Einführung in die Biostatistik

Vitaly Belik

Institut für Veterinär-Epidemiologie und Biometrie, FU Berlin

Oct 06, 2020

Kontakt

Prof. Dr. Vitaly Belik

Fachbereich Veterinärmedizin

Institut für Veterinär-Epidemiologie und Biometrie

Juniorprofessor

Leitung Arbeitsgruppe Systemmodellierung

Adresse	Königsweg 67 Raum 104 14163 Berlin
Telefon	+49 30 838 61129
Fax	+49 30 838 4 61129
E-Mail	vitaly.belik@fu-berlin.de
Homepage	Working Group Modelling

Werdegang

2004	MSc Physics / Biochemical Physics, Moscow Lomonosov University
2000-2001	Physics, HU Berlin, unterstützt durch Siemens AG
2004-2013	MPI für Dynamik und Selbstorganisation, Göttingen
2008	Dr. rer. nat. in Theoretischer Physik, Georg-August-Universität Göttingen
2010-2012	Massachusetts Institute of Technology, Cambridge, MA, USA
2013-2015	Gastwissenschaftler am Helmholtz-Zentrum für Infektionsforschung, Braunschweig
2014-2016	TU Berlin
2016-	Professor (W1) FU Berlin, AG Systemmodellierung, Institut für Veterinär-Epidemiologie und Biometrie

Neue Medien – Persönlichkeitsrechte - Copyright

- Keine unautorisierten Foto-, Ton- und Video-Aufzeichnungen während der Vorlesung
- Kein unberechtigtes Weiterverbreiten oder “ins Netz stellen” von Inhalten aus der Veranstaltung
- Klare Kennzeichnung von Zitaten
- Quellenabgabe aller “fremder” Materialien

Was ist Statistik / Biostatistik (Biometrie)?

- Was sind Ihre Erwartungen?
- Go to http://menti.com

Was ist Statistik?

- Welche Daten soll man zur Beantwortung einer gegebenen Aufgabenstellung ermitteln?
- Wie viel Daten soll man ermitteln?
- Auf welche Art soll man das Untersuchungsmaterial auswählen?
- Wie soll mans eine Untersuchungsdaten ermitteln?
- Wie sollen die gewonnenen Daten geordnet werden?
- Wie sollen die Daten beschrieben und übersichtlich dargestellt werden?
- Wie wertet man die Daten aus?
- Welche Schlüsse lassen sich ziehen?
- Wie zuverlässig sind die getroffenen Aussagen?
- Welche weiterführenden Fragestellungen haben die Ergebnisse aufgeworfen?

Was ist Statistik? (1)

1. Aufgabenstellung. Nach präziser Formulierung der Fragestellung muss eine geeignete Wahl von Merkmalengetroffen,eine Mess-bzw. Beobachtungsmethode festgelegt und ein Versuchsplan aufgestellt werden.
2. Datengewinnung. Gewinnung des Untersuchungsmaterials (Ziehen der Stichprobe) und Ausführungder Messungen bzw. Beobachtungen an diesem Material.
3. Datenverarbeitung. Das gewonnene Datenmaterial muss graphisch und rechnerisch aufbereitet werden, dann sind Schlüsse von der Stichprobe auf die Grundgesamtheit zu ziehen; diese werden anschließend geprüft und interpretiert.

Was ist Statistik? (2)

Statistik

ist eine wissenschaftliche Disziplin, deren Gegenstand die Entwicklung und Anwendung von Methoden zur Datenerhebung, -beschreibung und -analyse sowie der Beurteilung der Ergebnisse ist. Dabei unterscheidet man:

Was ist Statistik? (2)

Statistik

ist eine wissenschaftliche Disziplin, deren Gegenstand die Entwicklung und Anwendung von Methoden zur Datenerhebung, -beschreibung und -analyse sowie der Beurteilung der Ergebnisse ist. Dabei unterscheidet man:

Deskriptive (beschreibende) Statistik:

Methoden zur Auswertung und übersichtlichen Darstellung und Zusammenfassung von Daten.

Was ist Statistik? (2)

Statistik

ist eine wissenschaftliche Disziplin, deren Gegenstand die Entwicklung und Anwendung von Methoden zur Datenerhebung, -beschreibung und -analyse sowie der Beurteilung der Ergebnisse ist. Dabei unterscheidet man:

Deskriptive (beschreibende) Statistik:

Methoden zur Auswertung und übersichtlichen Darstellung und Zusammenfassung von Daten.

Induktive (schliessende) Statistik:

Methoden zum Treffen von vernünftigen Entscheidungen im Falle von Unsicherheit bzw. Risiko. “Den Zufall in den Griff bekommen”. “Sicherheit über Unsicherheit gewinnen”.

Biostatistik oder Biometrie

Biostatistik (Biometrie)

angewandte Statistik zur Beschreibung, Modellierung und Beurteilung biologisch-naturwissenschaftlicher Phänomene.

Biostatistik oder Biometrie

Biostatistik (Biometrie)

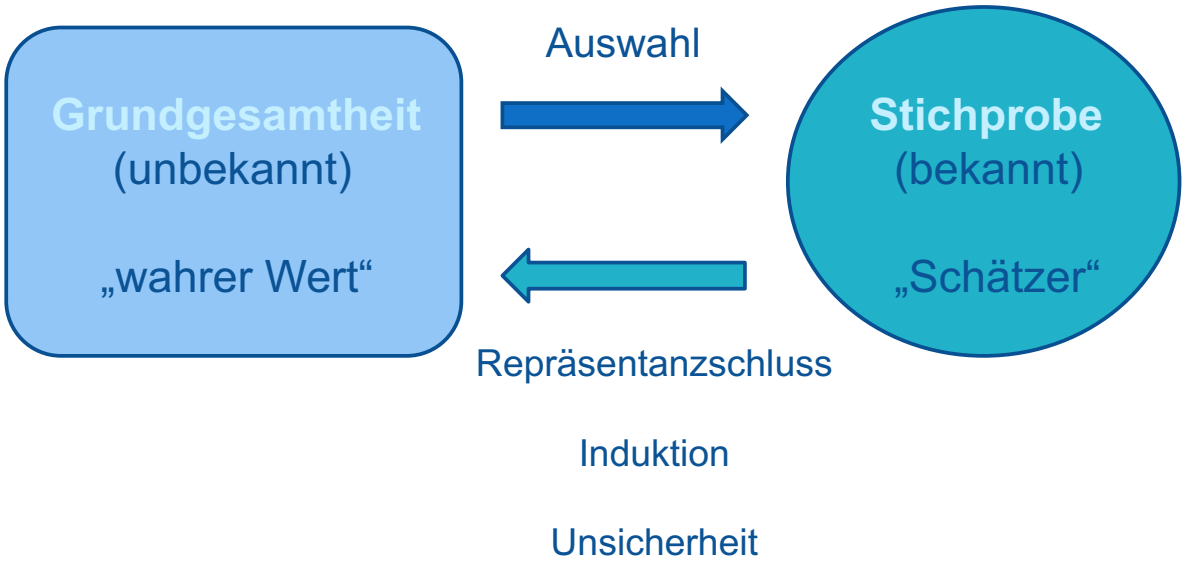
angewandte Statistik zur Beschreibung, Modellierung und Beurteilung biologisch-naturwissenschaftlicher Phänomene.

Beispiele

- Wie sicher ist das Ergebnis eines Diagnosetests zur Bestimmung einer Erkrankung?
- Wie viele Versuche müssen durchgeführt werden, um Verbesserung eines Produktes zu gewährleisten?

Deskriptive Statistik wird manchmal als *explorative* und Schliessende Statistik als *konfirmatorische Datenanalyse* bezeichnet.

Daten und Zufall



[Grafik: M. Doherr]

Statistische Schlüssweise

- Schätzen der unbekannten Parameter der Grundgesamtheit. “Finde eine Größe aus den Daten der Stichprobe, die”möglichst nah" an der unbekannten Wirklichkeit ist."
- Angabe von Konfidenzintervallen (Vertrauensbereichen). “Gebe basierend auf den Daten der Stichprobe ein Intervall an, das den wahren Wert (Populations-Parameter) mit einer gewissen Wahrscheinlichkeit überdeckt.”
- Entscheiden mittels eines statistischen Tests, ob anhand der Daten der Stichprobe eine Aussage über einen Parameter der Grundgesamtheit (bspw. Unterschied eines Mittelwertes zwischen Gruppen) wahr oder falsch ist.

Lernziele des Kurses

Ziel des Kurses is es Ihnen die wichtigsten statistischen Methoden zur Planung und Auswertung der Versuche und Daten aus wissenschaftlicher Studien zu vermitteln.

Sie sollen die Notwendigkeiten, Möglichkeiten und Grenzen grundlegender statistischer Analysen verstehen und selbst einfache statistische Berechnungen durchführen können.

Falls nötig, sollen Sie in der Lage sein bei einer statistischen Beratung, Ihr Anliegen sicher zu kommunizieren.

Kursübersicht

Nr	Date	Topic
1	06.10.20	Einführung. Arten von Daten.
2	13.10.20	Deskriptive Statistik
3	20.10.20	Schliessende Statistik. Parameterschätzung. Konfidenzeintervalle.
4	27.10.20	Hypothesentest, p-value
5	03.11.20	Zusammenhänge in Daten (kategoriell und kontinuierlich)
6	10.11.20	Grundlagen von Modellbildung. Lineare Regression. Modeldiagnostik
7	17.11.20	Generalized linear models. Mit einer unabhängigen und mehreren Variablen
8	24.11.20	Vergleich von zwei Mittelwerten ANOVA (einfache)
9	01.12.20	ANOVA 1 (mehrfache)
10	08.12.20	ANOVA 2 (mit Messwiederholungen)
11	15.12.20	Gemischte Modelle
12	05.01.21	Logistische Regression
13	12.01.21	Cluster- und Diskriminanzanalyse
14	19.01.21	Nichtparametrische Tests
15	26.01.21	Survival analysis
16	02.02.21	Elemente der Versuchsplanung
17	09.02.21	Konsultationen

Organisatorisches

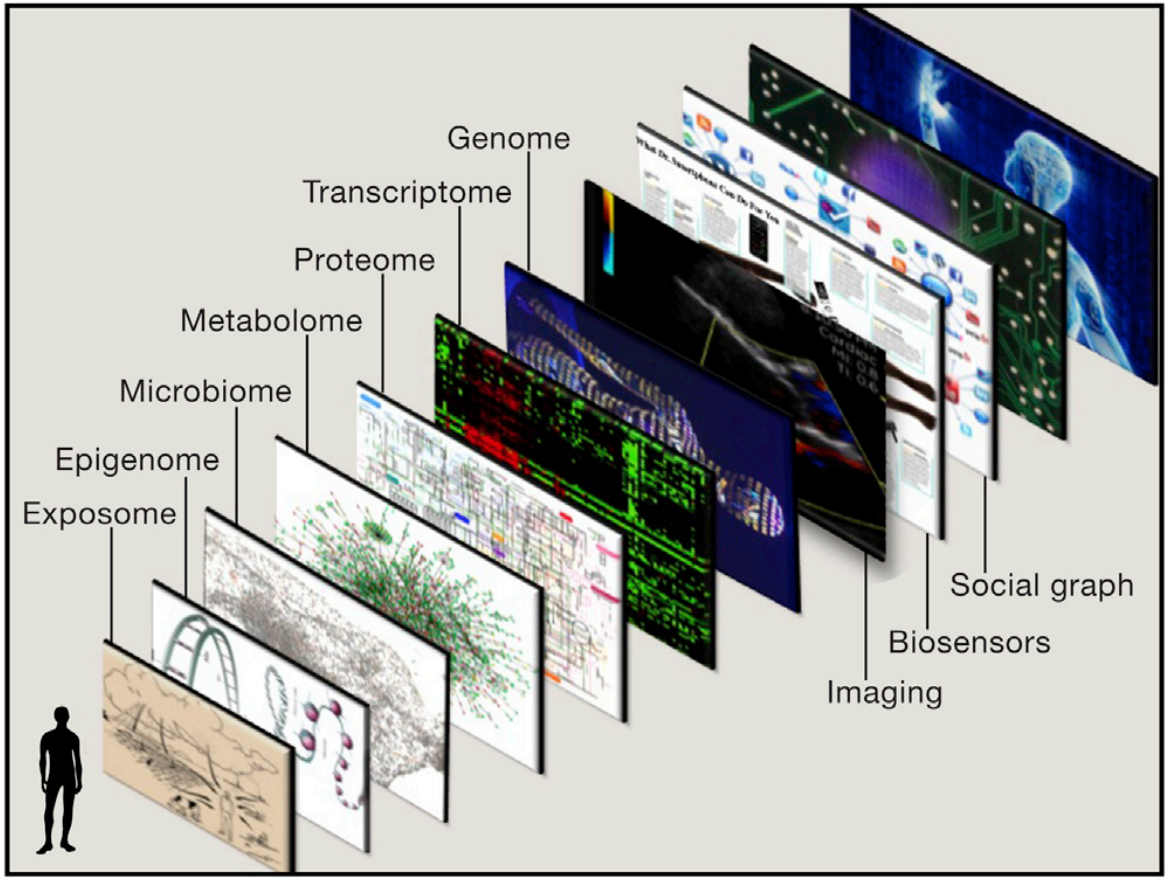
- Der Kurs besteht aus *Vorlesungen* und praktischen *Übungen*
- Für die Übungen wird Programmiersprache *R* (*RStudio*) benutzt
- Am Ende des Kurses ist eine *Klausur* vorgesehen
- Alternative ist es möglich ein *Projekt* zu bearbeiten (z.B. Analyse von Twitter)

Literatur

1. W. Köhler et al. *Biostatistik. Eine Einfürung für Biologen und Agrarwissenschaftler*
2. R. Kabacoff. *R in Action*
3. M. Crawley. *The R Book*
4. A. Field. *An Adventure in Statistics: The Reality Enigma*
5. A. Field *Discovering Statistics Using R*

Daten

In letzter Zeit mit der rasanten Entwicklung der ausgefallenen Sensoren (IoT), Rechner- und Speicherkapazitäten werden sehr viele Daten prozudiert.

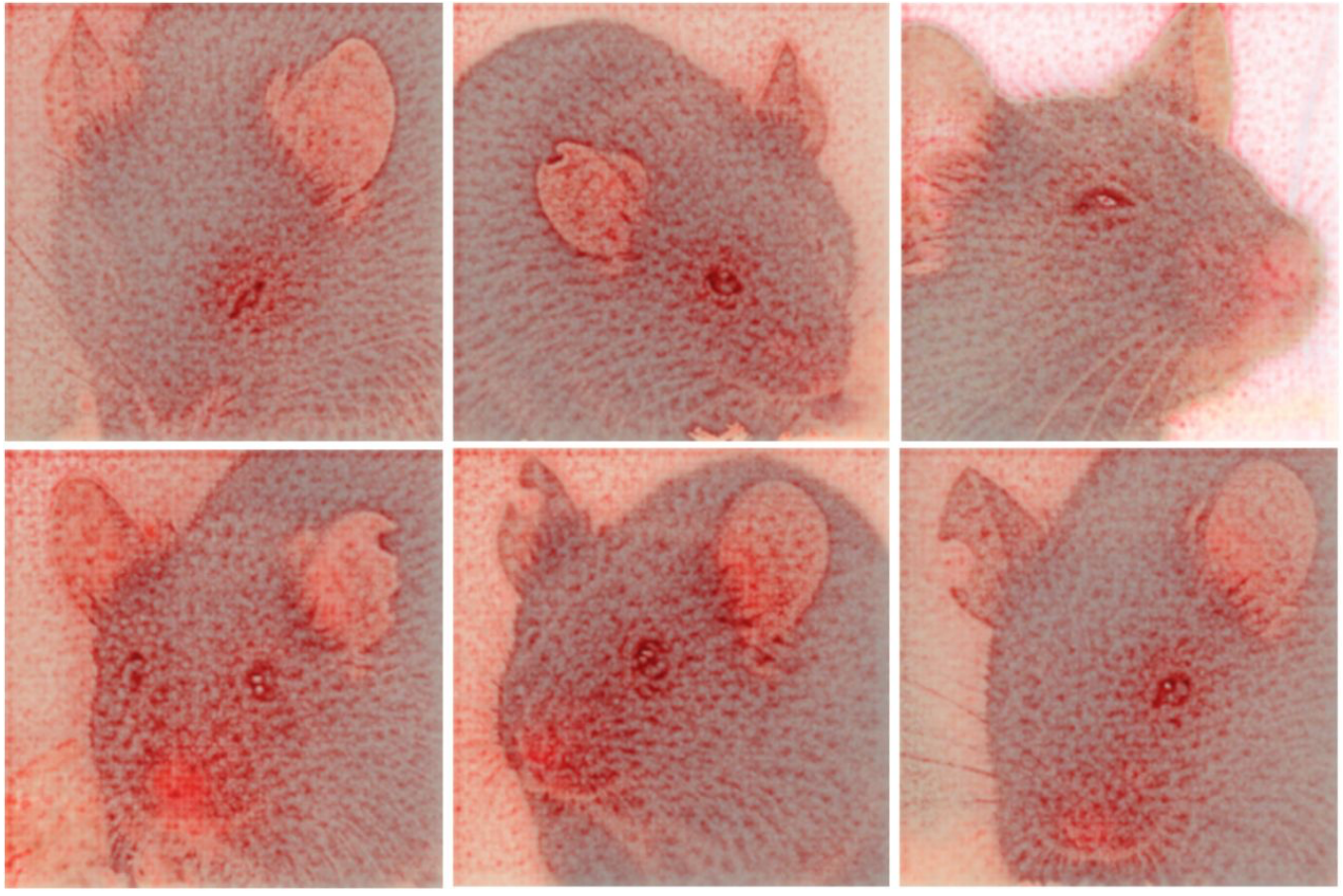


[Topol, 2014]

Daten sind heterogen

Bilder

Feature visualization von Mäuserbildern

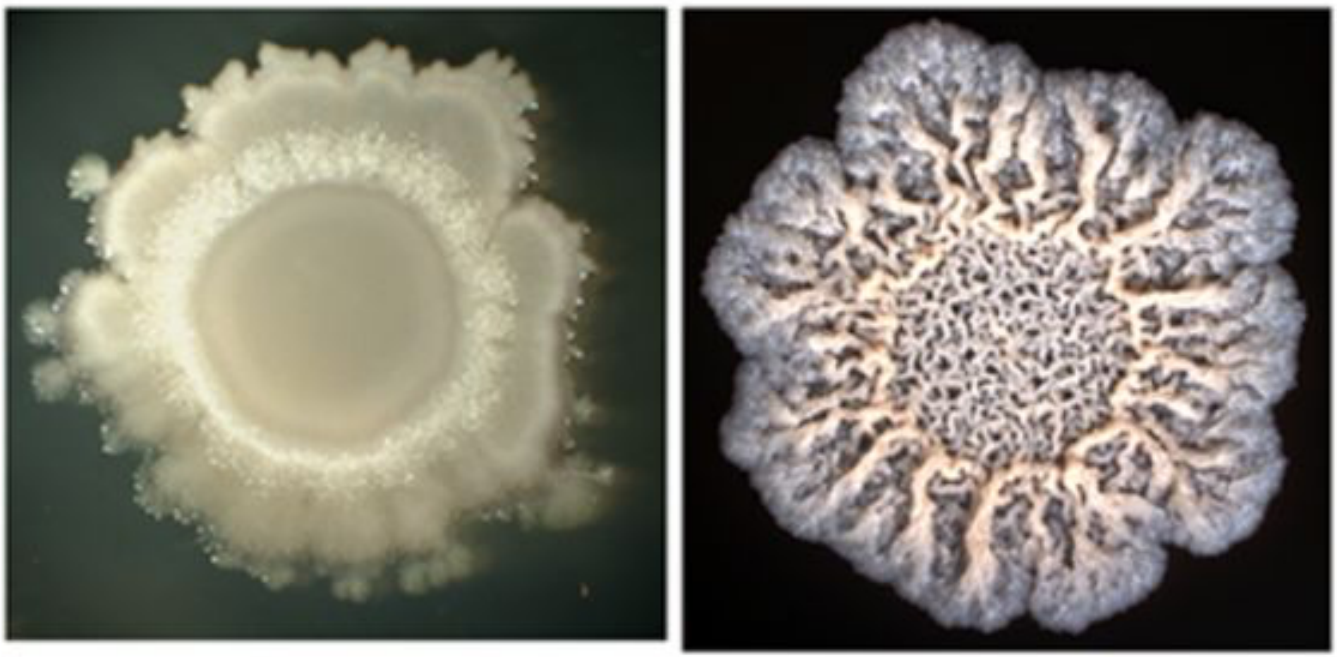


[https://doi.org/10.1371/journal.pone.0228059]

Daten sind heterogen (1)

Bilder

Biofilm von *B. subtilus*

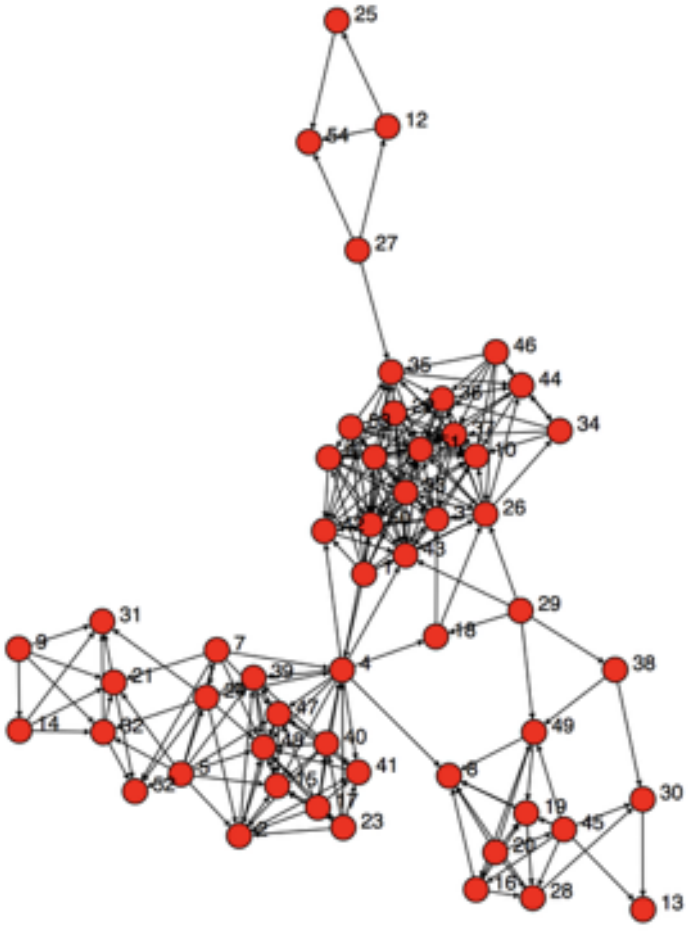


[https://doi.org/10.1128/JB.00028-13]

Daten sind heterogen (2)

Netzwerke

Kontaktnetzwerk von Tieren



[Daten: Thomas Selhorst]

Maschinelles Lernen

Große Mengen von heterogenen Daten (Big Data) verlangen nach entsprechenden Werkzeugen für die Datenanalyse. Dabei können, ausser klassischen statistischen Methoden, das *maschinelle Lernen* (z.B. *künstliche Neuronale Netze*) sehr hilfreich sein.

Maschinelles Lernen

Große Mengen von heterogenen Daten (Big Data) verlangen nach entsprechenden Werkzeugen für die Datenanalyse. Dabei können, ausser klassischen statistischen Methoden, das *maschinelle Lernen* (z.B. *künstliche Neuronale Netze*) sehr hilfreich sein.

Es stellt sich sogar die Frage, ob sich die Versuchsplanung und Datenanalyse nicht von einer Maschine erledigen lässt.

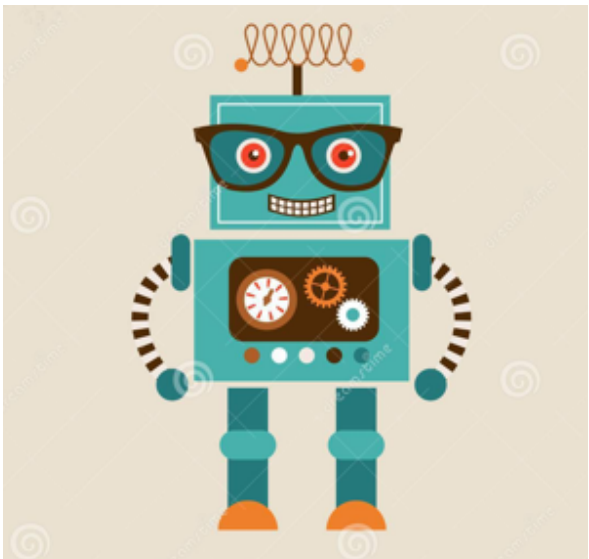
Roboscientist

SCIENCE VOL 324 3 APRIL 2009 **85**

The Automation of Science

Ross D. King,[1]* Jem Rowland,[1] Stephen G. Oliver,[2] Michael Young,[3] Wayne Aubrey,[1] Emma Byrne,[1] Maria Liakata,[1] Magdalena Markham,[1] Pınar Pir,[2] Larisa N. Soldatova,[1] Andrew Sparkes,[1] Kenneth E. Whelan,[1] Amanda Clare[1]

The basis of science is the hypothetico-deductive method and the recording of experiments in sufficient detail to enable reproducibility. We report the development of Robot Scientist "Adam," which advances the automation of both. Adam has autonomously generated functional genomics hypotheses about the yeast *Saccharomyces cerevisiae* and experimentally tested these hypotheses by using laboratory automation. We have confirmed Adam's conclusions through manual experiments. To describe Adam's research, we have developed an ontology and logical language. The resulting formalization involves over 10,000 different research units in a nested treelike structure, 10 levels deep, that relates the 6.6 million biomass measurements to their logical description. This formalization describes how a machine contributed to scientific knowledge.



Zurück zu eigentlichen Biostatistik!

Daten als Tabelle

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive	alone
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN	Southampton	no	False
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	Cherbourg	yes	False
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	NaN	Southampton	yes	True
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	C	Southampton	yes	False
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	NaN	Southampton	no	True
5	0	3	male	NaN	0	0	8.4583	Q	Third	man	True	NaN	Queenstown	no	True
6	0	1	male	54.0	0	0	51.8625	S	First	man	True	E	Southampton	no	True
7	0	3	male	2.0	3	1	21.0750	S	Third	child	False	NaN	Southampton	no	False
8	1	3	female	27.0	0	2	11.1333	S	Third	woman	False	NaN	Southampton	yes	False
9	1	2	female	14.0	1	0	30.0708	C	Second	child	False	NaN	Cherbourg	yes	False
10	1	3	female	4.0	1	1	16.7000	S	Third	child	False	G	Southampton	yes	False
11	1	1	female	58.0	0	0	26.5500	S	First	woman	False	C	Southampton	yes	True
12	0	3	male	20.0	0	0	8.0500	S	Third	man	True	NaN	Southampton	no	True
13	0	3	male	39.0	1	5	31.2750	S	Third	man	True	NaN	Southampton	no	False
14	0	3	female	14.0	0	0	7.8542	S	Third	child	False	NaN	Southampton	no	True
15	1	2	female	55.0	0	0	16.0000	S	Second	woman	False	NaN	Southampton	yes	True

Merkmale

Individuen oder Untersuchungsobjekte, die einer Erhebung / Untersuchung zu Grunde liegen, d.h. an / von denen Daten gesammelt werden, bezeichnet man als statistische Einheit, Merkmalsträger oder Untersuchungseinheiten.

Die Eigenschaften, die hinsichtlich des Untersuchungsziels an der statistischen Einheit untersucht werden, heißen *Merkmale*.

Studierendendaten

- Geschlecht, Körpergröße, Geburtsjahr
- Stadtnah oder ländlich aufgewachsen
- Wunsch, nach dem Studium in einem bestimmten Unternehmen zu arbeiten

Auswahl geeignete Merkmale

Objektivität

Die Ausprägung der zu ermittelneden Merkmales ist unabhängig von der Person des Auswerters eindeutig festszustellen.

Reliabilität

Das Merkmal gestattet reproduzierbare Mess- (bzw. Beobachtungs-) Ergebnisse, bei Wiederholung liegen also gleiche Resultate vor. Statt Reliabilität spricht man auch von “Zuverlässligkeit”.

Validität

Der Merkmal in seinen Ausprägungen spiegelt die für die Fragestellung wesentlichen Eigenschaften wider. Auch “Gültigkeit” oder “Aussagekraft” genannt.

Charakterisierung von Merkmalen

quantitative Merkmale:

Untersuchungseinheiten unterscheiden sich im absoluten (Zahlen-) Wert. - z.B. Alter, Gewicht, Temperatur, Anzahl Keime, Betriebsgröße, Schadstoffgehalt, ...

Charakterisierung von Merkmalen

quantitative Merkmale:

Untersuchungseinheiten unterscheiden sich im absoluten (Zahlen-) Wert. - z.B. Alter, Gewicht, Temperatur, Anzahl Keime, Betriebsgröße, Schadstoffgehalt, ...

qualitative Merkmale:

Untersuchungseinheiten unterscheiden sich in ihrer Ausprägung (Art) - z.B. Geschlecht, Name, Befund, Rasse, Therapie, Haltungsform, Region, ...

Skalenniveaus von Merkmalen

nominale (qualitative) Skala:

die Werte unterliegen keiner Rangfolge und sind nicht vergleichbar.

- z.B. Name, Geschlecht, Rasse, Haltungsform, Therapieform, pathologische Klassifikation

Skalenniveaus von Merkmalen

nominale (qualitative) Skala:

die Werte unterliegen keiner Rangfolge und sind nicht vergleichbar.

- z.B. Name, Geschlecht, Rasse, Haltungsform, Therapieform, pathologische Klassifikation

ordinale (qualitativ oder semiquantitative) Skala:

die Werte unterliegen einer Rangfolge, aber die Abstände zwischen den Werten der Skala lassen sich nicht interpretieren.

- z.B. Bewertung (Bonituren, Noten), Gesundheitszustand, Grad der Belastung mit Keimen (-, +, ++, +++)

Skalenniveaus von Merkmalen

nominale (qualitative) Skala:

die Werte unterliegen keiner Rangfolge und sind nicht vergleichbar.

- z.B. Name, Geschlecht, Rasse, Haltungsform, Therapieform, pathologische Klassifikation

ordinale (qualitativ oder semiquantitative) Skala:

die Werte unterliegen einer Rangfolge, aber die Abstände zwischen den Werten der Skala lassen sich nicht interpretieren.

- z.B. Bewertung (Bonituren, Noten), Gesundheitszustand, Grad der Belastung mit Keimen (-, +, ++, +++)

metrische (quantitative) Skala:

die Werte unterliegen einer Rangfolge und die Abstände zwischen den Werten der Skala lassen sich interpretieren.

- z.B. Gewicht, Betriebsgröße, Keimzahlen, ...

Skalenniveaus von Merkmalen

nominale (qualitative) Skala:

die Werte unterliegen keiner Rangfolge und sind nicht vergleichbar.

- z.B. Name, Geschlecht, Rasse, Haltungsform, Therapieform, pathologische Klassifikation

ordinale (qualitativ oder semiquantitative) Skala:

die Werte unterliegen einer Rangfolge, aber die Abstände zwischen den Werten der Skala lassen sich nicht interpretieren.

- z.B. Bewertung (Bonituren, Noten), Gesundheitszustand, Grad der Belastung mit Keimen (-, +, ++, +++)

metrische (quantitative) Skala:

die Werte unterliegen einer Rangfolge und die Abstände zwischen den Werten der Skala lassen sich interpretieren.

- z.B. Gewicht, Betriebsgröße, Keimzahlen, ...

Skalenniveaus von Merkmalen (1)

Es wird auch unterschieden zwischen

Intervallskala

Die Abstände zwischen Merkmalsausprägungen lassen sich vergleichen.
Die Skalla ist kontinuierlich.

- z.B. Temperatur in Grad Celsius

Verhältnisskala

Nicht nur die Differenz, sondern auch der Quotient aus zwei Messwerten darf verwendet werden.

