrund der Haupteffekte der Arbeitsbedingung "nüchtern" und des Geschlechts "weiblich" einen Durchschnitt $\bar{x} + \hat{a}_n + \hat{b}_w = 150 + 11.39 + (-0.99) = 160.72$ erwarten. Der tatsächliche Zellendurchschnitt liegt aber bei 148 und damit um 12.72 Punkte unterhalb des Werts, der aufgrund der Haupteffekte zu erwarten wäre. Also beträgt der Wechselwirkungseffekt für diese Zelle -12.72.

Das kann man auch folgendermaßen formulieren: Der Wechselwirkungseffekt ist derjenige Effekt, der von einer Kombination zweier Stufen (z.B. von der Kombination der Arbeitsbedingung "nüchtern" mit dem Geschlecht "weiblich") zusätzlich zu den zugehörigen "Haupteffekten" ausgeht.

<< *Allgemeiner Begriff der Wechselwirkung*

Der hier behandelte Begriff der Wechselwirkung in der Varianzanalyse ist ein Spezialfall eines allgemeinen Begriffs der Wechselwirkung. Allgemein spricht man von einer Wechselwirkung, wenn die Beziehung zwischen zwei oder mehreren Variablen sich ändert in

Abhängigkeit von den Werten, die eine dritte Variable annimmt.

Ein weiteres Beispiel für eine Wechselwirkung - außerhalb der Varianzanalyse: Empirische Untersuchungen haben ergeben, daß (zumindest in einigen Schultypen) die korrelative Beziehung zwischen dem Interesse der Schüler und ihren Leistungen um so enger ist, je weniger starr, pedantisch ("rigide") die Schüler sind.⁹⁵ Die Beziehung zwischen den Variablen "Interesse" und "Leistung" ändert sich also in Abhängigkeit von Werten, die die Variable "Rigidität" annimmt.

>>

e) Allgemeines Modell der Zweiweg-Varianzanalyse

Wir haben bisher definiert:

$$\mu_r - \mu = a_r$$

$$\mu_s - \mu = b_s$$

$$\mu_z - (\mu + a_r + b_s) = c_z$$

Außerdem gibt es auch in der Zweiweg-Varianzanalyse einen "Fehler". Er ist auch hier definiert als die Abweichung eines Meßwerts vom zugehörigen Gruppen- (also Zellen-) Mittel:

$$x_i - \mu_z = e_i$$

Setzen wir nun für a_r und b_s in der dritten Gleichung die in der ersten bzw. zweiten Gleichung angegebenen Ausdrücke ein und addieren dann die vier Gleichungen, so erhalten wir auf der linken Seite:

$$\mu_r - \mu + \mu_s - \mu + \mu_z - (\mu + \mu_r - \mu + \mu_s - \mu) + x_i - \mu_z$$

Wenn wir diesen Ausdruck ausrechnen (wobei sich viele Glieder gegenseitig aufheben) und daneben die Summe der rechten Seite unserer vier Gleichungen schreiben, so erhalten wir:

$$x_i - \mu = a_r + b_s + c_z + e_i$$

In Worten: Die Abweichung eines Meßwerts x_i vom Gesamtmittel μ setzt sich aus vier Komponenten zusammen: Dem Effekt der zugehörigen Zeile (Reihe), dem Effekt der zugehörigen Spalte, dem Wechselwirkungseffekt der Zelle und dem "Fehler". Diese Abweichungen der Meßwerte x_i vom Gesamtmittel μ machen aber die Varianz der x -Werte aus (die Gesamtvarianz ist ja die durchschnittliche quadrierte Abweichung der Meßwerte vom Gesamtmittel).

Entsprechend lösen wir in der Zweiweg-Varianzanalyse dann auch die Varianz in vier Komponenten auf: In die Varianz zwischen den Zeilen, die Varianz zwischen den Spalten, die auf Wechselwirkung zurückzuführende Varianz und die Varianz innerhalb der Zellen ("Fehlervarianz").

Addiert man auf beiden Seiten der letzten Gleichung den Wert μ , so erhält man:

⁹⁵Psychologische Interpretation: Die allgemeine Erwartung, daß Interesse und Leistung eng zusammenhängen, trifft bei rigiden Schülern weniger zu, weil sie (vielleicht aufgrund eines starken "Über-Ich"?) ihren Arbeitseinsatz weniger von Interessen abhängig machen und weil sie andererseits weniger dazu tendieren, ihre Interessen zu verändern, wenn Leistungen und Interessen nicht übereinstimmen - eben weil sie starr, rigide sind.

$$x_i = \mu + a_{r(i)} + b_{s(i)} + c_{z(i)} + e_i.$$

Dabei sind $r(i)$, $s(i)$ und $z(i)$ die Reihe, die Spalte und die Zelle, zu denen der Meßwert x_i gehört.

Genau so wie in der Einweg-Varianzanalyse ist auch hier die Modellgleichung keine Zusatzvoraussetzung für die Tests; sie ergibt sich aus der Definition der Effekte. (Wollen Sie es als Übung selbst herleiten?)

4) Die Zusatzvoraussetzungen der Zweifweg-Varianzanalyse

a) Unabhängige Stichproben

(Darauf wird in einem späteren Abschnitt genauer eingegangen).

b) Normalverteilung und Varianzhomogenität

Die den verschiedenen Zellen entsprechenden Populationverteilungen müssen alle *Normalverteilungen* sein und *gleiche Varianz* haben. Die Normalitätsannahme und die Homogenitätsannahme verlieren an Gewicht bei großen bzw. gleich großen Stichproben.

c) Orthogonalität

Die sogenannte *Orthogonalität*. Zur Erläuterung des damit Gemeinten betrachten wir die Tabelle 17. Es gilt die Proportion:

$$N_{11} : N_{12} = N_{21} : N_{22} = N_{31} : N_{32} = N_{.1} : N_{.2}$$

In Zahlen:

$$18 : 24 = 15 : 20 = 12 : 16 = 45 : 60$$

Allgemein muß gelten:

$$N_{11} : N_{12} : N_{13} : \dots = N_{21} : N_{22} : N_{23} : \dots = N_{31} : N_{32} : N_{33} : \dots = \text{usw.} = N_{.1} : N_{.2} : N_{.3} : \dots$$

In Worte gefaßt: Innerhalb jeder Zeile muß das Verhältnis der Besetzungshäufigkeiten der Zellen gleich sein.

Bisher haben wir den Begriff der Orthogonalität auf einen Vergleich der Häufigkeitsverhältnisse innerhalb der verschiedenen Zeilen gestützt. Genauso kann man aber auch die

Häufigkeitsverhältnisse innerhalb der Spalten zugrundelegen. In unserem Beispiel gilt

$$N_{11} : N_{21} : N_{31} = N_{12} : N_{22} : N_{32} = N_{1.} : N_{2.} : N_{3.},$$

bzw. in Zahlen

$$18 : 15 : 12 = 24 : 20 : 16 = 42 : 35 : 28.$$

In Worte gefaßt: Auch innerhalb jeder Spalte muß das Verhältnis der Besetzungshäufigkeiten der Zellen gleich sein.

Es genügt aber, diese Proportionalität in einer der beiden Richtungen zu überprüfen; denn wenn sie in der einen Richtung zutrifft, gilt sie auch für die andere.

Eine derartige Proportionalität von Häufigkeiten kennen wir schon als Nullhypothese des χ^2 -Tests für stochastische Unabhängigkeit. Der Unterschied läßt sich folgendermaßen charakterisieren:

- Beim χ^2 -Test stellen die Häufigkeiten, für die lt. H_0 diese Proportionalität gelten soll, die eigentlichen Daten dar. Ob die Proportionalität tatsächlich zutrifft, ist die eigentliche empirische Frage.
- In der Zweiweg-Varianzanalyse ist die Proportionalität als Zusatzvoraussetzung durch den Versuchsplan zu gewährleisten. Die empirische Fragestellung bezieht sich dann auf die Effekte der Zugehörigkeit zu den Gruppen auf Meßwerte in einer völlig andere Variable (in unserem Fall die in Tabelle @ angegebenen Leistungsmaße).

Wenn man diesen Unterschied einmal klargestellt hat, kann man die Orthogonalitäts-Forderung auch folgendermaßen formulieren:

Allgemeine Beschreibung: Die Zusatzvoraussetzung der Orthogonalität verlangt, daß für die Besetzungshäufigkeiten der Zellen die gleichen Proportionalitäten gelten wie für die f_e -Werte im χ^2 -Test für stochastische Unabhängigkeit.

Auch eine verkürzte Formeldarstellung ergibt sich aus diesem Vergleich. Die allgemeine Formel für f_e -Werte (vgl. @) läßt sich folgendermaßen übertragen: Für jede Zeile (Reihe) r und jede Spalte s muß für die Besetzungshäufigkeit N_{rs} der von dieser Zeile und dieser Spalte gebildeten Zelle gelten:

$$N_{rs} = N_{r.} \cdot N_{.s} / N.$$

<<

Eine Begründung für die Orthogonalitäts-Forderung ergibt sich aus Interpretationsschwierigkeiten in Fällen, in denen die Orthogonalität verletzt ist. Nehmen wir z.B. an, wir hätten unser

bisher betrachtetes Experiment ohne die Arbeitsbedingung "altmodisch" und mit folgenden Zell-Besetzungen durchgeführt:

$$N_{11} = 10; N_{12} = 10; N_{21} = 10; N_{22} = 40.$$

Falls wir die gleichen Zellendurchschnitte erhalten hätten (vgl. Tabelle 19), dann ließen sich die Gesamtdurchschnitte für Zeilen und Spalten aus den Zellendurchschnitten und den N-Werten der Zellen nach der Formel für Durchschnitte zusammengefaßter Gruppen (S. @) berechnen:

	Frauen	Männer	Gesamt
nüchtern	$\bar{x}_{11} = 148$	$\bar{x}_{12} = 172$	$\bar{x}_{.1} = 160$
hübsch	$\bar{x}_{21} = 168$	$\bar{x}_{22} = 146$	$\bar{x}_{.2} = 150.4$
gesamt	$\bar{x}_{.1} = 158$	$\bar{x}_{.2} = 151.2$	$\bar{x} = 153.1$

Tabelle 19: Zellenmittelwerte des "nicht-orthogonalen Versuchsplans"

Übungsaufgabe: Überprüfen Sie die Richtigkeit!

Wir sehen: Die "Gesamt"-Durchschnitte haben sich geändert, da sich die "relativen Stichprobengrößen" geändert haben und da die dritte Arbeitsbedingung ("altmodisch") fehlt. Entsprechend haben sich auch die Effekte geändert.

Allgemein: Die Höhe der Effekte hängt vom Verhältnis der Stichprobengrößen in den Zellen ab und davon, welche weiteren Stufen der einzelnen Faktoren vertreten sind.

Für uns ist dabei aber eines vor allem interessant: Wir finden jetzt Unterschiede zwischen den Zeilen und zwischen den Spalten, die vorher nicht in diesem Ausmaß da waren: Unter Arbeitsbedingung "hübsch" wird bemerkenswert schlechter gearbeitet als unter Arbeitsbedingung "nüchtern", und die Männer sind im Durchschnitt schlechter als die Frauen. Aber es gibt die Möglichkeit, daß einer der beiden Effekte nur eine sekundäre Konsequenz des anderen ist: Es wäre denkbar, daß unter Arbeitsbedingung "hübsch" nur deshalb schlechter gearbeitet wird, weil 40 der 50 Vpn, die unter dieser Bedingung getestet wurde, Männer sind (die laut Spaltendurchschnitt schlechter sind), während unter Arbeitsbedingung "nüchtern" gleich viele Männer wie Frauen arbeiten. Anders formuliert: Der unter Arbeitsbedingung "hübsch" besonders hohe Anteil von Männern (80% gegenüber 50%) hat den Durchschnitt dieser Arbeitsbedingung gesenkt. Es wäre denkbar, daß sonst gar kein Unterschied zwischen den Zeilen bestünde.

Daneben wäre auch die folgende Interpretation denkbar: In Wirklichkeit bestehen keine Unterschiede zwischen Männern und Frauen. Von den 50 Männern haben aber 40 unter der - laut Zeilendurchschnitt - ungünstigeren Arbeitsbedingung "hübsch" gearbeitet. Also: Die Männer sind vielleicht nur deshalb im Durchschnitt schlechter, weil sie vor allem unter der ungünstigeren Arbeitsbedingung antraten.

Schließlich besteht eine dritte Möglichkeit:

In Wirklichkeit bestehen keine Zeilen- und keine Spalteneffekte, sondern nur die Wechselwirkung. Da aber die Arbeitsbedingung "hübsch" für die Männer die ungünstigere ist und da gerade diese Zelle besonders stark besetzt (mit 40 Vpn im Gegensatz zu je 10 Vpn bei den übrigen Zellen), wird der Zeilendurchschnitt von "hübsch" und der Spaltendurchschnitt der Männer künstlich gesenkt.

Allgemein: Ist die Orthogonalitätsvoraussetzung verletzt, so kann es vorkommen, daß der eine Effekt die Existenz (oder auch das Fehlen) des anderen vortäuscht. Es läßt sich mathematisch beweisen, daß so etwas nicht vorkommt, wenn die Orthogonalitätsvoraussetzung erfüllt ist.

Es gibt spezielle Verfahren, mit denen man eine Zweiweg-Varianzanalyse auch bei

Verletzung der Orthogonalitätsvoraussetzung durchführen kann. Die Ergebnisse derartiger Varianzanalysen sind aber schwer zu interpretieren. Deshalb ist es besser, schon im Stadium der Versuchsplanung für Orthogonalität zu sorgen, soweit das möglich ist.

>>

d) Eine letzte Voraussetzung

<<

Eine letzte Voraussetzung der Zweiweg-Varianzanalyse in der hier vorgestellten Form ist es, daß es sich nicht um ein sogenanntes "Modell mit Zufallsfaktoren" und auch nicht um ein "gemischtes Modell" handelt. Was darunter zu verstehen ist, wird in einem späteren Abschnitt (9 cb) behandelt.

>>

5) Grundlagen und Durchführung der Berechnungen in der Zweiweg-Varianzanalyse.

a) Die Zerlegung der Abweichungen vom Gesamtdurchschnitt

Wir können die 6 Gruppen (Zellen) unseres Versuchsplans zunächst einmal so betrachten, als handele es sich um 6 Gruppen in einer Einweg- Varianzanalyse. Dann läßt sich jede Abweichung eines Meßwerts vom Gesamtdurchschnitt zerlegen in eine Abweichung des Meßwerts vom zugehörigen Zellendurchschnitt und in die Abweichung des Zellendurchschnitts vom Gesamtdurchschnitt. Für unsere erste Vp heiße das z.B.:

$$\begin{array}{rcll} 142 - 150.32 & = & (142 - 148) & + & (148 - 150.32) \\ \text{Abweichung eines} & & \text{Abweichung des Meßwerts} & & \text{Abweichung des} \\ \text{Meßwerts vom} & = & \text{vom zugehörigen} & + & \text{Zellendurchschnitts vom} \\ \text{Gesamtdurchschnitt} & & \text{Zellendurchschnitt} & & \text{Gesamtdurchschnitt} \end{array}$$

Das wesentlich Neue bei der Zweiwegvarianzanalyse ist nun, daß wir die Abweichung eines Zellendurchschnitts vom Gesamtdurchschnitt weiter zerlegen können in 3 Komponenten:

- Die Abweichung des zugehörigen Zeilendurchschnitts vom Gesamtdurchschnitt (und das ist die Schätzung des Zeileneffekts),
- die Abweichung des zugehörigen Spaltendurchschnitts vom Gesamtdurchschnitt (=Schätzung des Spalteneffekts) und in
- die Abweichung des Zellendurchschnitts von dem Wert, der ohne Wechselwirkung allein aufgrund des zugehörigen Zeilen- und Spalteneffekts zu erwarten wäre (=Schätzung des Wechselwirkungseffekts).

Da wir auf diese Komponenten im folgenden häufiger zurückgreifen, ist es umständlich, immer von "Schätzung des ...-Effekts" zu sprechen. Andererseits sind aber diese Effekte aufgrund

der wahren Durchschnitte definiert und nicht aufgrund der Stichprobendurchschnitte. Um Verwechslungen zu vermeiden, wollen wir vereinbaren: Wenn wir den Namen eines Effekts in Apostrophe (') einschließen, dann ist jeweils der aufgrund der Stichprobendurchschnitte geschätzte Effekt gemeint. Und um auch für die Abweichung eines Zelldurchschnitts vom Gesamtdurchschnitt einen kurzen Namen zu haben, nennen wir sie sinngemäß 'Zelleneffekt'.

