

Merkposten zu Statistik II,

Albrecht Iseler

Fassung SS 2005

B. Wahrscheinlichkeitstheorie

Anwendung der Wahrscheinlichkeitstheorie in der Psychologie:

- Wahrscheinlichkeitsaussagen über das Erleben und Verhalten
- Inferenzstatistik

I. Die Begriffe "Wahrscheinlichkeit" und "Zufall"

1. Wahrscheinlichkeit

a) 3 Wahrscheinlichkeitsbegriffe

Objektiver Wahrscheinlichkeitsbegriff: Zu erwartende relative Häufigkeit, wobei diese um so genauer eintritt, je mehr "Versuche".

Subjektiver Wahrscheinlichkeitsbegriff: Ausmaß der Gewißheit, mit der wir das Eintreten eines Ereignisses erwarten.

Axiomatischer Wahrscheinlichkeitsbegriff: Wahrscheinlichkeit ist eine Zahl von 0 bis 1, die einem Ereignis zugeordnet ist und die bestimmte Axiome erfüllt (vgl. Abschnitt @B.II.1.c).

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.

3. Modelle zur Veranschaulichung von Wahrscheinlichkeit und Zufall

a) Das Urnenmodell

Wahrscheinlichkeit einer Kugelart = Anteil dieser Kugelart an der Gesamtheit aller Kugeln.

Unterschied mit und ohne Zurücklegen.

b) Das Venn-Diagramm

Wahrscheinlichkeit eines Ereignisses = Anteil der dem Ereignis zugeordneten Fläche an der Gesamtfläche

4. Bedingte Wahrscheinlichkeiten

a) Der Begriff der bedingten Wahrscheinlichkeit

$$p(A|B) = \frac{f_{A \wedge B}}{f_B} = \frac{P_{A \wedge B}}{p_B} = \frac{\text{Verbundwahrscheinlichkeit}}{\text{Randwahrscheinlichkeit der Bedingung}}$$

Für $f_{A \wedge B}$ und f_B können z.B. entsprechende Kugelzahlen im Urnenmodell bzw. Flächen im Venn-Diagramm stehen.

b) Unabhängige und abhängige Ereignisse

Zwei Ereignisse A und B heißen unabhängig, wenn die Wahrscheinlichkeit des Ereignisses A unabhängig von Bedingungen über das Ereignis B ist.

Oder, was dasselbe ist: Wenn die bedingte Wahrscheinlichkeit $p(A|B)$ gleich der Randwahrscheinlichkeit $p(A)$ ist.

II. Kombinatorik

1. Wahrscheinlichkeitssätze

a) der Multiplikationssatz

Für die Wahrscheinlichkeit, daß die Ereignisse A und B beide eintreffen (Verbundwahrscheinlichkeit) gilt immer der erweiterte Multiplikationssatz:

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

Sind die Ereignisse A und B unabhängig, so gilt auch der (einfache) Multiplikationssatz für unabhängige Ereignisse:

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

Dieser läßt sich auf mehr als zwei Ereignisse ausdehnen, wenn sie voneinander vollständig unabhängig sind:

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

b) Der einfache Additionssatz

Wenn sich zwei Ereignisse gegenseitig ausschließen, gilt der einfache Additionssatz:

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

Dieser läßt sich auf beliebig viele Ereignisse ausdehnen, wenn sie sich gegenseitig ausschließen:

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

c) Die Axiome der mathematischen Wahrscheinlichkeitstheorie

Die über die Definition des axiomatischen Wahrscheinlichkeitsbegriffs (s.o.) hinausgehenden Inhalte dieses Abschnitts sind "Für Spezialisten".

d) Der erweiterte Additionssatz

Schließen sich die Ereignisse A und B nicht aus, verwenden wir für die "oder"-Verbindung den erweiterten Additionssatz:

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

e) Negation

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

f) Das Bayes-Theorem

Das Bayes Theorem verwendet man zum "Umdrehen" von bedingten Wahrscheinlichkeiten, d.h. wenn wir $p(B|A)$ suchen und $p(A|B)$ kennen oder umgekehrt. Wenigstens eine der Gleichungen für das Bayes-Theorem zu kennen, ist nützlich, aber nicht unbedingt erforderlich.

2. Anzahlen

a) Permutationen

$N!$ ist das Produkt aller ganzen Zahlen von 1 bis N

Zusatzdefinition: $0! = 1$

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

b) Teilsequenzen

Die Zahl der Möglichkeiten, aus N verschiedenen Elementen eine Teilsequenz von r Elementen

zu ziehen, ist $\frac{N!}{(N-r)!}$

c) Untergruppen (Kombinationen ohne Reihenfolge)

Die Zahl der möglichen Untergruppen von r verschiedenen Elementen aus N verschiedenen Elementen ist gleich $\binom{N}{r}$. Der dabei verwendete "Binomialkoeffizient" $\binom{N}{r}$ ist definiert als

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

III Wahrscheinlichkeitsverteilungen

1. Eindimensionale diskrete Wahrscheinlichkeitsverteilungen

a) Der Begriff der diskreten Wahrscheinlichkeitsverteilung.

Eine diskrete Wahrscheinlichkeitsverteilung besteht in einer Zuordnung von Wahrscheinlichkeiten zu den Klassen einer diskreten Skala. Diese Zuordnung kann durch eine Tabelle oder durch eine Formel erfolgen.

Die Größe, für deren Werte man Wahrscheinlichkeiten angibt, nennt man eine "zufällige Größe" oder "Zufallsvariable".

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

Neben dem Begriffspaar "Häufigkeitsverteilungen - Wahrscheinlichkeitsverteilungen" gibt es auch das weitgehend gleichbedeutende Begriffspaar "empirische Verteilungen - theoretische Verteilungen".

Die Verbindung zwischen beiden Arten von Verteilungen wird deutlich, wenn man die Wahrscheinlichkeitsangaben der Wahrscheinlichkeitsverteilung als erwartete relative Häufigkeiten interpretiert und andererseits aus der Beschreibenden Statistik (Häufigkeitsverteilungen) diejenigen Begriffe und Formeln heranzieht, die mit relativen Häufigkeiten arbeiten.

c) Kumulative Wahrscheinlichkeitsverteilung und Zentile

Die kumulative Wahrscheinlichkeit $p_{cum k}$ entspricht der kumulativen relativen Häufigkeit. Sie ist die Wahrscheinlichkeit, mit der die von der Wahrscheinlichkeitsverteilung erfaßte zufällige Größe in die k-te Klasse oder darunter fällt.