- z.B. Temperatur in Kelvin, Länge in Zentimetern

Skalenniveaus von Merkmalen (2)

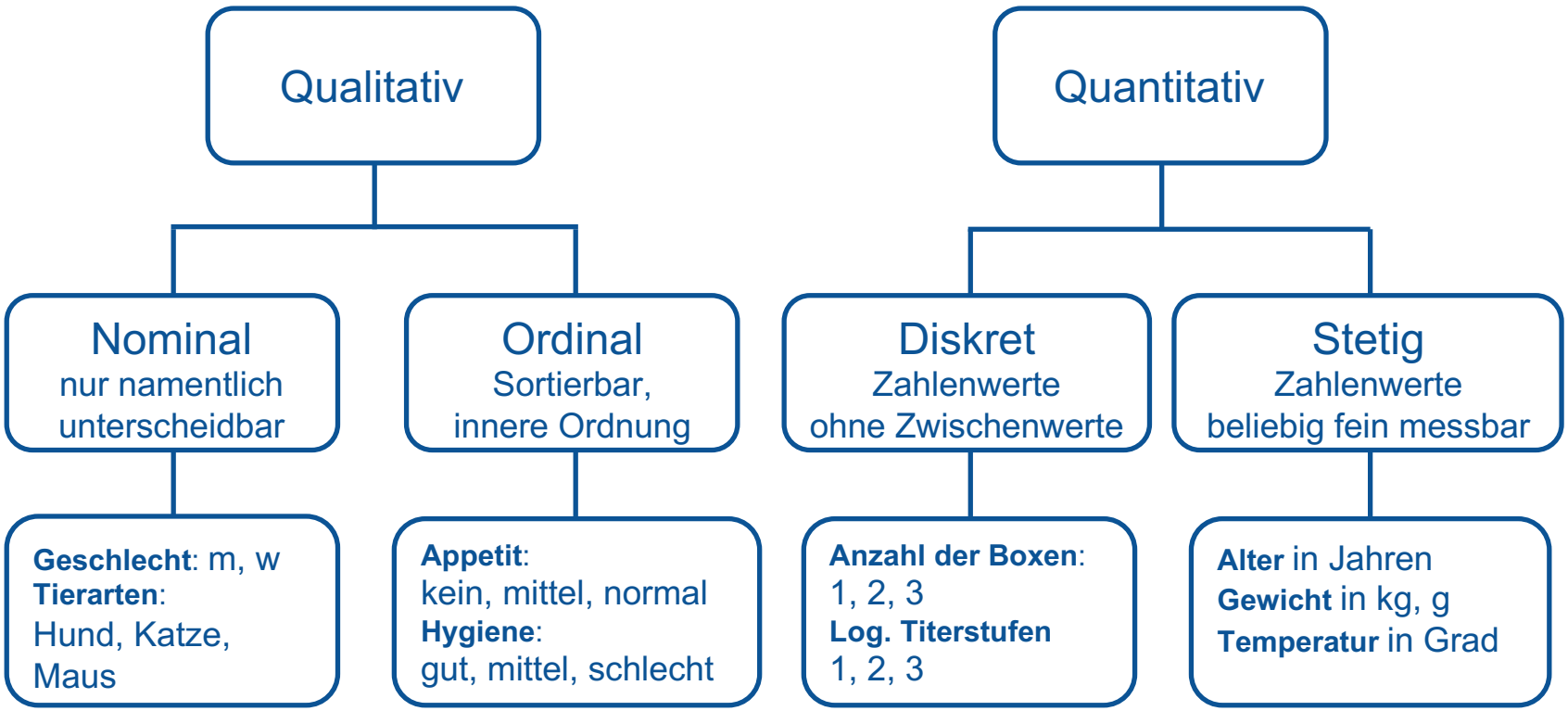
Die statistischen Auswertungsmöglichkeiten sind vom Skalenniveau abhängig, weil auf höherem Niveau mehr Information festgehalten und ausgewertet werden kann, als bei niedrigeren Skalierungen.

Skalenniveaus von Merkmalen (2)

Die statistischen Auswertungsmöglichkeiten sind vom Skalenniveau abhängig, weil auf höherem Niveau mehr Information festgehalten und ausgewertet werden kann, als bei niedrigeren Skalierungen.

Debei soll den Aufwand für den zusätlichen Informationgewinn berücksichtigt werden.

Skalenniveaus von Merkmalen (3)



Nicht jede Zahl ist eine Zahl.

Häufig werden Daten verschlüsselt, um die anschliessende Datenverarbeitung zu erleichtern

- Schulnoten: 1, 2, 3, 4, 5, 6 (ordinal)
- Testergebnis: 1, 0 (nominal)
- Kreiskennziffern: 3253, 3351 (nominal)
- Zuchtbuch-Nummern: 0511572 (nominal)

Einfache Datendarstellungen

Deskriptive Statistik

Vitaly Belik

Institut für Veterinär-Epidemiologie und Biometrie, FU Berlin

Oct 13, 2019

Was haben wir letztes mal gelernt?

http://menti.com

Wiederholung

In der letzen Vorlesung haben wir über die Aufgaben der (Bio)statistik gersprochen.

- Deskriptive (beschreibende) und Schliessende (induktive) Statistik
- Grundgesamtheit und Stichprobe
- Skalenniveaus von Merkmalen

Deskriptive (beschreibende) Statistik

Methoden zur:

- *Auswertung*
- übersichtlichen *Darstellung*
- *Zusammenfassung* von Daten.

Deskriptive (beschreibende) Statistik(1)

Tabellen

Graphiken

Charakterischtische Maßzahlen

Tabellen

Table 1: Titanic dataset

Index	survived	pclass	sex	age	deck	fare	alone
0	0	3	male	22		7.2500	False
1	1	1	female	38	C	71.2833	False
2	1	3	female	26		7.9250	True
3	1	1	female	35	C	53.1000	False
4	0	3	male	35		8.0500	True
5	0	3	male	NA		8.4583	True
6	0	1	male	54	E	51.8625	True
7	0	3	male	2		21.0750	False
8	1	3	female	27		11.1333	False
9	1	2	female	14		30.0708	False

Tabellen(1)

Urliste

Die ungeordnete Form von Messungen (Beobachtungen) einer Untersuchung, die der Reihe nach zusammengestellt ist.

Table 2: Urliste

	1	2	3	4	5	6	7	8	9	10
age	22	38	26	35	35	NA	54	2	27	14

Tabellen(1)

Primäre Tafel oder geordnete Liste

Table 3: Geordnete Liste

	8	10	1	3	9	4	5	2	7
age	2	14	22	26	27	35	35	38	54

Häufigkeitstabelle

Table 4: Häufigkeitstabelle

age	freq
0.42	1
0.67	1
0.75	2
0.83	2
0.92	1
1	7
2	10
3	6
4	10
5	4
6	3
7	3
8	4
9	8
10	2
11	4
12	1
13	2
14	6
14.5	1
15	5
16	17
17	13
18	26
19	25
20	15
20.5	1
21	24
22	27
23	15

Stamm-Blatt-Diagramm (stem and leaf diagram)

```
The decimal point is 1 digit(s) to the right of the |

0 | 011111111111112222222222333333444444444
0 | 5555666777888999999
1 | 001111233444444
1 | 555555666666666666666677777777777778888888888888888888888888999999+7
2 | 000000000000000111111111111111111111111122222222222222222222222223+33
2 | 555555555555555555555556666666666666666666777777777777777788888888+27
3 | 000000000000000000000000111111111111111111222222222222222222333333+14
3 | 555555555555555555566666666666666666666667777777888888888899999999
4 | 000000000000111111112222222222223333344444444
4 | 555555555556666677777777788888888899999
5 | 00000000011111112222223444444444
5 | 5566666778888899
6 | 000011122223344
6 | 5556
7 | 001114
7 |
8 | 0
```

Man rundet die Ergebnisse auf Blatt-Genaugkeit

Graphiken

Balkendiagramm

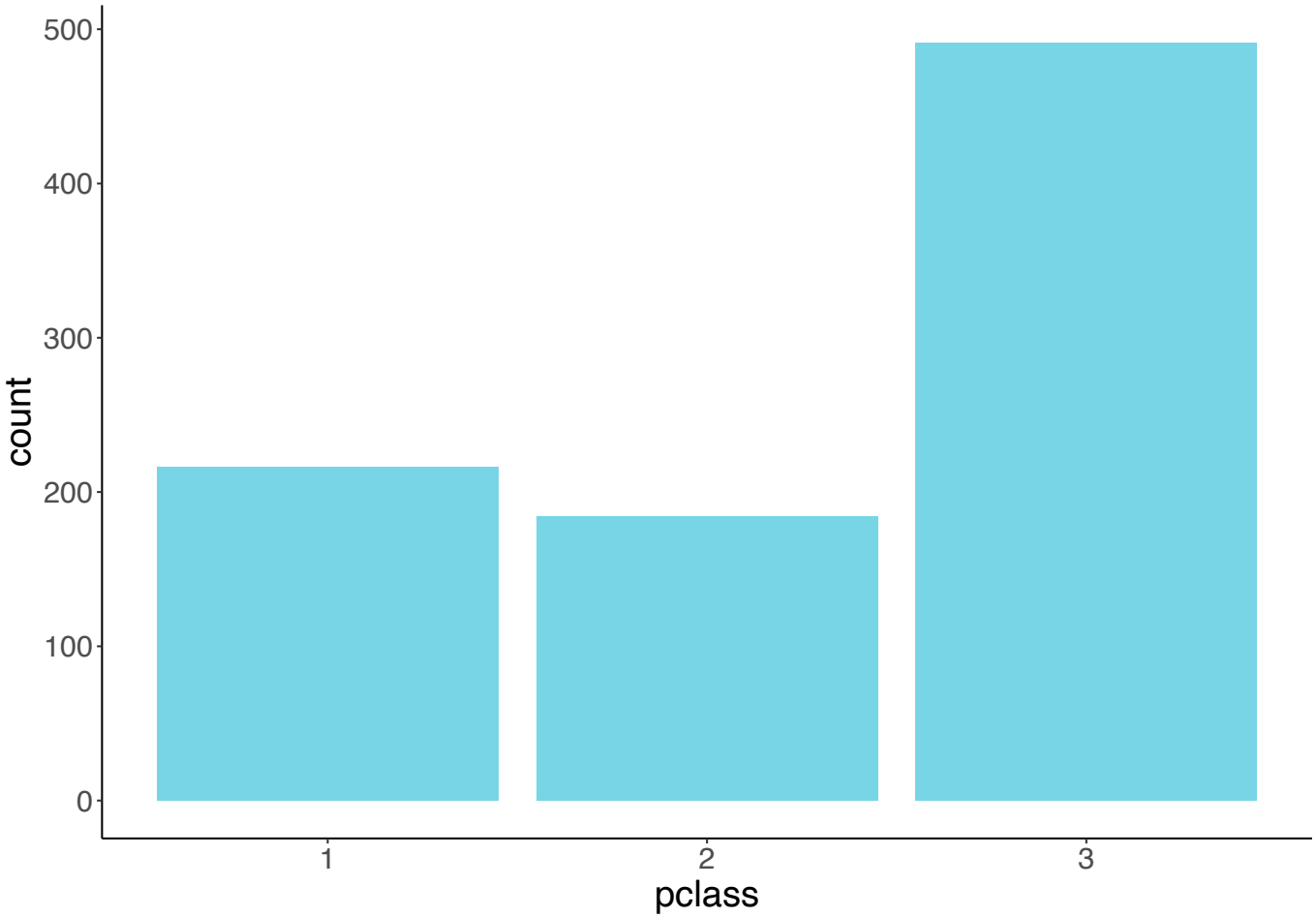


Figure 1: Balkendiagramm

Balkendiagramm (relative Häufigkeit)

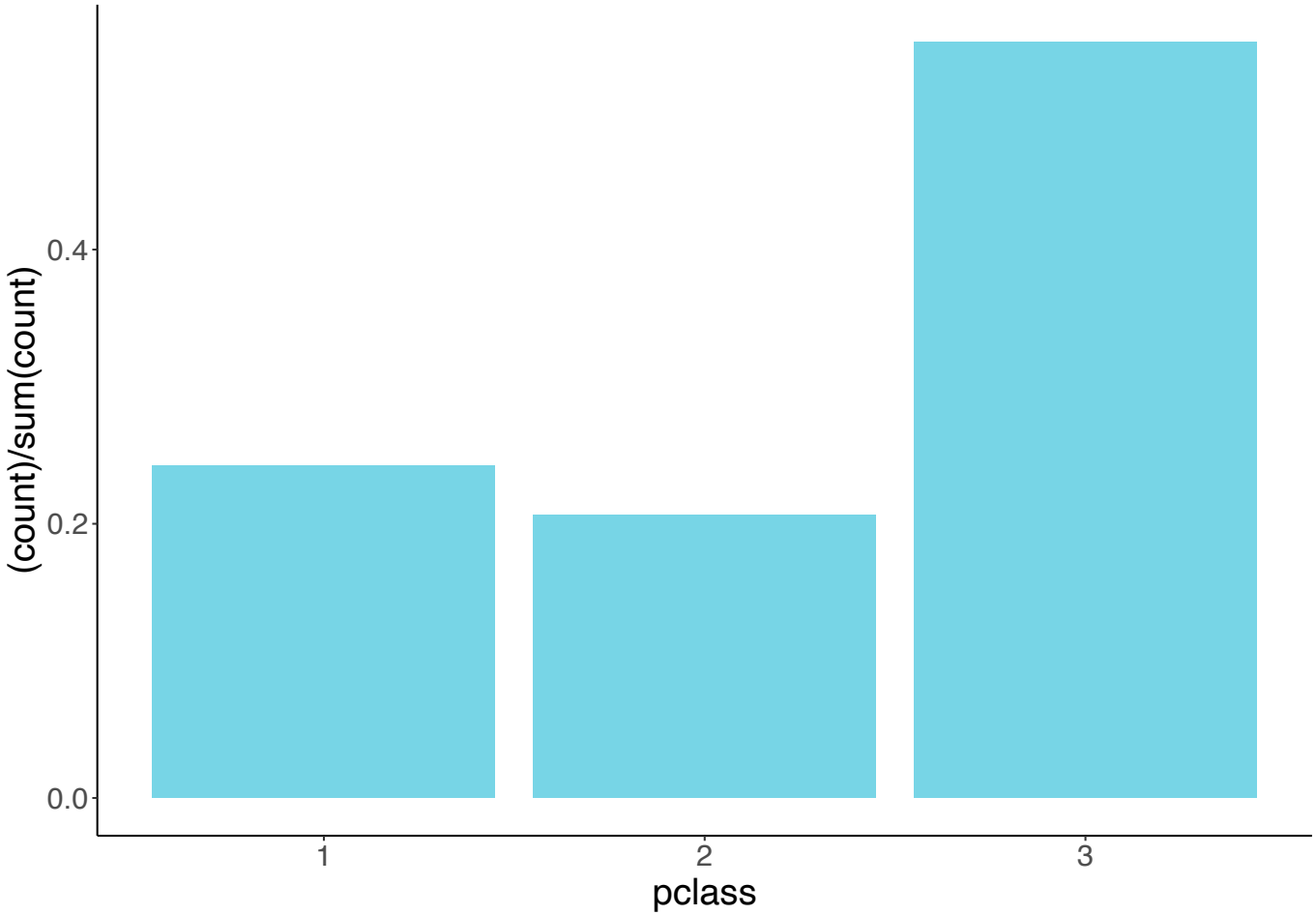


Figure 2: Balkendiagramm (relative Häufigkeit)

Komponenten-Balkendiagramm

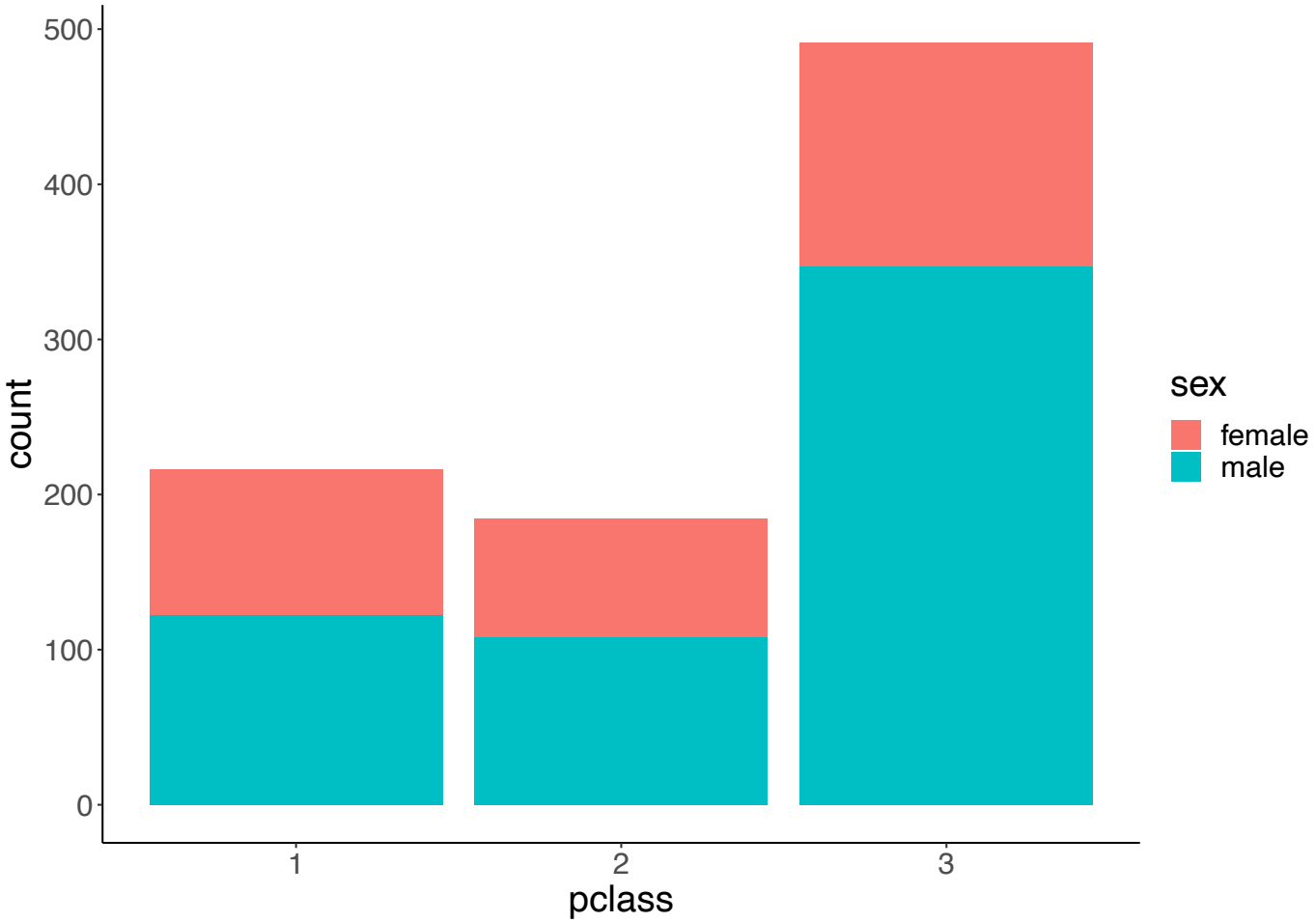


Figure 3: Komponenten-Balkendiagramm

Komponenten-Balkendiagramm (relative Häufigkeit)

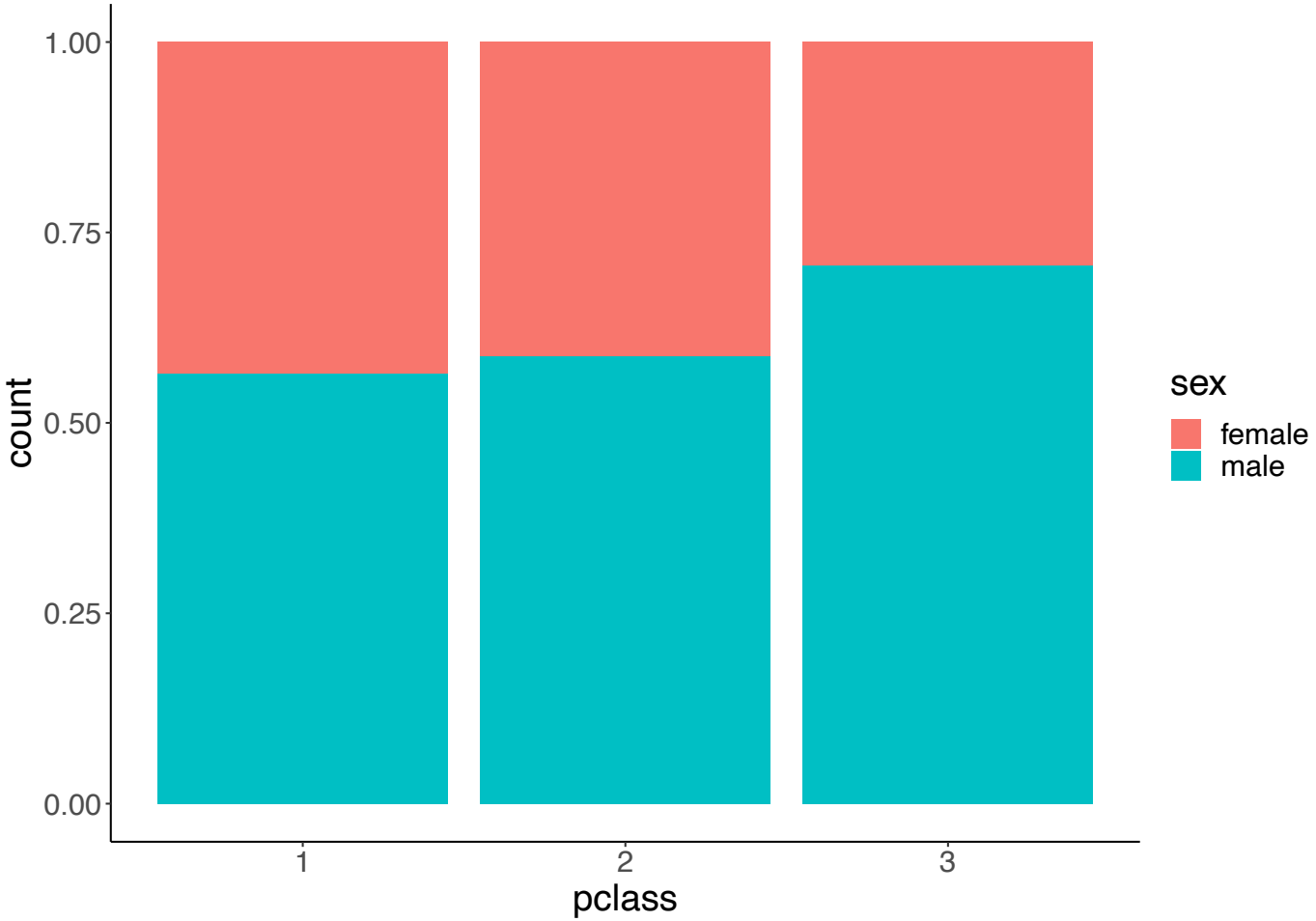


Figure 4: Komponenten-Balkendiagramm (relative Häufigkeit)

Komponenten-Balkendiagramm (relative Häufigkeit) (1)

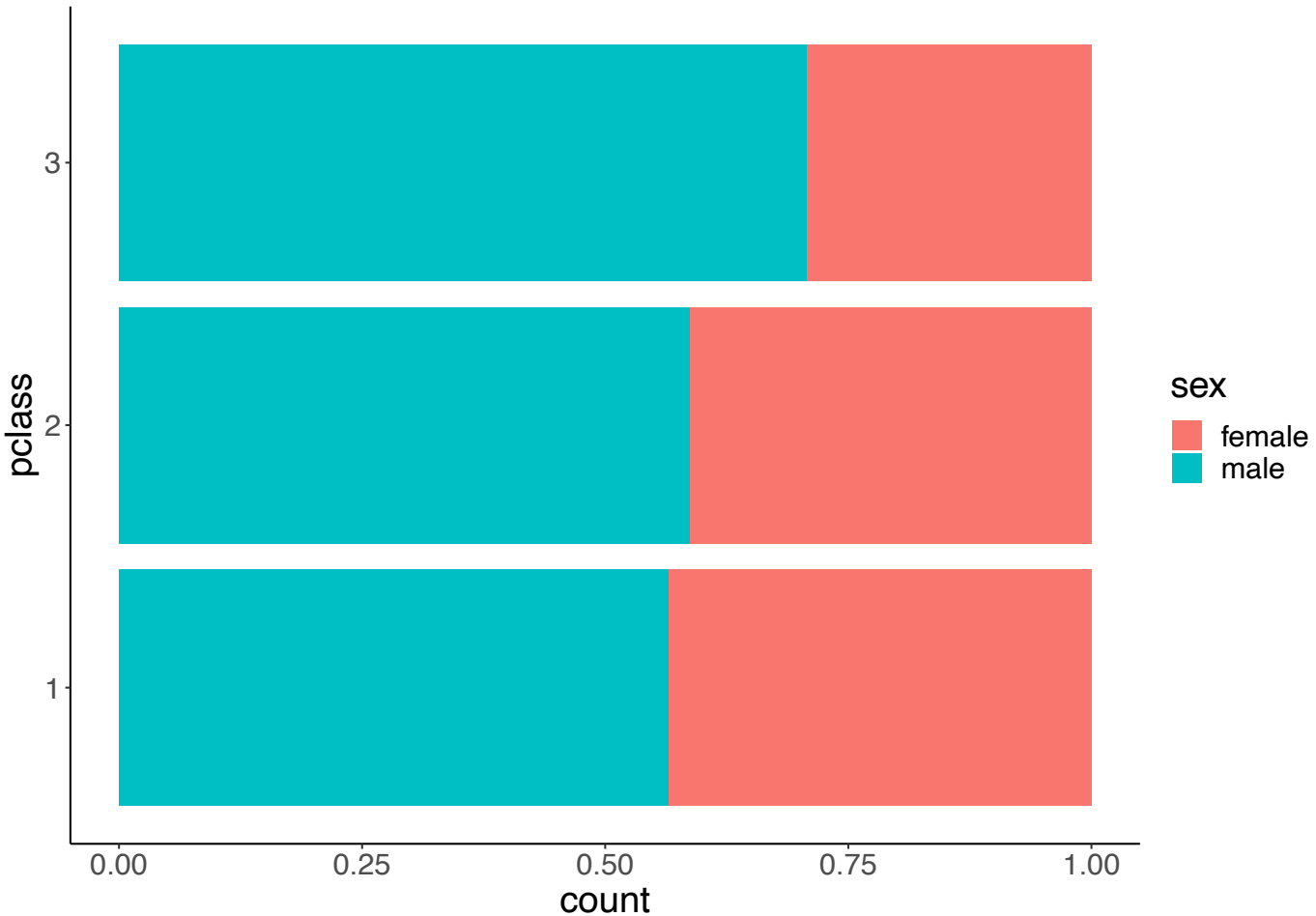


Figure 5: Komponenten-Balkendiagramm (relative Häufigkeit)

Daten auf der Zahlengerade

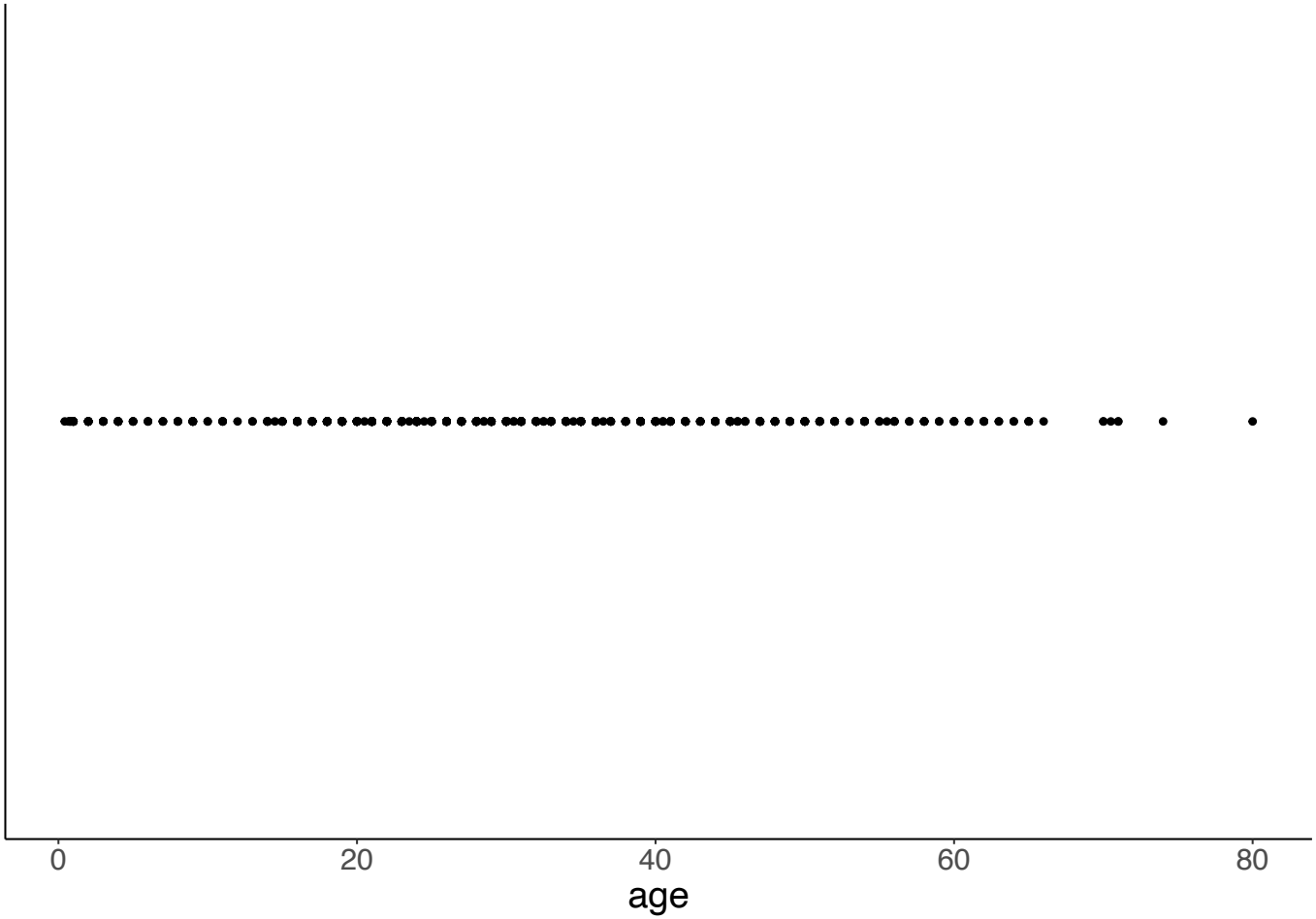


Figure 6: Daten auf der Zahlengerade

Histogramm

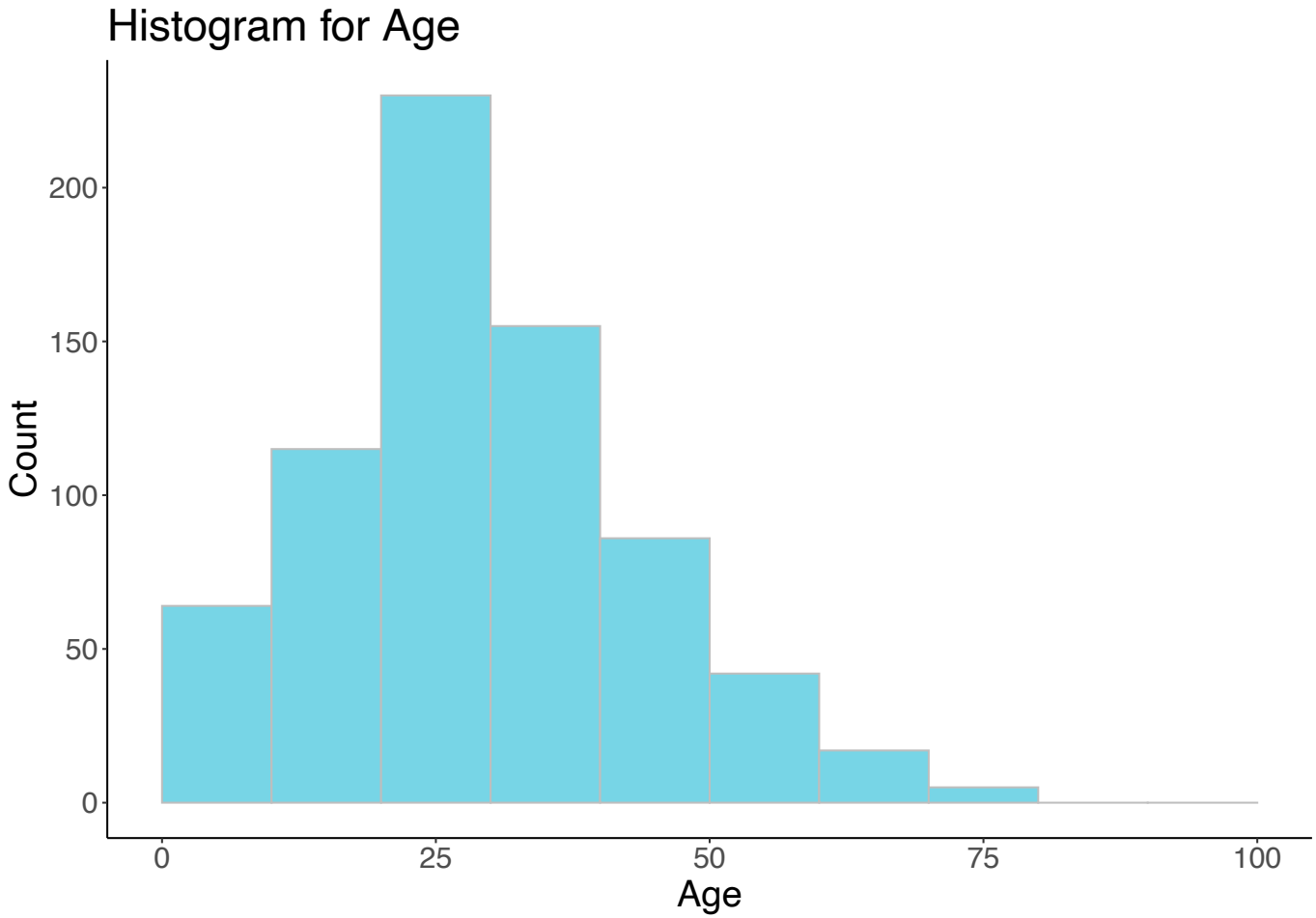


Figure 7: Histogramm

Histogramm (relative Häufigkeit)