Dann können wir die obige Zerlegung des 'Zelleneffekts' z.B. für die Gruppe "Frauen unter Arbeitsbedingung nüchtern" folgendermaßen hinschreiben:

$$(148-150.32) = (161.71-150.32) + (149.33-150.32) + (148-160.72)$$

$$\text{'Zelleneffekt'} = \text{'Zeileneffekt'} + \text{'Spalteneffekt'} + \text{'Wechselwirkungseffekt'}$$

Dabei ist 160.72 der Wert, der sich ohne Wechselwirkung allein aufgrund des 'Zeileneffekts' und des 'Spalteneffekts' als Durchschnitt für Frauen unter Arbeitsbedingung nüchtern ergeben müßte. Die Berechnung dieses Werts (vgl. S. @) zeigt, warum diese Zerlegung gültig ist: Er ergibt sich dadurch, daß zum Gesamtdurchschnitt der 'Zeileneffekt' und der 'Spalteneffekt' addiert wird. Die obige Gleichung läßt sich dann auch folgendermaßen schreiben:

$$\begin{aligned} & \text{Zelldurchschnitt - Gesamtdurchschnitt} \\ & = \text{'Zeileneffekt'} + \text{'Spalteneffekt'} \\ & + (\text{Zelldurchschnitt} - (\text{Gesamtdurchschnitt} + \text{'Zeileneff.'} + \text{'Spalteneff.'})) \end{aligned}$$

Und daß das nicht nur in unserem Fall, sondern ganz allgemein zutrifft, zeigt sich sofort, wenn man die Klammern in der letzten Zeile auflöst.

b) Die Zerlegung der Abweichungsquadratsumme

Zunächst können wir wieder - wie bei der Einweg-Varianzanalyse - die Quadratsumme der Abweichungen aller Meßwerte vom Gesamtdurchschnitt zerlegen in eine Quadratsumme innerhalb der Gruppe und eine Quadratsumme zwischen den Gruppen (Zellen). Letztere beruht auf den 'Zelleneffekten' und kann deshalb - ähnlich wie die 'Zelleneffekte' in 3 Komponenten zerlegt werden:

- Die Quadratsumme für Zeileneffekte ergibt sich, indem für jeden Meßwert der zugehörige 'Zeileneffekt' quadriert und aufsummiert wird.
- Die Quadratsumme für Spalteneffekte ergibt sich, indem für jeden Meßwert der zugehörige 'Spalteneffekt' quadriert und aufsummiert wird.
- Die Quadratsumme für Wechselwirkungseffekte ergibt sich, indem für jeden Meßwert der zugehörige 'Wechselwirkungseffekt' quadriert und aufsummiert wird.

Diese 3 Quadratsummen zusammen ergeben die Quadratsumme zwischen den Gruppen. Als Formel:

$$QS_{zw} = QS_{Zeilen} + QS_{Spalten} + QS_{Wechselwirkung}$$

<< Daß das so funktioniert, kann man ganz ähnlich wie bei der Zerlegung der totalen Quadratsumme in der Einwegvarianzanalyse zeigen. Wenn man für einen einzelnen Wert die Summe

'Zeileneffekt' + 'Spalteneffekt' + 'Wechselwirkungseffekt'

quadriert, dann ergeben sich nach der Formel

$$(a+b+c)^2 = a^2 + b^2 + c^2 + 2ab + 2ac + 2bc$$

außer den Quadraten der 'Effekte' auch doppelte Produkte, die sich aber bei einer Summation über alle Vpn aufheben. Was bei einer Summation über alle Vpn übrigbleibt (nach dem Muster der letzten 2 Zeilen von Tabelle V3, Seite @), sind die Summen der quadrierten 'Effekte', wobei natürlich für jede Vp die für sie zutreffenden 'Effekte' zu verrechnen sind. Und diese Summen sind die o.g. Quadratsummen.

>>

c) Vereinfachte Berechnung der Quadratsummen.

<<

Da im Grunde alles läuft wie bei der Einwegvarianzanalyse, gibt es auch vereinfachte Berechnungsmöglichkeiten für die Quadratsummen.

Wir bilden die folgenden Ausdrücke:

RQ, die "rohe Quadratsumme", ist wieder die Summe aller quadrierten Rohwerte. In unserem Beispiel:

$$RQ = 142^2 + 149^2 + \dots + 122^2 = 2409572.00$$

Zu jeder Zelle, jeder Zeile, jeder Spalte und für alle Meßwerte zusammen ("gesamt") bildet man ein Korrekturglied. Aus den bereits bei der Einweg-Varianzanalyse genannten Gründen berechnet man diese Korrekturglieder, indem man die entsprechende Meßwertsumme quadriert und dieses Quadrat durch die Zahl der zugrundeliegenden Meßwerte dividiert. Am einfachsten ist es, wenn man zunächst die Meßwertsummen für die Zellen berechnet (vgl. Tabelle S. @) und dann die Meßwertsummen für Zeilen, Spalten und "Gesamt" durch Zusammenfassung bildet.

Aus diesen Korrekturgliedern bildet man Summen, indem man "Korrekturglieder gleicher Art" zusammenfaßt:

SCR ist die Summe der zu den Zeilen ("Reihen") gebildeten Korrekturglieder:

$$SCR = (2664+4128)^2:42 + (2520+2920)^2:35 + (1536+2016)^2:28 = 2394491.47$$

SCS ist die Summe der zu den Spalten gebildeten Korrekturglieder:

$$SCS = (2664+2520+1536)^2:45 + (4128+2920+2016)^2:60 = 2372788.27$$

SCZ ist die Summe der zu den Zellen gebildeten Korrekturglieder:

$$SCZ = 2664^2:18 + 4128^2:24 + 2520^2:15 + 2920^2:20 + 1536^2:12 + 2016^2:16 = 2404592.00$$

C ist wieder das aufgrund der Gesamtsumme aller Meßwerte gebildete Korrekturglied:

$$C = (2664+4128+2520+2920+1536+2016)^2:105 = 2372711.01$$

Aus diesen verschiedenen Ausdrücken können wir nun die Quadratsummen berechnen:

$$QS_{tot} = RQ - C = 2409572.00 - 2372711.01 = 36860.99$$

QS_{tot} wird zerlegt in

$$QS_{in} = RQ - SCZ = 2409572.00 - 2404592.00 = 4980.00$$

$$QS_{zw} = SCZ - C = 2400592.00 - 2372711.01 = 31880.99$$

QS_{zw} wird weiter zerlegt in

$$QS_{Zeilen} = SCR - C = 2394491.47 - 2372711.01 = 21780.46$$

$$QS_{Spalten} = SCS - C = 2372788.27 - 2372711.01 = 77.26$$

$$\begin{aligned} QS_{Wechselw.} &= SCZ - SCR - SCS + C \\ &= 2400592.00 - 2394491.47 - 2372788.27 + 2372711.01 \\ &= 10023.27 \end{aligned}$$

Als Rechenprobe kann man überprüfen, ob

$$QS_{tot} = QS_{in} + QS_{zw}$$

und

$$QS_{zw} = QS_{Zeilen} + QS_{Spalten} + QS_{Wechselw.}$$

Man sollte sich aber darüber klar sein, daß Rechen- oder Rundungsfehler bei der Berechnung der den Quadratsummen zugrundeliegenden Ausdrücke RQ, SCR, SCS, SCZ und C von dieser Rechenprobe nicht bemerkt werden.

>>

d) Die Freiheitsgrade der Varianzquellen.

Wie in der Einweg-Variananalyse ist auch hier jeder Varianzquelle eine Freiheitsgradzahl zugeordnet. In den folgenden Formeln ist R die Zahl der Zeilen ("Reihen"), S die Zahl der Spalten und Z die Zahl der Zellen.

$$df_{tot} = N - 1 = 105 - 1 = 104$$

$$\begin{aligned}df_{in} &= N - Z = 105 - 6 = 99 \\df_{zw} &= Z - 1 = 6 - 1 = 5 \\df_{Zeilen} &= R - 1 = 3 - 1 = 2 \\df_{Spalten} &= S - 1 = 2 - 1 = 1 \\df_{Wechselw.} &= (R-1) \cdot (S-1) = (3-1) \cdot (2-1)=2\end{aligned}$$

Natürlich müssen wieder entsprechende Zerlegungen gelten wie für die Quadratsummen (Rechenprobe):

$$df_{tot} = df_{in} + df_{zw}$$

und

$$df_{zw} = df_{Zeilen} + df_{Spalten} + df_{Wechselw.}$$

Natürlich verläuft auch die Begründung für die Freiheitsgradzahlen ganz ähnlich wie in der Einwegvarianzanalyse. Für df_{tot} , df_{in} und df_{zw} können wir die Begründung für die entsprechenden Varianzquellen der Einwegvarianzanalyse unmittelbar übernehmen.

Für df_{Zeilen} können wir die Begründung für df_{zw} aus der Einweg- Varianzanalyse ebenfalls sinngemäß übernehmen. Nach den in Abschnitt FI4e entwickelten Prinzipien könnten wir aus den Zeilendurchschnitten den Gesamtdurchschnitt berechnen, und nach den gleichen Prinzipien gilt dann auch die Restriktion: Die mit den Vpn-Zahlen pro Zeile multiplizierten Zeileneffekte müssen zusammen 0 ergeben. Also sind nur R-1 Zeileneffekte "frei beweglich"; sind sie vorgegeben, dann ist der letzte festgelegt.

Entsprechendes gilt auch für die Spalten.

Für die Freiheitsgrade der Wechselwirkung ($df_{Wechselw.}$) können wir zunächst postulieren, daß sie zusammen mit df_{Zeilen} und $df_{Spalten}$ die Freiheitsgrade zwischen den Zellen (df_{zw}) ergeben müssen, und wie man leicht nachprüfen kann, tun sie dies dann, wenn sie nach der obigen Formel berechnet werden.

<< Wer es noch genauer wissen will, kann es auch aufgrund von Restriktionen überprüfen. Für die 'Wechselwirkungseffekte' gelten zwei Restriktionen - eine Zeilenrestriktion und eine Spaltenrestriktion.

Die Zeilenrestriktion lautet: Innerhalb einer jeden Zeile müssen die mit der jeweiligen Gruppengröße multiplizierten 'Wechselwirkungseffekte' der verschiedenen Zellen zusammen 0 ergeben. (Wem es wichtig ist, das nachzuprüfen, kann es versuchen oder den Beweis nachlesen, der diesem Abschnitt als Anhang beigefügt ist.)

Die Spaltenrestriktion lautet entsprechend: Innerhalb einer jeden Spalte müssen die mit der jeweiligen Gruppengröße multiplizierten 'Wechselwirkungseffekte' der verschiedenen Zellen zusammen 0 ergeben. (Der Beweis würde ganz entsprechend aussehen wie für die Zeilenrestriktion.)

Beide Restriktionen zusammen kann man aber auch folgendermaßen formulieren: Innerhalb der ersten Zeilen (außer der letzten) sind nur die 'Wechselwirkungseffekte' bis auf den letzten "frei beweglich"; der letzte in jeder Zeile ist dann wegen der Zeilenrestriktion durch die vorangehenden festgelegt. Und die 'Wechselwirkungseffekte' der letzten Zeile sind wegen der Spaltenrestriktion

festgelegt. Da wir aber insgesamt R Zeilen mit jeweils S 'Wechselwirkungseffekten' haben, können wir auch sagen: In R-1 Zeilen sind S-1 'Wechselwirkungseffekte' frei beweglich. Und daher ist die Freiheitsgradzahl für Wechselwirkung $(R-1) \cdot (S-1)$.

>>

<<

Anhang zu @d.: Beweis der Zeilenrestriktion für 'Wechselwirkungseffekte'

Es ist zu zeigen, daß innerhalb jeder Zeile die 'Wechselwirkungseffekte' nach Multiplikation mit der jeweiligen Gruppengröße zusammen 0 ergeben müssen. Sehen wir es uns für die erste Zeile unseres Versuchsplans an. Nach der Definition eines 'Wechselwirkungseffekts' gilt zunächst:

$$N_{11} \cdot \hat{c}_{11} = N_{11} \cdot (\bar{x}_{11} - (\bar{x} + \hat{a}_{\text{nüchtern}} + \hat{b}_{\text{Frauen}}))$$

$$N_{11} \cdot \hat{c}_{12} = N_{12} \cdot (\bar{x}_{12} - (\bar{x} + \hat{a}_{\text{nüchtern}} + \hat{b}_{\text{Männer}}))$$

Addieren wir beide Gleichungen, dann haben wir auf der linken Seite die Summe der mit den jeweiligen Gruppengrößen multiplizierten 'Wechselwirkungseffekte' - und die soll also 0 sein. Das sehen wir, wenn wir auch die rechten Seiten addieren. Um bei der Addition der rechten Seiten zusammengehörige Glieder besser zusammenfassen zu können, machen wir uns klar:

1. Nach unseren Regeln über Durchschnitte von zusammengefaßten Gruppen (vgl. Abschnitt FI4e) muß gelten:

$$N_{11} \cdot \bar{x}_{11} + N_{12} \cdot \bar{x}_{12} = (N_{11} + N_{12}) \cdot \bar{x}_1.$$

2. Nach der Definition eines 'Zeileneffekts' muß gelten

$$\bar{x} + \hat{a}_{\text{nüchtern}} = \bar{x}_1.$$

3. Nach dem, was wir in Abschnitt FI4e über die Restriktion der 'Effekte' in zusammengefaßten Gruppen gesagt haben, muß gelten:

$$N_{.1} \cdot \hat{b}_{\text{Frauen}} + N_{.2} \cdot \hat{b}_{\text{Männer}} = 0$$

Multiplizieren wir diese Gleichung mit N_{11} und dividieren sie durch $N_{.1}$, dann erhalten wir:

$$N_{11} \cdot \hat{b}_{\text{Frauen}} + (N_{11} \cdot N_{.2} : N_{.1}) \cdot \hat{b}_{\text{Männer}} = 0$$

Nun ist aber nach der Orthogonalitätsvoraussetzung (vgl. S. @)

$$N_{11} : N_{12} = N_{.1} : N_{.2}$$

oder - was dasselbe bedeutet (Multiplikation beider Seiten mit $N_{12} \cdot N_{.2} : N_{.1}$) -

$$N_{11} \cdot N_{.2} : N_{.1} = N_{12}$$

Setzen wir das in die zweite Gleichung von "3."@ ein, dann erhalten wir:

$$N_{11} \cdot \hat{b}_{\text{Frauen}} + N_{12} \cdot \hat{b}_{\text{Männer}} = 0$$

Wenn wir nun also die rechten Seiten der vor 1. stehenden zwei Gleichungen addieren und dabei von den bisher gefundenen Beziehungen Gebrauch machen, um zusammenpassende Ausdrücke zusammenzufassen, dann ergibt sich:

$$(N_{11}+N_{12}) \cdot \bar{x}_1 - (N_{11}+N_{12}) \cdot \bar{x}_1 - (N_{11} \cdot \hat{b}_{\text{Frauen}} + N_{12} \cdot \hat{b}_{\text{Männer}})$$

und das ist (unter Berücksichtigung der in 3. gefundenen Gleichung) 0, wie zu beweisen war.

>>

e) Die Signifikanztests in der Zweiweg-Varianzanalyse.

Mit einer Zweiweg-Varianzanalyse können wir 3 Nullhypothesen überprüfen:

- a. Die den Zeilendurchschnitten entsprechenden Populationsmittel sind alle gleich. Anders formuliert: In der Population sind alle Zeileneffekte 0.
- b. Die den Spaltendurchschnitten entsprechenden Populationsmittel sind alle gleich. Anders formuliert: In der Population sind alle Spalteneffekte 0.
- c. In der Population existieren keine Wechselwirkungen; die Wechselwirkungseffekte in der Population sind also alle 0.

Für jede dieser Nullhypothesen wird ein eigener F-Test als Signifikanztest durchgeführt. Wie in der Einweg-Varianzanalyse wird dabei die zu der jeweiligen Nullhypothese gehörende Quadratsumme durch ihre Freiheitsgradzahl dividiert. Das resultierende Mittelquadrat wird durch MQ_{in} dividiert, und der daraus resultierende F-Wert ist wieder "seinem Wesen nach einseitig" zu interpretieren.

Die Begründung ist ebenfalls dieselbe wie bei der Einweg-Varianzanalyse: MQ_{in} ist immer eine erwartungstreue Schätzung der lt. Homogenitätsvoraussetzung in allen Zellen gleichen Populationsvarianz. Trifft eine der Nullhypothesen zu, dann ist auch das dazu gehörende Mittelquadrat eine erwartungstreue Schätzung dieser Populationsvarianz, und daher ist der Quotient aus diesem Mittelquadrat und MQ_{in} in diesem Fall F-verteilt. Trifft dagegen eine Nullhypothese nicht zu, dann nimmt das entsprechende Mittelquadrat größere Werte an, weil auch noch die entsprechende "wahre Varianz" (zwischen den Zeilen oder zwischen den Spalten oder Wechselwirkungsvarianz) in das Mittelquadrat eingeht, und zwar umso mehr, je größer die Stichproben. Daher ergeben sich bei Ungültigkeit einer Nullhypothese für den F-Wert, der aus dem zu dieser Hypothese gehörenden Mittelquadrat berechnet wurde, größere Werte, als per Zufall zu erwarten ist. (Und die Werte in der F-Tabelle geben ja an, was - unter dem entsprechenden Signifikanzniveau - per Zufall zu erwarten ist.)