Entsprechend der kumulativen relativen Häufigkeit berechnet sich die kumulative Wahrscheinlichkeit als Summe der Wahrscheinlichkeit aller Klassen bis zur k-ten.

Der C_p (Centil P) ist der Skalenwert, dessen kumulative Wahrscheinlichkeit P% ist.

Beispiel: C_{50} = der Skalenwert, dessen kumulative Wahrscheinlichkeit 50% oder 0,5 ist. Bei natürlich diskreten Skalen sucht man denjenigen Wert der Größe x, bei dem diese kumulative Wahrscheinlichkeit erstmals erreicht oder unterschritten wird.

d) Kennziffern der zentralen Tendenz in diskreten Wahrscheinlichkeitsverteilungen

Median = C_{50} , wie in der beschriebenen Statistik.

Modalwert = der Wert mit der größten (Klassen-)Wahrscheinlichkeit.

Das Pendant zum Durchschnitt \bar{x} (arithmetisches Mittel) ist der Erwartungswert von x:

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

Wir multiplizieren also jeden X-Wert mit seiner Wahrscheinlichkeit und bilden die Summe dieser Produkte.

Entsprechend bilden wir auch den Erwartungswert einer von X abgeleiteten Größe $g(X)$:

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

Beispiel: $E(X^2)$, den Erwartungswert von X^2 , bilden wir, indem wir für jeden X-Wert x_k sein Quadrat mit der zugehörigen Wahrscheinlichkeit multiplizieren und die Summe dieser Produkte aufaddieren. Diesen Erwartungswert von X^2 können wir z.B. für die Varianzberechnung brauchen (s.u.).

e) Kennziffern der Variabilität in diskreten Wahrscheinlichkeitsverteilungen

Die Kennziffern der Variabilität in empirischen Verteilungen geben an, welches Ausmaß an Abweichungen vom Durchschnitt (oder anderen Kennziffern der zentralen Tendenz) man tatsächlich vorgefunden hat. Die entsprechenden Kennziffern in Wahrscheinlichkeitsverteilungen geben an, welches Ausmaß an Abweichungen von μ_x zu erwarten ist.

Am wichtigsten sind Varianz und Standardabweichung. Die Varianz einer empirischen Verteilung ist das arithmetische Mittel der quadrierten Abweichungen vom Durchschnitt \bar{x} . Entsprechend ist die Varianz einer Wahrscheinlichkeitsverteilung der Erwartungswert der quadrierten Abweichungen vom Erwartungswert μ_x der Verteilung.

In Formeln:

	Empirische Verteilungen	Wahrscheinlichkeitsverteilungen
Definition	$s_x^2 = aM[(X - \bar{x})^2]$ $= \sum_k p_k \cdot (x_k - \bar{x})^2$	$\sigma_x^2 = E[(X - \mu_x)^2]$ $= \sum_k p_k \cdot (x_k - \mu_x)^2$
Einfacher für Berechnungen	$s_x^2 = aM(X^2) - \bar{x}^2$	$\sigma_x^2 = E(X^2) - \mu_x^2$

Diese Formeln brauchen nicht auswendig gelernt zu werden. Es lohnt sich aber, sich um ein Verständnis dieser Formeln zu bemühen: Dann versteht man auch besser, was die Varianz einer Wahrscheinlichkeitsverteilung ist.

Die Standardabweichung einer Wahrscheinlichkeitsverteilung ist die Quadratwurzel der Varianz.

Standardisierung: Der standardisierte Wert $z_x = \frac{X - \mu_x}{\sigma_x}$ gibt an, wie viele

Standardabweichungen ein X-Wert über dem Erwartungswert μ_x liegt.

f) Schiefe, Exzeß.

Auch die Begriffe "Schiefe" und "Exzeß" lassen sich aus der beschreibenden Statistik übertragen. So hat z.B. die in Tabelle @2 (Abschnitt @B.III.1.a) angegebene Wahrscheinlichkeitsverteilung eine positive Schiefe (langer Schwanz nach rechts, wenn man sie aufzeichnet.)

g) Die Übertragbarkeit der Lehrsätze über Kennziffern von Verteilungen

Alle Lehrsätze über Kennziffern, insbesondere die über das Verhalten bei Skalentransformationen, lassen sich auf die Kennziffern von Wahrscheinlichkeitsverteilungen übertragen.

h) Kleinere Besonderheiten von Wahrscheinlichkeitsverteilungen

Für die Wahrscheinlichkeit bzw. kumulative Wahrscheinlichkeit schreibt man anstelle der auf die Klassennummer k bezogenen Symbole p_k und $p_{cum k}$ in Wahrscheinlichkeitsverteilungen oft $f(x)$ für die Wahrscheinlichkeit eines x -Wertes und $F(x)$ für sein kumulative Wahrscheinlichkeit.

Bei Wahrscheinlichkeitsverteilungen ist die Angabe der Verteilung über eine Formel noch üblicher als die Angabe über eine Tabelle.

Unter einer Verteilungsfamilie versteht man eine Gruppe von Wahrscheinlichkeitsverteilungen, die alle durch eine gemeinsame Formel darstellbar sind; in dieser Formel tauchen jedoch neben den Werten der zufälligen Größe (also dem x -Wert, dessen Wahrscheinlichkeit angegeben wird) noch weitere Größen auf, die durch einen Buchstaben angegeben sind. Bei jeder einzelnen Verteilung der entsprechenden Familie ist für den Buchstaben eine bestimmte Zahl einzusetzen.

Beispiele, die in den folgenden Abschnitten behandelt werden:

- Binomialverteilung: n = Zahl der Versuche
 p = Grundwahrscheinlichkeit
- Normalverteilung: μ = Erwartungswert
 σ = Standardabweichung

2. Die Binomialverteilung

a) Voraussetzungen und Formel der Binomialverteilung

Es wird eine Reihe von "Versuchen" durchgeführt, bei denen ein bestimmtes "kritisches Ereignis" auftreten kann. Die Zufallsgröße ist die Zahl der Versuche, bei denen das kritische Ereignis auftritt.

Diese Zufallsgröße hat unter den folgenden Bedingungen eine Binomialverteilung:

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.

b) Kennziffern der Binomialverteilung

Erwartungswert: $\mu_x = n \cdot p$

Varianz: $\sigma_x^2 = n \cdot p \cdot q$

3. Einige weitere eindimensionale diskrete Wahrscheinlichkeitsverteilungen.

"Für Spezialisten"

4. Mehrdimensionale diskrete Wahrscheinlichkeitsverteilungen

a) Der Begriff der mehrdimensionalen Wahrscheinlichkeitsverteilung

Eine 2-dimensionale Wahrscheinlichkeitsverteilung besteht in einer Zuordnung von Wahrscheinlichkeiten zu Meßwertpaaren bzw. zu den (2-dimensionalen) Klassen einer 2-dimensionalen Skala.