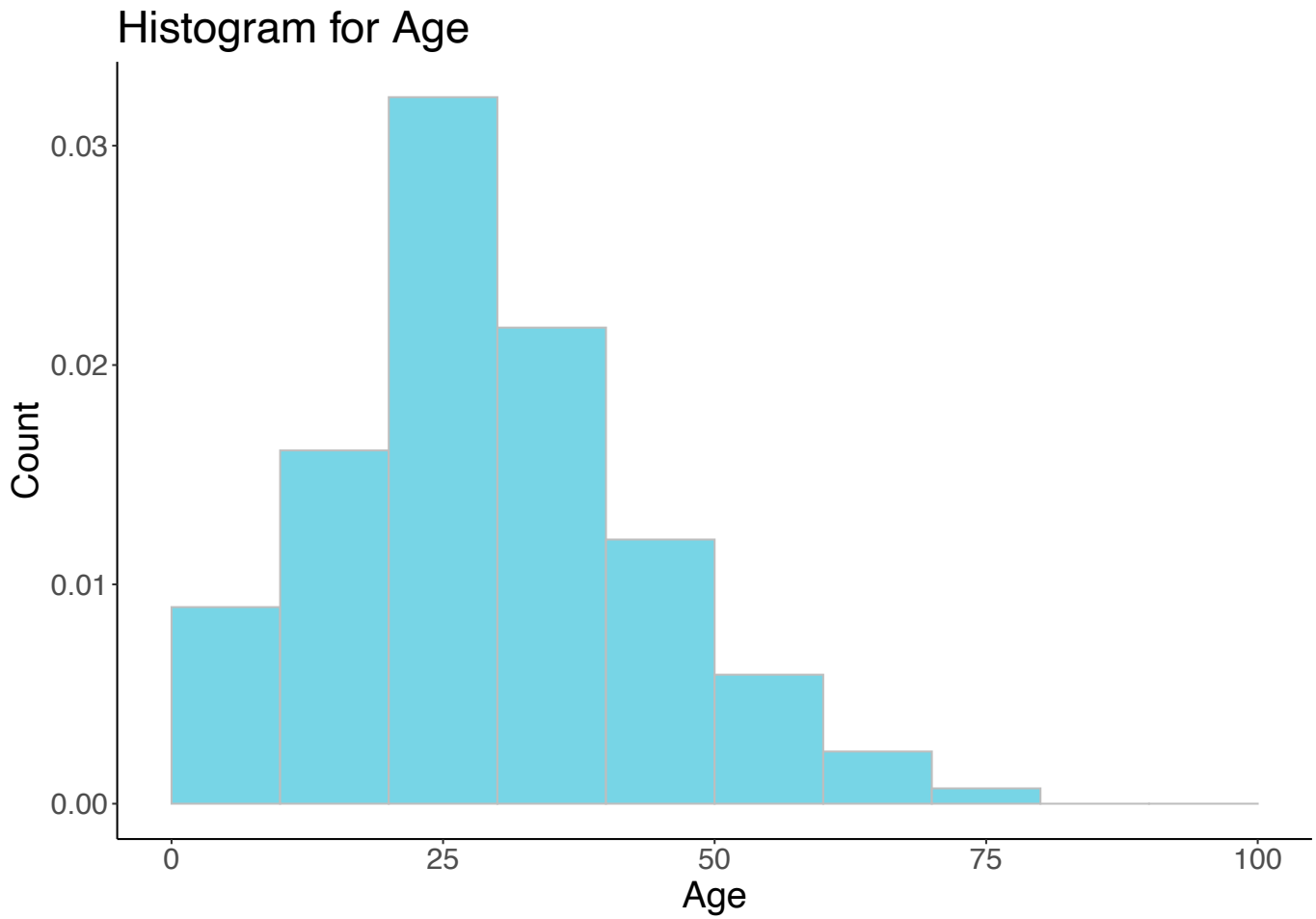


Figure 8: Histogramm (relative Häufigkeit)

Histogramm (relative Häufigkeit) (2)

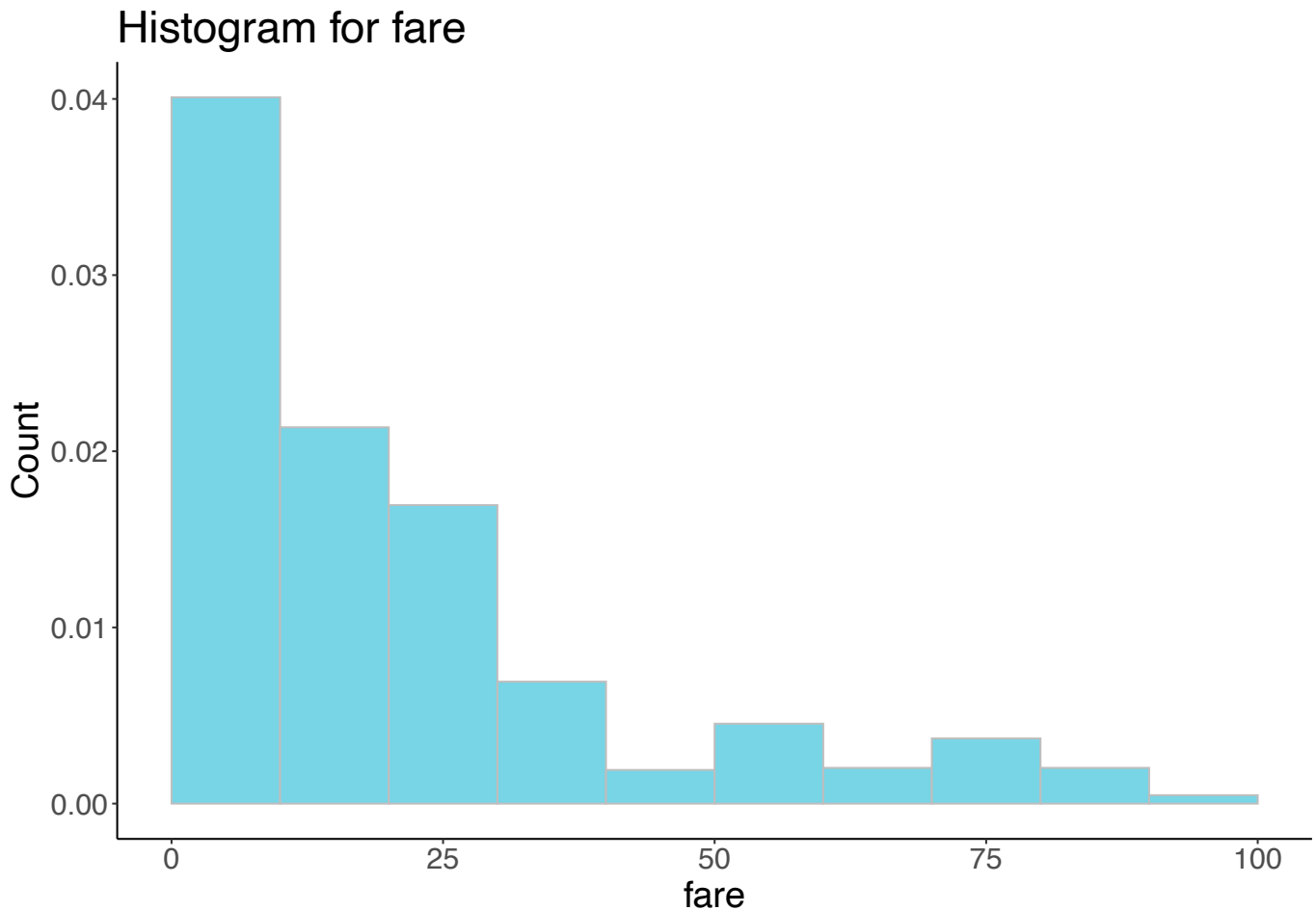


Figure 9: Histogram fare

Gruppenbildung

Klassenbreite b nach von Sturges

$$b = \frac{V}{1 + 3.32 \cdot \lg n} \approx \frac{V}{5 \lg n}$$

n - der Stichprobenumfang (Anzahl der Messwerte)

V - die Variationsbreite (Spannweite)

$\lg n$ - Zehnerlogarithmus von n

Histogramm (relative Häufigkeit) (3)

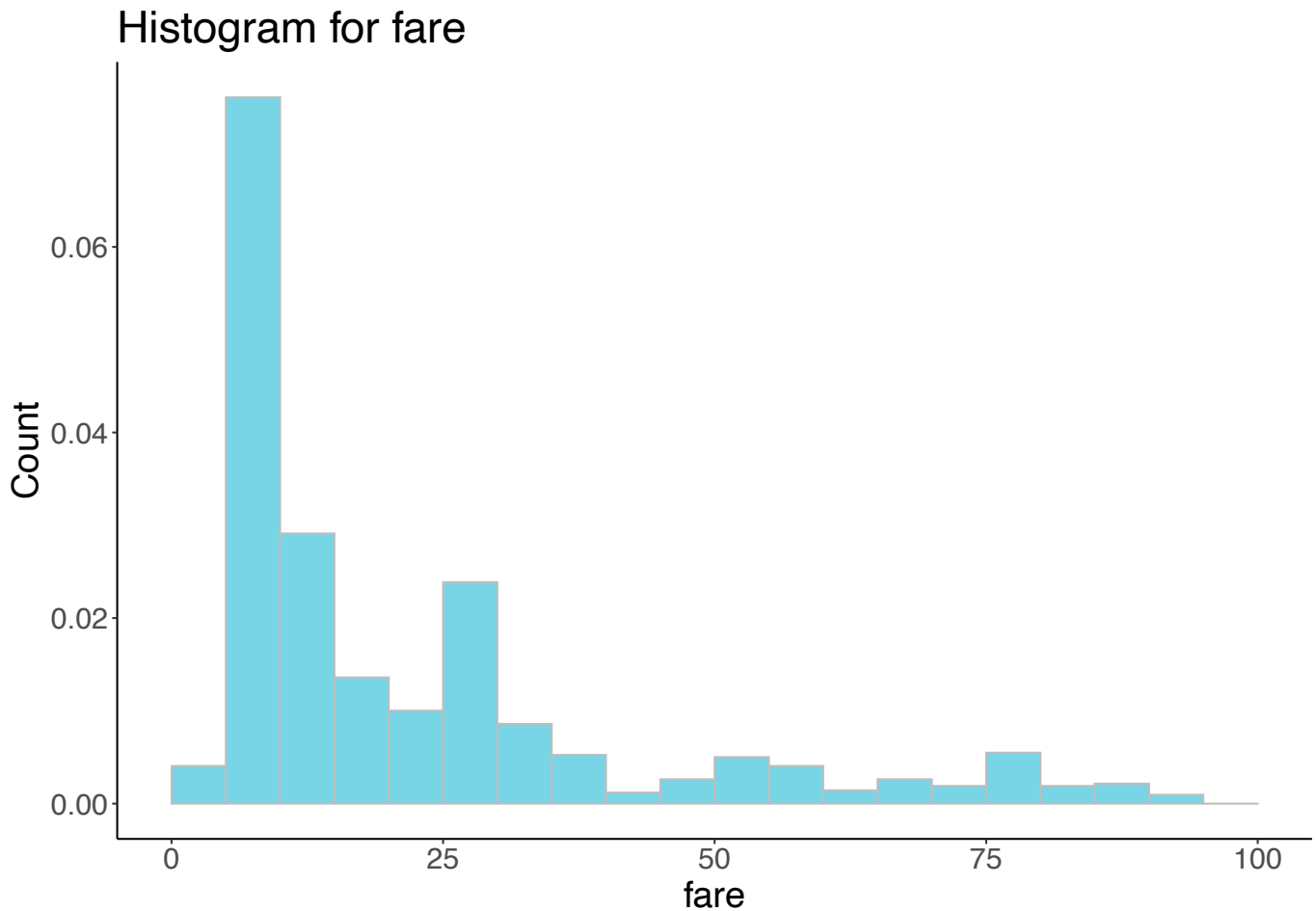


Figure 10: Histogram fare

Maßzahlen

Maßzahlen

Lageparameter

Streuungmaße

Lageparameter

Variationsbreite (Range)

(arithmetischer) Mittelwert

Median

Modus oder Modaler Wert

geometrischer Mittelwert

harmonischer Mittelwert

Streuungsmaße

Varianz und Standardabweichung

Quantile

Variationskoeffizient

Variationsbreite (Range)

$$V_x = \max(x) - \min(x)$$

(arithmetischer) Mittelwert

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

Nutzliche Eigenschaft

$$\sum_{i=1}^{N} cx_i = c \left(\sum_{i=1}^{N} x_i \right)$$

Der Mittelwert ist sehr empfindlich was extreme Werte betrifft

Median

Zentralwert

Ungerade Anzahl der Beobachtungen

$\left(\frac{n+1}{2}\right)$. Beobachtungswert

Gerade Anzahl der Beobachtungen

Mittelwert von $\left(\frac{n}{2}\right)$. und $\left(\frac{n}{2}+1\right)$. Beobachtungswerten

Der Median ist hauptsächlich bestimmt durch die Werte in der Mitte der Stichprobe und ist wenige abhängig von den extremen Werten

Modus oder Modaler Wert

Dichtemittel

Der häufigste Wert. Wenn alle Werte nur einmal vorkommen, gibt es keinen Modus.

Geometrischer Mittelwert

$$G = \sqrt[n]{x_1 \cdot \ldots \cdot x_n} = e^{\frac{1}{n} \sum_{i=1}^{n} \ln x_i}$$

wird z.B. für die Bestimmung der MIC benutz ($2^k c$, $k = 1, 2, \ldots$)

Harmonischer Mittelwert

$$H = \frac{1}{n}\left(\frac{1}{x_1} + \ldots + \frac{1}{x_n}\right)^{-1}$$

Vergleich vom Maßzahlen

Histogramm (relative Häufigkeit) (3)

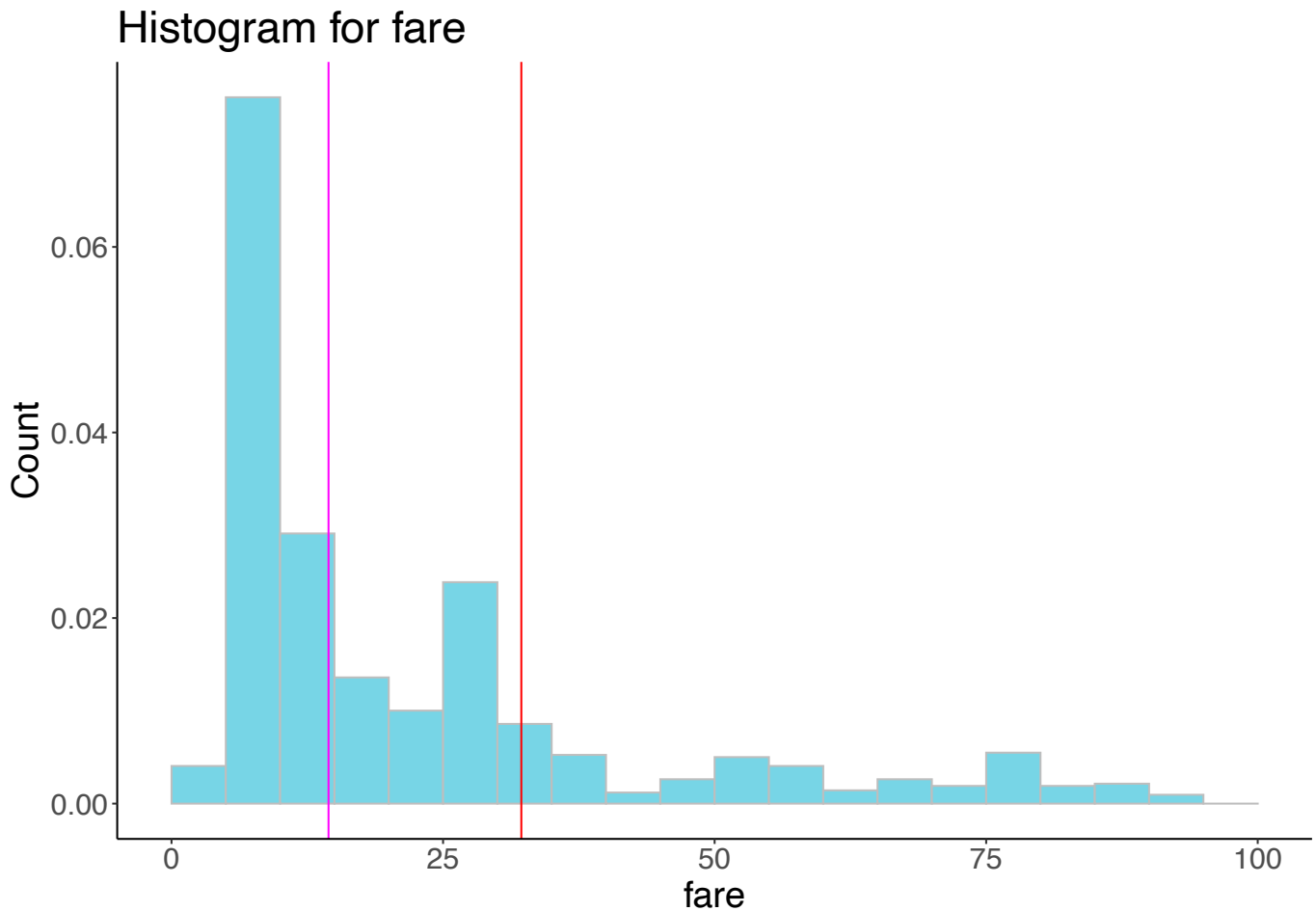


Figure 11: Histogram fare

Streuungsmaße

Varianz und Standardabweichung

Varianz

$$s_x^2 = \frac{1}{n-1} \sum_1^n (x_i - \bar{x})^2$$

$n - 1$ - Freiheitsgrad

Standardabweichung

$$s_x = \sqrt{\frac{1}{n-1} \sum_1^n (x_i - \bar{x})^2}$$

Der mittlere Fehler des Mittelwertes

Streuung von $\bar{x}$ um den wahren Mittelwert μ der Grundgesamtheit

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Variationskoeffizient

Das Verhältnis der Standardabweichung zum Mittelwert

$$\mathrm{CV} = \frac{s}{|\bar{x}|}$$

Erlaubt unambhängig vom Mittelwert die Streuung der Daten zu Vergleichen

Quantile

p. Perzentil

- $(k+1)$. Datenpunkt wenn $\frac{np}{100}$ nicht ganzzahlig ist $(k < \frac{np}{100}, k \in \mathbb{N})$
- Durchschnitt von k. und $(k+1)$. Datenpunkt wenn $\frac{np}{100}$ ganzzahlig ist

1. Quantil (Q_1) - Der Datenpunkt wo 25% Messpunkte unterhalb und 75% oberhalb liegen
2. Quantil (Q_3) - Der Datenpunkt wo 75% Messpunkte unterhalb und 25% oberhalb liegen
3. Quantil (Q_2) - Median

Interquartilsabstand und MAD

(Inter)quartilsabstand

$$IQR = Q_{0.75} - Q_{0.25}$$

Median-Abweichung (Mean Abslolute Deviation)

$$|x_i - Q_{0.5}|$$

Kumulative Verteilung

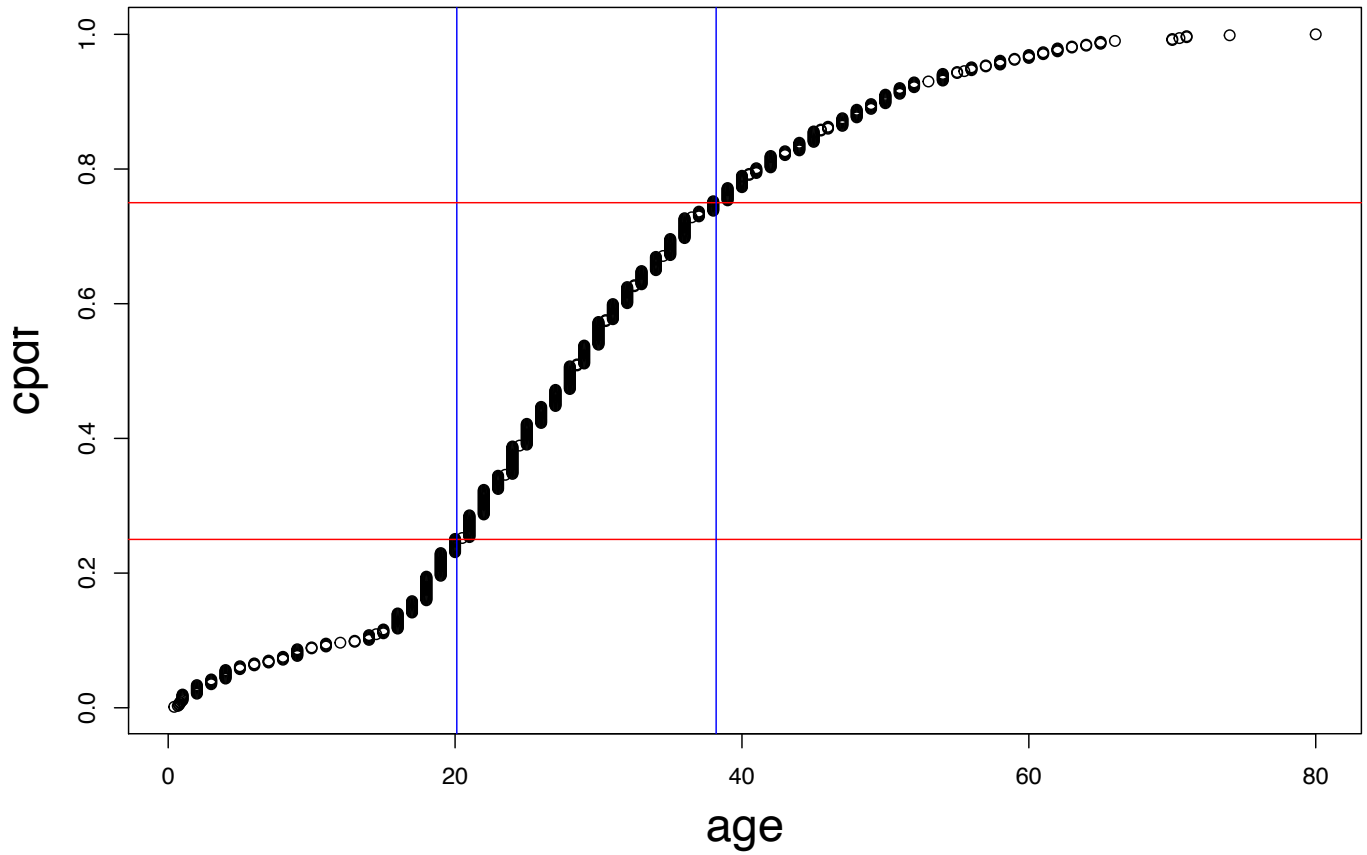
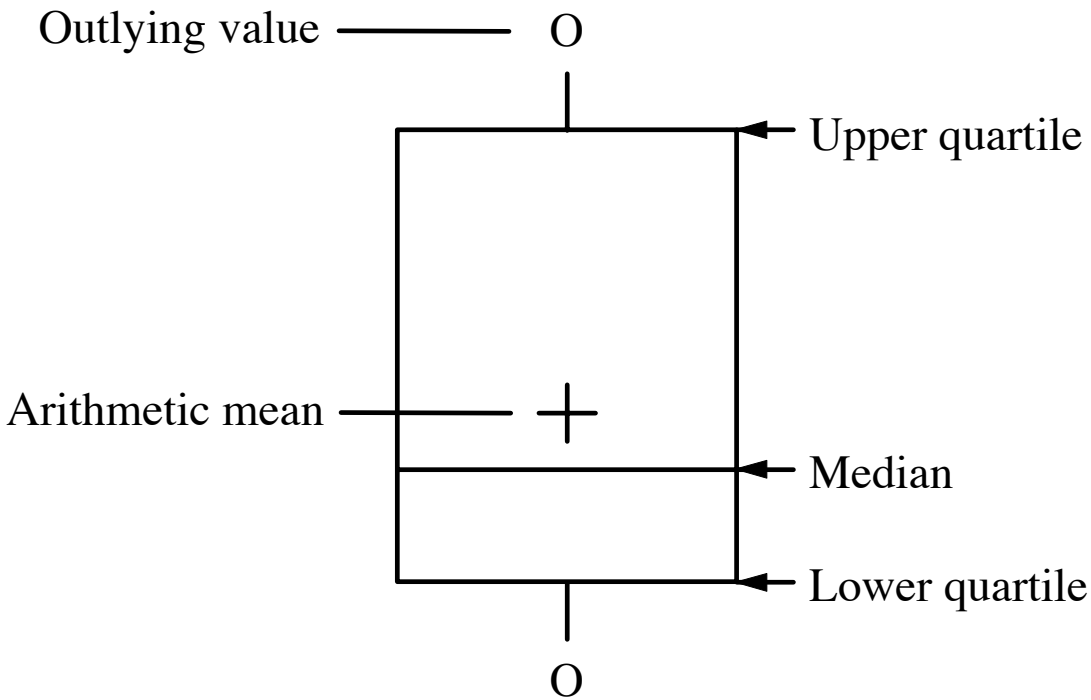