<< Genauer lassen sich mit ähnlichen Methoden, wie sie in Abschnitt @FI8 für die Einweg-

Varianzanalyse demonstriert wurden, die folgenden Gleichungen für die Erwartungswerte der Quadratsummen und Mittelquadrate in der Zweiweg-Varianzanalyse herleiten:

$$\begin{aligned} E(QS_{\text{Zeilen}}) &= (R-1) \cdot \sigma^2 + \sum_r N_r \cdot a_r^2 & E(MQ_{\text{Zeilen}}) &= \sigma^2 + \sum_r N_r \cdot a_r^2 / (R-1) \\ E(QS_{\text{Spalten}}) &= (S-1) \cdot \sigma^2 + \sum_s N_{.s} \cdot b_s^2 & E(MQ_{\text{Spalten}}) &= \sigma^2 + \sum_s N_{.s} \cdot b_s^2 / (S-1) \\ E(QS_{\text{Ww.}}) &= (R-1) \cdot (S-1) \cdot \sigma^2 + \sum_z N_z \cdot c_z^2 & E(MQ_{\text{Ww.}}) &= \sigma^2 + \sum_z N_z \cdot c_z^2 / ((R-1) \cdot (S-1)) \\ E(QS_{\text{in}}) &= (N-R \cdot S) \cdot \sigma^2 & E(MQ_{\text{in}}) &= \sigma^2 \end{aligned}$$

Dabei ist...

σ^2 = die (als identisch vorausgesetzte) Populationsvarianz innerhalb der Zellen.

N = Gesamtzahl aller Meßwerte (Vpn) in allen Gruppen zusammen.

R, S = Zahl der Zeilen ("Reihen") bzw. Spalten.

$N_r, N_{.s}, N_z$ = Zahl der Meßwerte in Zeile r bzw. Spalte s bzw. Zelle z.

a_r, b_s, c_z : Wahrer Haupt-Effekt der Zeile r bzw. Spalte s bzw. Wechselwirkungs-Effekt der Zelle z.

Wer die Herleitungen in Abschnitt @FI8 verstanden hat, mag selbst entscheiden, ob er die Übertragung des Ansatzes auf den Beweis der obigen Gleichungen selbst versuchen will. Ein Hinweis dazu: Die Formeln

$$E(RQ) = N \cdot \sigma^2 + \sum_g N_g \cdot \mu_g^2$$

und

$$E(C) = \sigma^2 + N \cdot \mu^2$$

für die Erwartungswerte der rohen Quadratsumme RQ und des Gesamt-Korrekturglieds C ergeben sich aus den gleichen Überlegungen wie bei der Einweg-Varianzanalyse. Die Herleitung der Formel

$$E(C_g) = \sigma^2 + N_g \cdot \mu_g^2$$

für das Korrekturglied der Gruppe g läßt sich ebenfalls sinngemäß übertragen, indem man Zeilen, Spalten und Zellen als "Gruppen" betrachtet. Unmittelbar einleuchtend ist das bei Zellen. Bei Zeilen und Spalten ist zu beachten, wie sich das wahre Zeilen- bzw. Spalten-Mittel aus den Zellenmitteln zusammensetzt, und dieser Wert ist dann als μ_g in die obige Formel für den Erwartungswert $E(C_g)$ einzusetzen.

>>

Die Ergebnisse der Berechnungen faßt man dann wieder in einer Tabelle zusammen. Für unser Experiment würde sie folgendermaßen aussehen:

Quelle	QS	df	MQ	F
Arbeitsbedingungen	21780.46	2	10890.23	216.51***
Geschlecht	77.26	1	77.26	1.54*
Arbeitsbed. × Geschlecht	10023.27	2	5011.64	99.63***
zwischen Zellen	31880.99	5		
innerhalb der Zellen	4980.00	99	50.30	
total	36860.99	104		

* : $p > .1$

*** : $p < .01$

Zur Gestaltung dieser Tabelle sehen wir (zusätzlich zu den schon bei der Einweg-Varianzanalyse behandelten Regeln):

1. Die Quellen werden bei der tabellarischen Zusammenfassung möglichst mit einem inhaltlichen Namen benannt ("Arbeitsbedingungen" und "Geschlecht" statt "Zeilen" und "Spalten"). Das erleichtert die Lesbarkeit.

2. Für die Wechselwirkung verbindet man die inhaltlichen Bezeichnungen der Faktoren, um deren Wechselwirkung es geht, mit einem \times -Zeichen (für "mal") und spricht von der "Wechselwirkung Arbeitsbedingungen mal Geschlecht". Wichtig ist das vor allem dann, wenn es - bei Varianzanalysen mit mehr als 2 Faktoren, wie wir sie im nächsten Kapitel kennenlernen - mehrere Wechselwirkungen gibt.

<< Genauer betrachtet, ist das Zeichen \times hier als Symbol für das "cartesische Produkt" zweier Mengen zu verstehen. Das cartesische Produkt der Menge der Arbeitsbedingungen und der Menge der Geschlechter besteht aus den "geordneten Paaren" (nüchtern, Frauen), (nüchtern, Männer), (hübsch, Frauen) usw., also aus allen Paaren, bei denen an erster Stelle eine Arbeitsbedingung und an zweiter Stelle ein Geschlecht steht. Bei der Wechselwirkung geht es ja gerade um die Effekte, die solche Bedingungskombinationen noch über die Effekte der beiden einzelnen Bedingungen hinaus haben.

>>

Nun zu den Ergebnissen. In unserem Experiment haben wir zunächst einen signifikanten Effekt der Arbeitsbedingungen: Die Nullhypothese, daß in der Population alle Arbeitsbedingungen im Durchschnitt zu gleichen Ergebnissen führen, kann mit extrem niedriger Irrtumswahrscheinlichkeit verworfen werden.

Allerdings ist es auch hier wie bei der Einweg-Varianzanalyse: Wenn wir die Nullhypothese verworfen, daß nicht alle Populationsmittel gleich sind, dann heißt das noch nicht, daß jedes von

jedem verschieden ist. Es können immer noch einzelne untereinander gleich sein. (In unserem Fall könnte es z.B. durchaus sein, daß die Populationsmittel für die Arbeitsbedingungen "hübsch" und "nüchtern" gleich sind.) Wenn man genaue Aussagen über den Unterschied zweier bestimmter Zeilenmittel in der Population machen wollte, müßte man wieder post-hoc-Tests anwenden.

Auch für die Wechselwirkung gibt es ein signifikantes F. Es liegt also eine signifikante Wechselwirkung vor.

Auch dazu ist aber eine Einschränkung erforderlich: Je nachdem, wie man die Wechselwirkung betrachtet, lautet die entsprechende Nullhypothese: "Die Effekte der verschiedenen Arbeitsbedingungen sind bei Männern und Frauen gleich", oder: "Die Effekte des Geschlechts sind bei allen Arbeitsbedingungen gleich". Ein signifikantes Ergebnis besagt auch bei der Wechselwirkung lediglich, daß die Nullhypothese nicht in vollem Umfang zutrifft: Nicht alle Effekte der Arbeitsbedingungen sind bei Männern und Frauen gleich - bzw. der Effekt des Geschlechts ist nicht bei allen Arbeitsbedingungen gleich. Es kann aber durchaus sein, daß der Effekt zweier Arbeitsbedingungen bei Männern und Frauen gleich ist - bzw. daß der Effekt des Geschlechts bei 2 Arbeitsbedingungen gleich ist. Auch hier gilt: Wenn man sich mit der globalen Ablehnung der Nullhypothese nicht begnügt, dann muß man post-hoc-Tests rechnen, um jede einzelne Aussage über Unterschiede, die man treffen will, zu überprüfen.

Nicht signifikant ist dagegen der Effekt des Geschlechts. Die Nullhypothese, daß Männer und Frauen gleich gut sind, kann nicht verworfen werden.

Auch dazu ist eine einschränkende Anmerkung erforderlich: Das Ergebnis gilt (soweit ein nicht signifikantes Ergebnis überhaupt eins ist) nur dann, wenn die Häufigkeit, mit der die verschiedenen Arbeitsbedingungen in der Population bei Männern und Frauen auftreten, die gleichen sind, wie in unserem Experiment, also

$$\text{nüchtern} : \text{hübsch} : \text{altmodisch} = 42 : 35 : 28 = 6 : 5 : 4$$

Würde dagegen z.B. in der Population wesentlich häufiger unter Arbeitsbedingung "hübsch" als unter Arbeitsbedingung "nüchtern" gearbeitet, dann wären die Frauen (die ja unter Arbeitsbedingung "hübsch" wesentlich besser sind als die Männer) auch insgesamt besser.

Ein entsprechender Gedanke gilt allgemein: Die Populationsmittel, über die eine Nullhypothese über einen Haupteffekt in der Zweiweg- Varianzanalyse eine Aussage macht, werden immer unter der Voraussetzung betrachtet, daß die Häufigkeiten, mit denen die Stufen des anderen Faktors besetzt sind, repräsentativ für die Population sind, um die es geht. Wenn - wie in unserem Fall - Wechselwirkungen vorliegen, dann kann von diesen Häufigkeiten das Auftreten oder Nicht-Auftreten eines Haupteffekts abhängen.

<< Auch hier sind "halbseitige" Entscheidungsregeln möglich (vgl. S.@).
Beispiel: Hätten wir die Entscheidungsregel formuliert, daß wir die Spalten- H_0 nur dann verwerfen, wenn wir ein genügend hohes F erhalten und wenn außerdem die Frauen einen höheren Durchschnitt erzielen als die Männer, so könnten wir zur Ermittlung der Irrtums-

wahrscheinlichkeit die Überschreitungswahrscheinlichkeit des F -Wertes halbieren.

>>

6) *Die Varianzanalyse mit mehr als zwei Faktoren*

a) Ein Experiment mit drei Faktoren

Ein neu entwickeltes Schlafmittel soll darauf untersucht werden, welche Wirkungen es bei neurotischen und nichtneurotischen ("normalen") Frauen und Männern hat. Zu diesem Zweck werden zunächst 4 Gruppen von V_{pn} gebildet: 100 neurotische Frauen, 100 normale Frauen, 100 neurotische Männer und 100 normale Männer. Diese Gruppen werden nach Zufall in je zwei gleich große Hälften geteilt. Die erste erhält vor dem Schlafen eine Dosis des Schlafmittels, die zweite eine Placebo-Tablette.

Was sind die Faktoren und Stufen?

Wir haben drei Faktoren: Den Faktor "Geschlecht" mit den Stufen "Frauen" und "Männer", den Faktor "Neurotizismus" mit den Stufen "neurotisch" und "normal" und den Faktor "Droge" mit den Stufen "Schlafmittel" und "Placebo".

Nehmen wir einmal an, es ergeben sich bei dem Experiment die in Tabelle @20 zusammengefaßten Durchschnitte der Schlafzeit in Stunden.

	Frauen		
	neurotisch	normal	Gesamt
Schlafmittel	6.90	8.40	7.65
Placebo	6.90	6.20	6.55
Gesamt	6.90	7.30	7.10
	Männer		
	neurotisch	normal	Gesamt
Schlafmittel	7.00	6.90	6.95
Placebo	6.40	7.30	6.85
Gesamt	6.70	7.10	6.90
	Frauen und Männer zusammengefaßt		
	neurotisch	normal	Gesamt
Schlafmittel	6.95	7.65	7.30
Placebo	6.65	6.75	6.70
Gesamt	6.80	7.20	7.00

Tabelle 20: Durchschnitte der Schlafzeit in Stunden

b) Die Effekte

ba) Die Haupteffekte

Bestimmen Sie die Haupteffekte!

Wenn wir den Effekt des Geschlechts "weiblich" schätzen wollen, dann vergleichen wir den Durchschnitt aller Frauen mit dem Gesamtdurchschnitt:

$$\hat{a}_w = 7.1 - 7.0 = 0.1$$

Der Effekt des Geschlechts "weiblich" besteht also schätzungsweise in einer Hebung der Schlafzeit um durchschnittlich 0.1 Stunde. - Ähnlich würde der Effekt des Geschlechts "männlich" zu berechnen sein.

Schätzen Sie auch von den übrigen Faktoren den Effekt der ersten Stufe!

$$\hat{b}_{ne} = 6.8 - 7.0 = -0.2$$

Den Effekt der Stufe "neurotisch" schätzen wir als eine Senkung der Schlafzeit um durchschnittlich 0.2 Stunden.

$$\hat{c}_s = 7.3 - 7.0 = 0.3$$

Den Effekt des Schlafmittels schätzen wir also als eine Hebung der Schlafzeit um durchschnittlich 0.3 Stunden. (Beachten Sie: Der Begriff des Effekts eines Faktors bedeutet in der

Varianzanalyse Vergleich mit dem Gesamtmittel! Würden wir den Effekt des Schlafmittels durch Vergleich des Schlafmitteldurchschnitts (7.3) mit dem Placebo-Durchschnitt (6.7) ermitteln und dann schätzen, der Effekt des Schlafmittels bestünde in einer Hebung der Schlafzeit um 0.6 Einheiten, so würde das auch einen Sinn ergeben, würde aber nicht dem varianzanalytischen Effekt-Begriff entsprechen.)

Bei der Bestimmung des Effekts der Stufe "weiblich" haben wir die Stufen der Faktoren "Neurotizismus" und "Droge" zusammengefaßt.

Allgemein haben wir bei der Berechnung des Effekts einer Stufe eines Faktors die Stufen der zwei übrigen Faktoren zusammengefaßt. Die so ermittelten Effekte nennt man genauso wie in der Zweiweg-Varianzanalyse die Haupteffekte.

bb) Die Wechselwirkungen erster Ordnung

Faßt man dagegen nur die Stufen *eines* Faktors zusammen, so entsteht eine Tabelle, wie wir sie aus der Zweiweg-Varianzanalyse kennen.

Beispiel: Faßt man die Stufen "Männer" und "Frauen" zusammen, so entsteht die letzte Tabelle aus Tabelle 20. Wir sehen an dieser Tabelle folgendes: Der Unterschied zwischen Neurotikern und Normalen ist unter Schlafmittel erheblich größer als unter Placebo.

Welcher statistische Begriff fällt ihnen dazu ein?

Es handelt sich um eine Wechselwirkung. Bekanntlich hat jede Wechselwirkung zwei Aspekte.

Wie würde der andere Aspekt lauten?

Der Effekt des Schlafmittels ist bei den Neurotikern geringer (0.15) als bei den Normalen (0.45).

Wie groß ist der Wechselwirkungseffekt für das Feld "Schlafmittel-Neurotiker"?

wäre keine Wechselwirkung vorhanden, so würden wir Additivität der Effekte erwarten:

$$\bar{x}_{s,ne} = \bar{x} + \hat{b}_{na} + \hat{c}_a = 7.0 + (-0.2) + 0.3 = 7.1$$

Der tatsächliche Durchschnitt beträgt 6.95. Wir schätzen also den Wechselwirkungseffekt der Stufenkombination "Schlafmittel-Neurotiker" als Senkung der Schlafzeit um 0.15 Stunden. Verwenden wir zur Kennzeichnung des Effekts den Buchstaben g:

$$\hat{g}_{s,ne} = 6.95 - 7.1 = -0.15$$

Übungsaufgabe: Schätzen Sie aufgrund der gleichen Tabelle die Wechselwirkungseffekte in den übrigen Zellen.

Ergebnis: $\hat{g}_{s,no} = \hat{g}_{p,ne} = 0.15$; $\hat{g}_{p,no} = -0.15$

Bei all diesen Wechselwirkungseffekten handelt es sich um eine Wechselwirkung von Droge und Neurotizismus. Man spricht auch von der "Droge x Neurotizismus-Wechselwirkung" (lies: Droge mal Neurotizismus...).

Daneben gibt es aber noch weitere Wechselwirkungen: Die "Geschlecht x Droge-Wechselwirkung" und die "Geschlecht x Neurotizismus-Wechselwirkung".

Können Sie sich darunter etwas vorstellen? Wenn ja, führen Sie die gleichen Überlegungen und Berechnungen durch, die oben für die "Droge x Neurotizismus-Wechselwirkung"

auftraten.

Wenden wir uns der Geschlecht x Droge-Wechselwirkung zu. Hierzu fassen wir die Stufen des Faktors "Neurotizismus" zusammen. Es entsteht Tabelle 21.

	Frauen	Männer	Gesamt
Schlafmittel	7.65	6.95	7.30
Placebo	6.55	6.85	6.70
Gesamt	7.10	6.90	7.00

Tabelle 21

Wie kommt man zu dieser Tabelle?

Antwort: Vergleichen Sie die letzten Spalten der drei Tabellen in Tabelle 21 mit den drei Spalten von Tabelle 20! Bei Tabelle 20 bedeutete das Wort "Gesamt" über der letzten Spalte ja "Durchschnitt nach Zusammenfassung von Neurotikern und Normalen".

Führen Sie jetzt die Überlegungen zur Wechselwirkung durch!

Wir sehen: Der Effekt des Geschlechts ist auf den verschiedenen Stufen des Faktors "Droge" verschieden. Anderer Aspekt: Der Drogeneffekt ist auf den verschiedenen Stufen des Faktors "Geschlecht" (also bei Männern und Frauen) verschieden.

Berechnen und verbalisieren Sie einen Wechselwirkungseffekt; verwenden Sie das Symbol d !

$$\hat{d}_{w,s} = \bar{x}_{w,s} - (\bar{x} + \hat{a}_w + \hat{c}_s) = 7.65 - (7.0 + 0.1 + 0.3) = 0.25$$

Wir schätzen den Wechselwirkungseffekt der Stufenkombination "Frauen-Schlafmittel" als eine Hebung der Schlafzeit um durchschnittlich 0.25 Stunden.