Aus einer 2-dimensionalen Wahrscheinlichkeitsverteilung können wir die Randverteilungen von x und y herausziehen, indem wir die Eintragungen in den Zeilen der Tabelle entweder senkrecht oder waagrecht addieren. Diese Randverteilungen stellen die "Randwahrscheinlichkeiten" bestimmter x - bzw. y -Werte dar.

Für die Wahrscheinlichkeit eines bestimmten Meßwertpaares x,y schreibt man auch $f(x,y)$.

b) Bedingte Verteilung

Eine bedingte Wahrscheinlichkeitsverteilung ist eine Zuordnung von bedingten Wahrscheinlichkeiten zu den Klassen einer Variablen, gegeben eine Bedingung für die andere Variable. Beispiel: Bedingte Wahrscheinlichkeiten von x , gegeben $y = 0$. Man schreibt z.B. $f(x | y=0)$.

Den Erwartungswert einer bedingten Verteilung nennt man einen "bedingten Erwartungswert".

c) Weitere Parallelen von 2-dimensionalen Wahrscheinlichkeits- und Häufigkeitsverteilungen

Genauso 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 Kovarianz, eine Produkt-Moment-Korrelation (und andere Korrelationskoeffizienten), Regressionsgeraden und im Zusammenhang damit Schätzfehlervarianzen, Standardschätzfehler sowie einen Schätzungseffekt.

d) Unabhängige und unkorrelierte zufällige Größen

Definition: Zwei zufällige Größen sind stochastisch unabhängig, wenn die Angabe von Bedingungen über den Wert der einen zufälligen Größe nichts an der Wahrscheinlichkeitsverteilung der anderen ändert, wenn also die entsprechenden bedingten Verteilungen gleich den zugehörigen Randverteilungen sind.

Definition: Zwei zufällige Größen sind unkorreliert, wenn ihre Kovarianz (und damit auch ihre Produkt-Moment-Korrelation) 0 ist.

Lehrsatz: Sind zwei zufällige Größen unabhängig, dann sind sie auch unkorreliert.

Dagegen gilt die Umkehrung dieses Lehrsatzes nicht: Zwei Zufallsgrößen können unkorreliert und trotzdem abhängig (also nicht unabhängig) sein.

IV Kontinuierliche Wahrscheinlichkeitsverteilungen

1) Grundlegende Begriffe und Beziehungen

a) Kumulative Wahrscheinlichkeiten und Ogive

Die kumulative Verteilungsfunktion einer Zufallsgröße ist durch die Gleichung

$$F(x) := p(X \leq x)$$

definiert. Die graphische Darstellung dieser kumulativen Wahrscheinlichkeiten wird auch hier als Ogive bezeichnet. Bei kontinuierlichen Wahrscheinlichkeitsverteilungen hat die Ogive keine Sprungstellen.

Bei einer kontinuierlichen Wahrscheinlichkeitsverteilung ist die Wahrscheinlichkeit, daß eine Zufallsgröße X einen ganz bestimmten Wert x hat, für jedes x null.

Differenzenregel für Intervallwahrscheinlichkeiten: Bei einer Zufallsvariablen X mit einer kontinuierlichen Wahrscheinlichkeitsverteilungen 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.

b) Die Dichtekurve

Man kann die "Dichtekurve" zunächst als Grenzfall eines Histogramms für immer feinere künstlich diskrete Skalen auffassen.

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.

Spezialfall der Flächenregel: Die kumulative Wahrscheinlichkeit $F(x)$ eines Wertes x ist gleich der gesamten Fläche unter der Dichtekurve bis x .

Außerdem ist die Dichtekurve eine graphische Darstellung der im folgenden Abschnitt definierten Dichtefunktion.

c) Näheres zur Bedeutung der Wahrscheinlichkeitsdichte

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.

2) Definition von Kennziffern kontinuierlicher Wahrscheinlichkeitsverteilungen aufgrund eines verallgemeinerten Erwartungswert-Begriffs

a) Annäherung an das Problem

Die Definitionen von Kennziffern, die auf Centilen beruhen, können aus diskreten Wahrscheinlichkeitsverteilungen und Häufigkeitsverteilungen übernommen werden. Für Erwartungswert-basierte Kennziffern ist ein verallgemeinerter Erwartungswert-Begriff erforderlich, da die Wahrscheinlichkeiten einzelner x -Werte null sind.

Schwerpunktregel: Der Erwartungswert einer Zufallsgröße mit kontinuierlicher Wahrscheinlichkeitsverteilung ist der Abszissenwert des Schwerpunkts der Fläche unter der Dichtekurve.

b) Der Erwartungswert nichtnegativer Zufallsvariablen

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 $+\infty$.

Dieser Erwartungswert-Begriff ist nicht nur auf kontinuierliche Zufallsvariablen anwendbar. Bei einer Anwendung dieser Definition auf diskrete Wahrscheinlichkeitsverteilungen kommt genau der nach den bisherigen Regeln bestimmte Erwartungswert heraus.

c) Mathematische Details zur Definition des Erwartungswerts nichtnegativer Zufallsvariablen
"Für Spezialisten."

d) Der Erwartungswert einer nichtnegativen, von einer Zufallsvariablen X abgeleiteten Größe $g(X)$

Der Erwartungswert $E[g(X)]$ wird ähnlich wie $E(X)$ aufgrund immer feinerer Abrundungen und immer größerer Kappungsgrenzen definiert.

e) Der Erwartungswert einer von einer Zufallsvariablen X abgeleiteten Größe $g(X)$ mit negativen Werten

Kann eine abgeleitete Größe $g(X)$ auch negative Werte haben, dann wird der Erwartungswert durch Zerlegung der Funktion g in einen "Positivteil" und einen "Negativteil" definiert.

f) Bestimmung von Erwartungswerten mit Mitteln der Integralrechnung
"Für Spezialisten"

3) *Die Normalverteilung*

Eine Normalverteilung entsteht häufig dann, wenn ein Merkmal durch Zusammenwirken sehr vieler Einzelfaktoren entsteht. Genauso wie "die Binominalverteilung" ist auch "die Normalverteilung" ein Sammelbegriff für eine ganze Familie von unendlich vielen Verteilungen. Die einzelnen Mitglieder dieser Familie unterscheiden sich in ihrem Erwartungswert und ihrer Varianz. Die Normalverteilung hat die Form einer um den Erwartungswert symmetrischen Glocke. Die Wendepunkte dieser "Gauß'schen Glockenkurve" liegen eine Standardabweichung unterhalb und eine Standardabweichung oberhalb des Mittelwertes.

a) Die Standardnormalverteilung

Eine zentrale Stellung innerhalb der Normalverteilungen nimmt diejenige Normalverteilung ein, deren Erwartungswert $\mu_x = 0$ und deren Varianz $\sigma_x^2 = 1$ ist. Da in dieser Normalverteilung alle x -Werte gleich den zugehörigen standardisierten Werten (also den z -Werten) sind, nennt man diese Normalverteilung auch die Standardnormalverteilung. In der Standardnormalverteilung ist also jeder x -Wert auch ein z -Wert.