Figure 12: Kumulative Verteilung

Boxplot


Outlying value
O
Upper quartile
Arithmetic mean
Median
Lower quartile
O

Boxplot

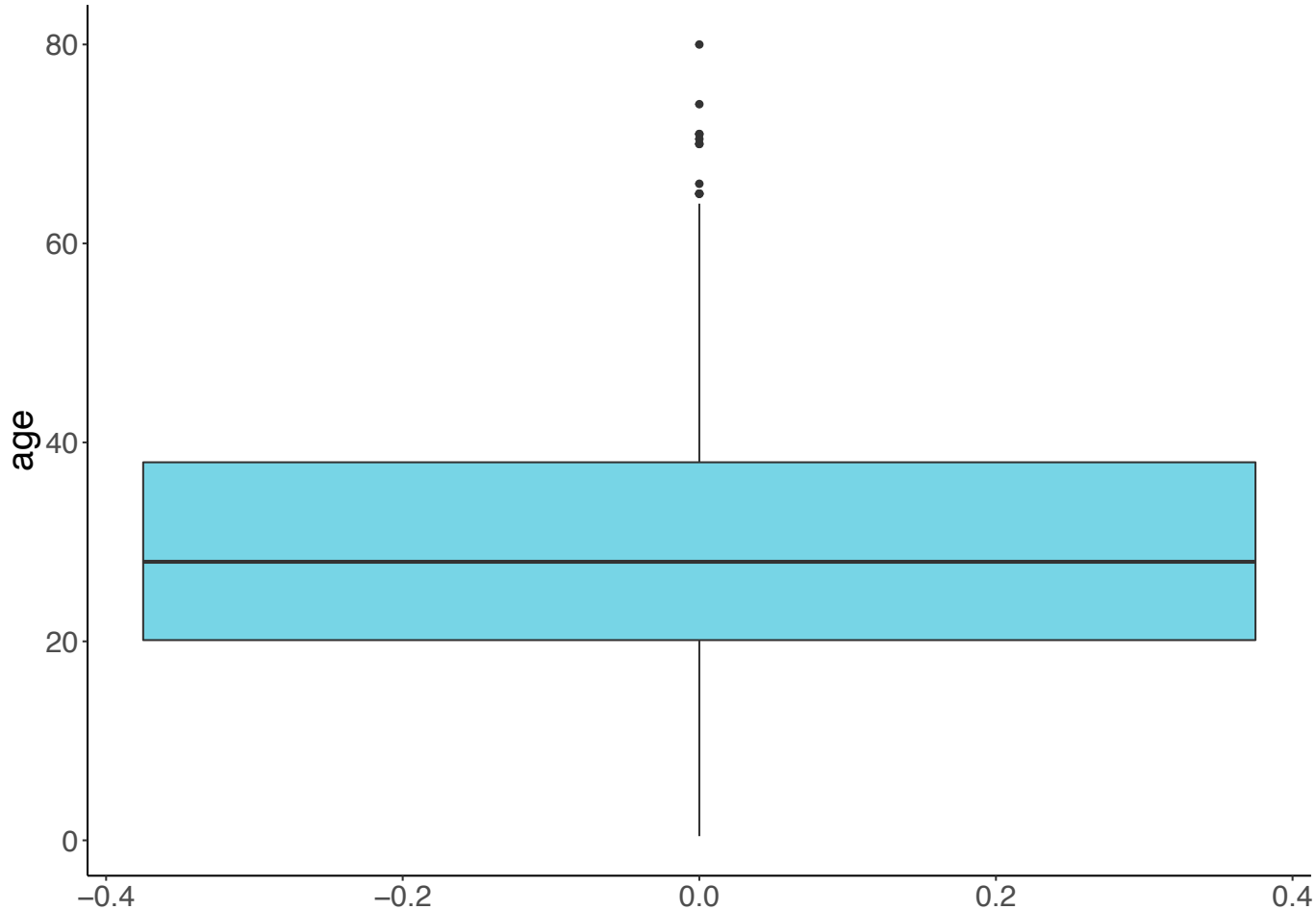


Figure 13: Boxplot

Boxplot

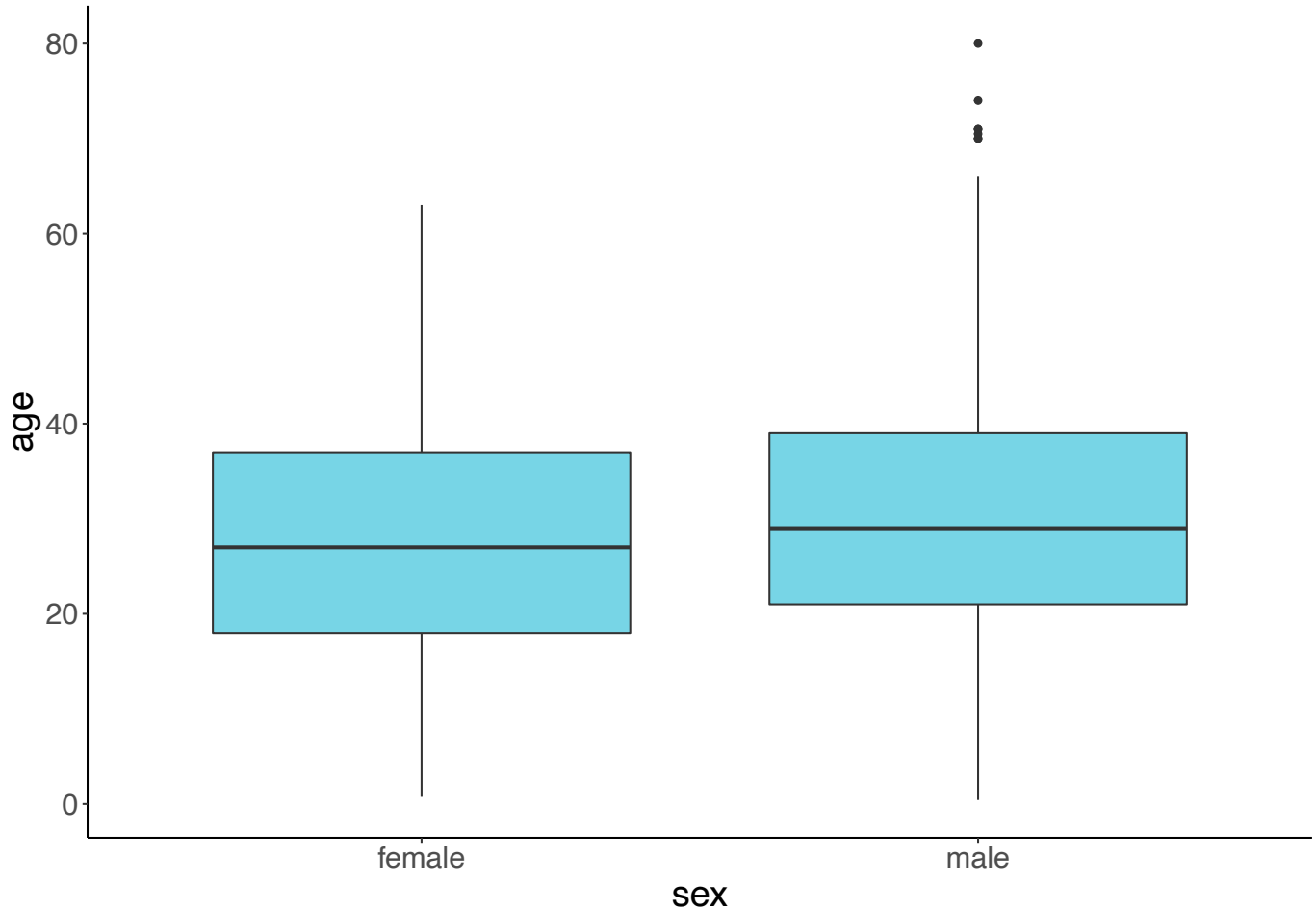
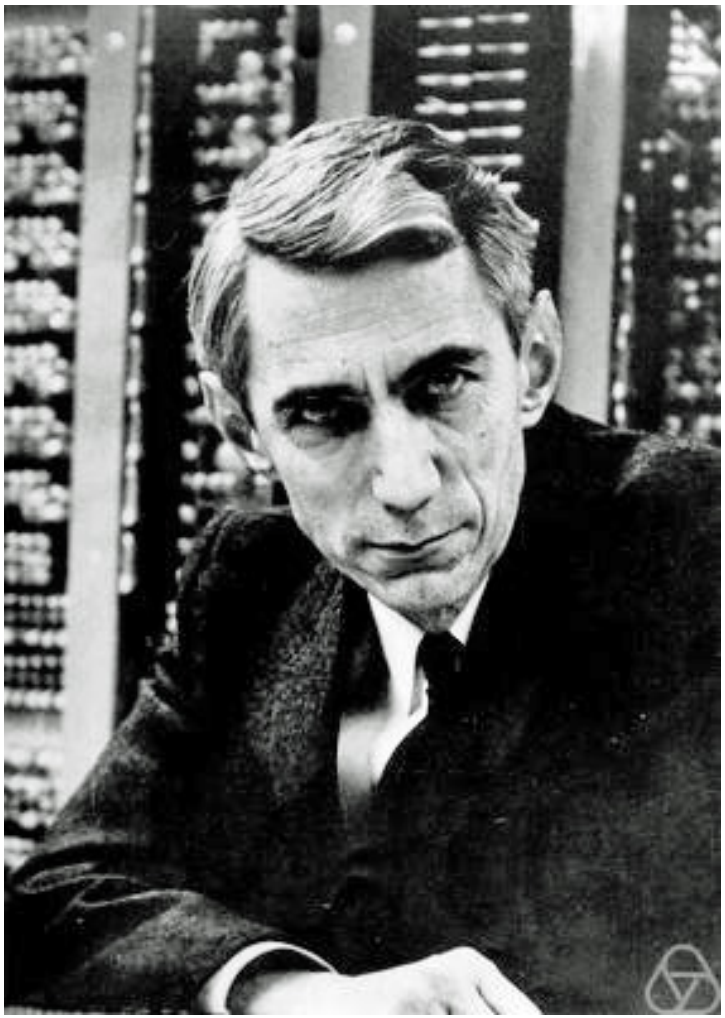


Figure 14: Boxplot

Diversitätsindizes

$$S = \sum_i p_i \ln p_i$$



Inferential Statistics

VB

Institut für Veterinär-Epidemiologie und Biometrie, FU Berlin

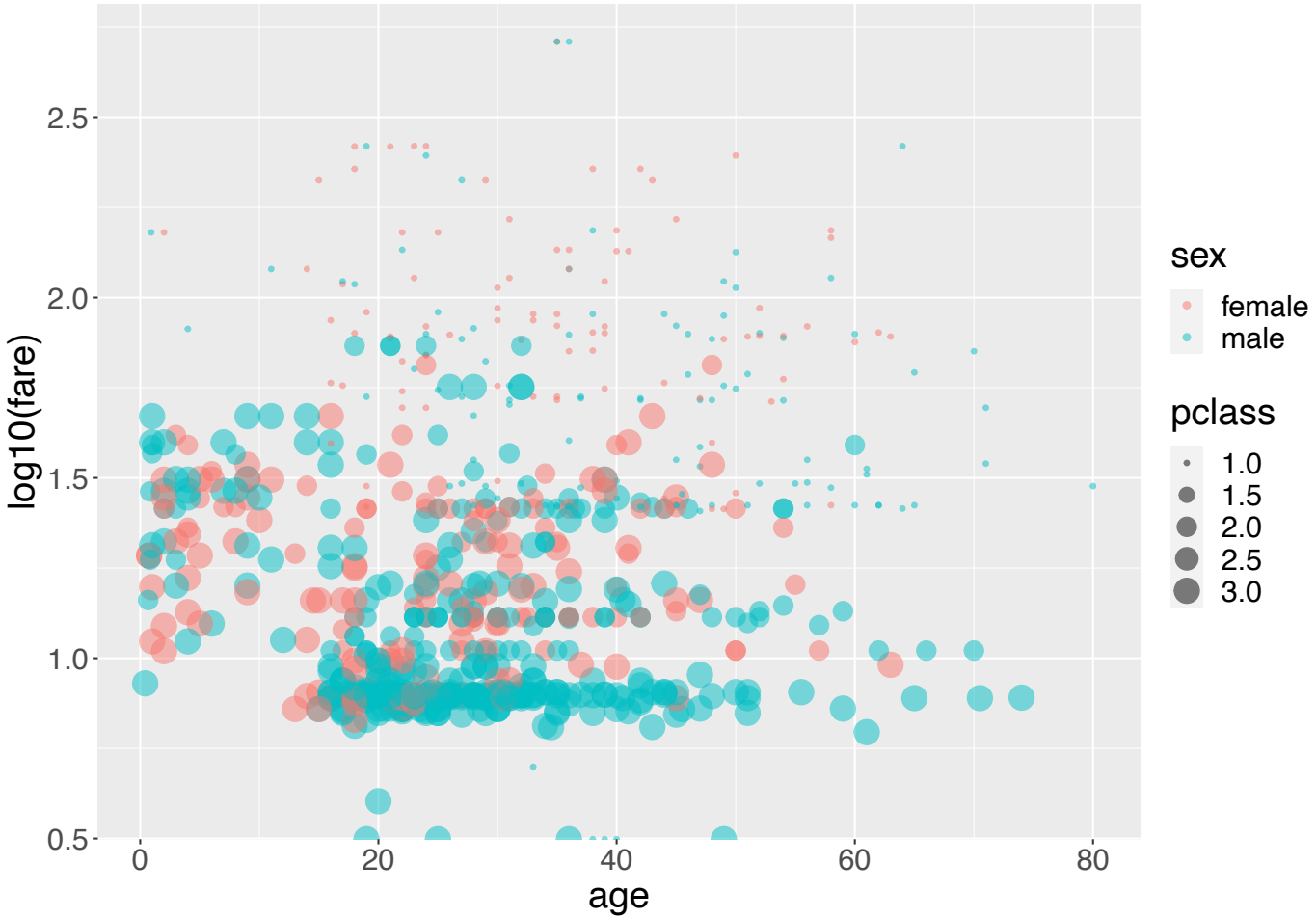
10/20/2020

Wiederholung

In der letzen Vorlesung haben wir über die über Beschreibende Statistik gersprochen.

- Charakteristische Maßzahlen
- Darstellungsforment: Tabellen, Histogram, Boxplot

Bivariate Daten



Was haben wir letztes mal gelernt?

Schliessende Statistik

Schliessende Statistik

- Test-Theorie
- Intervallschätzung

Schätzwerte und Tests

Basierend auf Schtichprobe, *schätz* man die Parameter der Grundgesamtheit.

Beispiel

Langjährige Beobachtungen (Grundgesamtheit): $\rho = 48$ der Neugeborenen sind Mädchen.

Erhebung aus drei Krankenhäusern (Stichprobe): 680 Geburten mit $\hat{\rho} = 51$ Mädchen.

Ist die Erhöhung *zufällig* oder nicht und aufgrund einer Ursache *signifikan*t?

Test-Theorie

Die Test-Theorie stellt eine Verbindung zwischen Stichtproben und Grundgesamtheit.

Es wird geprüft, aufgrund von Stichprobenwerten, ob gewisse *Hypothesen* über die Grundgesamtheit wahr sind oder nicht.

Es soll entschieden werden ob eine Hypothese beizubehalten oder zu verwerfen ist.

Hypothesen H_0 und H_1

Nullhypothese ist normalerweise die Behauptung, dass eine Behandlung oder Maßnahme in einem Versuch *keine Auswirkungen* hat und jegliche Unterschiede zwischen den Messwerten nur durch *Zufall* entstanden sind.

Beispiel

Man hat die Hypothese, dass die normalverteilte Grundgesamtheit einen wahren Mittelwert von $\mu = 18$ hat. Der aus einer Stichprobe ermittelte Mittelwert beträgt $\hat{\mu} = 19.5$.

- *Nullhypothese* H_0: der wahre Mittelwert μ ist gleich dem theoretischen Wert
- *alternative Hypothese* H_1: der wahre Mittelwert μ ist *nicht* gleich dem theoretischen Wert

Fehler 1. Art und 2. Art

Die Tests können nur *statistische* Aussagen über die "Wahrheit" der Hypothesen machen. Dabei können Fehler von zwei Arten vorkommen.

α-Fehler (1. Art)

Durch den Test wird die *Nullhypothese verworfen, obwohl sie in Wirklichkeit richtig* ist (*false negatives*).

β-Fehler (2. Art)

Die *Nullhypothese wird beibehalten, obwohl sie in Wirklichkeit falsch* ist (*false positives*).

α-Fehler

Die Wahrscheinlichkeit des α-Fehlers können wir für den Test selbst festlegen. Bekannt auch als *Signifikanzniveau* α, α-*Risiko*, *Irrtumswahrscheinlichkeit* α.

Beispiel

Eine Äpfellieferung wird auf Qualität geprüft und darf nicht mehr als 15% von Obst schlechter Qualität enthalten. Von jeder Palette (*Grundgesamtheit*) werden 10 Äpfel (*Stichprobe*) untersucht.

- H_0: die Untersuchte Palette ist "gut", i.e. enthält höchstens 15% schlechte Äpfel
- H_1: die Untersuchte Palette ist nicht "gut", i.e. enthält mehr als 15% schlechte Äpfel

α-Fehler(1)

Anzahl schlechtere Äpfel	i	0	1	2	3	4	5	≥ 6
WS für genau i schlechtere Äpfel	$P(i)$	0.197	0.347	0.276	0.130	0.040	0.009	0.001
WS für höchstens i schlechtere Äpfel	$\sum P(i)$	0.197	0.544	0.820	0.950	0.990	0.999	1.000

Figure 1: Wahrscheinlichkeiten berechnet nach Binomialverteilung mit $k = 10$ Ziehungen und $p = 0.15$. [Köhler et al.]

Bei Richtigkeit unserer Nullhypothese mit einer Wahrscheinlichkeit von 0.95 finden wir höchstens 3 Äpfel schlechter Qualität pro Stichprobe. Nur in 5% der Fällen würden wir in der Stichprobe mehr als 3 schlechte Äpfel vorfinden.

Wenn wir bereit sind in 5% der Fälle die Nullhypothese abzulehnen, würden wir ein α-Risiko von $\alpha = 5\%$ akzeptieren.

α-Fehler(2)

In der Stichprobe von Umfang $k = 10$ werden i schlechtere Äpfel vorgefunden.

- Ist $i \leq K$, wird H_0 angenommen.
- Ist $i > K$, wird H_0 abgelehnt.

Dabei wird $K = 3$ *kritischer Wert* genannt.

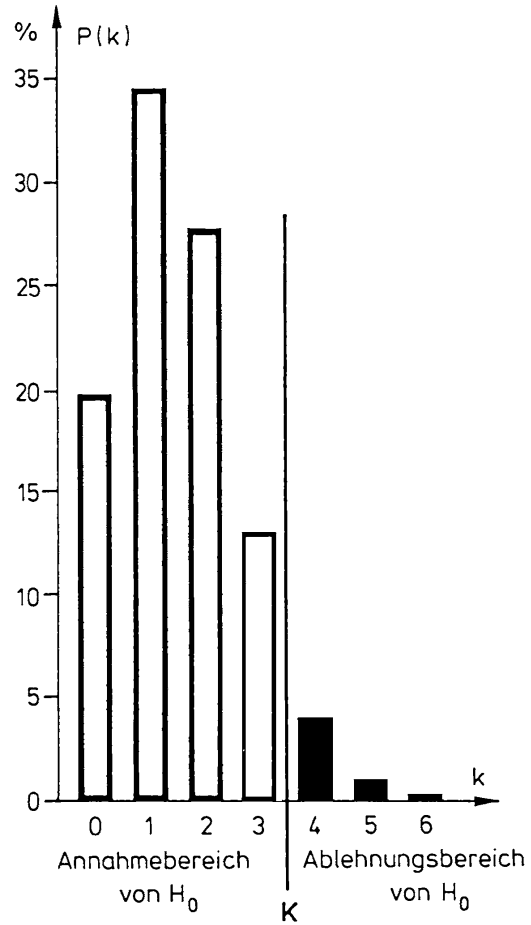


Figure 2: Wahrscheinlichkeiten berechnet nach Binomialverteilung mit $k = 10$ Ziehungen und $p = 0.15$. [Köhler et al.]

β-Fehler

Die *Nullhypothese wird beibehalten, obwohl sie in Wirklichkeit falsch* ist (*false positives*).

Beispiel

Falls die Äpfelpaletten 50% Äpfel minderer Qualität anstatt 15% hätten und unser Testverfahren mit $K = 3$ würde "schlechte" Paletten als "gute" akzeptieren, wäre das ein Fehler 2. Art.

i	0	1	2	3	4	5	6	7	≥ 8
$P(i)$	0.001	0.010	0.044	0.117	0.205	0.246	0.205	0.117	0.055
$\sum P(i)$	0.001	0.011	0.055	0.172	0.377	0.623	0.828	0.945	1.000

Figure 3: Wahrscheinlichkeiten berechnet nach Binomialverteilung mit $k = 10$ Ziehungen und $p = 0.5$. [Köhler et al.]

Mit Wahrscheinlichkeit von 0.172 würden wir die schlechten Paletten als gut akzeptieren, weil $\sum_j^3 P(j) = 0.172$ für $K = 3$.

β-Fehler(1)

Wir konnten die Größe des β-Fehlers nur berechnen, indem wir unterstellten, der wahre Anteil schlechterer Äpfel beträge $p = 50\%$.
Meist kennt man den wahren Wert von p nicht. Daher ist β unbekannt und im Falle der Beibehaltung von H_0 weiss man nicht wie groß die Wahrscheinlichkeit ist, dass die beibehaltene H_0 falsch ist.

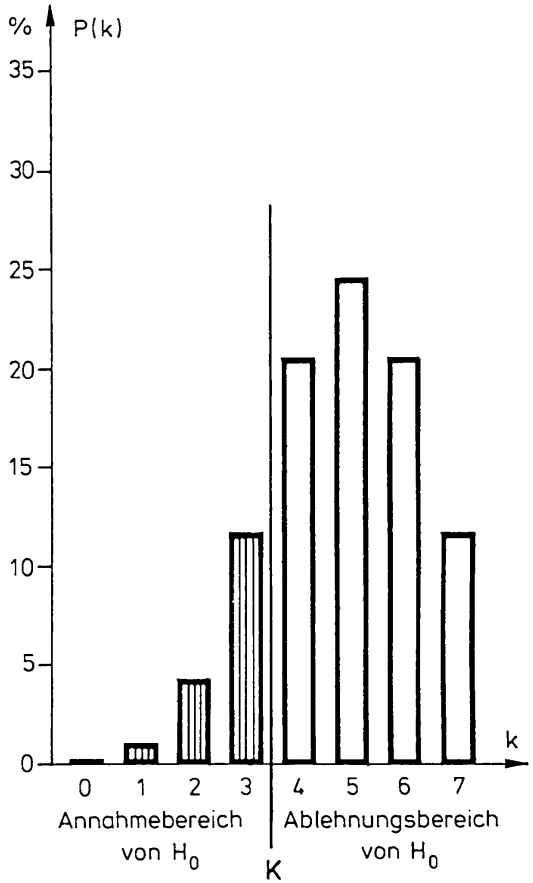


Figure 4: Wahrscheinlichkeiten berechnet nach Binomialverteilung mit $k = 10$ Ziehungen und $p = 0.5$. [Köhler et al.]

Größere Stichproben verkleinern β

Beispiel

Blutdruckmedikament soll geprüft werden. Stichprobe besteht aus 160 Personen vor (u_i) und nach (b_i) der Behandlung. Insgesamt 320 Blutdruck-Werte. Man bildet die Differenzen $d_i = b_i - u_i$ und deren Mittelwert $\bar{d}$.

Das Medikament hat eine Wirkung wenn $\bar{d}$ signifikant von null abweicht und nicht zufällig. Mit $\bar{d}$ soll geklärt werden ob der wahre Wert δ gleich oder ungleich null ist.

- $H_0(\delta = 0)$
- $H_1(\delta \neq 0)$

Größere Stichproben verkleinern β (1)

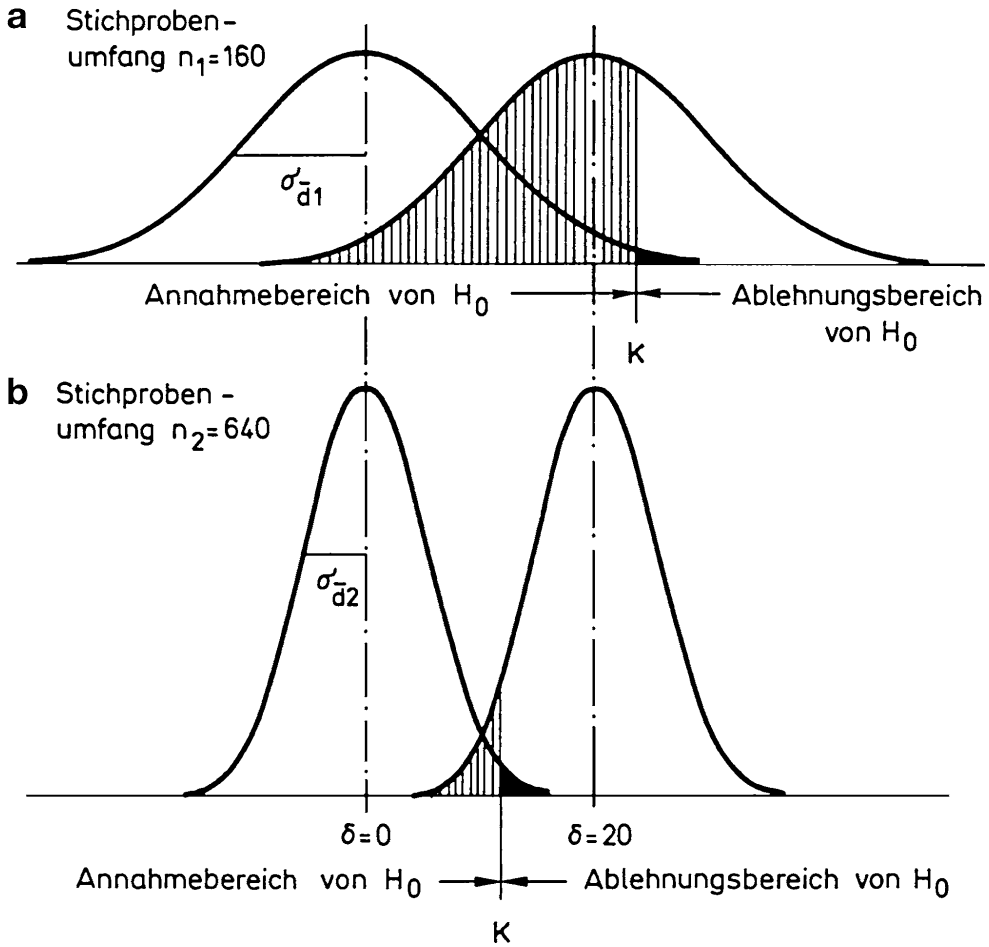


Figure 5: Mögliche Entscheidungen beim Testen.[Köhler et al.]

Um β-Fehler darstellen zu können mussten wir die Hypothese konkretesieren ($\delta = 20$).

		Wahrer Sachverhalt	
		$\delta = 0$	$\delta \neq 0$
Entscheidung des Tests	Annehmen von $H_0(\delta = 0)$	*richtige Entscheidung:* wahrer Sachverhalt stimmt mit Testergebnis überein. Wahrscheinlichkeit: $(1 - \alpha)$	*falsche Entscheidung:* $H_1(\delta \neq 0)$ wäre richtig, Testergebnis führt aber zu $H_0(\delta = 0)$. (Fehler 2. Art) Wahrscheinlichkeit: β
	Annehmen von $H_1(\delta \neq 0)$	*falsche Entscheidung:* $H_0(\delta = 0)$ wäre richtig, Testergebnis führt aber zu $H_1(\delta \neq 0)$. (Fehler 1. Art) Wahrscheinlichkeit: α	*richtige Entscheidung:* wahrer Sachverhalt stimmt mit Testergebnis überein. Wahrscheinlichkeit: $(1 - \beta)$

Figure 6: Mögliche Entscheidungen beim Testen.[Köhler et al.]

Schätzer

Punktschätzer

Der Parameter (z.B. μ oder σ) wird aus einer Stichprobe berechnet und als Schätzwert für die Grundgesamtheit angegeben.

Intervalschätzer

Man gibt ein ganzes Intervall als mögliche Wert des Schätzers. Man nennt solche Intervalle *Vertrauensbereiche* oder *Konfidenzintervalle*.

Konfidenzintervalle

- Das Konfidenzintervall enhält mit der *vorgegebenen* Wahrscheinlichkeit $(1 - \alpha)$ den wahren Parameterwert.
- Je größer $(1 - \alpha)$ desto größer das Konfidenzintervall.

Konfidenzintervalle(1)

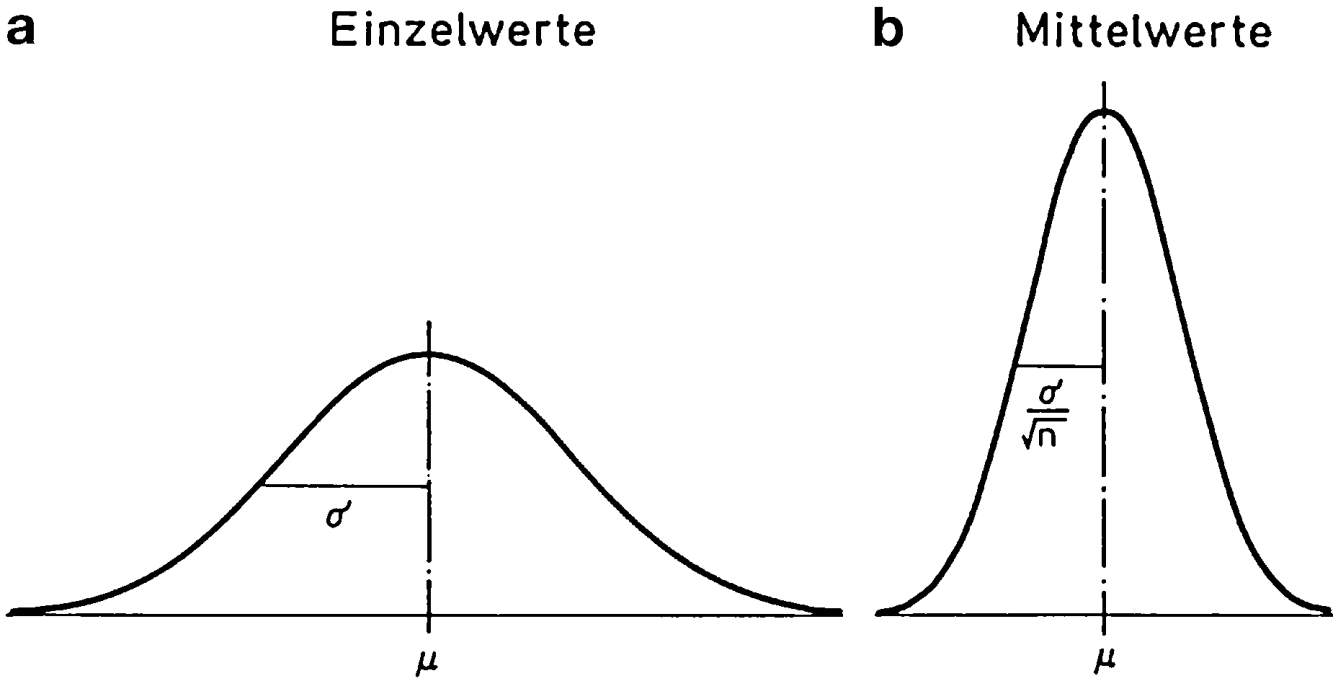


Figure 7: (a) Verteilung der Einzelwerte x der Grundgesamtheit. (b) Verteilung der Mittelwerte $\bar{x}$ von Stichproben des Umfangs n. Die Standardabweichung ist σ. [Köhler et al.]

Konfidenzintervalle für μ bei bekannter Varianz σ

Wenn die Varianz σ bekannt ist, kann man wie folgt das Intervall bestimmen, wo ungefähr 95% der $\bar{x}$-Werte liegen ($\alpha = 0.05$):

$$\mu - 2 \cdot \frac{\sigma}{\sqrt{n}} \lesssim \bar{x} \lesssim \mu + 2 \cdot \frac{\sigma}{\sqrt{n}}$$

$$\bar{x} - 2 \cdot \frac{\sigma}{\sqrt{n}} \lesssim \mu \lesssim \bar{x} + 2 \cdot \frac{\sigma}{\sqrt{n}}$$

Freihetsgrade

Anzahl der Messwerte minus Anzahl der aus den Messwerten ermittelten Parameter

Konfidenzintervalle für μ bei Normalverteilung

Fragestellung: finde das Intervall, dass den wahren Mittelwert μ mit der Sicherheitswahrscheinlichkeit $(1-\alpha)$ enthält.

Voraussetzung: Die Grundgesamtheit is normalverteilt mit *unbekannten* μ und σ.

$$s_x = \sqrt{\frac{1}{1-n}\left(\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}\right)}$$

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}}$$

$(1-\alpha)$-Konfidenzintervall

$$[\bar{x} - t_{\text{Tab}}(n-1,\alpha)\cdot s_{\bar{x}}; \bar{x} + t_{\text{Tab}}(n-1,\alpha)\cdot s_{\bar{x}}]$$

$t(m,\alpha)$ ist Studentsche t-Verteilung mit m Freiheitsgraden

t-Verteilung

Mehrkmal ist t-verteilt, wenn die Varianz des Merkmals unbekannt ist und mit der Stichprobenvarianz geschätzt werden muss.

$$f_m(x) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{m\pi}\Gamma\left(\frac{m}{2}\right)}\left(1+\frac{x^2}{m}\right)^{-\frac{m+1}{2}}$$

für $-\infty \leq x \leq \infty$

Beispiel

```
data <- c(341, 345, 338, 339, 340, 343, 341, 343, 341, 328, 343, 347, 337, 348, 339)
t.test(data, conf.level = 0.95)$conf.int
```

```
## [1] 338.2368 343.4965
## attr(,"conf.level")
## [1] 0.95
```

Konfidenzintervalle für die Differenz von μ_x und μ_y bei Normalverteilung

Fragestellung: finde das Intervall, dass den Betrag der Differenz der wahren Mittelwerte μ_x und μ_y der Grundgesamtheiten X uns Y mit der Sicherheitswahrscheinlichkeit $(1 - \alpha)$ enthält.

Voraussetzung: Beide Grundgesamtheiten sind normalverteilt mit gelichen *unbekannten* Varianzen. Die Stichproben sind unabhängig.

$$s_D = \sqrt{\frac{(n_X - 1) \cdot s_x^2 + (n_Y - 1) \cdot s_y^2}{n_X - n_Y - 2}}$$

$$s_{\bar{D}} = S_D \sqrt{\frac{n_X + n_Y}{n_X \cdot n_Y}}$$

Konfidenzintervalle für die Differenz von μ_x und μ_y bei Normalverteilung(1)

$(1-\alpha)$-Konfidenzintervall

$$[|\bar{x}-\bar{y}| - t_{\text{Tab}}(n_X+n_Y-2,\alpha)\cdot s_{\bar{D}}; |\bar{x}-\bar{y}| + t_{\text{Tab}}(n_X+n_Y-2,\alpha)\cdot s_{\bar{D}}]$$

Beispiel

```
data1 <- c(341, 345, 338, 339, 340, 343, 341, 343, 341, 328, 343, 347, 337, 348, 339)
data2 <- data1 + 10 + runif(length(data1))*20
t.test(data2, data1, conf.level = 0.95)$conf.int
```

```
## [1] 15.43699 25.19541
## attr(,"conf.level")
## [1] 0.95
```

Statistische Tests

Intervalldaten

Ordinaldaten

Nominale Daten

Statistische Tests

Vitaly Belik

Institut für Veterinär-Epidemiologie und Biometrie, FU Berlin

10/27/2019

Was haben wir letztes mal gelernt?

In der letzen Vorlesung haben wir über die über *Schliessende Statistik* gersprochen.

Test-Teorie

Punkt- und Intervallschätzer

t-Test (Vergleich zwei Mittelwerte)

Hypothesen H_0 und H_1

Nullhypothese ist normalerweise die Behauptung, dass eine Behandlung oder Maßnahme in einem Versuch *keine Auswirkungen* hat und jegliche Unterschiede zwischen den Messwerten nur durch *Zufall* entstanden sind.