Besteht auch eine "Geschlecht x Neurotizismus-Wechselwirkung"?

	neurotisch	normal	Gesamt
Frauen	6.90	7.30	7.10
Männer	6.70	7.10	6.90
Gesamt	6.80	7.20	7.00

Tabelle 22

Wir sehen in Tabelle 22 (vgl. dazu die "Gesamt"-Zeilen der Tabelle 20): Der Effekt beider Geschlechtsstufen ist auf den beiden Stufen des Faktors "Neurotizismus" gleich. Es besteht also keine Geschlechts x Neurotizismus-Wechselwirkung. Auch das Fehlen dieser Wechselwirkung hat einen zweiten Aspekt.

Können Sie ihn nennen?

Der Effekt beider Neurotizismusstufen ist bei Männern und Frauen gleich groß.

In Formelsprache:

$$\hat{t}_{w,ne} = \bar{x}_{w,ne} - (\bar{x} + \hat{a}_w + \hat{b}_{ne}) = 6.9 - (7.0 + 0.1 + (-0.2)) = 0$$

Dieser Wechselwirkungseffekt ist also hier gleich null, und dasselbe würde sich auch bei den übrigen Zellen der letzten Tabelle ergeben.

Wir haben manchmal gesagt, daß es sich bei zwei Wechselwirkungsfeststellungen nur um zwei verschiedene Aspekte der gleichen Wechselwirkung handelt. In anderen Fällen geben wir den Wechselwirkungen Namen, um damit auszudrücken, daß es sich um zwei verschiedene Wechselwirkungen handelt.

Wann das eine, wann das andere?

Regel: Solange die beteiligten Faktoren die gleichen sind, handelt es sich um die gleiche Wechselwirkung.

Beispiel: Ob wir sagen "Die Drogeneffekte sind bei Neurotikern und Normalen verschieden" oder "Die Neurotizismuseffekte sind unter Schlafmittel und unter Placebo verschieden" - in beiden Fällen sind die Faktoren "Neurotizismus" und "Droge" an der Wechselwirkung beteiligt, und daher handelt es sich um die gleiche Wechselwirkung. Die Richtigkeit dieser Regel kann man unmittelbar aus der Berechnungstechnik der Effekte ersehen: Wir würden in beiden Fällen die gleiche Formel einsetzen, um die Höhe der einzelnen Wechselwirkungseffekte zu schätzen.

An den bisher behandelten Wechselwirkungen waren jeweils zwei Faktoren beteiligt. Solche Wechselwirkungen nennt man Wechselwirkungen erster Ordnung. Wir haben gesehen: Um die einzelnen Wechselwirkungen voneinander zu unterscheiden, benennt man jede einzeln, indem man die beiden an ihr beteiligten Faktoren mit einem x-Zeichen verbindet.

bc) Die Wechselwirkung zweiter Ordnung

Kehren wir noch einmal kurz zu unserer Ausgangstabelle Tabelle@ zurück. Wir hatten da die Droge x Neurotizismus-Wechselwirkung anhand der Tabelle bestimmt, in der Männer und Frauen zusammengefaßt sind und hatten uns gefragt: Ist der Effekt des Neurotizismus auf den zwei Stufen des Faktors "Droge" (also in der Schlafmittel und Placebo-Gruppe) gleich? Untersuchen Sie die gleiche Frage für Männer und Frauen getrennt und vergleichen Sie die Ergebnisse!

Bei den Männern stellen wir fest: Unter Schlafmittel schlafen die Neurotiker länger, unter Placebo die Normalen. Bei den Frauen ist es umgekehrt: Unter Schlafmittel schlafen die Normalen länger, unter Placebo die Neurotikerinnen. Die Droge x Neurotizismus-Wechselwirkung ist also bei Männern und Frauen verschieden. In einem solchen Fall, in dem die Wechselwirkung zweier Faktoren auf den verschiedenen Stufen eines dritten Faktors verschieden ist, spricht man von einer Wechselwirkung zweiter Ordnung.

Ein weiteres Beispiel: Nehmen wir einmal an, es hätte die folgenden Durchschnitte gegeben (Tabelle 23)

	Frauen		
	neurotisch	normal	Gesamt
Schlafmittel	6.90	7.10	7.00
Placebo	6.60	7.00	6.80
Gesamt	6.75	7.05	6.90
	Männer		
	neurotisch	normal	Gesamt
Schlafmittel	6.90	7.00	6.95
Placebo	5.90	6.90	6.40
Gesamt	6.40	6.95	6.675

Tabelle 23

Läge auch hier noch die besprochene Wechselwirkung zweiter Ordnung vor?

Bei den Frauen wie auch bei den Männern besteht eine Droge x Neurotizismus-Wechselwirkung: Der Effekt des Schlafmittels ist bei den Neurotikern stärker als bei den Normalen. Diese Wechselwirkung erster Ordnung ist aber bei den Männern offenbar stärker als bei den Frauen. Folglich liegt auch bei diesen Daten eine Wechselwirkung zweiter Ordnung vor. Genau so, wie jede Wechselwirkung erster Ordnung zwei Aspekte hat, so hat jede Wechselwirkung zweiter Ordnung drei Aspekte: Es ließe sich auch zeigen, daß die Droge x Geschlechts-Wechselwirkung bei Neurotikern und Normalen nicht gleich ist. (Vorschlag: Überprüfen Sie anhand der ursprünglichen Daten.) Ebenso ist auch die Geschlecht x Neurotizismus Wechselwirkung unter Schlafmittel und Placebo nicht gleich.

Wieso handelt es sich nicht um drei verschiedene Wechselwirkungen, sondern um drei Aspekte der gleichen Wechselwirkung?

Weil in allen drei Fällen die gleichen Faktoren beteiligt sind, Geschlecht, Neurotizismus und Droge.

Eine andere Definition der Wechselwirkung zweiter Ordnung ermöglicht uns die Berechnung eines Wechselwirkungseffekts zweiter Ordnung. Bei der Zweiweg-Varianzanalyse sagten wir: Wenn es keine Wechselwirkung gäbe, dann müßte sich jeder Durchschnitt einer Stufenkombination (Zelle) ergeben, indem man die entsprechenden Haupteffekte zum Gesamtmittel addiert. Ebenso können wir hier sagen: wenn es keine Wechselwirkung zweiter Ordnung gäbe, dann müßte sich jeder Durchschnitt einer Kombination von Stufen aller drei Faktoren ergeben, indem man die entsprechenden Haupteffekte und die zugehörigen Wechselwirkungseffekte erster Ordnung zum Gesamtdurchschnitt addiert. Bspw. müßte beim Fehlen einer Wechselwirkung zweiter Ordnung für die Stufenkombination "neurotische Frau unter Schlafmittel" in den alten Daten gelten:

$$\bar{x}_{w,ne,s} = \bar{x} + \hat{a}_w + \hat{b}_{ne} + \hat{c}_s + \hat{d}_{w,s} + \hat{f}_{w,ne} + \hat{g}_{s,ne}$$

$$\begin{aligned} &= 7.0 + 0.1 + (-0.2) + 0.3 + 0.25 + 0 + (-0.15) \\ &= 7.3 \end{aligned}$$

Tatsächlich haben wir aber

$$\bar{x}_{w,ne,s} = 6.9$$

Dieser Wert liegt um 0.4 Stunden unterhalb des ohne Wechselwirkung zu erwartenden Wertes. Den Effekt der Wechselwirkung zweiter Ordnung für diese Zelle (Stufenkombination) können wir also schätzen als

$$\hat{h}_{w,ne,s} = 6.9 - 7.3 = -0.4$$

Übungsaufgabe: Schätzen Sie in gleicher Weise den Wechselwirkungseffekt für die Stufenkombination "normale Frauen unter Schlafmittel"!

Ergebnis: $\hat{h}_{w,no,s} = 0.4$

c) Signifikanztests

Selbstverständlich ist es auch in der Dreiweg-Varianzanalyse erforderlich, die einzelnen Effekte auf Signifikanz zu überprüfen, da auch hier in der Stichprobe auftretende Effekte durch Zufall entstanden sein können. Die "Totalquadratsummen" wird wieder unterteilt in eine "Quadratsumme zwischen den Gruppen" und eine "Quadratsumme innerhalb der Gruppen"; die erstere wird weiterhin unterteilt nach den besprochenen sieben Quellen (drei Haupteffekte, drei Wechselwirkungen erster Ordnung sowie die Wechselwirkung zweiter Ordnung). Jede dieser Quadratsummen hat eine Freiheitsgradzahl. Die Formeln für Quadratsummen und Freiheitsgrade sollen hier nicht näher besprochen werden, da man sie doch meistens in Handbüchern nachschlägt. Der Quotient aus Quadratsumme und Freiheitsgraden ergibt wieder ein Mittelquadrat. Die Signifikanz jedes Effekts überprüft man, indem man das entsprechende Mittelquadrat durch das "Mittelquadrat innerhalb" dividiert (manchmal dividiert man auch durch ein anderes Mittelquadrat, vgl. dazu Abschnitt 9d).

So erhält man einen F -Wert. Die Ergebnisse faßt man in einer Tabelle zusammen. Durch Anwenden der entsprechenden Formeln auf unser Beispiel würde sich Tabelle 24 ergeben.⁹⁶

⁹⁶Die Quadratsumme innerhalb kann man natürlich aus den Durchschnitten nicht berechnen: Sie ist ebenso wie diese frei erfunden!

Quelle	QS	d.f.	MQ	F
G (Geschlecht)	4.00	1	4.00	1.86
N (Neurotizismus)	16.00	1	16.00	7.44**
D (Droge)	36.00	1	36.00	16.74**
G x N	0.00	1	0.00	0.00
G x D	25.00	1	25.00	11.63**
D x N	9.00	1	9.00	4.19*
G x D x N	64.00	1	64.00	29.77**
zwischen Gruppen	154.00	7		
innerhalb Gruppen	842.80	392	2.15	
Insgesamt	996.80	399		

* : $p < 5\%$

** : $p < 1\%$

Tabelle 24

Bei der Signifikanzbeurteilung ist folgendes zu beachten:

Die Wahrscheinlichkeit, daß *irgendeiner* dieser sieben F -Werte fälschlicherweise signifikant wird, ist natürlich größer als 5%. Dieses Problem ergibt sich immer bei der Mehrweg-Varianzanalyse. Genau so wie eine mehrfache Anwendung des t -Tests (vgl. dazu F I 2) ist auch die Mehrweg-Varianzanalyse nicht geeignet, die Nullhypothese zu überprüfen, daß überhaupt keine Effekte bestehen, daß also die den acht Gruppenschnitten entsprechenden Populationsmittel alle gleich sind.

Wie würde man diese Hypothese am ehesten überprüfen?

Mit einer Einweg-Varianzanalyse mit acht Gruppen. Die Klärung einer derart unspezifischen Nullhypothese ist normalerweise nicht Zweck der Mehrweg-Varianzanalyse. Das hier besprochene Experiment könnte z.B. mit dem Ziel unternommen worden sein, zu überprüfen, ob das zur Diskussion stehende Schlafmittel bei Neurotikern und Normalen gleich gut wirkt.

Welche Varianzquelle wäre dann vor allem interessant?

Die Droge x Neurotizismus-Wechselwirkung. Bei der Schätzung der Effekte stellten wir fest, daß das Mittel bei Normalen besser wirkt als bei Neurotikern. Unter der genannten Fragestellung wäre also nur die folgende Nullhypothese von Interesse:

H_0 : Es besteht in der Population keine Droge x Neurotizismus Wechselwirkung.

Der entsprechende, auf dem 5%-Niveau signifikante F -Wert von 4.19 berechtigt uns zum Verwerfen dieser H_0 : Die aufgefundene Wechselwirkung dürfte kaum auf einem Stichprobenzufall beruhen. Man wird also das Schlafmittel vor allem an "Normale" verabreichen, während man bei Neurotikern eher ein anderes Mittel gibt, das eine umgekehrte Wechselwirkung zeigt.

Für diesen Schluß ist nur die Signifikanz eines F -wertes erforderlich. Es ergibt sich also für die Praxis hier gar nicht das Problem der gleichzeitigen Signifikanz mehrerer F -Werte: Die

Wahrscheinlichkeit, daß dieser eine F -Wert fälschlicherweise signifikant wird, ist 5%.

Es könnte aber auch sein, daß das Experiment auch unternommen wurde, um zu überprüfen, ob irgendwelche Wechselwirkungen der Droge mit einem der übrigen Faktoren auftritt. Von Interesse ist hier auch die Wechselwirkung zweiter Ordnung: Wir sehen, daß die Stufenkombination "normale Frauen unter Schlafmittel" über alle Haupteffekte und Wechselwirkungseffekte erster Ordnung hinaus zu einer weiteren Verlängerung der Schlafzeit um durchschnittlich 0.4 Stunden führte:

$$\hat{h}_{w,no,s} = 0.4 \text{ (vgl. Übungsaufgabe S.@)}$$

Bei den normalen Männern ergäbe sich dagegen eine ebenso hohe Senkung der Schlafzeit unter Schlafmittel:

$$\hat{h}_{m,no,s} = -0.4 \text{ (dieser Wert wurde bisher nicht berechnet)}$$

Das - wie wir gesehen haben - vor allem bei nichtneurotischen (normalen) Vpn wirksame Schlafmittel wäre demnach vor allem für nicht-neurotische Frauen zu empfehlen. Dieser Schluß baut aber gleichzeitig auf der Signifikanz der Droge x Neurotizismus-Wechselwirkung auf. Es wäre also erforderlich, die Signifikanz beider Wechselwirkungen zu überprüfen. Aber: Die Wahrscheinlichkeit, daß der eine oder der andere F -Wert fälschlicherweise (also trotz des Fehlens der Wechselwirkung in der Population) signifikant wird, ist größer als 5% - ganz entsprechend wie bei der mehrfachen Anwendung des t -Tests, bei der auch ein überhöhtes Risiko I obliegt (vgl. Abschnitt F.I.2). Eine moderne Methode, die aus dem Problem heraushilft, ist die Methode der multiplen Vergleiche (engl.: multiple comparisons), die hier nicht näher besprochen werden soll. Sie wird ziemlich eingehend bei Hays (1973) besprochen. Sie ist also u.a. dann angebracht, wenn die gleichzeitige Verwerfung mehrerer Nullhypothesen in der Mehrweg-Varianzanalyse erforderlich wird.

d) Varianzanalyse mit mehr als drei Faktoren

Die Zahl der Faktoren in der Varianzanalyse ist im Grunde unbegrenzt. Mit jedem weiteren Faktor treten mehr Effekte auf; insbesondere kommen Wechselwirkungen immer höherer Ordnung ins Spiel. Im Grunde geht aber alles so weiter wie bisher: Es gibt zunächst die Haupteffekte in jedem Faktor. Zur Berechnung faßt man die Stufen der übrigen Faktoren zusammen. Dann gibt es die Wechselwirkungen erster Ordnung, und zwar für jedes Paar von Faktoren extra. Man überprüft, ob die Effekte des einen Faktors auf den verschiedenen Stufen des anderen Faktors gleich sind, wenn man die Stufen der übrigen Faktoren zusammenfaßt (wir haben z.B. bei der Berechnung der "Droge x Neurotizismus-Wechselwirkung" die Stufen des an dieser Wechselwirkung nicht beteiligten Faktors "Geschlecht" zusammengefaßt.)

Genau so, wie eine Wechselwirkung erster Ordnung entsteht, indem die Effekte eines Faktors auf den Stufen eines anderen Faktors verschieden sind, so entsteht eine Wechselwirkung zweiter Ordnung, indem die Wechselwirkung zweier Faktoren auf den verschiedenen Stufen eines dritten Faktors verschieden ist. Und genau so geht es weiter. Eine Wechselwirkung entsteht,

indem die von drei Faktoren bestimmte Wechselwirkung zweiter Ordnung auf den verschiedenen Stufen eines vierten Faktors verschieden ist usw.

Mit zunehmender Faktorenzahl kommen also immer mehr Varianzquellen in die Varianzanalyse, und die Schwierigkeiten mit der mehrfachen Anwendung des F -Tests und der dadurch bedingten Überhöhung des Risikos I werden immer gewichtiger, wenn man die gleichzeitige Signifikanz mehrerer Varianzquellen fordert. Wir sahen aber schon bei der Dreiweg-Varianzanalyse: Oft ist vor allem eine Varianzquelle interessant, und das Experiment wird um dieser Varianzquelle willen durchgeführt. Dann ergibt sich das Problem nicht, da eigentlich nur ein F -Wert interessant ist.

Es ist eine Sache der Übung, beim Lesen entsprechender Experimente herauszufinden, welche Varianzquelle für bestimmte psychologische Fragestellungen relevant ist; andererseits ist es auch beim Planen von Experimenten erforderlich, sich vor der Datenerhebung klarzumachen, welche Faktoren notwendig sind und welche Varianzquelle die Frage, die auch experimentell untersucht werden soll, beantwortet. Diese Beziehung bestimmter psychologischer Fragen zu den statistischen Modellen der Varianzanalyse zu sehen, ist nicht mehr Aufgabe des Statistikkurses, sondern einmal ein Gegenstand des Experimental-Praktikums, vor allem aber - wie schon gesagt - eine Sache der Übung, die man am ehesten beim Lesen entsprechender Versuchsberichte in Zeitschriften erwirbt.

Es gibt noch viele weitere Anwendungsformen der Varianzanalyse, die alle eine spezielle Rechentechnik haben. Es würde zu weit führen, sie alle mit der gleichen Ausführlichkeit zu behandeln wie die bisherigen.