Bei der Benutzung von Tabellen der Standardnormalverteilung macht man Gebrauch von folgenden Eigenschaften:

1. Die Kurve der Standardnormalverteilung ist symmetrisch um den 0-Punkt.
2. Die gesamte Fläche unter der Kurve ist gleich 1.

In der Statistik hat sich das Symbol $\Phi(z)$ (sprich: Phi von z) für die kumulative Wahrscheinlichkeit eines z -Wertes bei Standardnormalverteilungen eingebürgert. Mit diesem Symbol könnten wir eine Regel für negative z -Werte formulieren:

$$\Phi(-z) = 1 - \Phi(z)$$

b) Weitere Normalverteilungsskalen

Besonders wichtig in der Psychologie sind die in der folgenden Tabelle zusammengestellten Normalverteilungsskalen:

Name der Skala	μ_x	σ_x
z-Skala	0	1
IQ-Skala	100	15
SW-Skala	100	10
T-Skala	50	10
C-Skala	5	2

Man kann die kumulativen Wahrscheinlichkeiten, die Überschreitungswahrscheinlichkeiten und die Wahrscheinlichkeiten von Intervallen bei Normalverteilung angeben, indem man die x -Werte zunächst in z -Werte transformiert und dann aufgrund der Standardnormalverteilungstabellen die entsprechenden Wahrscheinlichkeiten berechnet.

Wichtig: Bevor man die Skalenwerte in z -Werte übersetzt, muß man gegebenenfalls die auf der Basis der künstlich diskreten Skala formulierte Frage in eine kontinuierliche Skala übersetzen.

c) Der Zentrale Grenzwertsatz

Der "Zentrale Grenzwertsatz" lautet:

Bilden wir eine Summe von vielen stochastisch unabhängigen Größen, deren Wahrscheinlichkeitsverteilungen alle die gleiche Varianz haben, so ist die Wahrscheinlichkeitsverteilung dieser Summe approximativ eine Normalverteilung.

Diesen recht abstrakten Satz braucht man nicht auswendig zu kennen, wohl aber die wichtigsten Anwendungen: Die Normalverteilungs-Approximation der Binomialverteilung (s.u.) und der Wahrscheinlichkeitsverteilung des Stichprobendurchschnitts (vgl. Abschnitt @C.V).

Die zufällige Größe x in der Binomialverteilung kann man mit einer Normalverteilung approximieren, wenn n nicht zu klein und p nicht zu weit von 0.5 entfernt ist.

Faustregel: Die Binomialverteilung läßt sich mit einer Normalverteilung approximieren, wenn die Produkte $n \cdot p$ und $n \cdot q$ beide größer als 10 sind.

Als Mittelwert und Varianz dieser Normalverteilung nimmt man die aus den Formeln für die Binomialverteilung berechneten Werte. Da die Binomialverteilung eine diskrete Verteilung ist, muß man die "diskrete Frage" zunächst für eine kontinuierliche Skala umformulieren.

d) Die Interpretation von Verteilungsformen aufgrund des Zentralen Grenzwertsatzes

Sowohl das Vorliegen als auch das Nichtvorliegen einer Normalverteilung eines Merkmals kann man auf dem Hintergrund des Zentralen Grenzwertsatzes oft sinnvoll interpretieren.

e) Das Verhalten einer Normalverteilung bei Lineartransformation der x-Werte

Hat eine zufällige Größe x eine Normalverteilung, dann hat auch jede Lineartransformation dieser zufälligen Größe eine Normalverteilung.

4. Mehrdimensionale kontinuierliche Verteilungen

Aus diesem Abschnitt ist vor allem der Begriff der bivariaten Normalverteilung (Unterabschnitt @d) wichtig. Man braucht nicht alle Details der folgenden Definition zu kennen, sollte aber wissen, daß zu einer bivariaten Normalverteilung mehr gehört als eine Normalverteilung von X und Y .

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 $E(Y/ X = x)$.

C)Stichprobentheorie

I Grundbegriffe und Zielsetzung

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).

Stichproben können (zumal 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. Ziel der Stichprobentheorie ist es, anzugeben, welche Abweichungen der Stichprobenverteilung von der Populationsverteilung mit welcher Wahrscheinlichkeit zu erwarten sind.

II. Stichprobentechniken

Die von uns zu behandelnden Verfahren der schlußfolgernden Statistik gehen alle vom Modell der "Zufallsstichprobe mit Zurücklegen" aus, weil dieses Modell am einfachsten wahrscheinlichkeitstheoretisch zu analysieren ist. Die Wahrscheinlichkeitsverteilungen, die wir im Zusammenhang mit der Stichprobentheorie und der schlußfolgernden Statistik behandeln werden, sind aber beim Stichprobenziehen mit und ohne Zurücklegen praktisch identisch, solange der Umfang der Stichprobe nicht mehr als etwa 5 % der Populationsgröße beträgt. Allenfalls sind Stichprobenfehler beim Ziehen "ohne Zurücklegen" kleiner als beim Ziehen "mit Zurücklegen".

III. Beobachtungen als zufällige Größen

Grundprinzip der Stichprobentheorie nach dem Urnenmodell (mit Zurücklegen): 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. Versuchspersonen) einer Zufallsstichprobe gemacht werden, können als stochastisch unabhängig betrachtet werden.

IV. Der mathematische Stichprobenbegriff

Das im vorangehenden Abschnitt dargestellte Grundprinzip reicht aus, um die wichtigsten Fragestellungen der Stichprobentheorie zu beantworten. Daher wird der Begriff einer Stichprobe in der mathematischen Stichprobentheorie so definiert, daß er nur die folgende Annahme beinhaltet, wobei die Abkürzung "i.i.d." für 'independent, identically distributed (random variables)' steht:

i.i.d.-Annahme: Eine Stichprobe besteht aus N unabhängigen, identisch verteilten Zufallsvariablen. (Dabei bezeichnet N - wie üblich - die Stichprobengröße.)