- Die Test-Theorie stellt eine Verbindung zwischen Stichtproben und Grundgesamtheit.
- Es wird geprüft, aufgrund von Stichprobenwerten, ob gewisse *Hypothesen* über die Grundgesamtheit wahr sind oder nicht.
- Es soll entschieden werden ob eine Hypothese beizubehalten oder zu verwerfen ist.

Fehler I. und II. Art (α- und β-Fehler)

α-Fehler (1. Art)

Durch den Test wird die *Nullhypothese verworfen, obwohl sie in Wirklichkeit richtig* ist (*false negatives*).

β-Fehler (2. Art)

Die *Nullhypothese wird beibehalten, obwohl sie in Wirklichkeit falsch* ist (*false positives*).

Klassifikation der statistischen Tests

THE TABLE: Systematic of statistical tests and guide lines on how to choose a test.

Choose adequate combination based on row 1, columns A and B/C.

⇨ If possible, use non-parametric approaches

⇨ Examples of other statistical approaches: cluster, discriminant, principal component, path, time series analysis, ...

	A	B: explanatory variables		C: explanatory variables		
			Nonparametric tests		**Parametric tests**	
1			Outcome variable: ordinal scale		Outcome variable: interval or ratio scale possibly after transformation	Outcome variable: other distributions
			Residuals „symmetrical“		Residuals normally distributed	Residuals follow other distribution (e.g. Poisson, Binomial)
					linear models	**generalised linear models**
	Non-paired, independent fixed effects only	1 factor 2 levels	Mann-Whitney-U-test (Wilcoxon, rank sum test)	1 factor 2 levels	t-Test for independent data	🡻
		1 factor >2 levels	Kruskal-Wallis-test	≥1 factor >2 levels	ANOVA (analysis of variance, F-test)	🡻
		1 factor >2 ordered levels	Jonkheere-trend-test	≥1 factor >2 ordered levels	Analysis of variance with ordered factors or corresponding contrasts	🡻
		1 ordinally scaled	Spearman- , Kendall correlation	1 continuous	Pearson-correlation	–
				≥1 continuous	Regression	🡻
				≥1 any type in combination	Linear models[1]	Poisson-regression Logistic regression
					linear mixed-effects models	**Generalised linear mixed-effects models**
	dependent, repeated, nested additionally: random effects	1 factor 2 levels	Wilcoxon (signed rank test)	1 factor 2 levels	paired t-test	🡻
		1 factor >2 levels	Friedman-test	≥1 factor >2 levels	repeated measures, nested ANOVA	🡻
		1 factor >2 ordered levels	Page-Trend-test	≥1 factor >2 ordered levels	🡻	🡻
		1 ordinally scaled	No test if all data dependent	1 continuous	No test if all data dependent	🡻
				≥1 continuous	🡻	🡻
				≥1 any type in combination	Linear mixed-effects models[1]	Generalised linear mixed-Effects models
	Occurrences		χ^2-test, contingency table			Loglinear models

[1] all models listed above can be considered special cases, 🡻: choose model listed below

Figure 1: Test-Systematik [Quelle: L. Gygax]

Einseitige und zweiseitige Fragestellungen

Einseitige Fragestellung

Wenn schon *vor dem Versuch* feststeht, dass die Abweichung der Messgröße nur in *eine Richtung* möglich oder von Interesse ist. Dann prüft der Test nur, ob eine signifikante Abweichung in diese Richtung nachweisbar ist oder nicht.

- Beispiel. Überlebungsrate von Vieren nach der Bestralung mit Röntgenstrahlen verglichen mit der Kontrolle.

Zweiseitige Fragestellung

Kann in keine Richtung eine Veränderung ausgeschlossen werden, liegt eine *zweiseitige Fragestellung* vor.

- Beispiel. Ertrag nach der Bestralung der Pflanzensamen mit niederen Dosen Röntgenstrahlen verglichen mit der Kontrolle.

Einseitige und zweiseitige Fragestellungen (1)

Beim einseitigen Testen kann man bessere Signifikanzen nachweisen. Der einseitige Test darf aber nur dann angewandt werden, wenn aus theoretischen Erwägungen vor dem Versuch nur eine einseitige Veränderung von Interesse ist.

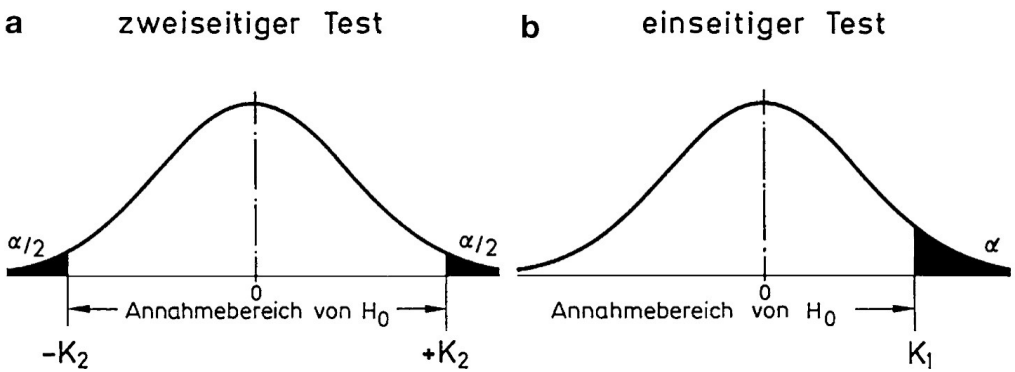


Figure 2: (a) Zweiseitiger Test. (b) einseitiger Test. [Quelle: Köhler et. al]

Teststatistiken

Zur Ermittlung von α und manchmal auch von β, hatten wir für unsere Testgröße basierend auf der Stichprobejeweils eine Wahrscheinlichkeitsverteilung verwendet. Man bezeichnet solche Testgrößen wie i (Äpfel) und $\bar{d}$ (Medikament) als *Prüfstatistiken* (Teststatistiken) und die zugehörigen Verteilungen als Prüfverteilungen.

Die Voraussetzung ist, dass die Stichprobe *zufällig* gezogen wird. Teststatistiken hängen von *Stichprobenumfang* ab. Daraus werden die *Freicheitsgrade* abgeleitet.

Die wichtigsten Verteilungen sind tabelliert oder werden mit Hilfe von Software berechnet.

Vorgehensweise bei statistischen Tests

1. Formulieren der zu überprüfenden Hypothese zu Grundgesamtheiten. Da wir beim Überprüfen der Hypothese nur auf die Stichproben zurückgreifen können, ist unsere Entscheidung fehlerhaft.

2. Wir legen die maximale Irrtumswahrscheinlichkeit in Form des *Signifikanzniveaus* α fest.

3. Wir wählen den *geeigneten Test*, der über unsere Hypothese entscheiden kann und zu den Gegebenheiten passt.

4. Wir berechnen die im Test *vorgeschriebenen* Teststatistiken und vergleichen sie mit geeigneten Tabellenwerten. Dabei entscheiden wir uns für die Beibehaltung oder Verwerfung der Nullhypothese.

5. Wenn wir den Test mit Hilfe der Software ausführen, bekommen wir den *P-Wert*. Falls

$P \geq \alpha \Rightarrow H_0$, die Nullhypothese wird beibehalten.

$P < \alpha \Rightarrow H_1$, die Alternativhypothese wird angenommen.

Unter der Bedinging, dass die Nullhypothese gilt, gibt der P-Wert due Überschreitungswahrscheinlichkeit P der aus den Daten berechneten Prüfstatistik.

Statistische Tests

Intervalldaten

Ordinaldaten

Nominale Daten

Test zu intervalskalierten (normalverteilten) Daten

Vergleich eines Mittelwertes mit dem theoretischen Wert (t-Test)

$$\hat{t} = \frac{|\bar{x} - \mu_T|}{s}\sqrt{n}$$

wird mit dem Tabellenwert t_{Tab} der t-Verteilung verglichen für gewünsches α und $FG = n - 1$.

$\hat{t} \leq t_{\text{Tab}} \Rightarrow H_0\,(\mu = \mu_T)$

$\hat{t} > t_{\text{Tab}} \Rightarrow H_1\,(\mu \neq \mu_T)$

n ist der Stichprobenumfang

s ist die Standardabweichung der Stichprobe

μ_T ist der theoretischer Mittelwert

$\bar{x}$ ist der arithmetischer Mittelwert der Stichporbe

Vergleich eines Mittelwertes mit dem theoretischen Wert (t-Test) (1)

Beispiel

Stichprobe von $n = 20$, $\bar{x} = 42.0$, und $s = 5.0$.

$\mu_T = 45.0$

$\hat{t} = 2.28$

$t_{\text{Tab}}(FG = 19; \alpha = 0.05) = 2.09$

$\hat{t} > t_{\text{Tab}} \rightarrow H_1(\mu \neq \mu_T)$

$\bar{x}$ weicht von μ_t signifikant ab.

Vergleich zweier Mittelwerte unabhängiger Stichproben (t-Test)

$$\hat{t} = \frac{|\bar{x} - \bar{y}|}{s_D} \sqrt{\frac{n_X \cdot n_Y}{n_X + n_Y}}$$

$$t_{\text{Tab}}(\alpha; FG = n_X + n_Y - 2)$$

Die gemeinsame Standardabweichung

$$s_D = \sqrt{\frac{(n_X - 1) \cdot s_x^2 + (n_Y - 1) \cdot s_y^2}{n_X + n_Y - 2}}$$

$\hat{t} \leq t_{\text{Tab}} \Rightarrow H_0 \, (\mu_X = \mu_Y)$

$\hat{t} > t_{\text{Tab}} \Rightarrow H_1 \, (\mu_X \neq \mu_Y)$

Vergleich zweier Mittelwerte unabhängiger Stichproben (t-Test) (1)

Beispiel

$n_X = 16,\ \bar{x} = 14.5,\ s_X^2 = 4$

$n_Y = 14,\ \bar{y} = 13.5,\ s_Y^2 = 3$

$s_D = 1.88$

$\hat{t} = 2.180$

$t_{\text{Tab}}(FG = 28; \alpha = 0.05) = 2.048$

$\hat{t} > t_{\text{Tab}} \Rightarrow H_1\,(\mu_X \neq \mu_Y)$

Die Mittelwerte der Stichproben sind signifikant verschieden

Vergleich zweier Mittelwerte verbundener Stichproben (t-Test)

Wenn die Annahme der Unabhängigkeit der Stichproben nicht erfüllt ist, spricht man über *verbundene Stichproben*. Z.B. wenn man dieselbe Gruppe von Individuen oder Objekten vor und nach einer Behandlung untersucht. Im Medikamenten-Beispiel haben wir uns gefragt, ob $\hat{d}$ signifikant *von null verschieden* ist.

Vergleich zweier Mittelwerte verbundener Stichproben (t-Test) (1)

$$\hat{t} = \frac{|\bar{d}|}{s_d}\sqrt{n}$$

$t_{\text{Tab}}(FG; \alpha)$

$\hat{t} \leq t_{\text{Tab}} \Rightarrow H_0(\delta = 0)$ oder $H_0\,(\mu_X = \mu_Y)$

$\hat{t} > t_{\text{Tab}} \Rightarrow H_1(\delta \neq 0)$ oder $H_1\,(\mu_X \neq \mu_Y)$

Der unverbundene t-Test bei verbundenen Stichproben seltener zu signifikanten Unterschieden führt als der verbundene (paarige or paired auf English) Test.

Vergelich zweier Mittelwerte verbundener Stichproben (t-Test) (2)

Beispiel

Baum	i	1	2	3	4	5	6	7	8
Jahr	X	36.0	31.5	34.0	32.5	35.0	31.5	31.0	35.5
Jahr	Y	34.0	35.5	33.5	36.0	39.0	35.0	33.0	39.5
Differenzen	d_i	2.0	−4.0	0.5	−3.5	−4.0	−3.5	−2.0	−4.0

Figure 3: Erträge in kg von acht Kirschbäumen in zwei Jahren. [Quelle: Köhler et. al]

Vergelich zweier Mittelwerte verbundener Stichproben (t-Test) (3)

$\bar{x} = 35.7$ $\bar{y} = 35.7$

$\bar{d} = -2.31$ $s_d = 2.33$ $n = 8$

Für verbundene Stichproben:

$\hat{t} = 2.8 > t_{\text{Tab}}(FG = 7; \alpha = 0.05) = 2.365 \Rightarrow H_1$

Es bestehen signifikante Mittelwertunterschiede.

Beim Ignorieren der Verbundenheit der Stichproben:

$\hat{t} = 2.07 < t_{\text{Tab}}(FG = 14; \alpha = 0.05) = 2.145 \Rightarrow H_0$

Vergleich zweier Varianzen (F-Test)

Der Test ist nach R.A. Fisher gennant und beschäftigt sich mit der Frage ob die Schätzwerte der Varianzen s_X^2 und s_Y^2 zwei Stichproben aus verschiedenen (normal verteilten) Grundgesamtheiten unterschiedlich sind.

Varianzquotient

$$\hat{F} = \frac{s_X^2}{s_Y^2}$$

dabei $s_X^2 > s_Y^2$.

$F_{\text{Tab}} = F_{n_Y-1}^{n_X-1}(\alpha)$

$\hat{F} \leq F_{\text{Tab}} \Rightarrow H_0\,(\sigma_X^2 = \sigma_Y^2)$

$\hat{F} > F_{\text{Tab}} \Rightarrow H_1\,(\sigma_X^2 \neq \sigma_Y^2)$

Vergleich zweier Varianzen (F-Test) (1)

Beispiel

$n_X = 16,\ \bar{x} = 14.5,\ s_X^2 = 4$

$n_Y = 14,\ \bar{y} = 13.5,\ s_Y^2 = 3$

$\hat{F} = 1.33$

$F_{13}^{15}(0.05) = 3.05$ (zweiseitiger Test)

$\hat{F} < F_{\text{Tab}} \Rightarrow H_0(\sigma_X^2 = \sigma_Y^2)$

Test zu ordinalskalierten Daten (nicht-parametrische Tests)

U-Test von Mann und Whitney (Wilcoxon-Rangsummen-Test)

Lagevergleich zweier unabhängiger Stichproben

- *Voraussetzung*: Die beiden Grundgesamtheiten sollen stetige Verteilungen von *gleichen Form* haben, die Stichproben seien unabhängig und die Daten mindestens ordinalskaliert.

Es wird eine gemeinsame Rangfolge der $(n_X + n_Y)$ Stichprobenwerte gebildet und daraus werden die Rangsummen R_X und R_Y berechnet.

Danach berechnet man

$$U_X = n_X \cdot n_Y + \frac{n_X(n_X+1)}{2} - R_X$$

$$U_Y = n_X \cdot n_Y + \frac{n_Y(n_Y+1)}{2} - R_Y$$

$\hat{U} = \min(U_X, U_Y)$ und $U_{\text{Tab}}(n_X, n_Y; \alpha)$

$\hat{U} \geq U_{\text{Tab}} \Rightarrow H_0$(Mediane gleich)

$\hat{U} < U_{\text{Tab}} \Rightarrow H_1$(Mediane verschieden)

U-Test von Mann und Whitney (Wilcoxon-Rangsummen-Test) (1)

Der *U*-Test hat geringere Voraussetzungen als *t*-Test und hat die *Effizienz* von 95. Effizienz ist das Verhältnis der Stichprobenumfänge, die in zwei verglichenen Tests zur selben *Güte* (auf English *power* $1 - \beta$ Wahrscheinlichkeit die Nullhypothese zu verwerfen, wenn H_1 erfüllt ist) führen. Also muss man den Stichprobenumfang von *U*-Test erhöhen um gleiche α- und β-Fehler wie im *t*-Tet zu haben.

Bei der Vergabe der Rangzahlen wird bei gleichen Werten (Bindungen oder *ties*) das arithmetische Mittel der zugehörigen Rangplätze vergeben. Bei zu vielen Bindungen benötigt $\hat{U}$ Korrekturen.

Falls man die Annahme der gleichen Verteilung fallen lässt vergleicht man mit *U*-Test die Unterschiede in den Verteilungen feststellen.

U-Test von Mann und Whitney (Wilcoxon-Rangsummen-Test) (2)

Beispiel

1955	Rangzahl						6		8	9	10	11		13		15	16	$R_1 = 88$
	Note X						2.2		2.5	2.9	3.0	3.2		3.8		4.2	4.5	
1975	Note Y	1.3	1.7	2.0	2.0	2.1		2.4					3.6		4.1			
	Rangzahl	1	2	3.5	3.5	5		7					12		14			$R_2 = 48$

Figure 4: Durchschnittsnoten und Ränge zweier Abitur-Jaurgänge. [Quelle: Köhler et. al]

$U_X = 12$ $U_Y = 52 \Rightarrow \hat{U} = 12$

$U_{\text{Tab}}(8, 8; 0.05) = 13$

$\hat{U} < U_{\text{Tab}} \Rightarrow H_1$

Die Unterschiede sind signifikant.

Wilcoxon-Test für Paardifferenz (Wilcoxon's signed-ranks test)

Legevergleich zweier verbundener Stichproben

Fragestellung: sind die Mediane zweier vernbundener Stichproben X und Y signifikant verschieden?

- *Voraussetzung*: Die beiden Grundgesamtheiten sollen stetige Verteilungen von *gleichen Form* haben, die Stichproben seien *verbunden* und die Daten mindestens ordinalskaliert.

Es werden die n Messwertdifferenzen $d_i = x_i - y_i$ $d_i \neq 0$ und die Ränge $r(|d_i|)$ gebildet.

Es werden die die Summen der Ränge der positiven Differenzen (W^+) und negativen differenzen (W^-) gebildet.

$\hat{W} = \min(W^+, W^-)$ und $W_{\text{tab}}(n; \alpha)$

$\hat{W} \geq W_{\text{Tab}} \Rightarrow H_0$(Mediane gleich) $\hat{W} < W_{\text{Tab}} \Rightarrow H_1$(Mediane verschieden)

Test zu nominallskalierten Daten (wird später behandelt)

Zusammenhänge in den Daten

VB

Institut für Veterinär-Epidemiologie und Biometrie, FU Berlin

11/03/2020

Wiederholung

Wiederholung

Zusammenhänge in kategoriellen Daten

Zusammenhänge in kategoriellen Daten

Um Zusammenhänge zwischen kategoriellen Daten festzustellen, können wir keine Mittelwerte oder ähnliche kontinuierliche Maßzahlen betrachten

Zusammenhänge in kategoriellen Daten

Um Zusammenhänge zwischen kategoriellen Daten festzustellen, können wir keine Mittelwerte oder ähnliche kontinuierliche Maßzahlen betrachten

Wir können nur Anzahl oder Häufgkeiten für unterschiedliche Kategorien untersuchen

Zusammenhänge in kategoriellen Daten (1)

- Pearsonsche Chi-Quadrat-Test (χ^2-Test)
- Exakter Test nach Fisher
- Yates-Korrektur
- G-Test
- Maßzahlen der Effektstärke

Kontingenztabelle (Beispiel)

Den 200 Katzen wir das *Tanzen* auf der Schur beigebracht indem sie enweder durch *Futter* oder *Aufmerksamkeit* belohnt werden. Nach einer Woche wird gezählt, wie viele Katzen erfolgreich trainiert werden konnten.

Kontingenztabelle (Beispiel)

Den 200 Katzen wir das *Tanzen* auf der Schur beigebracht indem sie enweder durch *Futter* oder *Aufmerksamkeit* belohnt werden. Nach einer Woche wird gezählt, wie viele Katzen erfolgreich trainiert werden konnten.

		Belohnung		
		Futter	Aufmerksamkeit	Gesamt
Tanzen	Ja	28	48	76
	Nein	10	114	124
	Gesamt	38	162	200

Pearsonsche Chi-Quadrat-Test (χ^2-Test)

Gibt es einen Zusammenhang zwischen der Anzahl der erfolgreich trainierten Katzen und Art der Belohnung?

Pearsonsche Chi-Quadrat-Test (χ^2-Test)

Gibt es einen Zusammenhang zwischen der Anzahl der erfolgreich trainierten Katzen und Art der Belohnung?

Pearsonsche Chi-Quadrat-Test (χ^2-Test)

Vergeleicht die beobachteten Häufigkeiten mit solchen, die man rein zufällig bekommen würde

Pearsonsche Chi-Quadrat-Test (χ^2-Test) (1)

$$\chi^2 = \sum \frac{(\text{observed}_{ij} - \text{model}_{ij})^2}{\text{model}_{ij}}$$

i ist die Reihe und j die Spalte in der Kontingenztabelle

Pearsonsche Chi-Quadrat-Test (χ^2-Test) (1)

$$\chi^2 = \sum \frac{(\text{observed}_{ij} - \text{model}_{ij})^2}{\text{model}_{ij}}$$

i ist die Reihe und j die Spalte in der Kontingenztabelle

$$\text{model}_{ij} = \frac{\text{row total}_i \times \text{column total}_j}{n}$$

für den Katzenfall $\chi^2 = 25.35$. Für das Signifikanzneveau von 0.05 entspricht χ^2-Statistik von 3.4. Da unser Wert höher ist, ist das Ergebnis signifikant.

Dabei $df = (r-1)(c-1)$ mit r - Anzahl der Reihen und c - Anzahl der Spalten.

Theoretische Kontingenztabelle (Beispiel)

Theoretische Kontingenztabelle (Beispiel)

		Belohnung		
		Futter	Aufmerksamkeit	Gesamt
Tanzen	Ja	14.44	61.56	76
	Nein	23.56	100.44	124
	Gesamt	38	162	200

Exakter Test nach Fisher

χ^2-Test basiert auf der Annahme, das die Abschätzung von χ^2 aus der Stichprobe der theoretischen χ^2-Verteilung entspicht.

Das gilt allerding nur für die Anzahl der Meßwerte für jede Kategorie größer als 5.

Für die kleinere Anzahl der Meßwerte pro Kategorie wird *Exakter Test nach Fisher* benutzt.

G-Test (The likelihood ratio)

Alternative zum Pearson-χ^2-Test

$$G = 2 \sum \text{observed}_{ij} \ln \left(\frac{\text{observed}_{ij}}{\text{model}_{ij}} \right)$$

G-Statistik ist auch nach χ^2 verteilt. Für den Beispiel mit den Katzen $G = 24.94$ und daher signifikant.

Yates-Korrektur

$$\chi^2 = \sum \frac{(|\text{observed}_{ij} - \text{model}_{ij}| - 0.5)^2}{\text{model}_{ij}}$$

Für den Katzenbeispiel $\chi^2 = 23.52$ und daher ist χ^2 kleiner geworden.

χ^2 in R

```
library(gmodels)
food <- c(28,10)
affection <- c(48,114)
catsTable <- cbind(food,affection)
CrossTable(catsTable, fisher = TRUE, chisq = TRUE, expected = TRUE, sresid = TRUE)
```

χ^2 in R (1) I

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |              Expected N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  200
##
##
##              |
##              |      food | affection | Row Total |
## -------------|-----------|-----------|-----------|
##         [1,] |        10 |       114 |       124 |
##              |    23.560 |   100.440 |           |
##              |     7.804 |     1.831 |           |
##              |     0.081 |     0.919 |     0.620 |
##              |     0.263 |     0.704 |           |
##              |     0.050 |     0.570 |           |
## -------------|-----------|-----------|-----------|
##         [2,] |        28 |        48 |        76 |
##              |    14.440 |    61.560 |           |
##              |    12.734 |     2.987 |           |
##              |     0.368 |     0.632 |     0.380 |
##              |     0.737 |     0.296 |           |
##              |     0.140 |     0.240 |           |
## -------------|-----------|-----------|-----------|
## Column Total |        38 |       162 |       200 |
##              |     0.190 |     0.810 |           |
## -------------|-----------|-----------|-----------|
```

χ^2 in R (1) II