In den folgenden Abschnitten soll kurz auf die speziellen Kennzeichen dieser weiteren Anwendungsformen eingegangen werden. Hat man nämlich einmal festgelegt, um welchen Typus eines Versuchsplans es sich handelt, dann kann man die weiteren Details in einem entsprechenden ausführlichen Buch (etwa Hays, @; Henning & Muthig, @; Winer, @) nachschlagen.

7) *Varianzanalyse mit korrelierten Stichproben*

a) Mehrere Messungen pro Vp

In den bisher besprochenen Versuchsplänen war jede Vp nur in einer Stichprobe mit einem Meßwert vertreten. Insofern entsprechen diese Versuchspläne dem t -Test für unabhängige Stichproben. Es gibt daneben in der Varianzanalyse auch Versuchspläne mit mehreren Messungen pro Vp. Diese entsprechen dem t -Test für korrelierte Stichproben.

Ein *Beispiel*: Wir wollen die Reaktionszeit von Vpn in drei verschiedenen Aufgaben untersuchen. Vor der Vp sind auf einem Brett mehrere Lampen montiert. Vor jeder Lampe ist eine Taste angebracht. Sobald eine Lampe aufleuchtet, soll die Vp möglichst schnell die zugehörige Taste drücken. Die Aufgaben unterscheiden sich durch die Zahl der Lampen und Tasten. Bei Aufgabe A sind es zwei Lampen und Tasten, bei Aufgabe B vier und bei

Aufgabe C acht; gemessen wird jeweils die Reaktionszeit, also die Zeit vom Aufleuchten der Lampe bis zum Tastendruck. Von jeder Vp werden 6 Messungen erhoben - je zwei bei jeder Aufgabe.

In Tabelle 25 sind die Ergebnisse der Reaktionszeit in Millisekunden angegeben. Rechts neben jedem Meßwertpaar steht der Durchschnitt der beiden Messungen (kursiver Wert).

Anmerkung: Die Werte sind frei erfunden! In einem tatsächlichen Experiment würden sich längere Reaktionszeiten ergeben.

Vp	A			B			C			Durchschnitt der Vp
1	63	67	65	70	68	69	68	72	70	68
2	75	69	72	71	75	73	73	75	74	73
3	79	81	80	80	82	81	86	84	85	82
4	83	81	82	87	81	84	88	84	86	84
5	70	66	68	80	68	74	78	76	77	73
6	64	66	65	74	72	73	79	71	75	71
7	79	81	80	81	85	83	87	85	86	83
8	80	80	80	80	82	81	83	87	85	82
9	62	64	63	67	61	64	67	63	65	64
10	63	57	60	64	62	63	66	72	69	64
Durchschnitt aller Vpn	71,5			74,5			77,2			74,4

Tabelle 25: Reaktionszeiten

Die ganze Tabelle 25 erinnert an eine Zweiweg-Varianzanalyse. Tatsächlich würde man diese auch rechnen. Wir haben zwei Faktoren: Den Faktor "Aufgabe" (mit den Stufen A, B und C) und einen Faktor "Vpn" (mit den Stufen 1 bis 10). Insgesamt haben wir also 30 Zellen mit je zwei Beobachtungen. Für den Experimentator war die Varianz "zwischen den Aufgaben" von besonderer Bedeutung: Je mehr Lampen, umso länger die Reaktionszeit. Für uns ist aber auch interessant, daß es noch eine Varianz zwischen den Vpn gibt: Einige Vpn sind durchweg schnell, andere sind in allen Aufgaben langsamer als die übrigen. Weiterhin findet sich auch eine Wechselwirkung "Vpn x Aufgaben":

Bei einigen Vpn (z.B. Vp 2 und Vp 9) zeigt sich nur ein geringer Aufgabeneffekt, bei anderen Vpn (z.B. Vp 5, 6 und 10) ist er wesentlich ausgeprägter. Die Effekte des Faktors "Aufgabe" sind also auf den Stufen des Faktors "Vpn" verschieden.

Schließlich gibt es eine Varianz innerhalb der Zellen: Nur ganz selten hat eine Vp bei beiden Durchführungen der gleichen Aufgabe die gleiche Reaktionszeit; meistens bestehen Unterschiede innerhalb der Zellen. Mit einer Zweiweg-Varianzanalyse könnte man berechnen, ob die drei erstgenannten Varianzquellen signifikant sind. Allerdings müßte man die *F*-Werte etwas anders

berechnen, als bisher besprochen wurde; diese Unterschiede tun aber im Augenblick nichts zur Sache.

Solche Versuchspläne mit wiederholten Messungen haben vor allem einen Vorteil: Eine erheblich größere Teststärke. Wir wollen einmal für einen Augenblick annehmen, daß unsere 60 Meßwerte aus einer Einweg-Varianzanalyse stammen würden, daß es sich also um 60 verschiedene Vpn handeln würde. In diesem Fall würden die ganzen Unterschiede innerhalb der Spalten zur "Fehlervarianz"; auch ohne Berechnung sehen wir, daß die Varianz innerhalb der Spalten viel größer ist als die minimalen Schwankungen innerhalb der Zellen. Eine große Fehlervarianz führt aber zu einem niedrigen F -Wert! Wir erzielen also mit der Einweg-Varianzanalyse nur schwer (nämlich mit vielen Vpn) große, d.h. signifikante F -Werte, auch wenn H_0 unrichtig ist.

Exakter betrachtet: In unserem Anspruchsniveau-Experiment zur Einweg-Varianzanalyse konnten wir einem im Vergleich zum Gruppenmittel hohen Wert nicht ansehen, ob er daran lag, daß die Vp in jedem Fall ein hohes Anspruchsniveau hat. Wir mußten alle Abweichungen vom Gruppenmittel zur "Varianz innerhalb" zählen.

Wo würden sich solche systematischen Unterschiede zwischen den Vpn, die nicht reine Meßfehler sind, sondern in Form einer allgemeinen erhöhten Reaktionsbereitschaft von der Vp mit ins Experiment gebracht werden, in unserem Reaktionszeit-Experiment niederschlagen?

In der Varianz "zwischen den Vpn".

In der Einweg-Varianzanalyse schlugen sich auch differentielle Effekte in der "Varianz innerhalb der Gruppen" nieder.

Wo tauchen sie hier auf?

Wenn ein differentieller Effekt vorliegt, wenn also der Effekt einer Stufe des Faktors "Aufgabe" bei verschiedenen Vpn verschieden ist, dann schlägt sich das in der Vpn x Aufgaben Wechselwirkung nieder.

Zusammengefaßt: In der Einweg-Varianzanalyse war die Varianz innerhalb der Gruppen ein Konglomerat aus systematischen Unterschieden zwischen den Vpn, differentiellen Effekten und eigentlichen Meßfehlern (d.h. Abweichungen der Vpn von ihrem "wahren Wert" unter der jeweiligen Bedingung). Bei unserem Versuchsplan mit wiederholten Messungen können wir diese Komponenten auseinanderziehen. Da wir von jeder Vp mehrere Meßwerte haben, können wir abschätzen wie groß ihre Meßfehler sind. Wir sehen nämlich, daß die Schwankungen innerhalb der Vpn minimal sind.

Von besonderer Bedeutung ist eine weitere Variante dieses Versuchsplans:

Will man lediglich die Signifikanz des Effekts "zwischen den Aufgaben" überprüfen, so genügt eine Beobachtung pro Zelle.

(In diesem Fall kann man natürlich keine "Varianz innerhalb der Zellen" bestimmen. Für die genannte Signifikanzprüfung ist diese auch nicht erforderlich: Da es sich um ein sogenanntes gemischtes Modell handelt (was man darunter versteht, wird später behandelt!), wird in den Nenner des F -Bruchs hier nicht das MQ_{in} , sondern das MQ_{ww} gesetzt. Dieser Versuchsplan mit

einer Beobachtung pro Zelle ist in der Literatur als "Kendall-Plan" bekannt. Da jede Vp unter jeder Bedingung nur einmal getestet wird, ist dieser Versuchsplan am ehesten dem *t*-Test für korrelierte Stichproben verwandt.

b) Randomisierte Blocks

Den *t*-Test für korrelierte Stichproben konnten wir nicht nur für Meßwertpaare derselben Vp verwenden; daneben kamen auch Messungen an Paaren in Frage, die sich schon vor dem Experiment als einander ähnlich herausgestellt haben (z.B. aufgrund anderer Tests). Ähnlich hier: Statt alle Messungen an jeder Vp durchzuführen können wir "Blocks" von Vpn bilden, die einander ähnlich sind.

Beispiel: Es soll untersucht werden, ob die Form einer Darbietung einer Denkaufgabe einen Einfluß auf die Lösungszeit hat.

Darbietungsform A: Vorlesen der Aufgabe

Darbietungsform B: Schriftliches Vorlegen der Aufgabe

Darbietungsform C: Die Aufgabe wird vorgelesen und gleichzeitig schriftlich vorgelegt.

Es ist nicht möglich, jede Vpn unter allen drei Bedingungen zu testen: Eine einmal gelöste Denkaufgabe ist für die betreffende Vp meistens keine mehr. Also: Die 12 Vpn werden aufgrund eines Denksporttests in Gruppen ("Blocks") von je etwa drei gleich guten Denksportlern aufgeteilt. Aus jedem Block kommt je eine Vp zu Darbietungsform A, eine zu B und eine zu C.

Ergebnisse der Lösungszeit in Sekunden in Tabelle 26 (auch hier gilt die Anmerkung von S.@).

	Darbietungsform			Durchschnittszeit der 3 Vpn des Blocks
	A	B	C	
Vpn-Block 1	67.9	68.5	43.6	60.0
Vpn-Block 2	84.5	72.6	63.7	73.6
Vpn-Block 3	103.4	90.5	82.6	92.6
Vpn-Block 4	120.8	123.7	98.4	114.3
Durchschnitt	94.1	88.8	72.1	

Tabelle 26: Blockdesign

Wir haben zwei Faktoren: "Darbietungsform" mit den Stufen A, B und C und "Blocks" mit den Stufen 1 bis 4. Da wir nur eine Beobachtung pro Zelle haben, handelt es sich um eine Art Kendall-Plan.

Wir hätten auch die 12 Vpn direkt, also ohne vorherige Bildung von Blocks, per Zufall den drei Darbietungsformen zuweisen können und hätten dann eine Einweg-Varianzanalyse mit drei

Gruppen A, B und C gehabt.

Nachteil: Die gesamte "Varianz innerhalb der Darbietungsform" wäre Fehlervarianz. Bei unserem Versuchsplan mit Blockbildung können wir dagegen feststellen:

Es besteht eine bemerkenswert hohe "Varianz zwischen den Blocks": Die Unterschiede innerhalb der Spalten sind zum großen Teil Unterschiede zwischen den Blocks; sie gehen vor allem darauf zurück, daß die Vpn des ersten Blocks am schnellsten sind, die Vpn des vierten Blocks am langsamsten, usw. Das liegt natürlich an der durch den Vortest bewerkstelligten Bildung von Blocks gleich guter Vp. Nach Abzug der "Varianz zwischen den Blocks" bleibt dann nur noch eine geringe Restvarianz innerhalb der Spalten. Diese kleine Restvarianz kommt dann in den Nenner des F -Bruchs in der Varianzanalyse, und somit wird dieser F -Bruch leichter groß. Wir erzielen also eher ein signifikantes F , falls H_0 falsch ist.

Anders ausgedrückt: Die Teststärke nimmt auch hier zu!

Wichtig: Wenn wir den t -Test für Paardifferenzen auf Vpn-Paare anwenden, dann müssen wir durch Zufall (z.B. mit Losen) entscheiden, welche Vp unter welcher Bedingung gemessen wird. Ausnahme: Die Vpn bringen die Zugehörigkeit zu den Gruppen als organismische Variable mit ins Experiment.

Beispiel: S.@ Erstgeborene - Zweitgeborene.

Ähnlich muß hier in jedem Block durch Zufall entschieden werden, welche Vp des Blocks unter Form A, welche unter Form B und welche unter Form C getestet wird. Man spricht daher auch von einem Versuchsplan mit "randomisierten Blocks" (engl.: "randomized blocks design").

8) Die modellhafte Einordnung von varianzanalytischen Versuchsplänen

Zu einem Bericht über ein varianzanalytisches Experiment gehört eine Beschreibung des Versuchsplans. Erst diese gibt dem Leser die Möglichkeit, sich darüber klarzuwerden, welchen genauen Aussagegehalt die F -Tests für die einzelnen Varianzquellen haben. Vom Inhalt dieser Beschreibung hängt es auch ab, welches varianzanalytische Modell man anlegt, d.h. welche Berechnungen anzustellen sind, bzw. (für die Praxis) in welchem Kapitel eines Handbuchs der Varianzanalyse man nachschlägt.

a) Ein Beispielexperiment

Was dabei alles zu beachten ist, wollen wir am Beispiel des folgenden Experiments behandeln. Die von den Vpn zu lösende Aufgabe läuft in der Fachliteratur unter dem Namen "Turm von Hanoi". Es handelt sich dabei um eine Art Brettspiel auf einem Brett mit drei Feldern. Gespielt wird mit fünf⁹⁷ kreisförmigen Scheiben, die alle verschieden groß sind. Die Aufgabe besteht darin, mit den Scheiben auf einem bestimmten der drei Felder einen Turm zu bauen, in welchem

⁹⁷Häufig wird das Spiel mit mehr Scheiben gespielt; in unserem Experiment jedoch mit fünf.

die Scheiben der Größe nach geordnet sind - die größte unten, die kleinste oben. Dabei sind die folgenden Spielregeln zu beachten: Liegen auf einem Feld mehrere Scheiben, dann müssen sie übereinander liegen. Bewegt werden darf immer nur eine Scheibe, und zwar darf sie nur auf eine größere Scheibe oder ein leeres Feld, nie aber auf eine kleinere Scheibe gelegt werden. Meßwert ist die Zahl der Züge, die die Vp braucht.

Die Schwierigkeit der Aufgabe hängt natürlich von der Ausgangsstellung ab. In unserem Experiment wird mit zwei Ausgangsstellungen A und B gespielt. Bei Stellung A liegen alle Scheiben auf dem ersten Feld; bei Stellung B liegen die zwei größten Scheiben auf dem ersten, die drei kleinsten auf dem zweiten Feld. Der Turm muß in beiden Fällen auf dem dritten Feld gebaut werden.

Jede Vp spielt beide Aufgaben zweimal. Die eine Hälfte der Vpn spielt zuerst zweimal mit Ausgangsstellung A und dann zweimal mit Ausgangsstellung B ("normale Reihenfolge"), die andere Hälfte der Vpn spielt in "umgekehrter Reihenfolge": zuerst zweimal B, dann zweimal A.

Vor diesen Hauptversuchen können die Vpn eine Viertelstunde lang sich mit der Technik des Spiels vertraut machen: Sie spielen mit einer für alle Vpn gleichen Ausgangsstellung, die weder mit A noch mit B etwas zu tun hat.

Von den 400 Psychologiestudenten eines Instituts, an dem das Spiel noch unbekannt ist, werden durch Zufall 20 zu Vpn bestellt. Sie werden rein nach Zufall weiterhin in vier Gruppen: N_a , N_b , U_a und U_b aufgeteilt. Die Vpn der Gruppen N_a und N_b spielen in "normaler Reihenfolge", die Vpn der Gruppen U_a und U_b in "umgekehrter Reihenfolge". Die Vpn der Gruppen N_a und U_a dürfen sich während des Spiels schriftliche Notizen machen, die Vpn der Gruppen N_b und U_b nur während der Vorübungsphase, aber nicht während der Hauptversuche. Sie finden die Ergebnisse (Zugzahlen) in Tabelle 27 zusammengefaßt.

Gruppe	Vp	Aufgabe A		Aufgabe B		Zeilen Ø
		1. Versuch	2. Versuch	1. Versuch	2. Versuch	
Na	1	90	76	82	77	81.25
	2	95	87	95	84	90.23
	3	81	85	82	79	81.75
	4	94	82	76	74	79.00
	5	75	75	67	58	68.75
	Ø Na	85	81	80.4	74.4	80.20
Nb	6	93	89	92	83	89.26
	7	88	81	83	76	82.00
	8	97	89	93	86	91.25
	9	110	98	104	108	105.00
	10	95	81	83	73	83.00
	Ø Nb	96.6	87.6	91	85.2	90.10
Ua	11	87	74	95	88	86.00
	12	95	79	100	89	90.75
	13	72	69	90	81	78.00
	14	76	68	84	87	78.76
	15	102	76	109	98	95.50
	Ø Ua	86.4	72.6	95.6	80.6	85.60
Ub	16	85	71	99	82	84.25
	17	102	89	109	105	101.25
	18	110	101	120	115	111.50
	19	103	98	116	109	106.50
	20	96	99	110	98	100.75
	Ø Ub	99.2	91.6	110.8	101.8	100.85

Durchschnitte bei Gruppenzusammenfassung:

Gruppe N = (Na + Nb)	90.8	84.3	85.7	79.8	85.15
Gruppe U = (Ua + Ub)	92.8	82.1	103.2	95.2	93.33
Gruppe a = (Na + Ua)	85.7	76.8	88.0	81.5	83.00
Gruppe b = (Nb + Ub)	97.9	89.6	100.9	93.5	95.48
alle vier Gruppen	91.8	83.2	94.45	87.5	89.24

Tabelle 27: Daten zum Turm-von-Hanoi-Experiment.