V. Wahrscheinlichkeitsverteilungen von Stichprobenkennziffern

Der Durchschnitt der Meßwerte ist eine zufällige Größe und hat eine Wahrscheinlichkeitsverteilung; er kann zufällig größer oder kleiner ausfallen. Der Erwartungswert

der Wahrscheinlichkeitsverteilung des Stichprobendurchschnitts \bar{x} ist gleich dem Populationsmittel μ_x :

$$E(\bar{x}) = \mu_x$$

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.

Die Wahrscheinlichkeitsverteilung des Stichprobendurchschnitts ist 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.

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.

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.

D. Parameterschätzung

I. Zielsetzung und Terminologie

Wir unterscheiden:

Parameter: Populationskennziffern, Kennziffern von Wahrscheinlichkeitsverteilungen.

Statistiken: Alle beschreibenden Stichprobenkennziffern, aber auch alle anderen Werte, die wir aus den Daten einer Stichprobe berechnen (z.B. die Summe der quadrierten Abweichungen vom Durchschnitt).

Schätzfunktionen: Statistiken, die zur Schätzung von Parametern herangezogen werden.

Als Symbol für eine Schätzfunktion nimmt man das Symbol des zu schätzenden Parameters mit einem "Dach" (^) darüber. Beispiel: $\hat{\mu}$ ist der geschätzte Wert von μ .

Auch Schätzfunktionen haben eine Wahrscheinlichkeitsverteilung: Sie können zufällig (insbesondere durch Zufälle bei der Stichprobenauswahl) größer oder kleiner ausfallen. Ziel der Theorie der Parameterschätzung ist es, Eigenschaften "guter" Schätzfunktionen zu beschreiben.

II. Kriterien guter Schätzfunktionen

1. Erwartungstreue

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).

Beispiele:

- Der Stichprobendurchschnitt ist eine erwartungstreue Schätzfunktion für das Populationsmittel. (Vgl. die Gleichung $E(\bar{x}) = \mu_x$ aus Abschnitt @C.V)
- Die Stichprobenvarianz unterschätzt tendenziell die Populationsvarianz: Der Erwartungswert der Wahrscheinlichkeitsverteilung dieser Kennziffer ist (um den Faktor $(N-1)/N$) kleiner als die Populationsvarianz. Als Schätzfunktion für die Populationsvarianz wäre die Stichprobenvarianz also verzerrt. Um das zu kompensieren, dividiert man die Summe der Abweichungsquadrate nicht durch die Stichprobengröße N , sondern durch $N - 1$, um eine erwartungstreue Schätzfunktion für die Populationsvarianz zu erhalten.

2. Minimalvarianz

Die Varianz der Wahrscheinlichkeitsverteilung einer Schätzfunktion nennt man die Schätzfehlervarianz; ihre Quadratwurzel ist der Standardschätzfehler bei der Schätzung des Parameters. Für eine gute Schätzfunktion ergibt sich 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.

3. Konsistenz

Zur Konsistenz gehört insbesondere, daß die Schätzfehlervarianz umso kleiner wird, je größer die Stichprobe wird, und für eine unendlich große Stichprobe null betragen würde.

Beispiel: Beim Stichprobendurchschnitt als Schätzfunktion für das Populationsmittel ergibt sich die Konsistenz aus dem Wurzel-N-Gesetz (bzw. der Formel für $\sigma_{\bar{x}}^2$, vgl. Abschnitt @C.V).

III. Konfidenzintervalle

Darstellung im Skript nach Abschnitt E, unmittelbar vor F: Varianzanalyse!

Ein Konfidenzintervall ist ein Bereich, innerhalb dessen wir einen Populationsparameter mit einer bestimmten Verlässlichkeit vermuten.

1. Konfidenzintervalle für das Populationsmittel

Die Grenzen des Konfidenzintervalls für das Populationsmittel μ_x berechnen sich nach der Formel $\bar{x} \pm t_{\text{tab}} \cdot \hat{\sigma}_{\bar{x}}$

Dabei ist t_{tab} der aus der t -Tabelle entnommene t -Wert, dessen zweiseitige Überschreitungswahrscheinlichkeit bei $N-1$ Freiheitsgraden gleich $1 - V$ (bzw. $100\% - V$) ist. Dabei ist V das "Verlässlichkeitsniveau" (z.B. 0.95 bzw. 95%).

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

$$\sigma_{z'} = \frac{1}{\sqrt{N - 3}}$$

3) Konfidenzintervalle bei Schätzungen aufgrund von Regressionsgleichungen

Grenzen des Konfidenzintervalls für die Schätzung von y aus x :

$$\hat{y} \pm z_{\text{tab}} \cdot \hat{\sigma}_{y.x}$$

mit

$$\hat{\sigma}_{y.x} = \text{geschätzter Standardfehler} = \hat{\sigma}_y \cdot \sqrt{1 - r_{xy}^2}$$

E. "Signifikanztests" oder das "Testen von Hypothesen"

I. Das Problem

Um zu überprüfen, ob eine 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 über den oder die Populationsparameter, auf die sich die Hypothese bezieht, die Nullhypothese (H_0), und ihre Negation die Alternativhypothese (H_1). 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 spricht dann für die Alternativhypothese.

II. Beispiele

Die Beispiele selbst gehören nicht zu den "Merkposten". Wichtig für Signifikanztests ist jeweils die Frage: 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 die H_0 richtig ist?

III. Das Entscheidungs-Modell der Signifikanzprüfung bei zweiseitigen Signifikanztests

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. Die Wahrscheinlichkeitsverteilung der Prüfgröße bei Gültigkeit von H_0 muß bekannt sein. Häufig ist allerdings diese Wahrscheinlichkeitsverteilung nur dann bekannt, wenn noch weitere Zusatzvoraussetzungen erfüllt sind, die selbst mit der Nullhypothese nichts zu tun haben.

Dann formulieren wir einen "kritischen Bereich" und verbinden ihn mit der folgenden Entscheidungsregel: Fällt nach Erhebung der Daten die Prüfgröße in den kritischen Bereich, so verwerfen wir H_0 und entscheiden uns für die H_1 . Fällt die Prüfgröße nicht in den kritischen Bereich, so rechnen wir weiterhin mit der Möglichkeit, daß H_0 richtig ist. H_0 wird aber nicht unter Verwerfung von H_1 akzeptiert, sondern lediglich als weitere Möglichkeit *neben* H_1 beibehalten.