```
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## ------------------------------------------------------------
## Chi^2 =  25.35569     d.f. =  1     p =  4.767434e-07
##
## Pearson's Chi-squared test with Yates' continuity correction
## ------------------------------------------------------------
## Chi^2 =  23.52028     d.f. =  1     p =  1.236041e-06
##
##
## Fisher's Exact Test for Count Data
## ------------------------------------------------------------
## Sample estimate odds ratio:  0.1519927
##
## Alternative hypothesis: true odds ratio is not equal to 1
## p =  1.311709e-06
## 95% confidence interval:  0.06086544 0.352389
##
## Alternative hypothesis: true odds ratio is less than 1
## p =  7.7122e-07
## 95% confidence interval:  0 0.3131634
##
## Alternative hypothesis: true odds ratio is greater than 1
## p =  0.9999999
## 95% confidence interval:  0.07015399 Inf
##
##
##
```

Odds-Ratio (Chancenverhältnis, relative Chance, Quotenverhältnis)

Maßzahl der Effektstärke im Falle von Assoziation zwischen Variablen (measure of associations)

$$\text{odds}_{\text{dancing after food}} = \frac{\text{had food and danced}}{\text{had food but didn't dance}} = 28/10$$

$$\text{odds}_{\text{dancing after affection}} = \frac{\text{had affection and danced}}{\text{had affection but didn't dance}} = 48/114$$

$$\text{odds ratio} = \frac{\text{odds}_{\text{dancing after food}}}{\text{odds}_{\text{dancing after affection}}} = 6.65$$

Odds-Ratio

odds ratio $= 6.57\,(2.84, 16.43)$ bedeutet, dass die Belohung mit Futter um 6.65 die Chance erhöht, der Katze das Tanzen auf der Schnur beizubringen, verglichen mit der Belohnung durch Zuwendung.

Odds-Ratio

odds ratio $= 6.57\,(2.84, 16.43)$ bedeutet, dass die Belohung mit Futter um 6.65 die Chance erhöht, der Katze das Tanzen auf der Schnur beizubringen, verglichen mit der Belohnung durch Zuwendung.

Generell gilt: - odds ratio > 1 die Chance auf Erfolg ist größer durch die Futter-Belohnung, als durch die Zuwendung

Odds-Ratio

odds ratio $= 6.57\,(2.84, 16.43)$ bedeutet, dass die Belohung mit Futter um 6.65 die Chance erhöht, der Katze das Tanzen auf der Schnur beizubringen, verglichen mit der Belohnung durch Zuwendung.

Generell gilt: - odds ratio > 1 die Chance auf Erfolg ist größer durch die Futter-Belohnung, als durch die Zuwendung

- odds ratio $= 1$ sowohl die Futter-Belohnung als auch durch die Zuwendung hat die gleiche Chance auf Erfolg

Odds-Ratio

odds ratio $= 6.57\,(2.84, 16.43)$ bedeutet, dass die Belohung mit Futter um 6.65 die Chance erhöht, der Katze das Tanzen auf der Schnur beizubringen, verglichen mit der Belohnung durch Zuwendung.

Generell gilt: - odds ratio > 1 die Chance auf Erfolg ist größer durch die Futter-Belohnung, als durch die Zuwendung

- odds ratio $= 1$ sowohl die Futter-Belohnung als auch durch die Zuwendung hat die gleiche Chance auf Erfolg
- odds ratio < 1 die Chance auf Erfolg ist kleiner durch die Futter-Belohnung, als durch die Zuwendung

Kontingenzkoeffizient

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

n - Stichprobenumfang

Kontingenzkoeffizient

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

n - Stichprobenumfang

Der Nachteil von C ist das der immer kleiner als 1 ist und hängt von der Form der Kontingenztabelle. Daher betrachten man den *korrigierten Kontingenzkoeffizient*

$$C = \frac{C}{C_{\max}} = \sqrt{\frac{\chi^2 \cdot m}{(\chi^2 + n)(m-1)}}$$

m - $\min(r, c)$

Mosaic plots I

Datensatz über Zullasungen in UC Berkeley für 6 größte Fachbereiche im Jahr 1973

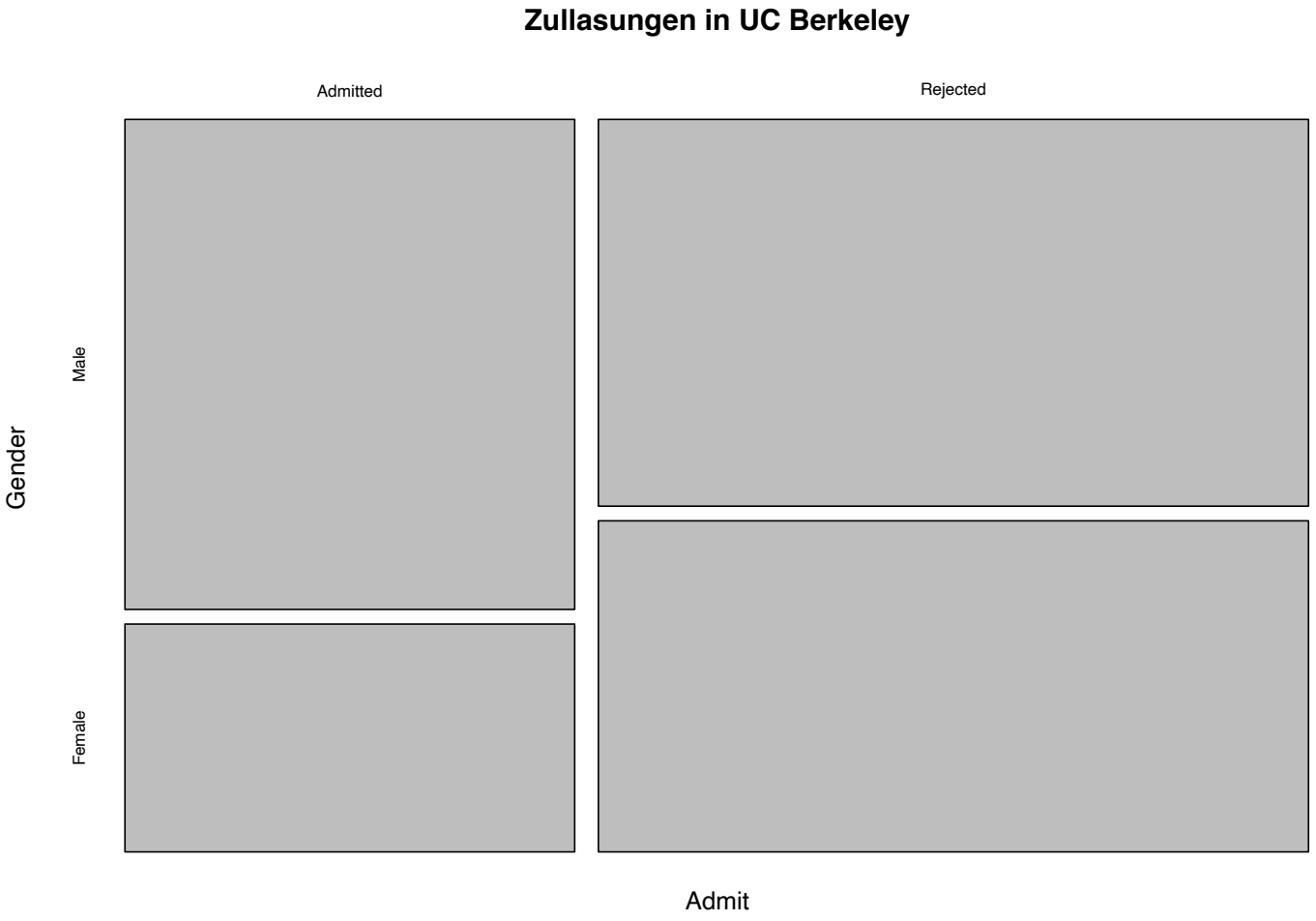
```
## , , Dept = A
##
##           Gender
## Admit      Male Female
##   Admitted  512     89
##   Rejected  313     19
##
## , , Dept = B
##
##           Gender
## Admit      Male Female
##   Admitted  353     17
##   Rejected  207      8
##
## , , Dept = C
##
##           Gender
## Admit      Male Female
##   Admitted  120    202
##   Rejected  205    391
##
## , , Dept = D
##
##           Gender
## Admit      Male Female
##   Admitted  138    131
##   Rejected  279    244
##
## , , Dept = E
##
##           Gender
## Admit      Male Female
```

Mosaic plots II

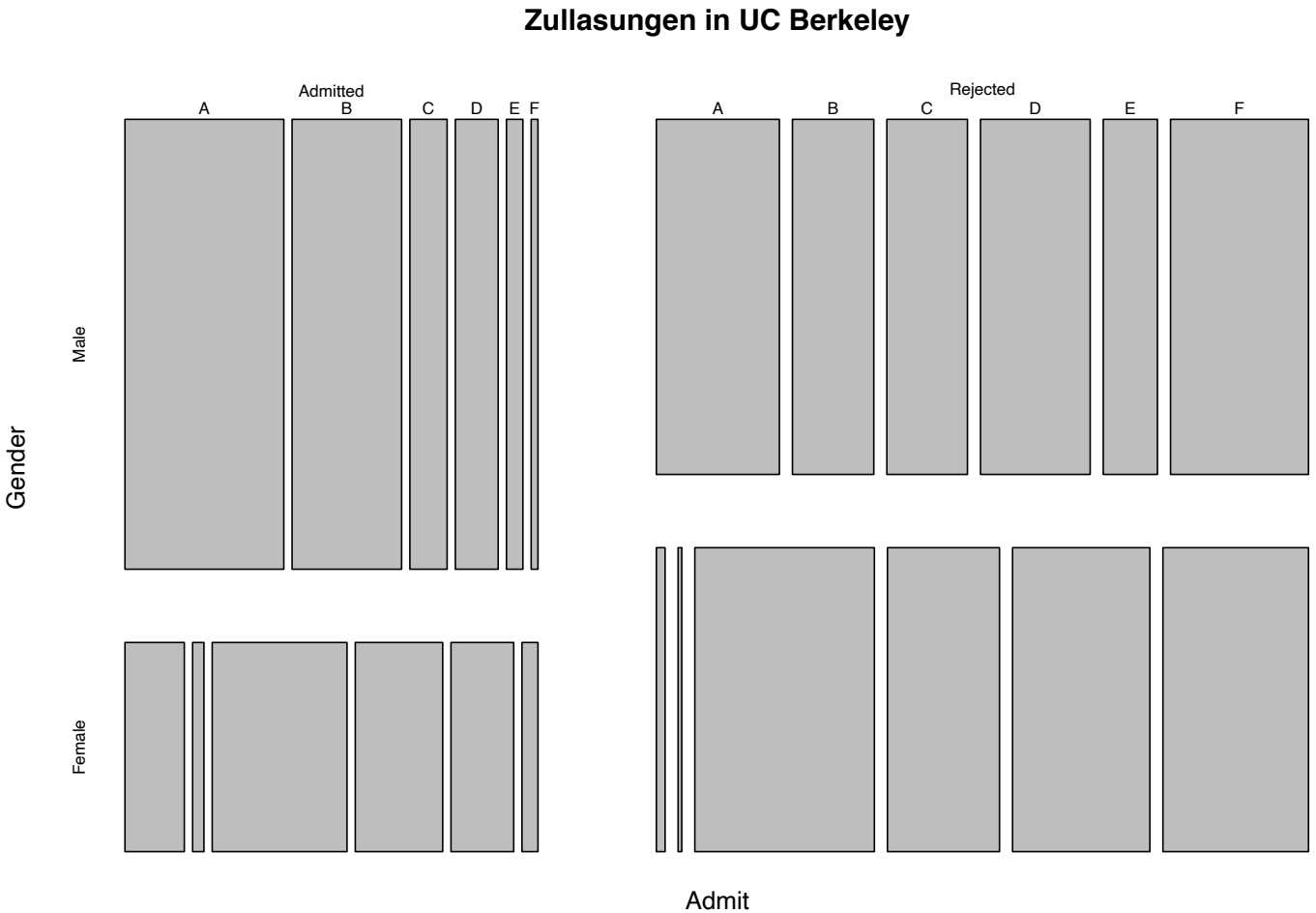
```
##   Admitted   53     94
##   Rejected  138    299
##
## , , Dept = F
##
##           Gender
## Admit      Male Female
##   Admitted   22     24
##   Rejected  351    317
```

	Male	Female
Admitted	1198	557
Rejected	1493	1278

Simpson-Paradoxon (1)



Simpson-Paradoxon (2)



Simpson-Paradoxon (3)

Es scheint, dass die Bewertung verschiedener Gruppen unterschiedlich ausfällt, je nachdem ob man die Ergebnisse der Gruppen kombiniert oder nicht. Dieses Phänomen tritt oft bei statistischen Auswertungen in den Sozialwissenschaften und in der Medizin auf [*Wikipedia*]

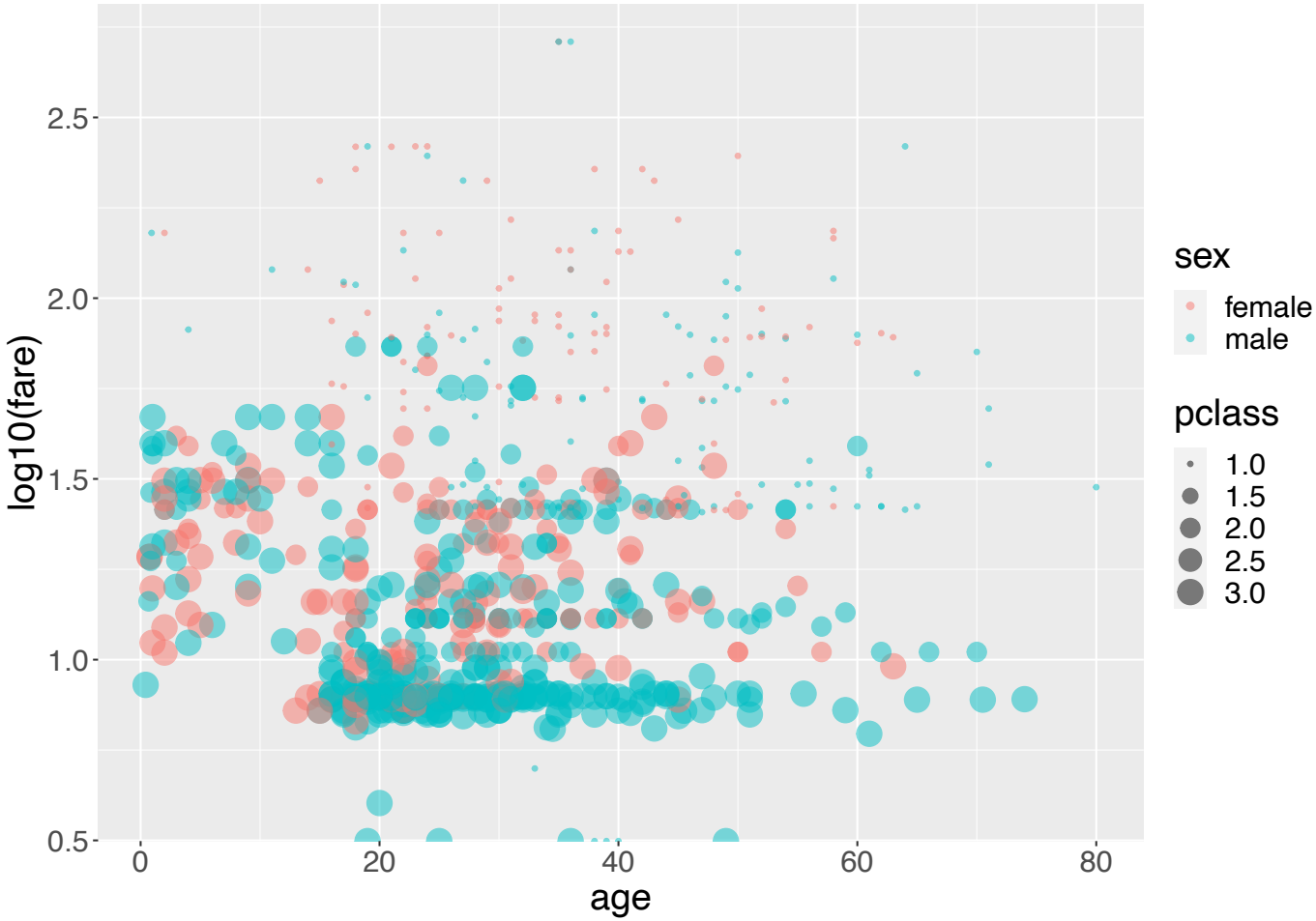
Simpson-Paradoxon (3)

Es scheint, dass die Bewertung verschiedener Gruppen unterschiedlich ausfällt, je nachdem ob man die Ergebnisse der Gruppen kombiniert oder nicht. Dieses Phänomen tritt oft bei statistischen Auswertungen in den Sozialwissenschaften und in der Medizin auf [*Wikipedia*]

Liegen je nach Beurteilungsweise deutlich unterschiedliche Ergebnisse vor, kann dies auf nicht erfasste Einflussfaktoren zurückgeführt werden. Wollen Auswertende mögliche Fehlschlüsse vermeiden, müssen sie diese Einflussfaktoren finden, soweit sie vorhanden sind. Das Vorliegen eines Simpson-Paradoxons kann hier als Indikator dienen. [*Wikipedia*]

Zusammenhänge in kontinuierlichen Daten

Bivariate Daten



Zusammenhänge in kontinuierlichen Daten

- Pearson-Korrelationskoeffizient
- Spearman-Rangkorrelationskoeffizient

Pearson-Korrelationskoeffizient

Variance

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

Covariance

$$\mathrm{cov}(x, y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{n - 1}$$

Pearson-Korrelationskoeffizient

$$r = \frac{\operatorname{cov}(x, y)}{s_x s_y} = \frac{\sum (x - \bar{x})(y - \bar{y})}{(n-1) s_x s_y}$$

Korregierter z-score

$$z_r = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right)$$

Sein Standardfehler

$$SE_{z_r} = \frac{1}{\sqrt{n-3}}$$

Konfidenzintervalle

$$z_r \pm (1.96 \times SE_{z_r}), \text{ Transformation: } r = \frac{e^{2z_r} - 1}{e^{2z_r} + 1}$$

Korrelationskoeffizient (Beispiel)

```
#fig.show = 'hide'
#results='hide'
library(ggplot2)
df <- read.table('/Users/vitaly/Lehre/WS1920/Beuth/4_Statistische_Tests_and_Regression/Data/Data\ files/Album\ Sales\ 1.dat', s
cor(df)
```

```
##           adverts     sales
## adverts 1.0000000 0.5784877
## sales   0.5784877 1.0000000
```

```
cor.test(df$adverts,df$sales)
```

```
##
##  Pearson's product-moment correlation
##
## data:  df$adverts and df$sales
## t = 9.9793, df = 198, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.4781207 0.6639409
## sample estimates:
##       cor
## 0.5784877
```

Spearman-Korrelationskoeffizient (Beispiel)

```r
cor(df, method = "spearman")
```

```
##           adverts     sales
## adverts 1.0000000 0.5541557
## sales   0.5541557 1.0000000
```

```r
cor.test(df$adverts, df$sales, alternative = "greater", method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  df$adverts and df$sales
## S = 594444, p-value < 2.2e-16
## alternative hypothesis: true rho is greater than 0
## sample estimates:
##       rho
## 0.5541557
```

Kendall-tau (Beispiel)

```
cor(df, method = "kendall")
```

```
##           adverts     sales
## adverts 1.0000000 0.3985301
## sales   0.3985301 1.0000000
```

```
cor.test(df$adverts, df$sales, alternative = "greater", method = "kendall")
```

```
##
##  Kendall's rank correlation tau
##
## data:  df$adverts and df$sales
## z = 8.2362, p-value < 2.2e-16
## alternative hypothesis: true tau is greater than 0
## sample estimates:
##       tau
## 0.3985301
```

Wahrscheinlichkeitsverteilungen

Normalverteilung oder Gauß-Verteilung

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] = \mathcal{N}(\mu, \sigma)$$

μ — Mittelwert

σ — Standardabweichung

Normalverteilung oder Gauß-Verteilung (1)

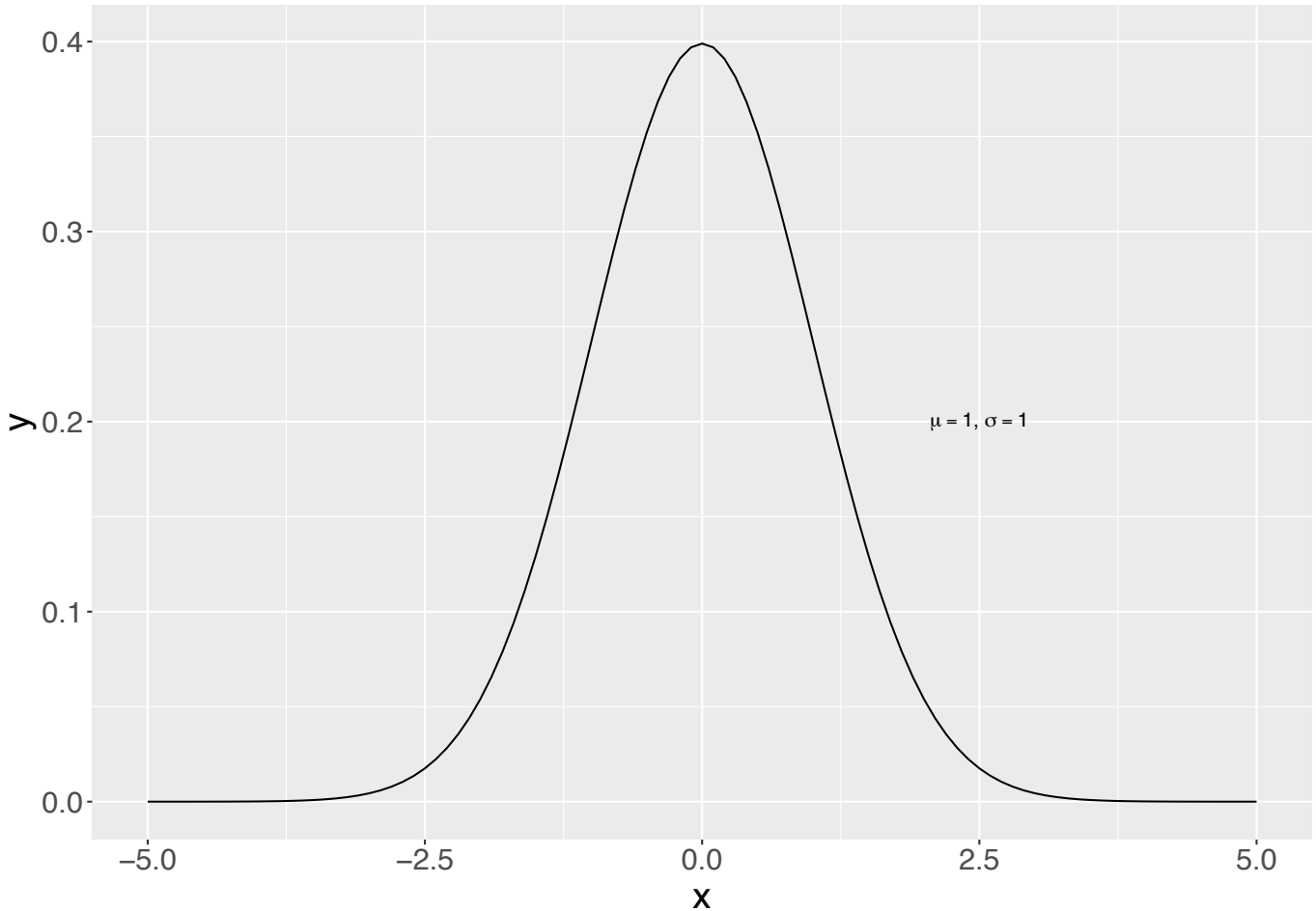


Figure 1: Normalverteilung

Binomiale Verteilung

$$p(k|n,p) = \binom{n}{k} p^k (1-p)^{n-k}$$

n - Anzahl der Versuche

p - Erfolgs- oder Trefferwahrscheinlichkeit

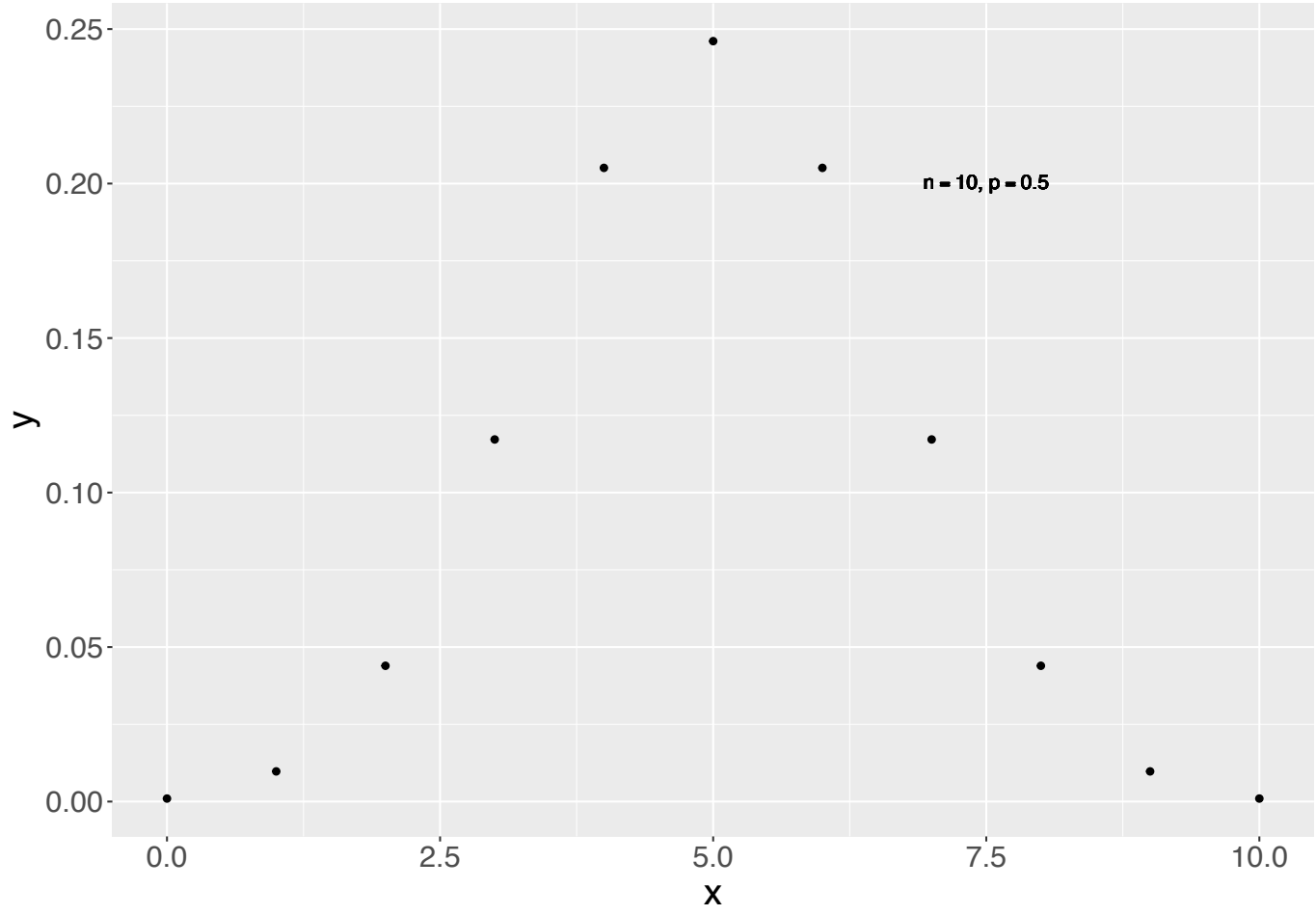


Figure 2: Binomiale Verteilung

Poisson-Verteilung

$$P_\lambda(k) = \frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}$$

λ Erwartungswert und gleichzeitig die Varianz der Poisson-Verteilung

Poisson-Verteilung (1)

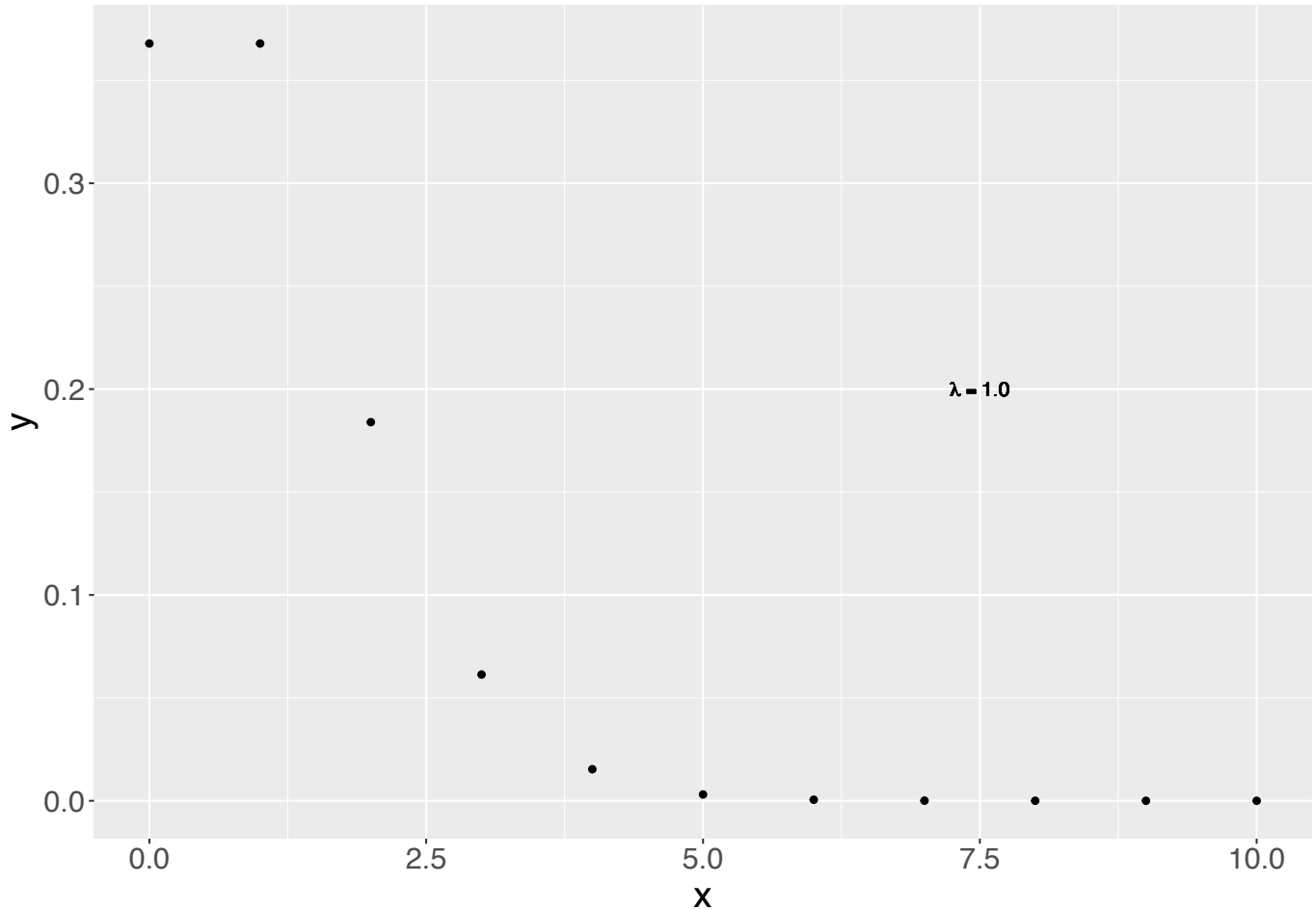


Figure 3: Poisson-Verteilung

Introduction into Linear Regression

Vitaly Belik

Institute for Veterinary Epidemiology and Biostatistics

March 11, 2019

(Lineare) Regression

$$Y = f(X_1, ..., X_n)$$

Regression ist ein weit gefasster Begriff für eine Reihe von Methoden, die zur Vorhersage einer *Antwort*-Variable (oder einer *abhängigen, resultierenden*) aus einer oder mehreren *Prädiktor*-Variablen (*unabhängigen, erklärenden*) verwendet werden.

Ziele der Regression

- *Bestimmung* der erklärenden Variablen, die sich auf die Antwortvariable beziehen
- *Beschreibung* der Form der Beziehung
- *Bereitstellen* einer Gleichung für die *Vorhersage* der Antwortvariablen aus den erklärenden Variablen

Aus der Regression konnte keine *Kausalität* direkt abgeleitet werden!

Regression example (1)

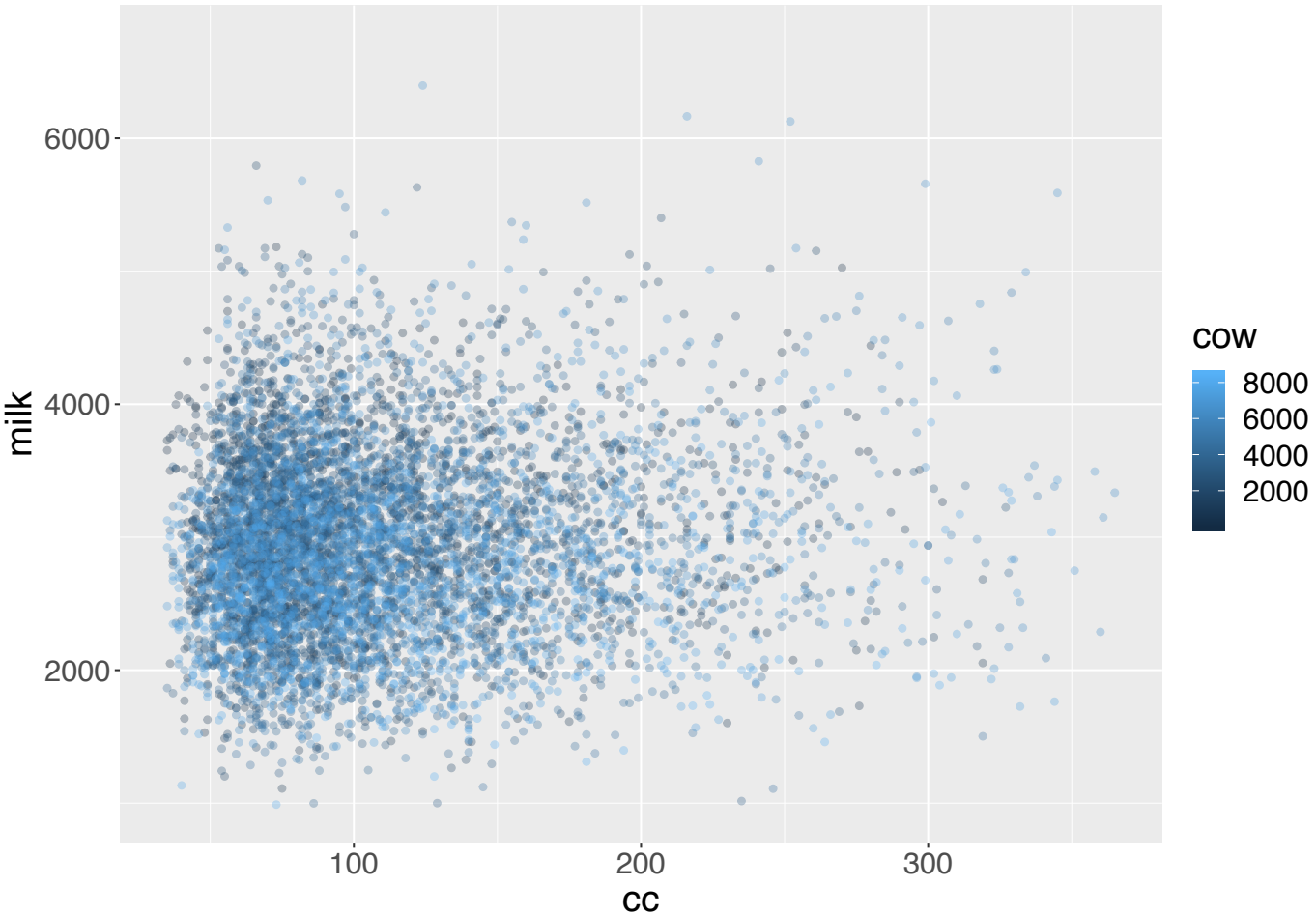


Figure 1: Impact of CC interval (interval from calving to conception) on Milk volume. 9383 lactation records in 42 year-round calving herds. http://projects.upei.ca/ver/data-and-samples/.

Regression example (2)

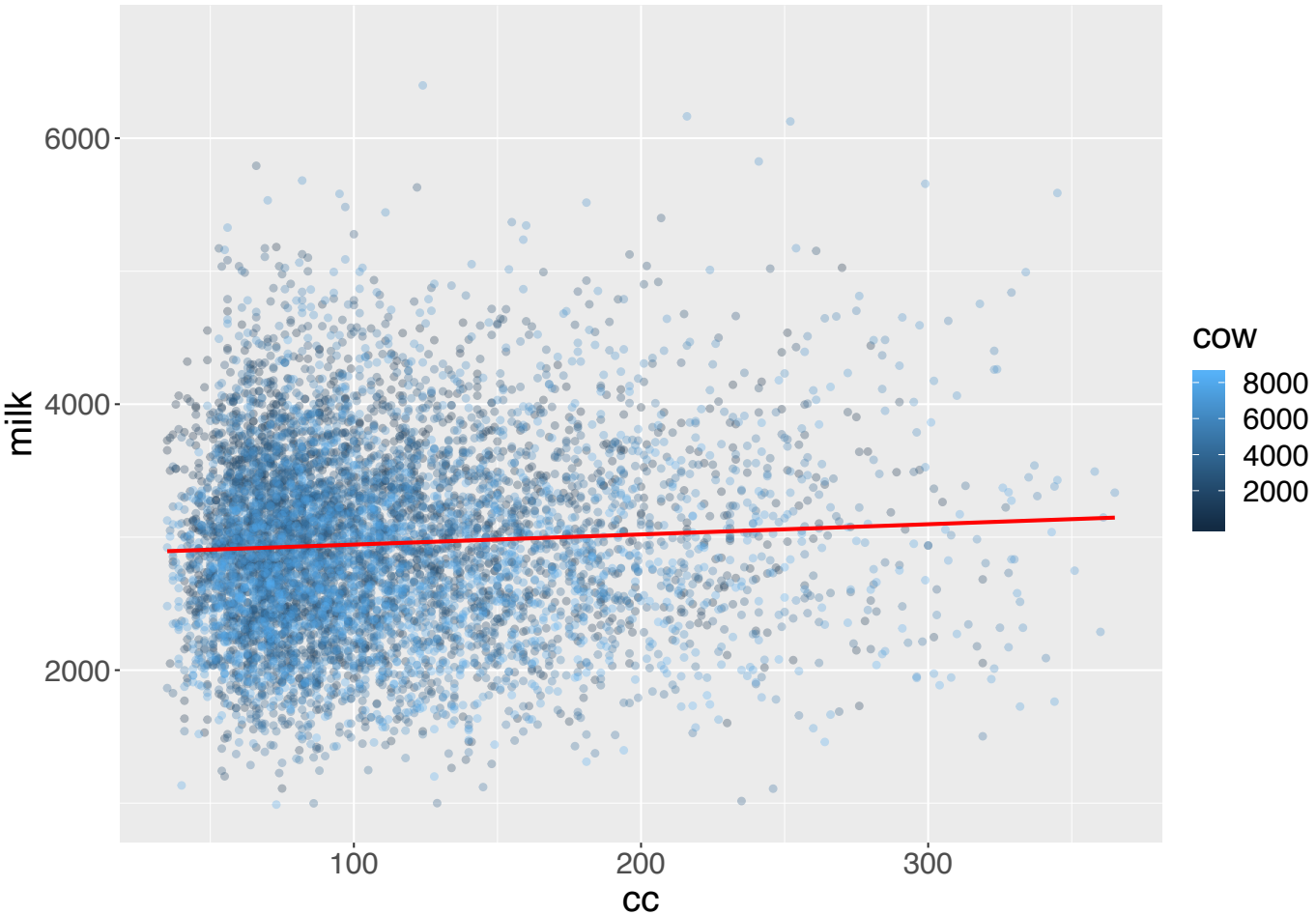


Figure 2: Impact of CC interval (interval from calving to conception) on Milk volume. 9383 lactation records in 42 year-round calving herds. http://projects.upei.ca/ver/data-and-samples/.

Regression example (3)

```
library(ggplot2)
load("~/Lehre/WS1819/Spring_School/ver-master/data/ver2_data_R/daisy2.rdata")
ggplot(daisy2, aes(x = cc, y = milk120, colour = cow)) +
theme(text = element_text(size=18)) +
geom_point(alpha = 0.3) +
labs(x = "cc", y = "milk") +
geom_smooth(fill=NA,color="red",size=1,method="lm")
```

Regression example: model output

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

```
model <- lm(milk120~cc, data = daisy2)
summary(model)
```

```
##
## Call:
## lm(formula = milk120 ~ cc, data = daisy2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2031.7  -490.9   -35.4   432.1  3434.5
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2867.6913    18.9499 151.330  < 2e-16 ***
## cc             0.7652     0.1526   5.014 5.47e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 705.1 on 7032 degrees of freedom
##   (2349 observations deleted due to missingness)
## Multiple R-squared:  0.003562,   Adjusted R-squared:  0.00342
## F-statistic: 25.14 on 1 and 7032 DF,  p-value: 5.471e-07
```

Regression example: ANOVA table

```
anova(model)
```

```
## Analysis of Variance Table
##
## Response: milk120
##             Df     Sum Sq  Mean Sq F value    Pr(>F)
## cc           1   12495528 12495528  25.137 5.471e-07 ***
## Residuals 7032 3495627193   497103
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANOVA table

$$\hat{Y} = \beta_0 + \beta_1 X$$

To assess how much information the variable X contains about the variable Y, we consider how much of the *sum of squares* (SS) of the variable Y could be explained by the knowledge of the variable X.