Ø steht für "Durchschnitt"

b) Überlegungen zu den einzelnen Faktoren

ba) Feststellung der Faktoren

Selbstverständlich muß aus jeder varianzanalytischen Versuchsbeschreibung hervorgehen, welche Faktoren mit welchen Stufen vorliegen. Insbesondere ist zu beachten, ob es auch einen Faktor "Vpn" bzw. "Vpn-Blocks" gibt (vgl. Abschnitt 8!).

Welche Faktoren und welche Stufen liegen in unserem Experiment vor?

Zunächst der Faktor "Ausgangsstellung" mit den Stufen "Erstversuch" und "Zweitversuch". Jede Vp hat bei jeder Ausgangsstellung zwei Versuche. Zunächst wäre man versucht, hier genau so zu verfahren wie im Reaktionszeitexperiment aus Abschnitt 8 und die beiden Versuche zur gleichen Aufgabenstellung als zwei Beobachtungen der gleichen Zelle zu betrachten. Das ist aber nicht zulässig, und zwar aus psychologischen Gründen. Während im Reaktionszeitexperiment bei entsprechender Vorbereitung der Vpn keine nennenswerten Lerneffekte auftreten, ist dies bei Problemlöseaufgaben wie der hier gestellten der Fall. Beobachtungen innerhalb der gleichen Zelle müssen aber austauschbar sein, d.h. sie müssen aus der gleichen Populationsverteilung bzw. Wahrscheinlichkeitsverteilung stammen. Das kann man im Reaktionszeitexperiment annehmen; in unserem Experiment sinkt der Erwartungswert der Zugzahl, da die Vpn lernen. Wenn aber der Erwartungswert sich ändert, dann kann es sich nicht mehr um einen Meßwert aus der gleichen Wahrscheinlichkeitsverteilung handeln. Also können wir "Erstversuch" und "Zweitversuch" nicht als Beobachtungen der gleichen Zelle auffassen. Es bleibt der genannte Ausweg, einen eigenen Faktor "Versuchsnummer" mit den Stufen "Erstversuch" und "Zweitversuch" anzunehmen.

Die insgesamt vier Versuche jeder Vp sind also nicht vier Stufen eines Faktors; vielmehr entstehen sie durch die Kreuzung von zwei Faktoren mit je zwei Stufen.

Weniger problematisch sind die Faktoren Reihenfolge mit den Stufen "normal (n)" und "umgekehrt (U)" sowie der Faktor "Notizen" mit den Stufen "erlaubt (A)" bzw. "nur während der Vorübungsphase erlaubt (B)".

Gibt es bei diesem Experiment auch einen Faktor Vpn?

Ja. Faustregel: Immer, wenn mehrere Messungen pro Vpn vorliegen, gibt es einen Faktor Vpn.

Gibt es einen Faktor "Vpn-Block"?

Nein. Es ist in der Versuchsbeschreibung nichts davon gesagt, daß solche Blocks von gleich guten Vpn gebildet worden sind. Die Vpn sind den Gruppen direkt per Zufall zugewiesen worden.

Wie hätte eine Bildung von "Blocks" hier ausgesehen?

Man hätte aufgrund eines Vortests (z.B. aufgrund der Leistung in der für alle Vpn gleichen Vorübungsphase) "Blocks" von jeweils vier gleich guten Vpn bilden können und aus jedem Block per Zufall je eine Vp den vier Gruppen N_a , N_b , U_a und U_b zugewiesen. In diesem Fall hätte man

zusätzlich zu dem Faktor Vpn auch noch einen Faktor "Blocks".⁹⁸ Falls man solche "Blockbildung" vornimmt, muß in der Versuchsbeschreibung drauf hingewiesen werden. Da in der Versuchsbeschreibung nichts davon steht, ist anzunehmen, daß keine Blockbildung vorgenommen wurde.

Mit diesen Überlegungen sind die Faktoren und ihre Stufen beschrieben. Wir sehen: Meistens waren die Angaben aus der Versuchsbeschreibung ausreichend, um Faktoren aufzufinden. In einem Fall (Faktor "Versuchsnummer") waren jedoch zusätzlich zur Versuchsbeschreibung theoretische Überlegungen erforderlich, um das angemessene Faktorenmodell zu ermitteln.

bb) "Fixe" oder "zufällige" Effekte?

Ein noch größeres Gewicht haben theoretische Überlegungen bei einer zweiten Frage, die ebenfalls für jeden Faktor einzeln zu untersuchen ist: Handelt es sich um einen Faktor mit "fixen" oder mit "zufälligen" Effekten? D.h.: Sind die Stufen, die der Faktor im Experiment hat, alle Stufen, über die eine Aussage gemacht werden soll, oder sind sie als eine Stichprobe aus einer Population zu betrachten, über die eine Aussage gemacht werden soll? Im ersten Fall haben wir fixe Effekte, im zweiten Zufallseffekte. Zur Veranschaulichung des damit Gemeinten wollen wir auf ein einfacheres Experiment zurückgreifen, nämlich auf das mit den verschiedenen Arbeitsbedingungen ("hübsch", "nüchtern" und "altmodisch").

Nehmen wir an, die Fragestellung hätte gelaute: Unterscheidet sich die durchschnittliche Arbeitsleistung unter diesen drei Arbeitsbedingungen? Es soll also nur über die Arbeitsleistung unter diesen drei Arbeitsbedingungen eine Aussage gemacht werden.

Wäre dann der Faktor Arbeitsbedingung ein Faktor mit fixen oder mit zufälligen Effekten?

Da nur über die drei im Experiment auftretenden Stufen eine Aussage gemacht werden soll, handelt es sich beim Faktor Arbeitsbedingung um einen Faktor mit fixen Effekten.

Die Fragestellung könnte aber auch lauten: Es gibt im Grunde sehr viele mögliche Arbeitsbedingungen. Wir können diese unmöglich alle in unser Experiment aufnehmen. Wir betrachten unsere drei Arbeitsbedingungen als Zufallsstichprobe aus der Vielzahl aller möglichen Arbeitsbedingungen. Anhand dieser drei ausgewählten Stufen wollen wir nun die Frage beantworten: Unterscheidet sich die durchschnittliche Arbeitsleistung unter den verschiedenen Arbeitsbedingungen, aus denen unsere drei Stufen eine Auswahl sind?

Wieso handelt es sich beim Faktor Arbeitsbedingung jetzt um einen Faktor mit Zufallseffekten?

Weil die drei Stufen, die der Faktor Arbeitsbedingung im Experiment hat, nicht alle Stufen sind, über die eine Aussage gemacht werden soll, sondern als Stichprobe aus dieser Gesamtheit gelten. Diese Betrachtungsweise ist natürlich nur sinnvoll, wenn es sich tatsächlich um eine

⁹⁸Jede Vp hätte, wie im bisherigen Versuchsplan, insgesamt vier Versuche in der Hauptphase gehabt: zwei auf Problem A und zwei auf Problem B.

Zufallsstichprobe der möglichen Arbeitsbedingungen handelt, etwa dann, wenn die drei Arbeitsbedingungen durch mehrfaches Würfeln oder durch Losen aus der Gesamtheit aller Arbeitsbedingungen, über die eine Aussage gemacht werden soll, als Stichprobe gezogen wurde.

Warum spricht man wohl in einem Fall von "Fix-Effekten", im anderen von "Zufallseffekten"?

Bei der ersten Fragestellung ist schon vor der Durchführung des Experiments festgelegt ("fix"), mit welchen Effekten der Faktor "Arbeitsbedingung" ins Experiment eingeht: Mit den Effekten a_h , a_n und a_a . Die Höhen dieser Effekte sind zwar unbekannte, aber doch feststehende ("fixe") Populationsparameter. Bei der zweiten Fragestellung hängt es dagegen vom Zufall ab, mit welchen Effekten der Faktor auftritt: Es könnten ganz andere Stufen ins Experiment kommen und damit auch ganz andere Effekte, die nicht nur dem Namen sondern auch der Höhe nach von den drei genannten verschieden sind.

In diesem Sinn sind die Effekte des Faktors "Arbeitsbedingung" im einen Fall fixe (wenn auch unbekannte) Populationsparameter, im anderen Fall zufällige Größen.

Vereinfachend nennt man einen Faktor mit fixen Effekten auch Fix-Faktor und einen Faktor mit Zufallseffekten Zufallsfaktor.

Den Faktor Geschlecht kann man nicht sinnvoll als Zufallsfaktor betrachten: Es gibt keine Population von Geschlechtern, aus denen die zwei Geschlechter "männlich" und "weiblich" eine Stichprobe sind! Trotzdem muß auch bei der Frage nach der Varianz zwischen Geschlechtern zunächst untersucht werden, ob der Faktor Arbeitsbedingung als Fix-Faktor oder als Zufallsfaktor zu betrachten. Die Fragestellung könnte lauten: Unterscheiden sich die durchschnittlichen Leistungen von Männern und Frauen in den drei Arbeitsbedingungen, die wir untersuchen?

Stellen Sie für beide Fragestellungen fest, ob der Faktor "Arbeitsbedingung" ein Fix-Faktor oder ein Zufallsfaktor ist!

Im ersten Fall sollen eventuelle Aussagen über Unterschiede im Leistungsdurchschnitt von Männern und Frauen nur für die drei im Experiment vorhandenen Stufen des Faktors Arbeitsbedingung gelten. Folglich ist der Faktor "Arbeitsbedingung" bei dieser Fragestellung ein Fix-Faktor. Er ist dagegen ein Zufallsfaktor, wenn Aussagen über Unterschiede in der durchschnittlichen Leistungsfähigkeit nicht nur für diese drei Arbeitsbedingungen allein gemacht werden, sondern auf die Gesamtheit aller Arbeitsbedingungen, aus denen unsere drei eine Stichprobe darstellen, übertragen werden sollen.

Der Faktor "Geschlecht" ist dagegen aus den genannten Gründen bei beiden Fragestellungen ein Fix-Faktor: Es gibt immer nur den Unterschied der Durchschnitte von Männern und Frauen und nicht Durchschnittsunterschiede in einer Population von Geschlechtern, aus denen unsere zwei Geschlechter eine Stichprobe sind.

Formulieren Sie auch die Frage nach der Wechselwirkung zweifach, so daß der Faktor "Arbeitsbedingung" im einen Fall ein Fix-Effekt im anderen Fall ein Zufallseffekt ist!

Fix-Effekt: Unterscheiden sich die Geschlechtseffekte unter diesen drei Arbeitsbedingungen

(Anderer Aspekt: Sind die Effekte dieser drei Arbeitsbedingungen bei Männern und Frauen verschieden)?

Zufallseffekt: Unterscheiden sich die Geschlechtseffekte unter den vielen Arbeitsbedingungen, aus denen unsere drei eine Stichprobe sind (Anderer Aspekt: Sind die Effekte der vielen Arbeitsbedingungen bei Männern und Frauen verschieden)?

Wir haben am Beispiel eines relativ einfachen varianzanalytischen Experiments gesehen, wie viele Varianten von Fragestellungen es gibt, die jeweils zu verschiedenen Einstufungen der Faktoren als Fix- oder Zufallsfaktoren führen. Es würde natürlich zu weit führen, die noch viel größere Vielfalt von Fragestellungen zu verfolgen, die sich aus unserem Experiment zum "Turm von Hanoi" ergäbe. Daher nur ein Beispiel mit zwei Fragestellungen. Das Beispiel demonstriert auch, daß der wichtigste "Zufallsfaktor" in der Psychologie wohl der Faktor Vpn ist, sofern es einen solchen Faktor gibt. Meistens soll eine Aussage nicht nur über die tatsächlich untersuchten Personen gemacht werden, sondern über eine "Population" von Personen, aus der die untersuchten Personen eine Zufallsstichprobe darstellen.

Vergleichen wir die drittletzte und die vorletzte Zahl der letzten Spalte unserer Datentabelle, dann sehen wir: Der Gesamtdurchschnitt aller Versuche mit Notizen beträgt 83.00 Züge; ohne Notizen werden im Durchschnitt 95.48 Züge benötigt. Anscheinend braucht man mit Notizen im Durchschnitt weniger Züge. (Zur Signifikanzprüfung würde man die "Zwischen-Varianz" des Faktors Notizen heranziehen)!

1. Fragestellung: Besteht bei den 400 Studenten, aus denen unsere 20 eine Stichprobe sind, der genannte Effekt des Notizenmachens auf die Zugzahl, wenn man ihnen diese beiden Ausgangsstellungen vorlegt, wenn man weiterhin durch Los entscheidet, in welcher der beiden möglichen Reihenfolgen die Aufgaben vorgelegt werden und wenn jede Vp zwei direkt aufeinanderfolgende Versuche hat, deren Zugzahlen beide berücksichtigt werden?

Welche Faktoren sind Fix-Faktoren, welche Zufallsfaktoren?

Die Faktoren "Aufgabe" und "Versuchsnummer" sind eindeutig Fix-Faktoren: Es wird nach dem "Notizeneffekt" nur für diese beiden Aufgaben und nur für Erst- und Zweitversuche gefragt. Beim Faktor "Notizenmachen" könnte man bei etwas anderer Fragestellung und Versuchsplanung einen Zufallsfaktor annehmen. Es gäbe ja viele weitere mögliche Stufen des Faktors "Notizen" außer unseren Stufen a (Notizen generell erlaubt) und b (Notizen in der Vorübungsphase erlaubt, in den Hauptversuchen verboten). es wäre eine Stufe c denkbar, bei der die Vpn nur nach jedem zehnten Versuch Notizen machen dürfen, oder eine Stufe d, bei der die Vpn in jedem Versuch nur bis zum 30. Zug Notizen machen dürfen und von da an aus dem Kopf spielen müssen usw.

Würde die Frage lauten: Bestehen irgendwelche Einflüsse der verschiedenen möglichen Grade der Notizerlaubnisse, aus denen zufällig diese zwei (a und b) herausgegriffen sind, dann hätten wir einen Zufallsfaktor.

Gefragt ist aber nach "dem genannten Effekt des Notizenmachens", d.h. nach dem Unterschied von a und b. Die Stufen, über die eine Aussage gemacht werden soll, sind alle im Experiment repräsentiert. Also handelt es sich um einen Fix-Faktor.

Beim Faktor "Reihenfolge" verführt die Zuweisung durch Los zur Annahme eines Zufallsfaktors. Entscheidend für die Frage "Zufallsfaktor oder Fix-Faktor" ist aber nicht, ob die Vpn den Stufen (hier: Reihenfolge) des fraglichen Faktors durch Los zugewiesen werden, sondern ob alle zur Diskussion stehenden Stufen im Experiment vertreten sind oder ob nur eine Stichprobe der Stufe vorhanden ist. Ersteres ist der Fall: Bei zwei Aufgaben gibt es nur zwei mögliche Reihenfolgen, wenn die Versuche zu jeder Aufgabe direkt aufeinander folgen (und nur danach ist gefragt); beide Reihenfolgen sind im Experiment vertreten. Also ist auch der Faktor "Reihenfolge" ein Fix-Faktor.

Schließlich der Faktor Vpn. Gefragt ist nach dem Notizeneffekt bei den 400 Studenten, aus denen die 20 Vpn, die Stufen unseres Vpn-Faktors, eine Stichprobe sind. Damit handelt es sich um einen Zufallsfaktor.

2. Fragestellung: Besteht bei den 400 Studenten, aus denen unsere 20 eine Stichprobe sind, der genannte Notizeneffekt, wenn man ihnen irgendwelche zwei Ausgangsstellungen vorlegt, wenn man weiterhin durch Los entscheidet.... (usw. wie bei der ersten Fragestellung)?
Welche Faktoren sind jetzt zufällig und welche fix?

An der Formulierung der Fragestellung hat sich nur hinsichtlich der Ausgangsstellung etwas geändert. Jetzt wird nicht mehr nach dem Notizeneffekt nur bei den zwei Ausgangsstellungen A und B, sondern bei irgendwelchen zwei Ausgangsstellungen gefragt. A und B sind jetzt eine Stichprobe aus den vielen möglichen Ausgangsstellungen, nach denen gefragt wird. Der Faktor ist damit ein Zufallsfaktor. Die übrigen sind geblieben wie bei der ersten Fragestellung.

Die Anwendung des Modells der Zufallseffekte auf dem Faktor "Ausgangsstellung" wäre natürlich nur dann sinnvoll, wenn man die tatsächlichen Stufen "A" und "B" durch Zufall aus den vielen möglichen Ausgangsstellungen herausgesucht hätte.

Die ganze Unterscheidung von Fix-Effekten und Zufallseffekten erscheint zunächst reichlich theoretisch. Es läßt sich aber mathematisch beweisen, daß die von uns besprochenen Rechen-techniken und Zusatzvoraussetzungen der Varianzanalyse im allgemeinen nur bei Fix-Effekten ausreichend sind. Bei Zufallseffekten können insbesondere andere Mittelquadrate als das MQ_{in} in den Nenner der F -Brüche kommen; außerdem kommen weitere Zusatzvoraussetzungen ins Spiel.