Die Formulierung des kritischen Bereichs muß nach diesem Entscheidungsmodell vor Erhebung der Daten erfolgen!

2) Die Wahl des kritischen Bereichs

a) Beispiele

b) Fehlertypen und Risiken

Einen Fehler erster Art ("Fehler I") machen wir, wenn wir die Nullhypothese verwerfen, obwohl sie richtig ist.

Einen Fehler zweiter Art ("Fehler II") machen wir, wenn wir die Nullhypothese beibehalten, obwohl 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 .

c) Die Abhängigkeit der Risiken von der Wahl des kritischen Bereichs.

Die Höhe beider Risiken hängt von der Wahl des kritischen Bereichs ab. Wir stehen in dem Dilemma, entweder ein hohes Risiko I oder ein hohes Risiko II in Kauf zu nehmen. Es gibt nur zwei Prinzipien, aus diesem Dilemma herauszukommen:

- 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; s.u. @d)

- 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 (s.u. @e).

d) Näheres zum Prinzip des konservativen Testens

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. Wir behaupten damit aber noch nicht, daß dieses Gesetz nicht zuträfe. Insofern ist der Fehler I schlimmer als ein Fehler II.

Wir wählen daher unseren kritischen Bereich so, daß das Risiko I niedrig ist. Meistens wählt man ein Risiko I (eine "Irrtumswahrscheinlichkeit") von 5 %. Man spricht von signifikanten Abweichungen von der Nullhypothese nur, wenn die Irrtumswahrscheinlichkeit höchstens 5 % beträgt. Wenn H_1 besonders stark von allen bisherigen Erfahrungen abweicht, wählt man ein "Signifikanzniveau" von 1 % oder weniger. 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 besonderen Schaden anderer Art) verbunden ist. - Ergebnisse, die auf dem 1 % Niveau signifikant sind, nennt man auch sehr signifikant.

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.

$$\text{Teststärke} = 1 - \beta.$$

Die Teststärke hängt von folgenden Faktoren ab:

1. Je stärker die Population von der Nullhypothese abweicht, um so stärker ist der Test.
2. Je strenger die Entscheidungsregel, je geringer also das Risiko I, das wir in Kauf zu nehmen bereit sind, um so schwächer ist der Test.
3. Je größer die Stichprobe, um so stärker wird der Test.

f) Einseitige und zweiseitige Entscheidungsregeln

(Dieses Thema wird im Skript anläßlich von Beispielen zu einzelnen Signifikanztests abgehandelt.)

Bei einer zweiseitigen Entscheidungsregel führen Abweichungen von der Nullhypothese in beiden möglichen Richtungen zum Verwerfen der Nullhypothese.

Eine einseitige Entscheidungsregel beinhaltet nur für eine der möglichen Abweichungen von der Nullhypothese ein Verwerfen der Nullhypothese.

Die bei der einseitigen Entscheidungsregel zur Anwendung kommende einseitige Überschreitungswahrscheinlichkeit ist halb so groß wie die zweiseitige Überschreitungswahrscheinlichkeit des jeweiligen Tabellenwerts. Damit ist eine Zunahme der Teststärke verbunden.

Einseitige Entscheidungsregeln kommen dann zur Anwendung, wenn vor dem Experiment klar ist, daß nur eine der möglichen Abweichungen von der Nullhypothese in Frage kommt oder wenn sich aus einer der möglichen Abweichungen die gleichen Handlungskonsequenzen ergeben würden wie aus einem Zutreffen der Nullhypothese.

IV. Einzelne Signifikanztests

Es sind in der Statistik sehr viele Techniken zur Prüfung von Hypothesen entstanden. Sie unterscheiden sich insbesondere in folgenden Merkmalen:

- 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?

1. Signifikanztests zur Prüfung von Hypothesen über einzelne Parameter

a) Der t-Test für einen einzelnen Mittelwert

Dieser Test überprüft, ob ein Populationsmittel einen bestimmten Wert (μ_0) hat:

$$H_0: \mu_x = \mu_0$$

Zusatzvoraussetzung: Normalverteilung der X-Werte in der Population.

Wichtigstes Anwendungsgebiet: Als t-Test für Paardifferenzen.

b) Signifikanzprüfungen für einzelne Korrelationskoeffizienten

ba) Der t-Test für die Signifikanz einer Korrelation

Überprüft, ob die Populationskorrelation null ist.

$$H_0: \rho_{xy} = 0$$

bb) Signifikanzprüfung über Tabellen

In einigen Büchern findet man Tabellen, in denen die niedrigsten lt. obigem t-Test signifikant von

0 verschiedenen Korrelationskoeffizienten zusammengestellt sind.

bc) Signifikanzprüfung über die Fisher'sche z' -Transformation

Er überprüft die Hypothese, daß ρ_{xy} einen bestimmten Wert ρ_0 hat:

$$H_0: \rho_{xy} = \rho_0.$$

Zusatzvoraussetzung: Bivariate Normalverteilung von x und y in der Population; $N \geq 50$.

Dieser Signifikanztest basiert auf einer Transformation der Korrelation in eine andere Skala (die z' -Skala), die von $-\infty$ bis $+\infty$ geht.

2) Signifikanztests zum Vergleich zweier Populationsparameter anhand von unabhängigen Stichproben

Zwei Stichproben heißen unabhängig, wenn jeder Meßwert in der einen Stichprobe stochastisch unabhängig von der Gesamtheit der Meßwerte der anderen Stichprobe ist.

Gegenbeispiel: Korrelierte Stichproben (s.u. @E.IV.3.)

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: Normalverteilung der Meßwerte, unabhängige Stichproben.

Exkurs: Der Begriff der Freiheitsgrade

In einer Stichprobe der Größe N sind nur $N - 1$ Abweichungswerte (Abweichungen vom Stichprobendurchschnitt) "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.

Bei zweiseitiger Alternativhypothese berechnen wir beim F -Test:

$$F = \frac{\text{größere Varianzschätzung}}{\text{kleinere Varianzschätzung}}$$

Die Irrtumswahrscheinlichkeit beim Verwerfen von H_0 ist dann doppelt so groß wie die in der Tabelle berücksichtigte einseitige Überschreitungswahrscheinlichkeit des F -Wertes!!

b) Der t -Test für den Vergleich zweier Populationsmittel bei homogener Varianz und unabhängigen Stichproben

Überprüft, ob zwei Populationsmittel sich um einen bestimmten Betrag Δ_0 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.