Table 1: Decomposition of sums of squares in regression model with k predictor variables [after Dohoo, p. 327]

Source of variation	Sum of squares	Degrees of freedom	Mean square	F-test
Model	$SSM = \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2$	$dfM = k$	$MSM = \frac{SSM}{dfM}$	$\frac{MSM}{MSE}$
Error	$SSE = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2$	$dE = n - (k+1)$	$MSE = \frac{SSE}{dfE}$	
Total	$SST = \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2$	$dfT = n - 1$	$MST = \frac{SST}{dfT}$	

$MSE = \sigma^2$ and σ is called *standard error of prediction*.

F-test to assess the model significance

H_0: $\beta_1 = \beta_2 ... \beta_k = 0$, $\beta_0 \neq 0$

H_1: at least some of the coefficients $\beta_i \neq 0$, $i \neq 0$

- Variables are chosen in a way maximizing F-statistic
- F-test has a straightforward meaning if independent variables are manipulated treatments in a controlled experiment
- Caution if applied to observational variables (influenced by the number of variables, their correlations and sample size)

t-test (with $n - (k + 1)$) degrees of freedom (1)

H_0: $\beta_i = \beta^* \; e.g. \beta^* = 0$

H_1: $\beta_i \neq \beta^* \; e.g. \beta^* = 0$

$$t = \frac{\beta_i - \beta^*}{SE(\beta_i)}$$

$SE(\beta_i)$ is a *standard error of the esimated coefficient*.

In the case of single predictor

$$SE(\beta_1) = \sqrt{\frac{MSE}{SSX(\beta_1)}},$$

where $SSX_1 = \sum_{i=1}^{n} \left(X_{1,i} - \bar{X_1}\right)^2$ is *sum of square of the variable* X.

t-test (with $n - (k + 1)$) degrees of freedom (2)

- Variables are chosen in a way maximazing t-statistic ensuring its significance
- t-test has a straightforward meaning only if independent variables are manipulated treatments in a controlled experiment
- Caution is required if applied to observational variables (influenced by the number of variables, their correlations and sample size)

Intervals for prediction (1)

Two sources of variation of prediction estimates:

- from the estimation of regression parameters (SE)
- from the variation associated with a new observation x^*(variation about the regression equation for the mean)

Error of the mean

For simple regression (single predictor) *error of the mean* of a large number of new observations

$$SE_{\text{mean}}(Y|x^*) = \sigma\sqrt{\frac{1}{n} + \frac{\left(x^* - \bar{X}\right)^2}{SSX}}$$

Intervals for prediction (1)

Standard error for a new observation

For simple regression (single predictor)

$$SE_{\text{obs}}(Y|x^*) = \sigma\sqrt{1 + \frac{1}{n} + \frac{\left(x^* - \bar{X}\right)^2}{SSX}}$$

95% Confidence interval

$$95\%CI = Y \pm t_{0.05}(SE)$$

Coefficient of determination, R^2

- R^2 describes the amount of variance in the outcome variable expained by the predictor variables.
- R^2 is a squared *correlation coefficient* between the predicted and observed Y-values

$$R^2 = \frac{SSM}{SST} = 1 - \frac{SSE}{SST}$$

Adjusted R^2

Because R^2 alwasy increases with the number of variables, *adjusted* R^2 is considered

$$R^2_{\text{adjusted}} = 1 - \frac{MSE}{MST}$$

Akaike Information Criterion (AIC)

In estimating the amount of information lost by a model, AIC deals with the trade-off between the goodness of fit of the model and the simplicity of the model. In other words, AIC deals with both the risk of overfitting and the risk of underfitting.

- One should prefer a model with smaller AIC

```
model1 <- lm(milk120~cc, data = daisy2)
AIC(model1)
```

```
## [1] 112227.5
```

```
model2 <- lm(milk120~parity, data = daisy2)
AIC(model2)
```

```
## [1] 104745.6
```

Another simple regression

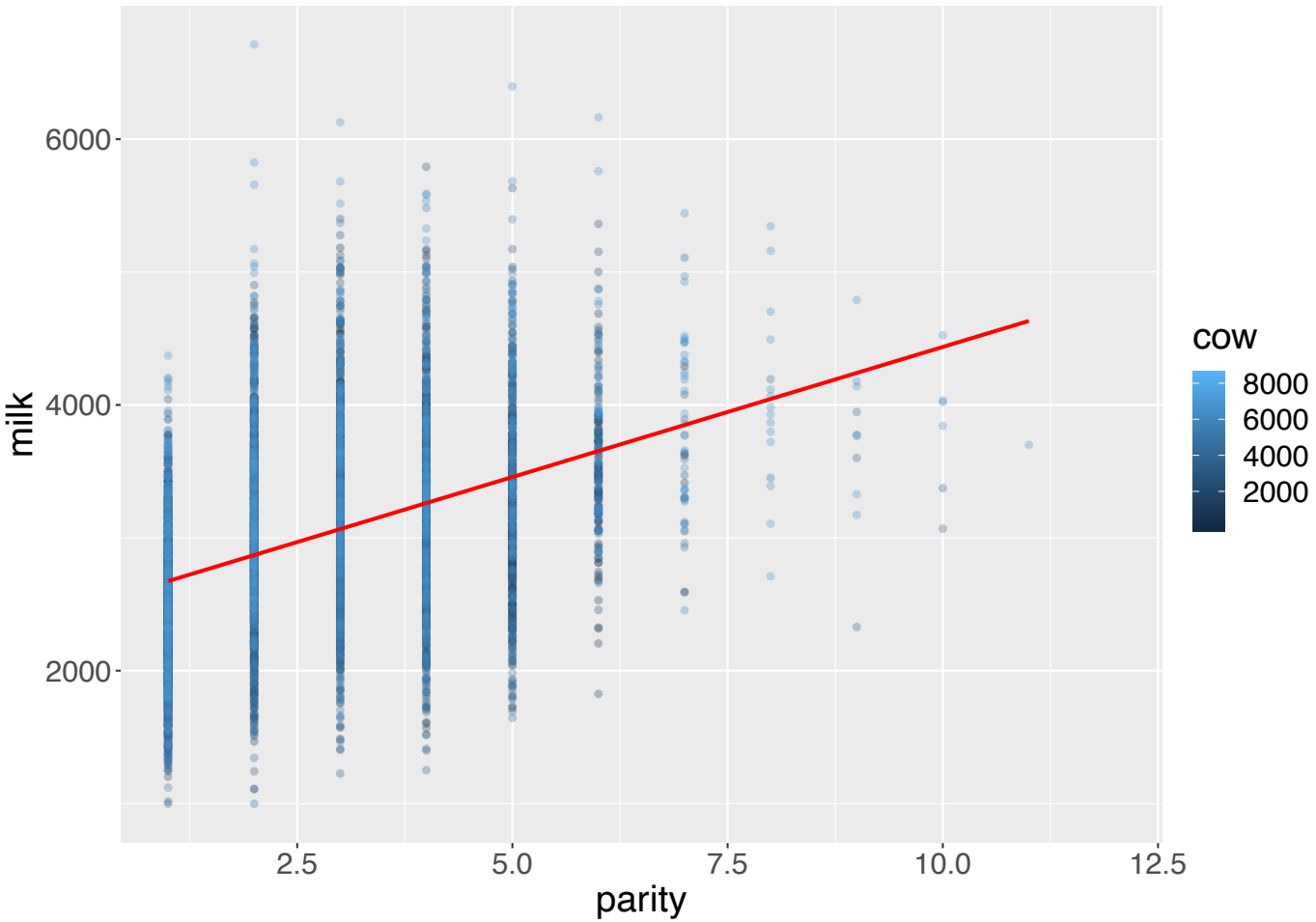


Figure 3: Impact of parity on Milk volume. 9383 lactation records in 42 year-round calving herds. http://projects.upei.ca/ver/data-and-samples/.

Another simple regression: model output

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

```
model <- lm(milk120~parity, data = daisy2)
summary(model)
```

```
##
## Call:
## lm(formula = milk120 ~ parity, data = daisy2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2010.3  -454.6   -36.3   413.5  3842.9
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2478.286     17.176  144.28   <2e-16 ***
## parity       195.807      5.385   36.36   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 667.8 on 6608 degrees of freedom
##   (2773 observations deleted due to missingness)
## Multiple R-squared:  0.1667, Adjusted R-squared:  0.1666
## F-statistic:  1322 on 1 and 6608 DF,  p-value: < 2.2e-16
```

Another simple regression: ANOVA table

```
anova(model)
```

```
## Analysis of Variance Table
##
## Response: milk120
##             Df     Sum Sq   Mean Sq F value    Pr(>F)
## parity       1  589631568 589631568  1322.1 < 2.2e-16 ***
## Residuals 6608 2947090043    445988
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Another simple regression: prediction interval (1)

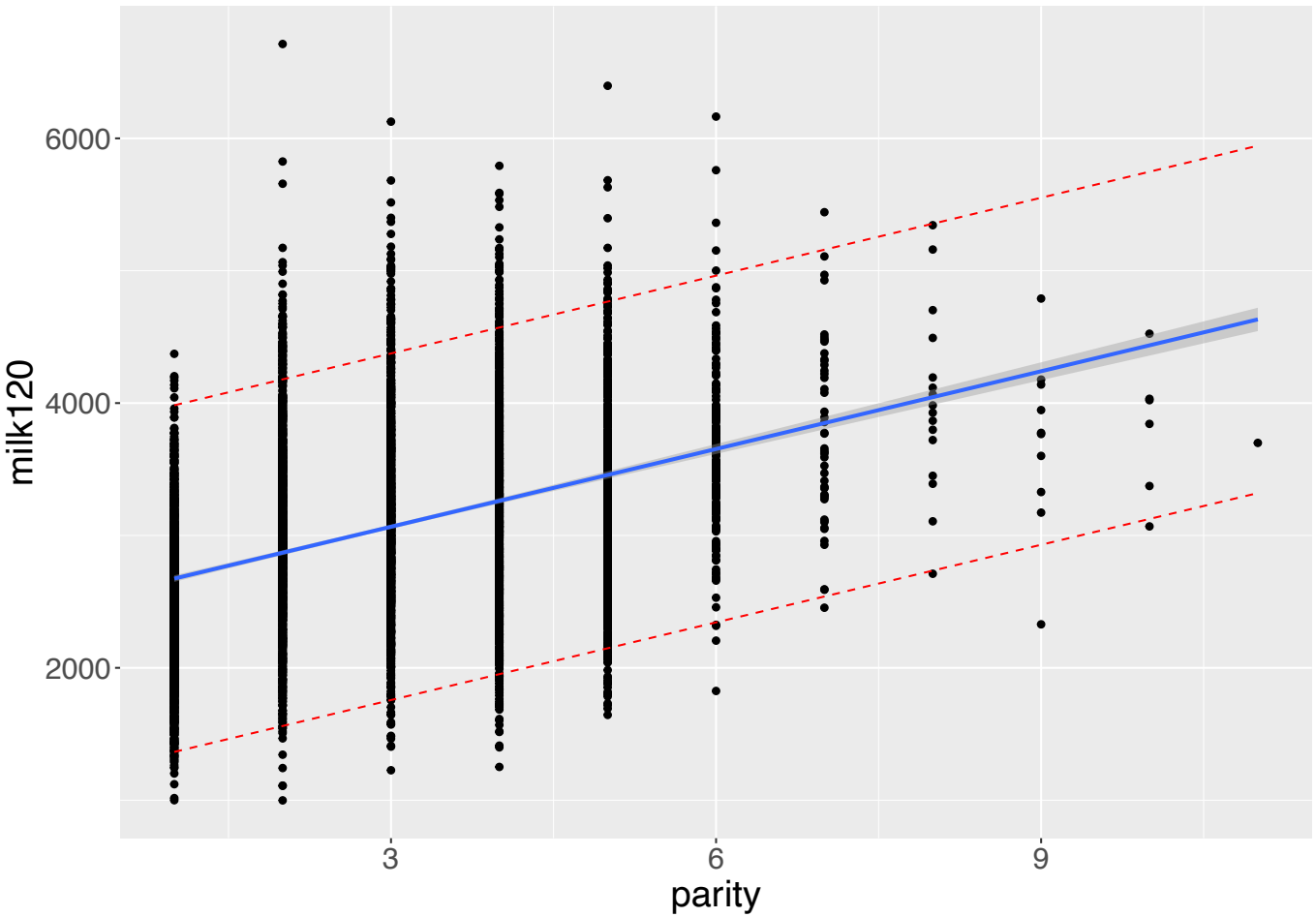


Figure 4: Prediction interval for the regression milk volume on parity.

Another simple regression: prediction interval (2)

```
temp_var <- predict(model, interval="prediction")
new_df0 <- na.omit(data.frame("milk120" = daisy2$milk120, "parity" = daisy2$parity))
new_df <- cbind(new_df0, temp_var)
ggplot(new_df, aes(x=parity, y=milk120))+
    geom_point() +
    geom_line(aes(y=lwr), color = "red", linetype = "dashed")+
    geom_line(aes(y=upr), color = "red", linetype = "dashed")+
    geom_smooth(method=lm, se=TRUE)
```

[Implementation idea from https://rpubs.com/Bio-Geek/71339]

Multiple Linear Regression

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \varepsilon$$

$$Y = \beta_0 + \sum_{i}^{n} \beta_i X_i + \varepsilon$$

Multiple regression 1: model output

```r
model <- lm(milk120 ~ parity + dyst + twin + rp + vag_disch, data = daisy2)
summary(model)
```

```
##
## Call:
## lm(formula = milk120 ~ parity + dyst + twin + rp + vag_disch,
##     data = daisy2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2014.4  -451.0   -32.9   412.0  3902.3
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2482.496     17.497 141.883   <2e-16 ***
## parity       195.788      5.406  36.220   <2e-16 ***
## dyst         -80.862     43.001  -1.880   0.0601 .
## twin           9.374     57.866   0.162   0.8713
## rp           -63.599     33.118  -1.920   0.0549 .
## vag_disch    123.049     53.271   2.310   0.0209 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 667.5 on 6604 degrees of freedom
##   (2773 observations deleted due to missingness)
## Multiple R-squared:  0.1681, Adjusted R-squared:  0.1675
## F-statistic: 266.9 on 5 and 6604 DF,  p-value: < 2.2e-16
```

Multiple regression 1: ANOVA table

```
anova(model)
```

```
## Analysis of Variance Table
##
## Response: milk120
##             Df     Sum Sq   Mean Sq   F value  Pr(>F)
## parity       1  589631568 589631568 1323.4967 < 2e-16 ***
## dyst         1    1414549   1414549    3.1751 0.07481 .
## twin         1       6718      6718    0.0151 0.90227
## rp           1    1140835   1140835    2.5607 0.10960
## vag_disch    1    2377030   2377030    5.3355 0.02093 *
## Residuals 6604 2942150910    445510
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Multiple regression 2: model output

```r
library(lubridate)
daisy2$date <- as.character(daisy2$calv_dt)
daisy2$date <- ymd(daisy2$date)
daisy2$mth <- month(daisy2$date)
daisy2$aut_calv <- with(daisy2, ifelse(mth %in% c(9:12), "fall", "other"))
daisy2$hs100 <- daisy2$herd_size / 100  # herd size scaled by dividing by 100
daisy2$hs100_ct <- daisy2$hs100 - mean(daisy2$hs100)  # centered

model1 <- lm(wpc ~ aut_calv + hs100_ct  + twin + rp + vag_disch + dyst + rp*vag_disch, data = daisy2)
summary(model1)
```

```
##
## Call:
## lm(formula = wpc ~ aut_calv + hs100_ct + twin + rp + vag_disch +
##     dyst + rp * vag_disch, data = daisy2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -78.30 -39.98 -16.77  26.03 254.11
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    73.5593     0.9815  74.948  < 2e-16 ***
## aut_calvother  -2.4442     1.2819  -1.907  0.05659 .
## hs100_ct        4.6369     0.5674   8.172 3.53e-16 ***
## twin           15.5310     4.9822   3.117  0.00183 **
## rp             12.1403     2.9645   4.095 4.26e-05 ***
## vag_disch       6.6528     5.7964   1.148  0.25111
## dyst            5.9930     3.3819   1.772  0.07642 .
## rp:vag_disch   -2.6061    10.5217  -0.248  0.80439
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 54.99 on 7462 degrees of freedom
##   (1913 observations deleted due to missingness)
## Multiple R-squared:  0.01417,    Adjusted R-squared:  0.01324
## F-statistic: 15.32 on 7 and 7462 DF,  p-value: < 2.2e-16
```

Multiple regression 2: ANOVA table

```
anova(model1)
```

```
## Analysis of Variance Table
##
## Response: wpc
##                Df   Sum Sq Mean Sq F value    Pr(>F)
## aut_calv        1    14341   14341  4.7421 0.0294644 *
## hs100_ct        1   193855  193855 64.1018 1.361e-15 ***
## twin            1    38759   38759 12.8165 0.0003458 ***
## rp              1    62035   62035 20.5129 6.015e-06 ***
## vag_disch       1     5460    5460  1.8054 0.1791075
## dyst            1     9662    9662  3.1950 0.0739055 .
## rp:vag_disch    1      186     186  0.0613 0.8043854
## Residuals    7462 22566410    3024
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interactions (1)

Previously we considered only the *main effects* of the predictor variables. This assumes the association between Y and X_i is the same at all levels (values) of X_j ($i \neq j$) and the association of X_j to Y is the same at all levels (values) of X_i.

If this assumption is violated and the effect of one predictor variable depends on the values of another predictor variable, the addition of *Interaction term* could significantly improve the model.

For continuous variables interaction could be included in the following way

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$$

Interactions, example 1 (1)

For interval from wait period to conception (wpc)

```
model <- lm(wpc~milk120*parity, data = daisy2)
summary(model)
```

```
##
## Call:
## lm(formula = wpc ~ milk120 * parity, data = daisy2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -73.17 -39.16 -16.11  25.41 247.09
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    65.3914728  6.0499361  10.809   <2e-16 ***
## milk120         0.0019932  0.0020838   0.957    0.339
## parity         -1.6645556  2.0620816  -0.807    0.420
## milk120:parity  0.0004894  0.0006480   0.755    0.450
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 53.34 on 5753 degrees of freedom
##   (3626 observations deleted due to missingness)
## Multiple R-squared:  0.002014,   Adjusted R-squared:  0.001494
## F-statistic:  3.87 on 3 and 5753 DF,  p-value: 0.008884
```

Interactions, example 1 (2)

```
anova(model)
```

```
## Analysis of Variance Table
##
## Response: wpc
##                   Df   Sum Sq Mean Sq F value   Pr(>F)
## milk120            1    31148 31147.8 10.9496 0.000942 ***
## parity             1      260   260.0  0.0914 0.762404
## milk120:parity     1     1623  1622.6  0.5704 0.450124
## Residuals       5753 16365322  2844.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
AIC(model)
```

```
## [1] 62130.22
```

Interactions, example 1 (3)

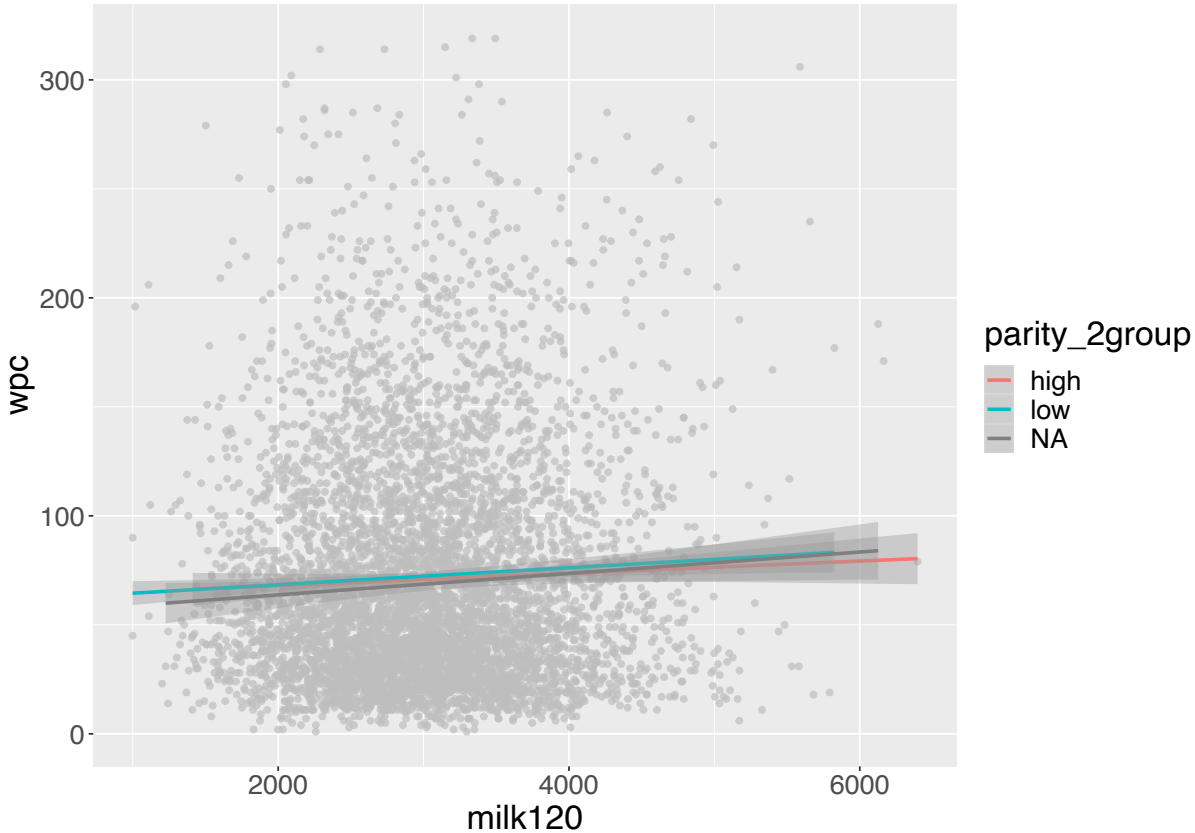


Figure 5: Impact of interaction between milk volume (milk120) and parity on the interval from waiting period to conception (wpc).

Interactions, example 1 (4)

```
library(dplyr)
d <- data.frame("milk120" = daisy2$milk120, "wpc"= daisy2$wpc , "parity" = daisy2$parity)
d = na.omit(d)
x <- d$parity
d$parity_2group <-
  case_when(x > median(x) ~ "high",
            x < median(x) ~ "low")
count(d, parity_2group)
d %>%
  ggplot() +
  aes(x = milk120, y = wpc, group = parity_2group, color = parity_2group) +
  geom_point(color = "grey", alpha = .7) +
  geom_smooth(method = "lm")
```

[Visualization idea from: https://sebastiansauer.github.io/vis_interaction_effects/]

Interactions, example 2 (1)

```r
daisy2$milk120k <- daisy2$milk120/1000 # scaling of the variables
model <- lm(wpc~as.factor(dyst)*milk120k, data = daisy2)
summary(model)
```

```
##
## Call:
## lm(formula = wpc ~ as.factor(dyst) * milk120k, data = daisy2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -72.84 -40.60 -16.48  26.00 254.17
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 67.7666     2.8723  23.594   <2e-16 ***
## as.factor(dyst)1            30.3670    15.4199   1.969   0.0490 *
## milk120k                     1.8437     0.9442   1.953   0.0509 .
## as.factor(dyst)1:milk120k   -8.1006     5.3236  -1.522   0.1281
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 55 on 7030 degrees of freedom
##   (2349 observations deleted due to missingness)
## Multiple R-squared:  0.0014, Adjusted R-squared:  0.0009739
## F-statistic: 3.285 on 3 and 7030 DF,  p-value: 0.01989
```

Interactions, example 2 (2)

```
anova(model)
```

```
## Analysis of Variance Table
##
## Response: wpc
##                             Df   Sum Sq Mean Sq F value  Pr(>F)
## as.factor(dyst)              1    13967 13966.8  4.6169 0.03169 *
## milk120k                     1     8844  8844.1  2.9235 0.08734 .
## as.factor(dyst):milk120k     1     7004  7004.3  2.3154 0.12815
## Residuals                 7030 21266817  3025.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
AIC(model)
```

```
## [1] 76343.14
```

Interactions, example 2 (3)

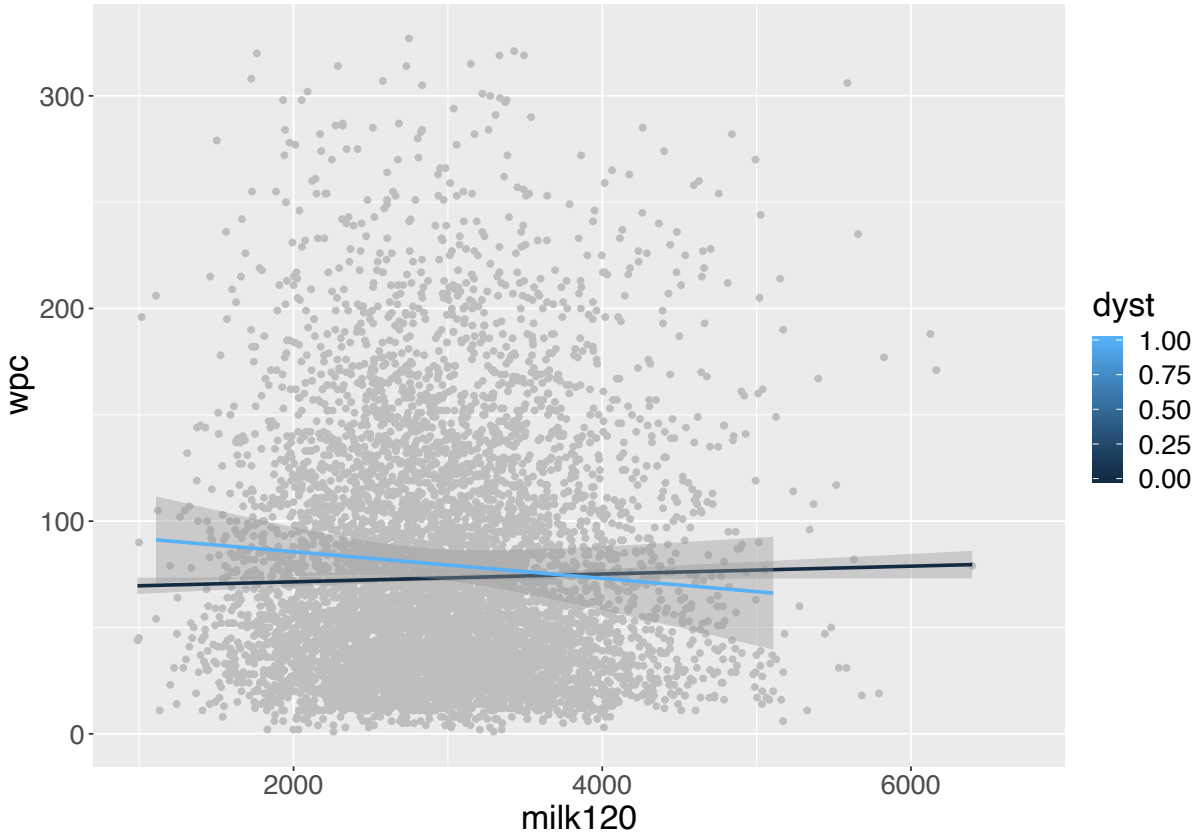


Figure 6: Impact of interaction between milk volume (milk120) and dystocia (dyst) on the interval from waiting period to conception (wpc).

Interactions, example 2 (4)

```
dyst_0 <- filter(daisy2, dyst == "0")
dyst_1 <- filter(daisy2, dyst == "1")
ggplot(daisy2) +
  aes(x = milk120, y = wpc, color = dyst) +
  geom_point(color = "grey") +
  geom_smooth(method = "lm", data = dyst_0) +
  theme(text = element_text(size = 20)) +
  geom_smooth(method = "lm", data = dyst_1)
```

Model diagnostics

Independence

Homoscedasticity

Normal distribution of residuals

Linearity

Independence

- usually is clear from the nature of the data
- lack of independence e.g. in case of multiple observations of a single animal or the same herd
- lack of independence if serial correlations are present like measurements during particular season

Homoscedasticity (1)

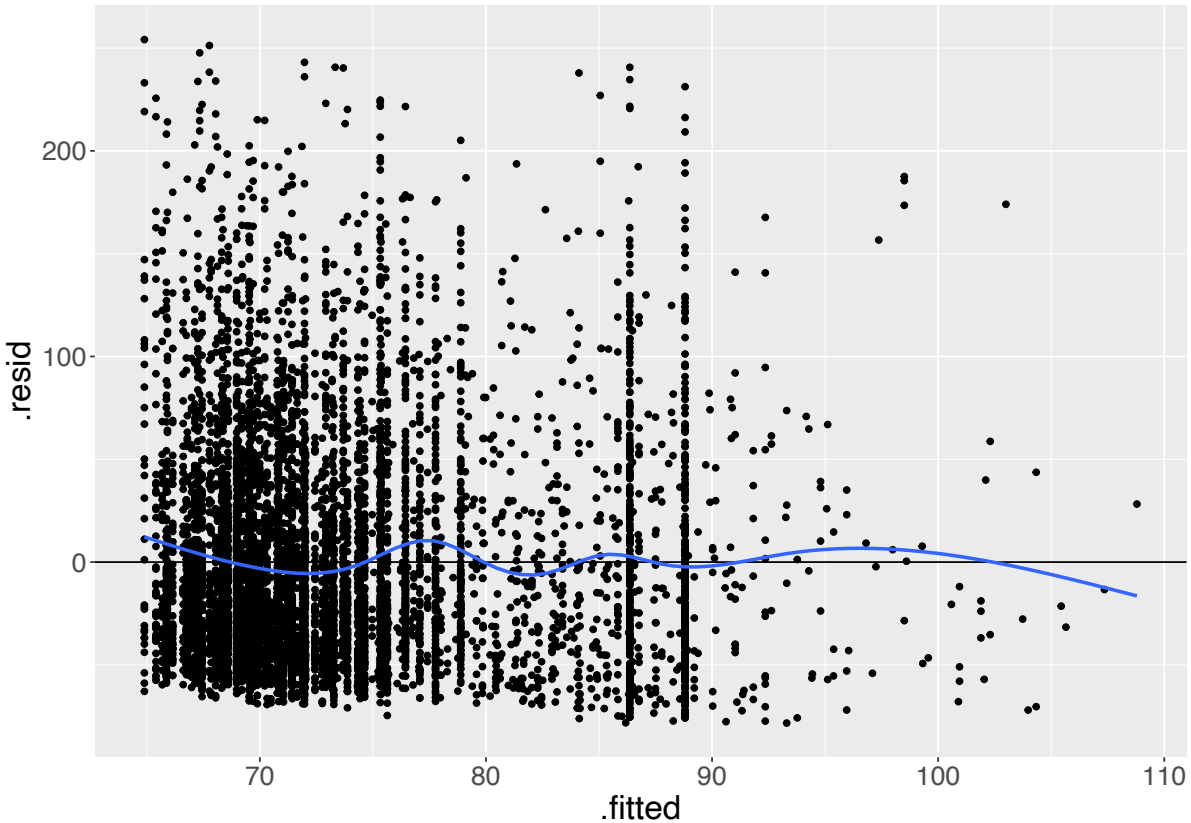


Figure 7: Evaluation of the homoscedasticity assumption for interval from waiting period to conception (wpc) regression.

Homoscedasticity (2)