Die Unterscheidung hängt - wie mehrfach betont - von der Fragestellung ab, aber auch davon, daß wirklich eine Zufallsstichprobe aus der Stufenpopulation gezogen werden muß, wenn man das Modell der Zufallseffekte verwenden will. Ersteres geht natürlich aus der Versuchsbeschreibung nicht hervor, sondern aus der vorhergehenden Ableitung der Hypothesen (Fragestellungen), dagegen sollte die Methode der Stufenauswahl beim Zufallsmodell im Versuchsbericht erwähnt werden.

c) Das Gesamtmodell

ca) Das Verhältnis der Faktoren zueinander

Die bisherigen Fragestellungen betrafen die einzelnen Faktoren und ihre Stufen. Daneben sind auch Feststellungen über das Verhältnis der einzelnen Faktoren zueinander von Bedeutung. Insbesondere ist nach der Orthogonalität sowie nach "Überkreuzung" oder "Schachtelung" der Faktoren zu fragen.

Wir müssen dazu zunächst den bisher nur für zwei Faktoren definierten Begriff der Orthogonalität erweitern: Orthogonalität liegt vor, wenn für jeden Faktor gilt, daß das Verhältnis der Zahl der Beobachtungen, die auf den verschiedenen Stufen eines Faktors gemacht werden, bei allen Stufenkombinationen gleich sind.

Beispiel aus dem Schlafexperiment zur Dreiweg-Varianzanalyse: Wir untersuchten zunächst für den Faktor "Neurotizismus": Das Verhältnis von Beobachtungen auf den Stufen "normal" und "neurotisch" ist bei der Stufenkombination "Männer unter Schlafmittel" 50 : 50; ebenso auch bei den Stufenkombinationen "Männer unter Placebo", "Frauen unter Schlafmittel" und "Frauen unter Placebo". damit ist die Forderung der Gleichheit der Verhältnisse der Zahl der Beobachtungen, die auf den Stufen des Faktors Neurotizismus gemacht werden, für alle vier Kombinationen von Stufen der zwei übrigen Faktoren überprüft. Was hier für den Faktor Neurotizismus festgestellt wurde, müßte auch für alle übrigen Faktoren überprüft werden. Es gibt allerdings eine sehr einfache

Regel: Werden in allen Zellen gleich viele Beobachtungen gemacht, dann liegt Orthogonalität vor.

Das ist im Schlafexperiment der Fall, aber auch in unserem Experiment zum "Turm von Hanoi".

Was wäre hier eine Zelle?

Eine Zelle ist - allgemein - durch die Kombination von Stufen aller Faktoren definiert. Das Feld "Gruppe N_b, Aufgabe A, Erstversuch" wäre eine solche Kombination von Stufen aller fünf Faktoren. In jeder solchen Zelle steckt eine Beobachtung, also gleiche Beobachtungszahl also Orthogonalität.

Die Orthogonalität geht indirekt aus der Versuchsbeschreibung hervor, wenn - was selbstverständlich ist - die Zahl der Beobachtungen für jede Zelle klar ist. Es ist jedoch zu fordern, daß man sich vor der Datenerhebung auch über die Orthogonalität Sicherheit verschafft und bei eventuellen Verletzungen der Orthogonalität (bei denen ja, wie schon erwähnt, besondere Rechentechniken notwendig werden) auf diese Verletzungen in der Versuchsbeschreibung ausdrücklich hinweist.

Neben der Orthogonalitätsfrage ist vor allem die Frage von Bedeutung, ob die Faktoren "vollständig überkreuzt" ("completely crossed") sind. Das heißt: Für jedes Faktorenpaar ist die Frage zu untersuchen, ob jede im Experiment vertretene Stufe des einen Faktors mit jeder im

Experiment vertretenen Stufe des anderen Faktors kombiniert auftritt. Bei unseren früheren Experimenten war das immer der Fall.

Beispiel: Bei dem Experiment mit den Arbeitsbedingungen: Jede Arbeitsbedingung (Stufe des einen Faktors) trat mit beiden Geschlechtern (Stufen des anderen Faktors) kombiniert auf. Es treten nicht beide Geschlechter mit allen nur denkbaren Stufen des Faktors "Arbeitsbedingung" kombiniert auf. Das ist für die Frage nach der "Überkreuzung" irrelevant. Entscheidend ist hier, ob beide Geschlechter mit allen *im Experiment vertretenen* Stufen des Faktors Arbeitsbedingung kombiniert auftreten. Wäre dagegen der Versuchsleiter ein Kavalier gewesen und hätte unter Arbeitsbedingung "altmodisch" nur Männer antreten lassen, dann wäre die "vollständige Überkreuzung" verlorengegangen. Den Tabellen zu diesem Experiment (z.B. Tabelle @) kann man unmittelbar den Sinn des Wortes "überkreuzt" entnehmen: Jede "Zelle" repräsentiert die Kombination einer Stufe des Faktors Arbeitsbedingung mit einer Stufe des Faktors Geschlecht; die Zelle entsteht an der Stelle der *Überkreuzung* der entsprechenden Zeile.

Im Schlafexperiment haben wir uns die Frage nach der "Überkreuzung" für drei Faktorenpaare zu stellen:

Wie ist das Ergebnis?

Beide Geschlechter treten mit beiden Stufen des Faktors "Neurotizismus" und mit beiden Stufen des Faktors "Droge" kombiniert auf. Der Faktor "Geschlecht" ist also mit den Faktoren "Neurotizismus" und "Droge" vollständig überkreuzt. Dies gilt aber auch für das Verhältnis der Faktoren "Neurotizismus" und "Droge": Schlafmittel und Placebo (Stufen des Faktors Droge) werden mit beiden Stufen des Faktors "Neurotizismus" kombiniert (d.h. beide Drogen werden an Neurotiker und an Normale verabreicht).

Für eine anschauliche Darstellung der Überkreuzung müßte man hier mit einer dreidimensionalen Tabelle arbeiten, etwa indem man die beiden Datentabellen für Männer und Frauen aus Abschnitt 7a übereinanderlegt.

Wie ist es in unserem Beispiel zum "Turm von Hanoi"?

Zunächst: Wie ist das Verhältnis der Faktoren "Ausgangsstellung" und "Versuchsnummer"?

Bei beiden Ausgangsstellungen gibt es einen Erstversuch und einen Zweitversuch. Alle Stufen des Faktors "Ausgangsstellung" werden also mit allen Stufen des Faktors "Versuchsnummer" kombiniert. Die beiden Faktoren sind also vollständig überkreuzt.

Wie ist das Verhältnis dieser beiden Faktoren zu den übrigen Faktoren?

Auf beiden Stufen des Faktors "Reihenfolge" und auf allen 20 Stufen des Faktors "Vpn" werden beide im Experiment vertretenen Ausgangsstellungen vorgelegt, und zwar mit Erstversuch und Zweitversuch. Die Faktoren "Ausgangsstellung" und "Versuchsnummer" sind also nicht nur miteinander, sondern auch mit den drei übrigen Faktoren vollständig überkreuzt.

Wie ist es mit dem Verhältnis der Faktoren "Reihenfolge" und "Vpn"?

Sie wären nur dann vollständig überkreuzt, wenn jede Vp das Experiment in normaler und einmal in umgekehrter Reihenfolge mitmachen würde, denn nur dann wäre jede Stufe des Faktors "Reihenfolge" mit jeder Stufe des Faktors "Vpn" kombiniert. Statt der Überkreuzung liegt hier

ein anderes Verhältnis vor.

Jede Stufe des Faktors Vpn, also jede Vp, wird nur in einer Reihenfolge getestet. Als Gegenstück zum Bild der Überkreuzung bietet sich hier das der Schachtelung an: So wie in einer Schachtel mehrere Gegenstände sein können, wobei aber ein Gegenstand nicht in mehreren Schachteln sein kann, so sind mehrere Vp ("Gegenstände") mit nur jeweils einer Stufe des Faktors "Reihenfolge" kombiniert ("hineingeschachtelt").

Allgemein wollen wir dann, wenn mehrere Stufen eines ersten Faktors jeweils zusammengebündelt mit nur einer Stufe eines zweiten Faktors kombiniert sind, dieses Verhältnis verkürzt zum Ausdruck bringen, indem wir sagen: Die Stufen des ersten sind in die Stufen des zweiten hineingeschachtelt (engl.: "nested").

Wie ist das Verhältnis des Faktors "Notizen" zum Faktor "Vpn"?

Mehrere Vpn (die Vpn 1-5 und 11-15) sind zusammengebündelt und nur mit der Stufe a ("Notizen erlaubt") kombiniert; die übrigen Vpn sind gemeinsam nur mit der Stufe b kombiniert. Die Stufen des Faktors "Vpn" sind also auch in die Stufen des Faktors "Notizen" hineingeschachtelt.

Wie ist schließlich das Verhältnis der Faktoren "Reihenfolge" und "Notizen" zueinander?

Auf beiden Stufen des Faktors "Notizen" wird in beiden Reihenfolgen getestet. Die Faktoren "Notizen" und "Reihenfolge" sind also vollständig überkreuzt.

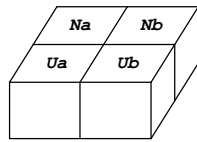
Eine adäquate, anschauliche Darstellung des Verhältnisses aller Faktoren zueinander wäre wegen der vielen Überkreuzungen nur in einem vierdimensionalen Raum möglich. Formulieren ließe sie sich folgendermaßen:

Die Vpn werden am besten dargestellt durch zweidimensionale Tabellen mit vier Feldern der folgenden Art@:

	Erstversuch	Zweitversuch
Ausgangsstellung A	90	76
Ausgangsstellung B	82	77

Tabelle @: Daten der Vp 1

Die vier Daten jeder Vp werden in eine solche Tabelle eingetragen, und diese Tabellen sind in insgesamt vier Schachteln verteilt (die vier Gruppen Na, Nb, Ua und Ub). Die Faktoren "Reihenfolge" (mit den Stufen N und U) und "Notizen" (mit den Stufen a und b), durch die sich vier Schachteln unterscheiden, stehen zueinander im Verhältnis der Überkreuzung (vgl. Abb.@).



Abbildung@: Schachteln mit überkreuzten Faktoren

Dieses Schachtelungsbild ist auch deshalb adäquat, weil jede V_p nur unter einer der vier Stufenkombinationen N_a , N_b , U_a und U_b getestet wird. Eine dieses Bild sinnvoll ergänzende Unterbringung der Faktoren "Aufgabenstellung" und "Versuchsnummer" wäre nur bei mehr als drei Dimensionen möglich.

Für eine Erweiterung des Begriffs der Schachtelung stellen wir uns eine andere Planung vor: Die dritte Gruppe sei - genau wie bisher - mit umgekehrter Reihenfolge der Aufgaben getestet worden, habe aber nur nach jedem zehnten Zug Notizen machen dürfen (Stufe c des Faktors Notizen). Die V_{pn} der vierten Gruppe - ebenfalls bei umgekehrter Reihenfolge - hätten in jedem Versuch nur bis zum 30. Zug Notizen machen dürfen und von da an aus dem Kopf spielen müssen (Stufe d). Die beiden letzten Gruppen würde man dann als U_c und U_d bezeichnen, während die beiden ersten Gruppen wie bisher N_a und N_b wären.

Wie wäre jetzt das Verhältnis der Faktoren "Reihenfolge", "Notizen" und "Vpn"?

Geändert hat sich vor allem das Verhältnis der Faktoren "Notizen" und "Reihenfolge": Es treten nicht alle vier Stufen des Faktors "Notizen" mit beiden Stufen des Faktors "Reihenfolge" kombiniert auf: Die Stufe a und b sind nur mit normaler Reihenfolge kombiniert, die Stufen c und d nur mit umgekehrter. Die Stufen des Faktors "Notizen" sind also in die Stufen des Faktors "Reihenfolge" hineingeschachtelt. Andererseits sind aber auch die Stufen des Faktors "Vpn" in die Stufen des Faktors "Notizen" hineingeschachtelt. Es ergibt sich das in Abbildung@ dargestellte Bild:

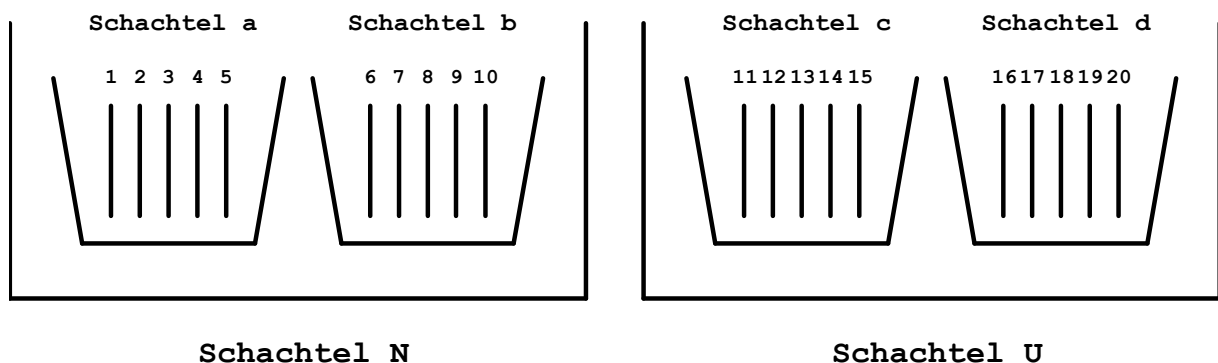


Abbildung @: Zweistufige Schachtelung

Die Stufen a, b, c und d des Faktors "Notizen" sind kleine Schachteln (mit schrägen Seiten), in die die V_{pn} (numerierte senkrecht Striche) hineingeschachtelt sind; jeder senkrechte Strich

steht dabei für die "Seitenansicht" einer Tabelle, in der die vier Daten der jeweiligen Vp so eingetragen sind, wie es Tabelle @28 für Vp 1 darstellt. Die kleinen Schachteln sind selbst wieder in große Schachteln (mit senkrechten Wänden) hineingeschachtelt.

Das wesentlich Neue gegenüber unserem ursprünglichen Beispiel ist hier das Auftreten einer zweistufigen Schachtelung.

Natürlich sind neben vollständiger Überkreuzung und Schachtelung noch weitere Verhältnisse der Kombination von Faktorstufen denkbar, die aber in der Praxis wenig bedeuten haben.

Ob es sich beim Verhältnis der Faktoren um Schachtelung oder vollständige Überkreuzung handelt, sollte aus jedem varianzanalytischen Bericht hervorgehen; allerdings wird man selten ausdrückliche Formulierungen finden. In unserer Versuchsbeschreibung (in Abschnitt 9 a) geht die vollständige Überkreuzung der Faktoren "Vpn", "Ausgangsstellung" und "Versuchsnummer" aus dem Satz hervor:

"Jede Vp spielt jede Aufgabe zweimal".

Die Schachtelung kommt zum Ausdruck in dem Satz:

"Sie (die Vpn) werden in vier Gruppen Na, Nb, Ua und Ub aufgeteilt" und in der darauffolgenden Beschreibung der vier verschiedenen Bedingungen, unter denen die vier Gruppen arbeiten.

Es ist schließlich fast selbstverständlich, daß in jeder durch die Überkreuzung bzw. Schachtelung der Faktoren entstehenden Zellen mindestens eine Beobachtung sein soll. Wenn also mehrere Faktoren alle paarweise überkreuzt sind, dann sollte auch jede denkbare Kombination von Stufen dieser Faktoren durch eine Zelle mit mindestens einer Beobachtung repräsentiert sein. Erst dann spricht man von vollständiger Überkreuzung dieser Faktoren.

Theoretisch denkbar sind Verletzungen dieser Regel bei mindestens drei überkreuzten Faktoren.

Würde beispielsweise in unserem Schlafmittelexperiment die Zelle "neurotische Männer unter Placebo" fehlen, dann wären zwar alle Faktoren *paarweise* überkreuzt, man würde aber nicht von einem "Versuchsplan mit vollständiger Überkreuzung aller Faktoren" sprechen.

Wollen Sie dies überprüfen?

Man könnte von einem Extremfall der Verletzung der Orthogonalität sprechen, indem man sagt, die Zahl der Beobachtungen in dieser Zelle sei gleich Null. Die Möglichkeiten, auch solche Fälle mit den Methoden für nicht-orthogonale Versuchspläne zu erfassen, sind aber äußerst beschränkt!

cb) Das mathematische Gesamtmodell

Unsere bisherigen Überlegungen zur formalen Erfassung von Versuchen befaßten sich vor allem mit der rein logischen Struktur des Versuchsplans. Der Abschluß der formalen Erfassung

ist die Formulierung eines mathematischen Modells.

Zunächst ist hier die Frage zu klären, ob es sich um "Modell I", "Modell II" oder um ein gemischtes Modell handelt.

Definitionen:

Von Modell I spricht man, wenn alle Faktoren Fix-Faktoren sind. Das Modell II liegt vor, wenn alle Faktoren als Zufallsfaktoren zu behandeln sind.

Sind die Faktoren teilweise Fix-Faktoren und teilweise Zufallsfaktoren (wie etwa in unserem Beispielerperiment zum Turm von Hanoi), so liegt ein gemischtes Modell vor.

Gleichsam die Krönung des Ganzen ist dann die Modellgleichung, in der die Zusammensetzung jedes Meßwertes aus Populationsmittel, Haupteffekten, Wechselwirkungen und Meßfehler formuliert wird. (vgl. z.B. die oberste Gleichung auf Seite@).