Exkurs: Das Problem der Zusatzvoraussetzungen bei Signifikanztests

a) Der traditionelle Weg: "Vortests"

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 . Es besagt nur, daß die Daten *auch* mit H_0 vereinbar sind.

b) Modernere Verfahren

ba) Skalentransformation

Die Normalverteilungsannahme wird für die Stichprobe erfüllt, indem man die Maßskala so transformiert, daß eine Normalverteilung entsteht.

bb) Versuchsplanerische Maßnahmen

Die Zusatzvoraussetzung der Varianzhomogenität verliert erheblich an Gewicht, wenn die Stichproben gleich groß sind.

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

Insbesondere sind hier die "verteilungsfreien" oder "parameterfreien" Verfahren zu erwähnen (s.u.).

c) Der z-Test für den Vergleich zweier Produktmomentkorrelationen

Überprüft, ob zwei Produkt-Moment-Korrelationen in der Population gleich sind:

$$H_0: \rho_1 = \rho_2.$$

Der Signifikanztest arbeitet mit der Fisher' schen z'-Transformation.

3) Signifikanztests zum Vergleich zweier Populationsparameter anhand von korrelierten Stichproben

Zwei Meßwertreihen (Stichproben) heißen korreliert, wenn jeweils einem Meßwert der einen Reihe ein Meßwert der anderen Reihe zugeordnet ist.

Beispiele:

- Die zwei Messungen x und y , für deren Populationsvergleich der Test durchgeführt wird, werden an den gleichen V_{pn} erhoben.
- Bildung von V_{pn} -Paaren: Meßwert x wird an der einen V_p erhoben, Meßwert y an der anderen, die der ersten in irgendeiner Weise ähnlich ist.

Vorteil der Versuchspläne mit abhängigen Stichproben: größere Teststärke, solange x und y positiv korrelieren.

a) Der t -Test für Paardifferenzen

Nullhypothese wie beim t -Test für unabhängige Stichproben. 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, dessen Zusatzvoraussetzung (Normalverteilung) sich auf die d -Werte bezieht.

4) Verteilungsfreie oder parameterfreie Signifikanztests

a) Allgemeines

Die folgenden Tests unterscheiden sich in mehreren Punkten von "klassischen":

1. Die Hypothesen sind unspezifischer. Die Alternativhypothese lautet beispielsweise: Es bestehen Unterschiede in der zentralen Tendenz (ohne spezifische Bezugnahme auf eine bestimmte Kennziffer der zentralen Tendenz). Vorteil: Möglichkeit der Anwendung bei Ordinalskalen
2. Weniger (meistens keine) Zusatzvoraussetzungen über die Populationsverteilung des Merkmals
3. Bei manchen Tests geringere Teststärke als bei den "klassischen".

Zu 3: Kompensation durch Vergrößerung der Stichprobe.

Das Effizienzverhältnis zweier Tests ist das umgekehrte Verhältnis der zur Erzielung gleicher Teststärke erforderlichen Stichprobengrößen

b) Der Vorzeichentest

Dieser Test ist nicht als solcher "Merkposten"; seine Darstellung dient lediglich dem Verständnis des Vorgehens. Er hat sogar eine besonders niedrige Teststärke!

c) Der Mann-Whitney U-Test

Der U-Test von Mann und Whitney überprüft - wie der t -Test für unabhängige Stichproben - Unterschiede in der zentralen Tendenz bei unabhängigen Stichproben; im Unterschied zum t -Test ist er auch bei Ordinalskalenniveau problemlos und braucht weder Normalverteilung noch Varianzhomogenität als Zusatzvoraussetzung.

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).

Bei Rangkopplungen gibt es spezielle Korrekturformeln, die das Ergebnis aber meist nur geringfügig verändern.

Bei hinreichend großen Stichproben kann die Verteilung der Prüfgröße U mit einer Normalverteilung approximiert werden.

d) Der χ^2 -Test nach Pearson

Der Test überprüft, ob die Abweichung beobachteter Häufigkeiten (f_o) von den Häufigkeiten, die unter einer bestimmten Nullhypothese zu erwarten wären (f_e) noch durch Zufall entstanden sein kann. Je nach Art dieser Nullhypothese unterscheidet man verschiedene Anwendungsformen, insbesondere die folgenden:

da) Der χ^2 -Test für die Güte der Anpassung ohne zusätzliche Parameterschätzung

Die Nullhypothese behauptet eine ganz bestimmte Wahrscheinlichkeitsverteilung (oder Populationsverteilung) der Beobachtungen auf die einzelnen Klassen.

db) Der χ^2 -Test für die Güte der Anpassung mit vorgeschalteter Parameterschätzung

Dieser Test überprüft, ob die Populationsverteilung zu einer bestimmten Familie gehört (z.B. Normalverteilung); die Parameter der Verteilungsfamilie werden von der Nullhypothese nicht spezifiziert, sondern aus den Daten geschätzt.

dc) Der χ^2 -Test für stochastische Unabhängigkeit in zweidimensionalen Verteilungen

Dieser Test überprüft, ob zwei kategoriale Merkmale stochastisch unabhängig sind. (D.h., daß die f_e -Werte, von denen die beobachteten Häufigkeiten laut Nullhypothese nur zufällig abweichen, in einer Zeile (bzw. Spalte) der Tabelle zueinander in dem gleichen Verhältnis stehen wie die entsprechenden Spaltensummen (bzw. Zeilensummen).

Hinweis: Die Merkposten zu den Konfidenzintervallen, die im Skript an dieser Stelle behandelt werden, stehen hier am Ende von Abschnitt D, wo sie eigentlich hingehören.

F Varianzanalyse

I Die Einweg-Varianzanalyse

1) Allgemeine Zielsetzung

Gleichzeitiger Vergleich mehrerer Durchschnitte mit unabhängigen Stichproben. Die Nullhypothese besagt, daß die entsprechenden Populationsmittel alle gleich sind.

2) Das Radikalitäts-Problem bei mehrfacher Anwendung des t-Tests

Würden wir die Nullhypothese durch mehrfache Anwendung des t-Tests überprüfen, so wäre das Risiko I überhöht!

3) Zusatzvoraussetzungen

- Unabhängige Stichproben.
- Die Populationsverteilungen, deren Mittel verglichen werden, müssen Normalverteilungen sein (Normalitätsannahme) und gleiche Varianzen haben (Homogenitätsannahme).

4) Die Aufspaltung der Varianz in Varianzquellen.

a) Die Zerlegung der Abweichungen vom Gesamtdurchschnitt.

Jede Varianz beruht auf der Abweichung der Meßwerte vom Durchschnitt.

Haben wir mehrere Gruppen, dann läßt sich für jeden Meßwert seine Abweichung vom Gesamtdurchschnitt additiv zerlegen in

- die Abweichung des Meßwerts vom zugehörigen Gruppenschnitt und
- die Abweichung des Gruppenschnitts vom Gesamtdurchschnitt.