Wichtig: In Abweichung von unserer bisherigen Praxis wählen die meisten Bücher in den Modellgleichungen für Fix-Effekte griechische und nur für Zufallseffekte lateinische Buchstaben. Wechselwirkungseffekte zählen dabei als Zufallseffekte, wenn mindestens einer der an der Wechselwirkung beteiligten Faktoren ein Zufallsfaktor ist.

d) Zusammenfassung

Zur modellhaften Einordnung von varianzanalytischen Versuchsplänen stellt man zunächst die Faktoren und ihre Stufen fest (ba) und überlegt sich, ob es sich um Fix-Faktoren oder Zufallsfaktoren handelt (bb). Weiterhin ist die Frage nach der Orthogonalität zu prüfen und zu untersuchen, ob vollständige Überkreuzung oder Schachtelung vorliegt (ca). Schließlich formuliert man ein entsprechendes mathematisches Modell.

Erst aus der Gesamtheit dieser Feststellungen geht hervor, welche Berechnungen anzustellen sind (insbesondere, welches Mittelquadrat in den Nenner der einzelnen F -Brüche kommt). Es ist sogar möglich, daß im Einzelfall eines bestimmten Modells kein adäquater Signifikanztest für bestimmte Varianzquellen existiert. Von den genannten Feststellungen hängt es aber auch ab, welche Zusatzvoraussetzungen die einzelnen Signifikanztests haben. Die Feststellungen sind daher nicht nur erforderlich, um das im Einzelfall richtige Berechnungsmodell zu finden (praktisch: um das richtige Kapitel in einem entsprechenden Handbuch zu finden), sondern auch zur Prüfung der Frage, ob ein bestimmter varianzanalytischer Versuchsplan überhaupt zu Daten führt, für die es eine angemessene Varianzanalyse zur Beantwortung der jeweiligen Forschungsfrage gibt.

Daher eine ganz wichtige *Regel*: Der gesamte Plan für die Auswertung eines Experiments muß erstellt sein, bevor man anfängt, Daten zu erheben! Sonst kann es leicht passieren, daß man mit viel Aufwand Versuche durchführt und im Nachhinein feststellen muß, daß es für die Fragestellung, auf die es einem ankommt, keinen brauchbaren Test gibt!

Und das gilt nicht nur für die Varianzanalyse!!!

Inhaltsverzeichnis

B. Wahrscheinlichkeitstheorie	- 1 -
I. Die Begriffe "Wahrscheinlichkeit" und "Zufall"	- 7 -
1) Wahrscheinlichkeit	- 7 -
a) Wahrscheinlichkeitsbegriffe	- 7 -
b) Gegenseitige Abwägung der drei Wahrscheinlichkeitsbegriffe	- 8 -
2) Der Begriff des Zufalls	- 10 -
3) Modelle zur Veranschaulichung von Wahrscheinlichkeit und Zufall	- 10 -
a) Das Urnen-Modell	- 10 -
b) Das Venn-Diagramm	- 13 -
c) Zur Bedeutung mechanischer Wahrscheinlichkeitmodelle in der Psychologie	- 15 -
d) Die Angabe von Wahrscheinlichkeiten durch "Odds"	- 16 -
4) Bedingte Wahrscheinlichkeiten	- 17 -
a) Der Begriff der bedingten Wahrscheinlichkeit	- 17 -
b) Unabhängige und abhängige Ereignisse	- 20 -
II. Kombinatorik	- 25 -
1) Wahrscheinlichkeitssätze	- 25 -
a) Der Multiplikationssatz	- 25 -
b) Der einfache Additionssatz	- 27 -
c) Die Axiome der mathematischen Wahrscheinlichkeitstheorie	- 27 -
d) Der erweiterte Additionssatz	- 28 -
e) Negation	- 29 -
f) Das Bayes-Theorem	- 29 -
g) Die Anwendung der Wahrscheinlichkeitssätze auf bedingte Wahrscheinlichkeiten	- 32 -
h) Erweiterung von Multiplikationssatz und Additionssatz bei mehr als zwei Ereignissen	- 32 -
i) Ein Beispiel für die kombinierte Anwendung mehrerer Wahrscheinlichkeitssätze	- 36 -
2) Anzahlen	- 41 -
a) Permutationen	- 41 -
b) Teilsequenzen	- 45 -
c) Untergruppen (Kombinationen ohne Reihenfolge)	- 47 -
III. Diskrete Wahrscheinlichkeitsverteilungen	- 50 -
1) Eindimensionale diskrete Wahrscheinlichkeitsverteilungen	- 50 -
a) Der Begriff der Wahrscheinlichkeitsverteilung - hergeleitet an einem Beispiel	- 50 -
b) Relative Häufigkeiten als Grundlage der Beziehungen zwischen empirischen Verteilungen und Wahrscheinlichkeitsverteilungen	- 53 -

c) Kumulative Wahrscheinlichkeitsverteilungen und Zentile	- 55 -
d) Kennziffern der Zentralen Tendenz in diskreten Wahrscheinlichkeitsverteilungen	- 57 -
e) Kennziffern der Variabilität in diskreten Wahrscheinlichkeitsverteilungen . . .	- 60 -
f) Schiefe, Exzeß und Momente	- 62 -
g) Die Übertragbarkeit der Lehrsätze über Kennziffern von Verteilungen	- 64 -
h) Kleinere Besonderheiten von Wahrscheinlichkeitsverteilungen	- 66 -
2) Die Binomialverteilung	- 68 -
a) Die Voraussetzungen und die Formel der Binomialverteilung	- 68 -
b) Kennziffern der Binomialverteilung	- 73 -
3) Einige weitere eindimensionale diskrete Wahrscheinlichkeitsverteilungen	- 74 -
4) Mehrdimensionale diskrete Wahrscheinlichkeitsverteilungen	- 75 -
a) Der Begriff der mehrdimensionalen diskreten Wahrscheinlichkeitsverteilung, dargestellt an einem Beispiel	- 75 -
b) Bedingte Verteilung	- 78 -
c) Weitere Parallelen von 2-dimensionalen Wahrscheinlichkeits- und Häufigkeitsverteilungen	- 79 -
d) Unabhängige und unkorrelierte zufällige Größen	- 85 -
IV Kontinuierliche Wahrscheinlichkeitsverteilungen	- 89 -
1) Grundlegende Begriffe und Beziehungen	- 89 -
a) Kumulative Wahrscheinlichkeiten und Ogive	- 90 -
b) Die Dichtekurve	- 96 -
c) Näheres zur Bedeutung der Wahrscheinlichkeitsdichte	- 102 -
d) Schlußbemerkung zur Vorgehensweise	- 110 -
2) Definition von Kennziffern kontinuierlicher Wahrscheinlichkeitsverteilungen aufgrund eines verallgemeinerten Erwartungswert-Begriffs	- 110 -
a) Annäherung an das Problem	- 111 -
b) Der Erwartungswert nichtnegativer Zufallsvariablen	- 115 -
c) Mathematische Details zur Definition des Erwartungswerts nichtnegativer Zufallsvariablen	- 125 -
d) Der Erwartungswert einer nichtnegativen, von einer Zufallsvariablen X abgeleiteten Größe $g(X)$	- 129 -
e) Der Erwartungswert einer von einer Zufallsvariablen X abgeleiteten Größe $g(X)$ mit negativen Werten	- 135 -
f) Bestimmung von Erwartungswerten mit Mitteln der Integralrechnung	- 138 -
3) Die Normalverteilung	- 138 -
a) Allgemeines	- 138 -
b) Die Standardnormalverteilung	- 144 -
c) Weitere Normalverteilungs-Skalen	- 150 -
d) Der Zentrale Grenzwertsatz	- 152 -
e) Die Interpretation von Verteilungen aufgrund des Zentralen Grenzwertsatzes	

.....	- 154 -
f) Das Verhalten einer Normalverteilung bei Skalentransformationen	- 155 -
4) Mehrdimensionale kontinuierliche Verteilungen	- 156 -
a) Der Begriff der gemeinsamen Dichte	- 157 -
b) Randverteilung und bedingte Verteilungen	- 160 -
c) Kennziffern in kontinuierlichen 2-dimensionalen Verteilungen	- 163 -
d) Bivariate und Multivariate Normalverteilungen	- 164 -
C. Stichprobentheorie	- 168 -
I. Grundbegriffe und Zielsetzung	- 168 -
II. Stichprobentechniken	- 170 -
III. Beobachtungen als zufällige Größen	- 171 -
IV. Der mathematische Stichprobenbegriff	- 172 -
V. Wahrscheinlichkeitsverteilungen von Stichprobenkennziffern	- 174 -
D. Parameterschätzung	- 176 -
I. Zielsetzung und Terminologie	- 176 -
II. Kriterien guter Schätzfunktionen	- 179 -
1. Erwartungstreue	- 179 -
2. Minimalvarianz	- 184 -
3. Konsistenz und weitere Kriterien	- 185 -
E "Signifikanztests" oder das "Testen von Hypothesen"	- 186 -
I. Das Problem	- 186 -
II. Drei Beispiele	- 187 -
1) Der Vorzeichentest	- 188 -
2) Der t-Test für korrelierte Stichproben	- 189 -
3) Der U-Test	- 194 -
III. Das Entscheidungs-Modell der Signifikanzprüfung bei zweiseitigen Signifikanztests	- 197 -
1) Entscheidungsregeln aufgrund kritischer Bereiche	- 197 -
2) Die Wahl des kritischen Bereichs	- 198 -
a) Beispiele	- 198 -
b) Fehlertypen und Risiken	- 200 -
c) Die Abhängigkeit der Risiken von der Wahl des kritischen Bereichs.	- 201 -
d) Näheres zum Prinzip des konservativen Testens	- 202 -
e) Das Risiko II und die Teststärke	- 204 -
f) Konsequenzen für die Praxis	- 206 -
IV Einzelne Signifikanztests	- 207 -
1) Signifikanztests zur Prüfung von Hypothesen über einzelne Parameter	- 208 -
a) Der t -Test für einen einzelnen Mittelwert	- 208 -

Exkurs: Einseitige Entscheidungsregeln	- 210 -
b) Signifikanzprüfungen für einzelne Korrelationskoeffizienten	- 212 -
ba) Der t -Test für die Signifikanz einer Korrelation	- 212 -
bb) Signifikanzprüfung über Tabellen	- 213 -
bc) Signifikanzprüfung über die Fisher'sche z' -Transformation	- 213 -
2) Signifikanztests zum Vergleich zweier Populationsparameter anhand von unabhängigen Stichproben	- 217 -
a) Der F -Test für den Vergleich zweier Varianzen	- 218 -
b) Der t -Test für den Vergleich zweier Populationsmittel bei homogener Varianz und unabhängigen Stichproben	- 222 -
Exkurs: Das Problem der Zusatzvoraussetzungen bei Signifikanztests	- 224 -
a) Der traditionelle Weg: "Vortests"	- 224 -
b) Andere Verfahren	- 224 -
ba) Skalentransformation	- 224 -
bb) Versuchsplanerische Maßnahmen	- 225 -
bc) Verwendung von Signifikanztests mit weniger Zusatzvoraussetzungen	- 225 -
bd) Abschätzung des Fehlers, der bei einer Verletzung der Zusatzannahmen auftritt	- 225 -
c) Der z -Test für den Vergleich zweier Produktmomentkorrelationen	- 226 -
3) Signifikanztests zum Vergleich zweier Populationsparameter anhand von korrelierten Stichproben	- 228 -
a) Der t -Test für Paardifferenzen	- 229 -
b) Die Überprüfung der Varianzgleichheit	- 229 -
4) Verteilungsfreie oder parameterfreie Signifikanztests	- 231 -
a) Allgemeines	- 231 -
b) Der Vorzeichentest	- 232 -
c) Der Mann-Whitney U-Test	- 233 -
ca) Definition und Berechnung der Prüfgröße U	- 234 -
cb) Die Signifikanzprüfung von U	- 234 -
cc) Die Behandlung von Rangkopplungen	- 236 -
cd) Teststärke und Zusatzvoraussetzungen beim U-Test	- 239 -
d) Der χ^2 -Test nach Pearson	- 240 -
da) Als χ^2 -Test für die Güte der Anpassung ohne zusätzliche Parameterschätzung	- 241 -
db) Der χ^2 -Test für die Güte der Anpassung mit vorgeschalteter Parameterschätzung	- 243 -
dc) Der χ^2 -Test für stochastische Unabhängigkeit in zweidimensionalen Verteilungen	- 246 -
dd) Schlußbemerkungen zu den χ^2 -Tests nach Pearson	- 252 -
Übersicht über Signifikanztests	- 254 -

<i>Fortsetzung</i> von: D. Parameterschätzung	- 257 -
III Konfidenzintervalle	- 257 -
1) Konfidenzintervalle für das Populationsmittel	- 257 -
2) Konfidenzintervalle für die Produkt-Moment-Korrelation ρ_{xy}	- 260 -
3) Konfidenzintervalle bei Schätzungen aufgrund von Regressionsgleichungen	- 260 -
4) Konfidenzintervalle und Signifikanztests	- 262 -
F Varianzanalyse	- 264 -
I Die Einweg-Varianzanalyse	- 264 -
1) Ein Experiment (nach Hays, 1973)	- 264 -
2) Das Radikalitäts-Problem bei mehrfacher Anwendung des t-Tests	- 267 -
3) Zusatzvoraussetzungen	- 268 -
4) Die Aufspaltung der Varianz in Varianzquellen.	- 268 -
a) Die Zerlegung der Abweichungen vom Gesamtdurchschnitt.	- 269 -
b) Die Zerlegung der Abweichungs-Quadratsumme.	- 270 -
c) Vereinfachte Berechnung der Quadratsummen.	- 274 -
d) Die Freiheitsgrade der verschiedenen Varianzquellen.	- 276 -
e) Gesamtdurchschnitt und Freiheitsgradzahl zwischen den Gruppen bei unterschiedlicher Gruppengröße.	- 277 -
5) Logik und Durchführung der Einwegvarianzanalyse.	- 279 -
6) "Effekte" und "Fehler"	- 282 -
a) Der Begriff des Effekts	- 283 -
b) Organismische und experimentelle Effekte	- 285 -
c) Differentielle Effekte	- 287 -
d) Der Begriff des Fehlers	- 288 -
e) Allgemeines Modell der Einweg-Varianzanalyse	- 290 -
7) Varianzanalytische Parameterschätzung als Teil der Korrelationsstatistik	- 290 -
8. Anhang: Herleitung der Erwartungswerte von Quadratsummen und Mittelquadraten in der Einweg-Varianzanalyse.	- 291 -
II. Die Mehrweg-Varianzanalyse	- 294 -
1) Ein Experiment zur Zweiweg-Varianzanalyse	- 294 -
2) "Faktoren" und "Stufen"	- 295 -
a) Begriffe	- 295 -
b) Die Unterscheidung von Zeilen und Spalten	- 296 -
c) Die Verwendung der Indices in der Zweiweg-Varianzanalyse	- 297 -
3) Effekte in der Zweiweg-Varianzanalyse	- 297 -
a) Die Durchschnitte	- 297 -
b) Effekte des Zeilenfaktors	- 298 -
c) Effekte des Spaltenfaktors	- 299 -
d) Wechselwirkung	- 299 -

e) Allgemeines Modell der Zweiweg-Varianzanalyse	- 304 -
4) Die Zusatzvoraussetzungen der Zweiweg-Varianzanalyse	- 305 -
a) Unabhängige Stichproben	- 305 -
b) Normalverteilung und Varianzhomogenität	- 305 -
c) Orthogonalität	- 305 -
d) Eine letzte Voraussetzung	- 308 -
5) Grundlagen und Durchführung der Berechnungen in der Zweiweg-Varianzanalyse.	- 308 -
a) Die Zerlegung der Abweichungen vom Gesamtdurchschnitt	- 308 -
b) Die Zerlegung der Abweichungssumme	- 309 -
c) Vereinfachte Berechnung der Quadratsummen.	- 310 -
d) Die Freiheitsgrade der Varianzquellen.	- 311 -
e) Die Signifikanztests in der Zweiweg-Varianzanalyse.	- 314 -
6) Die Varianzanalyse mit mehr als zwei Faktoren	- 318 -
a) Ein Experiment mit drei Faktoren	- 318 -
b) Die Effekte	- 319 -
ba) Die Haupteffekte	- 319 -
bb) Die Wechselwirkungen erster Ordnung	- 320 -
bc) Die Wechselwirkung zweiter Ordnung	- 322 -
c) Signifikanztests	- 324 -
d) Varianzanalyse mit mehr als drei Faktoren	- 326 -
7) Varianzanalyse mit korrelierten Stichproben	- 327 -
a) Mehrere Messungen pro Vp	- 327 -
b) Randomisierte Blocks	- 330 -
8) Die modellhafte Einordnung von varianzanalytischen Versuchsplänen	- 331 -
a) Ein Beispiexperiment	- 331 -
b) Überlegungen zu den einzelnen Faktoren	- 334 -
ba) Feststellung der Faktoren	- 334 -
bb) "Fixe" oder "zufällige" Effekte?	- 335 -
c) Das Gesamtmodell	- 339 -
ca) Das Verhältnis der Faktoren zueinander	- 339 -
cb) Das mathematische Gesamtmodell	- 343 -
d) Zusammenfassung	- 344 -

Korrekturen

In der Fassung SS 2003 wurden die in vorangehenden Semestern angefallenen Korrekturen eingearbeitet.

Die Fassung SS 2005 unterscheidet sich von der Fassung SS 2003 nur in der Seiteneinteilung.