Auf diesen beiden Abweichungen beruhen die "Varianzquellen": Die "Varianz innerhalb der Gruppen" beruht auf den Abweichungen der Meßwerte vom jeweiligen Gruppenschnitt, und die "Varianz zwischen den Gruppen" auf den Abweichungen der Gruppenschnitte vom Gesamtdurchschnitt.

b) Die Zerlegung der Abweichungs-Quadratsumme.

Auch die Summe der quadrierten Abweichungen vom Gesamtdurchschnitt - die "Gesamt-Quadratsumme" oder "totale Quadratsumme" (Abkürzung: $QS_{\text{tot.}}$) läßt sich additiv zerlegen in

- die Summe der quadrierten Abweichungen der Meßwerte vom jeweiligen Gruppenschnitt (Bezeichnung: "Quadratsumme innerhalb der Gruppen" oder auch einfach "Quadratsumme innerhalb", Abk.: QS_{in}) und

- die Summe der quadrierten Abweichungen der Gruppendurchschnitte vom Gesamtdurchschnitt (Bezeichnung: "Quadratsumme zwischen den Gruppen" oder "Quadratsumme zwischen", Abk.: QS_{zw}).

c) Vereinfachte Berechnung der Quadratsummen.

Bei der Varianz-Berechnung (Statistik I) zeigte sich, daß es einfacher ist, nicht die Definitionformel zugrundezulegen. Entsprechende Vereinfachungsformeln gibt es auch für die drei Quadratsummen in der Einweg-Varianzanalyse.

d) Die Freiheitsgrade der verschiedenen Varianzquellen.

Auch die $N - 1$ Freiheitsgrade für die Abweichungen der Meßwerte vom Gesamtdurchschnitt (df_{tot}) lassen sich additiv aufteilen in "Freiheitsgrade innerhalb der Gruppen (df_{in})" und "Freiheitsgrade zwischen den Gruppen (df_{zw})".

e) Gesamtdurchschnitt und Freiheitsgradzahl zwischen den Gruppen bei unterschiedlicher Gruppengröße.

Bei unterschiedlicher Größe der Gruppen berechnet sich der Gesamtdurchschnitt etwas anders als bei gleicher Größe; am Prinzip der Varianzzerlegung ändert das aber nichts.

5) Logik und Durchführung der Einwegvarianzanalyse.

Wenn wir die Quadratsummen zwischen den Gruppen und innerhalb der Gruppen durch die entsprechenden Freiheitsgrade dividieren, erhalten wir die "Mittelquadrate" MQ_{zw} und MQ_{in} . Diese werden mit einem F-Test verglichen, der überprüft, ob MQ_{zw} signifikant größer ist als MQ_{in} . Dieser F-Test überprüft, ob die vorgefundene Varianz zwischen den Gruppen signifikant größer ist, als es aufgrund der Varianz innerhalb der Gruppen "per Zufall" zu erwarten wäre.

6) "Effekte" und "Fehler"

a) Der Begriff des Effekts

Unter einem Effekt versteht man in der Varianzanalyse die Abweichung des zu einem Gruppenmittelwert gehörenden Populationsmittels vom Gesamtmittelwert. Der Effekt wird aufgrund der Abweichung des Gruppendurchschnitts vom Gesamtdurchschnitt geschätzt.

b) Organismische und experimentelle Effekte

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 einfach durch eine "experimentelle" Variable.

Bei experimentell manipulierten Variablen können wir die Kausalrichtung von Effekten eindeutig festlegen: Signifikante Unterschiede der Populationsmittel können nur auf die verschiedenen experimentellen Bedingungen zurückgehen, unter denen die Meßwerte erhoben wurden. Dagegen gelten bei organismischen Effekten die gleichen Einschränkungen hinsichtlich kausaler Interpretationen wie bei Korrelationen.

c) Differentielle Effekte

Von differentiellen Effekten spricht man, wenn ein Effekt nicht bei allen Vpn gleich ist.

Liegen differentielle Effekte vor, so sind die nach den Regeln aus Abschnitt a bestimmten Effekte die Populations- bzw. Stichproben-durchschnitte der individuellen Effekte.

d) Der Begriff des Fehlers

Der Meßwertfehler bei der i-ten Beobachtung ist die Differenz des i-ten Meßwerts vom zugehörigen (wahren) Gruppenmittel.

Die Varianz innerhalb der Gruppen beruht nicht nur auf eigentlichen Meßfehlern, sondern auch auf differentiellen Effekten und auf überdauernden Eigenschaften der Vpn.

II Die Mehrweg-Varianzanalyse

Bei der Mehrweg-Varianzanalyse kommt hinzu, daß sich die Meßwertgruppen nach verschiedenen Variablen (den "Faktoren") unterscheiden. Die einzelnen Ausprägungsformen der Faktoren nennt man auch "Stufen" des Faktors.

Der Effekt einer Stufe eines Faktors ist die Differenz zwischen dem entsprechenden Mittelwert und dem Gesamtmittel ("Haupteffekt").

Von einer Wechselwirkung spricht man in der Varianzanalyse, wenn die Effekte des einen Faktors auf den verschiedenen Stufen des anderen Faktors unterschiedlich sind.

Der mit einer Zelle verbundene Wechselwirkungseffekt ist die Differenz zwischen dem zu dieser Zelle gehörenden Mittelwert und demjenigen Mittelwert, der sich ergeben würde, wenn nur die entsprechenden Haupteffekte zum Gesamtmittel addiert würden.

Das Vorliegen einer Wechselwirkung zeigt sich im "Interaktionsdiagramm" dadurch, daß die Kurven für verschiedene Stufen eines Faktors nicht parallel sind.

Für die Berechnung einer Mehrweg-Varianzanalyse wird die Quadratsumme zwischen den Gruppen noch einmal zerlegt in Anteile, die auf die Haupteffekte zurückgehen sowie einen Anteil, der auf Wechselwirkung zurückgeht. Für jede dieser Quadratsummen gibt es eine Freiheitsgradzahl und ein Mittelquadrat (= Quadratsumme : Freiheitsgrade).

Ist das Mittelquadrat signifikant größer als MQ_{in} (F-Test), dann ist die entsprechende

Nullhypothese (keine Effekte in der Population) zu verwerfen.

Schlußbemerkung

Es gibt auch noch höhere Stufen der Varianzanalyse, bei denen noch weitere Probleme auftreten, die in diesem Kurs nicht mehr behandelt werden.