```
library(car)
ncvTest(model1)
library(lmtest)
bptest(model1)
ggplot(model1, aes(.fitted, .resid)) +
  geom_hline(yintercept = 0) +
  geom_point() +
  geom_smooth(se = FALSE)
```

```
library(car)
ncvTest(model1)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 18.21011, Df = 1, p = 1.9783e-05
```

```
library(lmtest)
bptest(model1)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model1
## BP = 19.601, df = 7, p-value = 0.006499
```

Normal distribution of residuals (1)

Q-Q plot

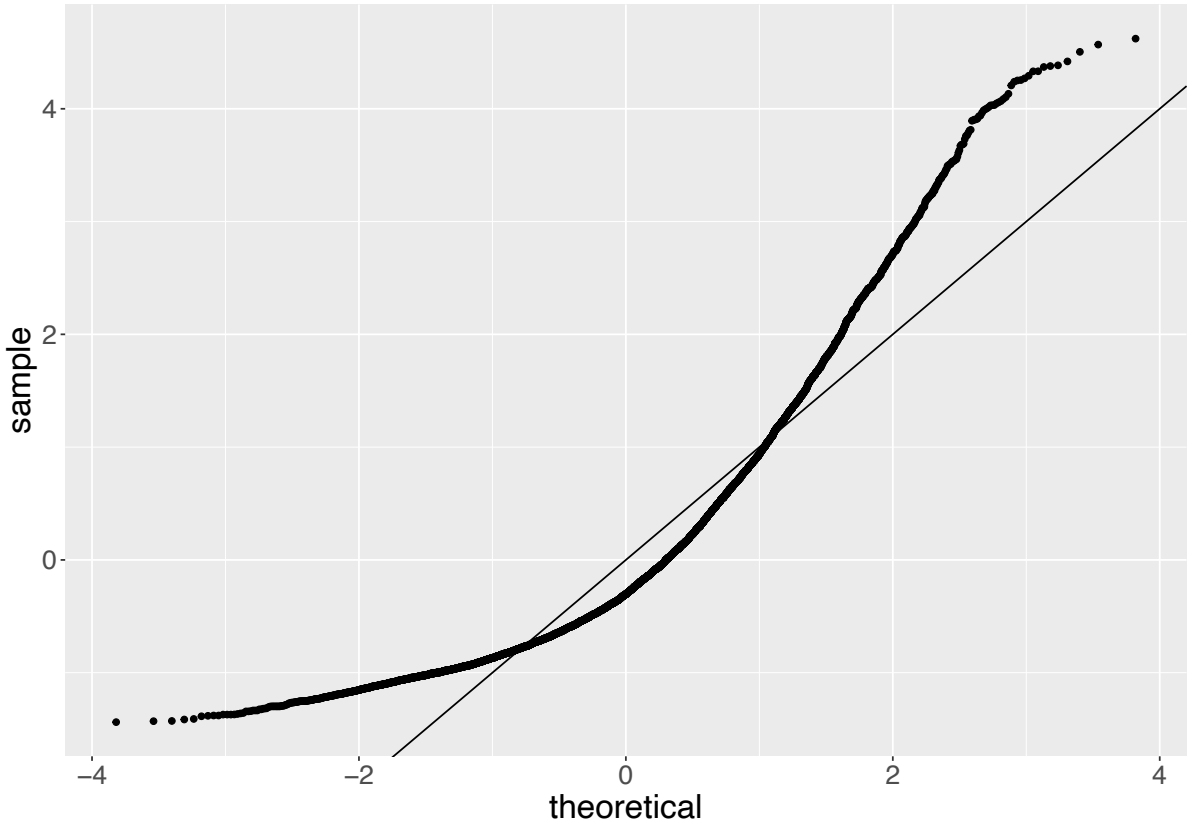


Figure 8: Evaluation of the normality of residuals assumption for interval from waiting period to conception (wpc) regression.

Normal distribution of residuals (2)

```
ggplot(model1, aes(sample = .stdresid)) +
  stat_qq() +
  theme(text = element_text(size = 20)) +
  geom_abline()
```

Shapiro-Wilk's test

p $<$ 0.05 indicates *non-normality*

```
shapiro.test(resid(model1)[0:5000])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(model1)[0:5000]
## W = 0.87175, p-value < 2.2e-16
```

Leverage

Given a residual $r_i = Y_i - \hat{Y}_i$ its variance is given as

$$\text{var}(r_i) = \sigma^2(1 - h_i),$$

with h_i called *leverage* of observation i. It indicates the potential of this observation to have a major impact on this model.

For a *simple regression*:

$$h_i = \frac{1}{n} + \frac{\left(X_i - \bar{X}\right)^2}{SSX}.$$

Leverage depends only on the predictor. Also for leverage holds

$$\frac{1}{n} < h_i < 1.$$

Residuals vs. leverage

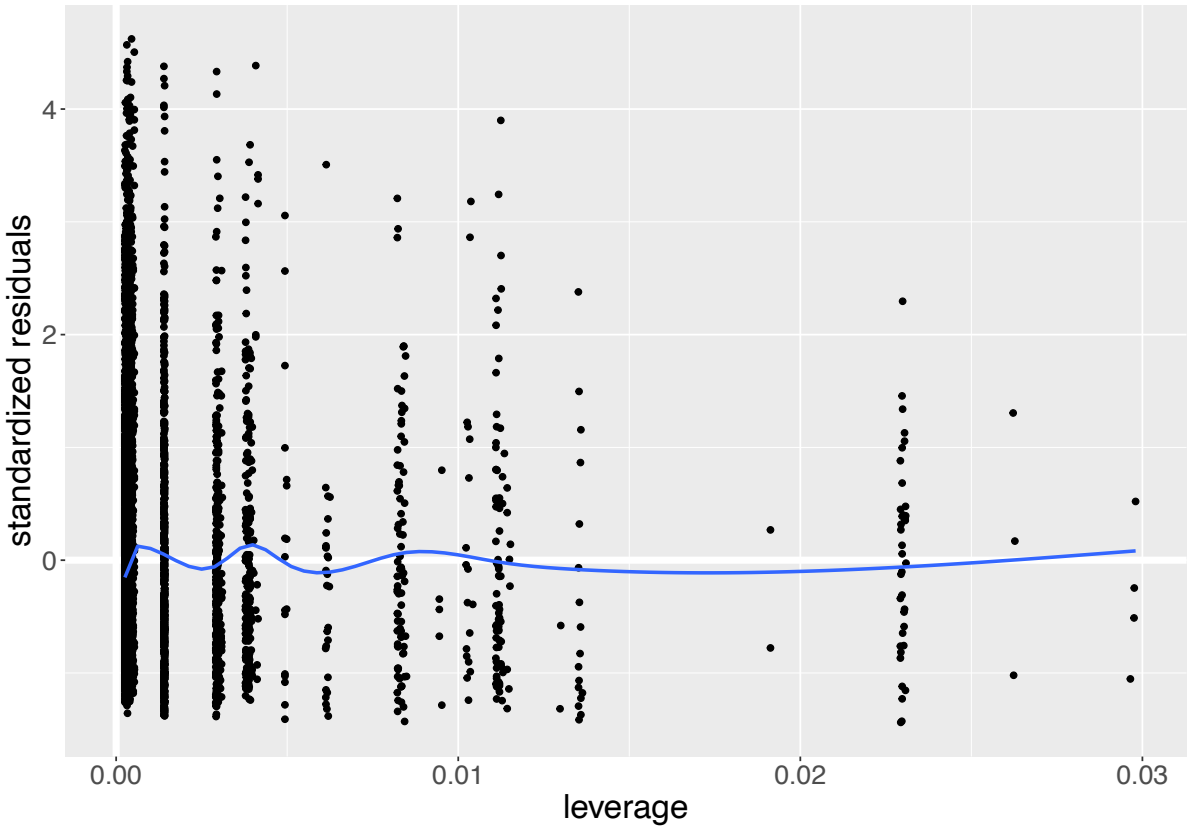


Figure 9: Residuals vs. leverage for interval from waiting period to conception (wpc) regression.

Residuals vs. leverage (2)

```
ggplot(model1, aes(.hat, .stdresid)) +
  geom_vline(size = 2, colour = "white", xintercept = 0) +
  geom_hline(size = 2, colour = "white", yintercept = 0) +
  geom_point() +
  geom_smooth(se = FALSE)
```

References

- Dohoo I, Martin W, Stryhn H, Veterinary Epidemiologic Research, 2d edition (2009)
- Kabacoff R, R in action, 2d edition (2015)
- Crawley M J, The R book (2007)

Modeldiagnostik

VB

Institut für Veterinär-Epidemiologie und Biometrie, FU Berlin

11/17/2020

Wiederholung

Modelldiagnostik

Unregelmäßigkeiten in den Daten oder falsche Angaben zu den Beziehungen zwischen den Prädiktoren und der Antwortvariablen können dazu führen, dass Sie sich für ein Modell entscheiden, das äußerst ungenau ist.

Einerseits können Sie daraus schließen, dass ein Prädiktor und eine Antwortvariable keine Beziehung zueinander haben, obwohl dies tatsächlich der Fall ist.

Auf der anderen Seite können Sie den Schluss ziehen, dass ein Prädiktor und eine Antwortvariable zusammenhängen, obwohl dies nicht der Fall ist!

Es kann auch vorkommen, dass Sie ein Modell erhalten, das in realen Umgebungen schlechte Vorhersagen mit erheblichen und unnötigen Fehlern macht.

Modelldiagnostik

Eine Reihe von Techniken, die als Regressionsdiagnose bezeichnet werden, bieten die erforderlichen Tools zur Bewertung der Angemessenheit des Regressionsmodells und können Ihnen dabei helfen, Probleme aufzudecken und zu beheben.

Modelldiagnostik (1)

- Normalverteilung von Residuen
- Unabhängigkeit
- Linearität
- Homoskedastizität
- Ausreißer
- Kollinearität

Normalverteilung von Residuen

Wenn die abhängige Variable für einen festen Satz von Prädiktorwerten normalverteilt ist, sollten die Residuenwerte normalverteilt sein (mit einem Mittelwert von 0).

Das Q-Q-Plot ist ein Wahrscheinlichkeitsdiagramm der standardisierten Residuen gegen die Werte die für die normalverteilten Residuen zu erwarten sind. Wenn Sie die Normalitätsannahme erfüllt haben, sollten die Punkte in diesem Diagramm auf der geraden 45-Grad-Linie liegen.

z-score

$$z = \frac{x - \bar{x}}{\hat{\sigma}}$$

Q-Q-Plot

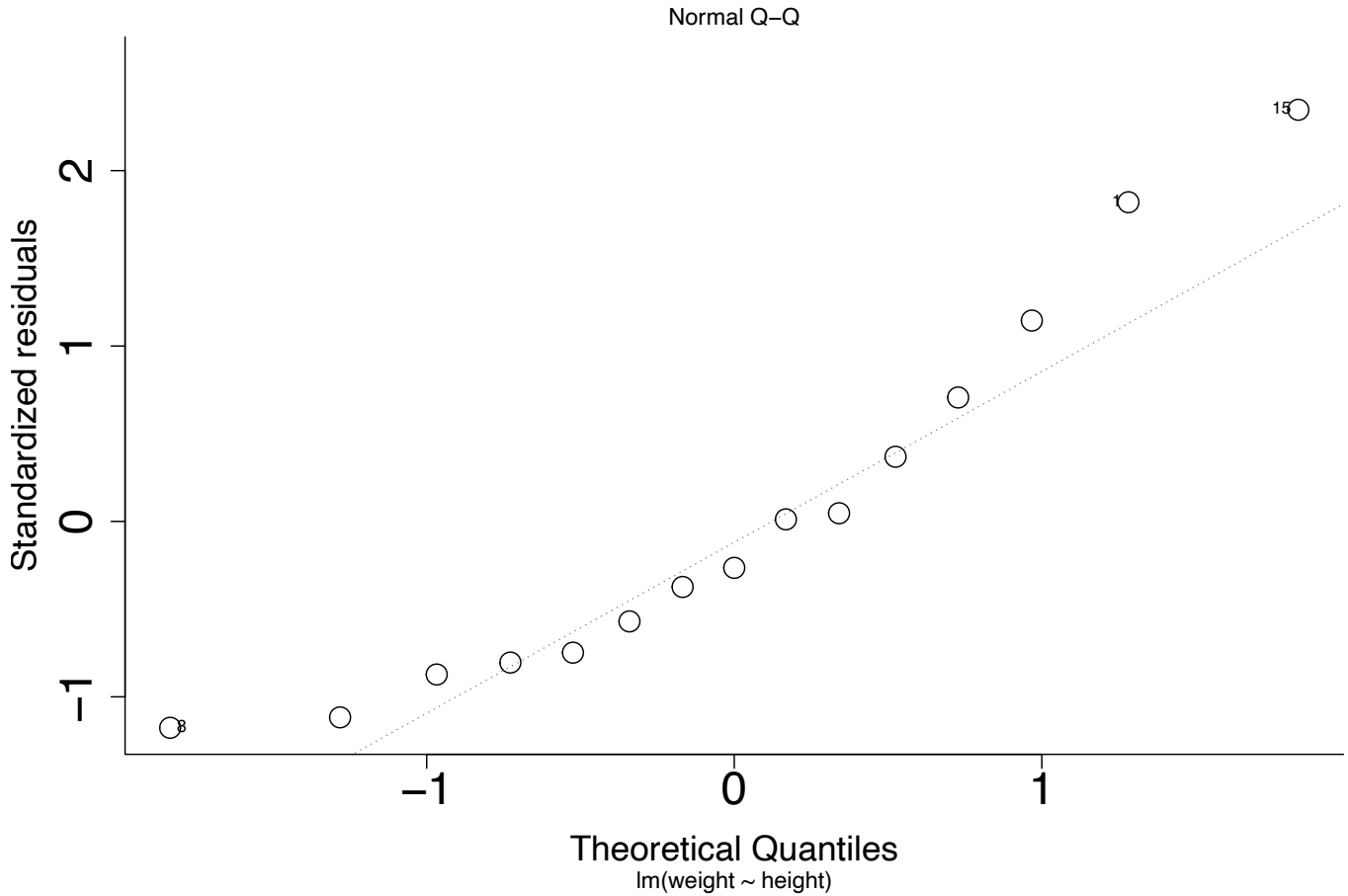


Figure 1: Modelldiagnostik für Gewicht vs Größe für Frauen

Quantil-Quantil-Diagramm (Q-Q plot)

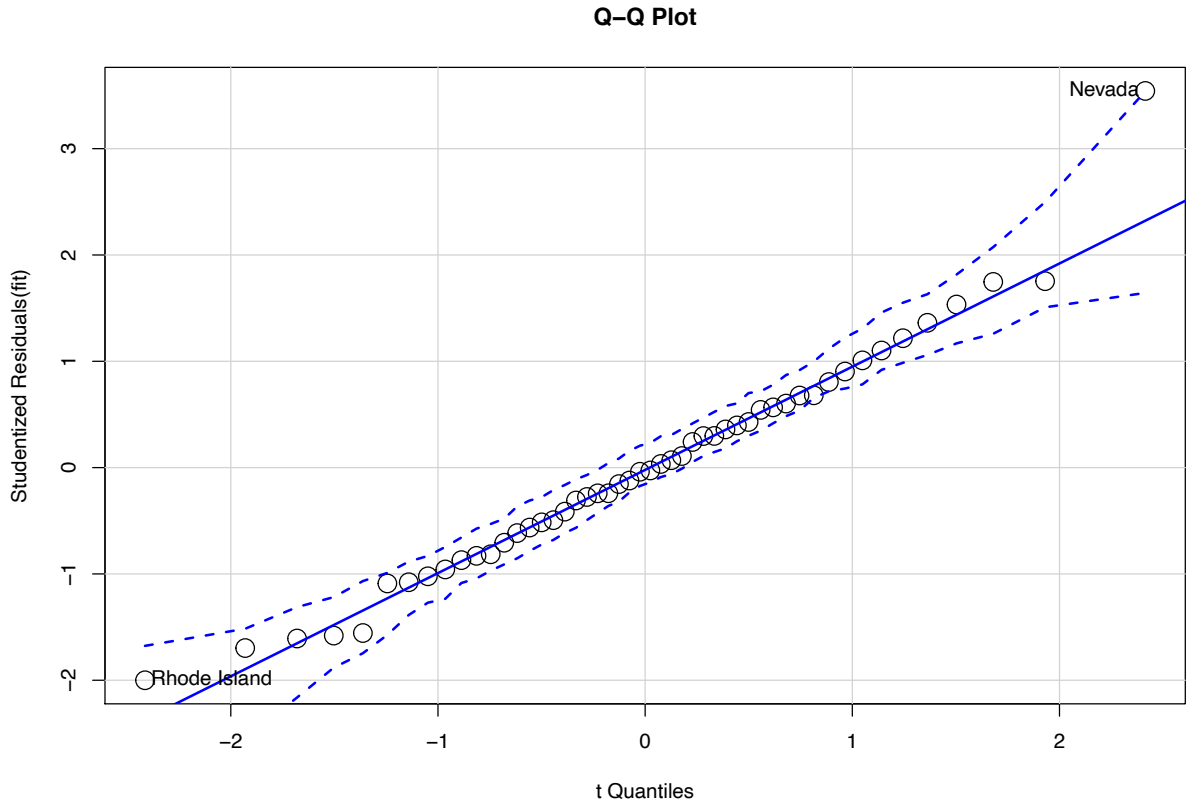


Figure 2: Modelldiagnostik auf Normalverteilung der Residuen

```
states["Nevada",]
```

```
##        Murder Population Illiteracy Income Frost
## Nevada   11.5        590        0.5   5149   188
```

```
fitted(fit)["Nevada"]
```

```
##   Nevada
## 3.878958
```

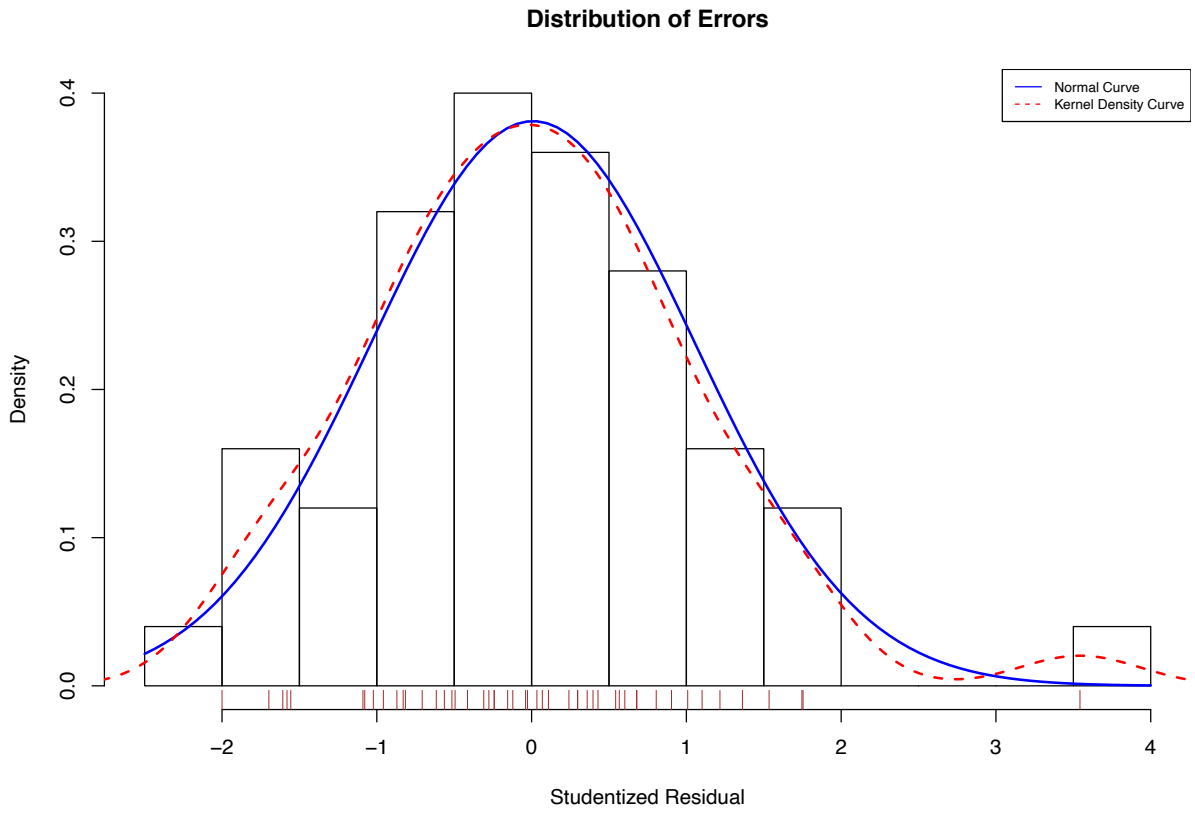


Figure 3: Modelldiagnostik auf Normalverteilung der Residuen

Quantil-Quantil-Diagramm (Q-Q plot) (1)

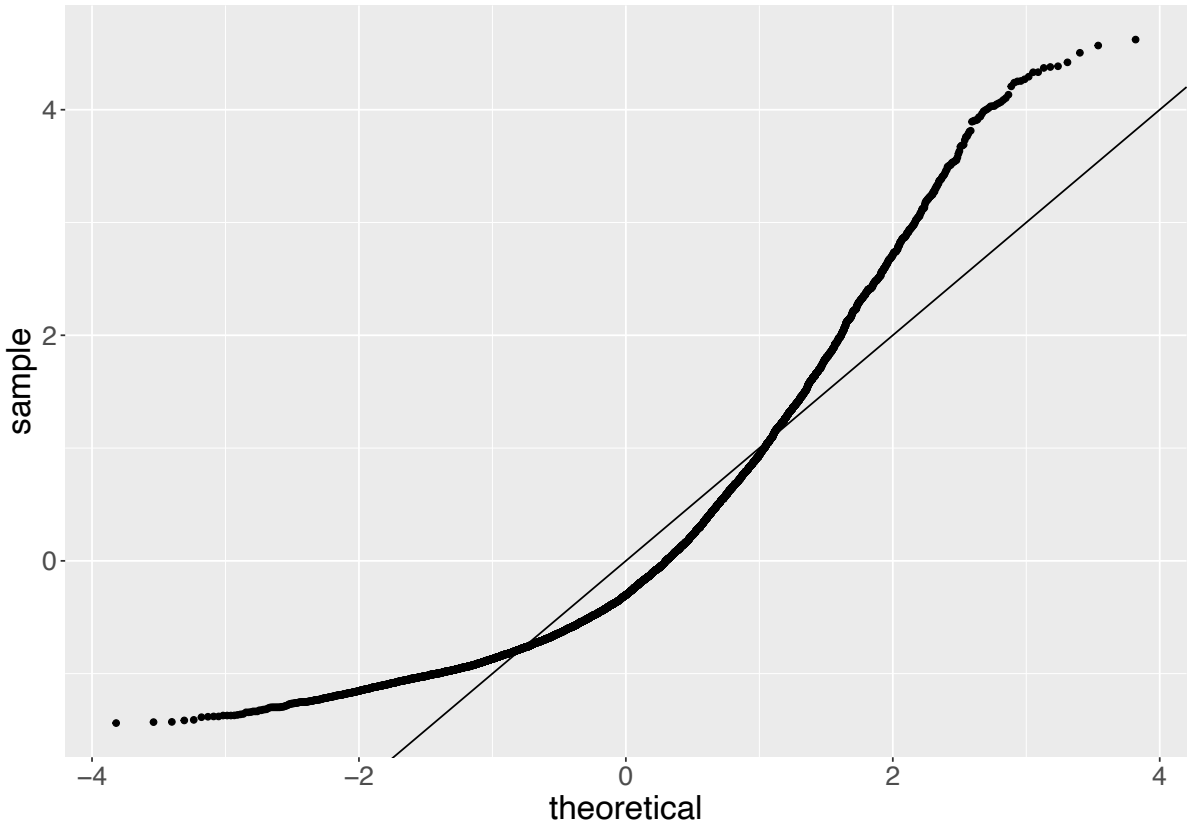


Figure 4: Evaluation of the normality of residuals assumption for interval from waiting period to conception (wpc) regression.

Normalverteilung von Residuen (1)

```
ggplot(model1, aes(sample = .stdresid)) +
  stat_qq() +
  theme(text = element_text(size = 20)) +
  geom_abline()
```

Shapiro-Wilk's test

p $<$ 0.05 indicates *non-normality*

```
states <- as.data.frame(state.x77[,c("Murder", "Population",
                        "Illiteracy", "Income", "Frost")])
fit <- lm(Murder ~ Population + Illiteracy + Income + Frost, data=states)
shapiro.test(resid(fit))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(fit)
## W = 0.98264, p-value = 0.6672
```

Unabhängigkeit

- ergibt sich in der Regel aus der Art der Daten
- mangelnde Unabhängigkeit, z. bei mehreren Beobachtungen eines einzelnen Tieres oder derselben Herde
- mangelnde Unabhängigkeit, wenn serielle Korrelationen wie Messungen während einer bestimmten Jahreszeit vorliegen

Unabhängigkeit (1)

Sie können nicht feststellen, ob die abhängigen Variablenwerte von diesen Darstellungen unabhängig sind. Sie müssen Ihr Verständnis dafür verwenden, wie die Daten gesammelt wurden. Es gibt keinen Grund zu der Annahme, dass das Gewicht einer Frau das Gewicht einer anderen Frau beeinflusst. Wenn Sie herausfinden, dass die Daten aus Familien stammen, müssen Sie möglicherweise Ihre Vermutung der Unabhängigkeit anpassen.

Durbin–Watson test

```
states <- as.data.frame(state.x77[,c("Murder", "Population",
                        "Illiteracy", "Income", "Frost")])
fit <- lm(Murder ~ Population + Illiteracy + Income + Frost, data=states)
durbinWatsonTest(fit)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1      -0.2006929      2.317691   0.222
##  Alternative hypothesis: rho != 0
```

Der nicht signifikante p-Wert (p = 0,288) deutet auf eine fehlende Autokorrelation und umgekehrt auf eine Unabhängigkeit von Fehlern hin. Der Verzögerungswert (in diesem Fall 1) gibt an, dass jede Beobachtung mit der im Datensatz daneben liegenden Beobachtung verglichen wird. Obwohl der Test für zeitabhängige Daten geeignet ist, gilt er weniger für Daten, die nicht auf diese Weise gruppiert wurden.

Linearität

Wenn die abhängige Variable linear mit den unabhängigen Variablen zusammenhängt, sollte es keine systematische Beziehung zwischen den Residuen und den vorhergesagten (dh angepassten) Werten geben. Mit anderen Worten, das Modell sollte alle in den Daten vorhandenen systematischen Abweichungen erfassen und nur zufälliges Rauschen zurücklassen.

Linearität (1)

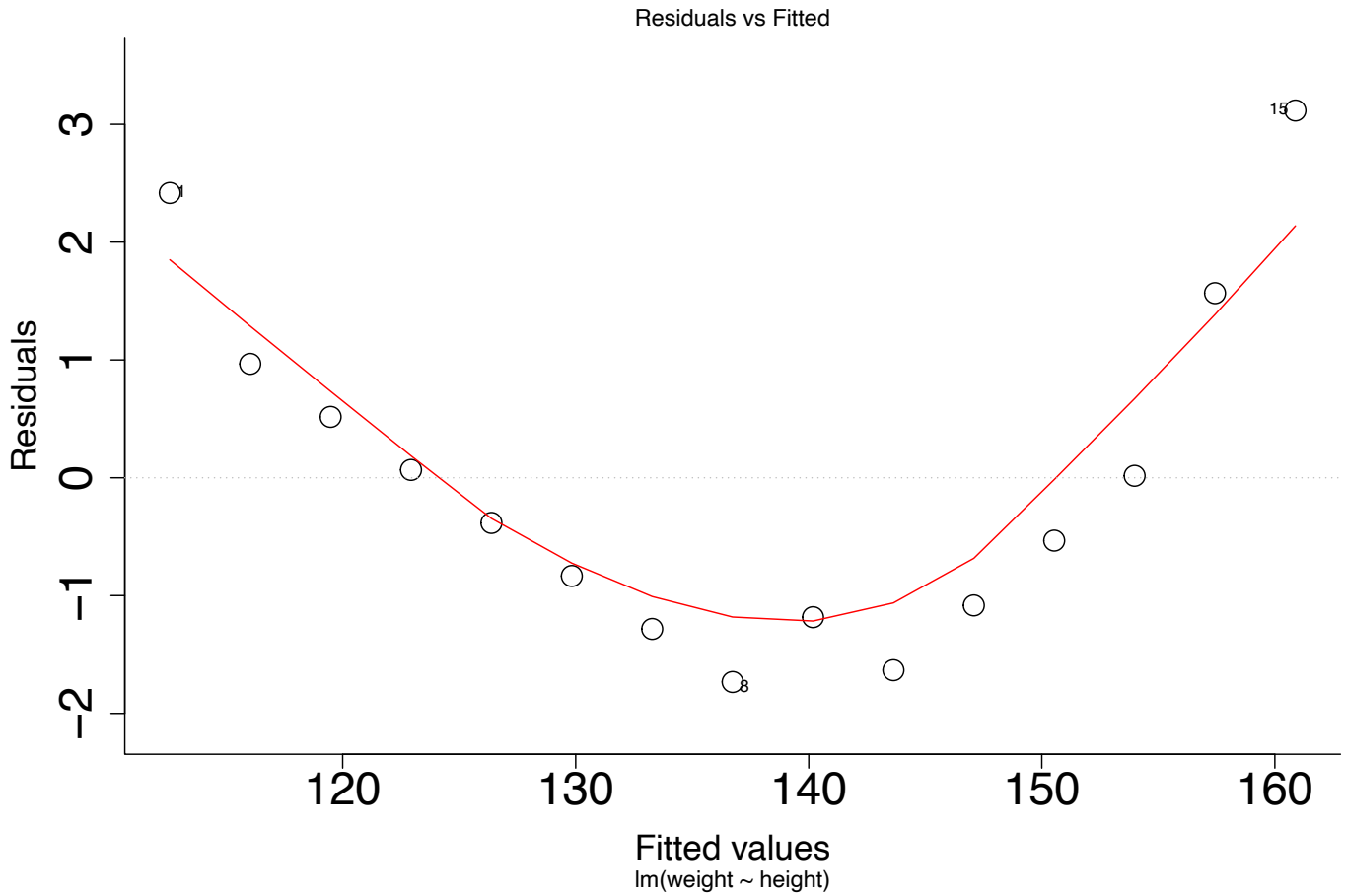


Figure 5: Modelldiagnostik auf Linearität für Gewicht vs Größe für Frauen

Linearität (2)

Im Diagramm Residuen vs. angepasste Werte sehen Sie deutliche Hinweise auf eine gekrümmte Beziehung, was darauf hindeutet, dass Sie der Regression möglicherweise einen quadratischen Term hinzufügen möchten.

Component plus residual plots

$$\text{Residuals} + \beta_i X_i \text{ versus } X_i$$

Die Nichtlinearität in einer dieser Darstellungen deutet darauf hin, dass Sie die funktionale Form dieses Prädiktors in der Regression möglicherweise nicht angemessen modelliert haben. In diesem Fall müssen Sie möglicherweise krummlinige Komponenten wie Polynome hinzufügen, eine oder mehrere Variablen transformieren (z. B. $\log(X)$ anstelle von X verwenden) oder die lineare Regression zugunsten einer anderen Regressionsvariante aufgeben.

Component plus residual plots (1)

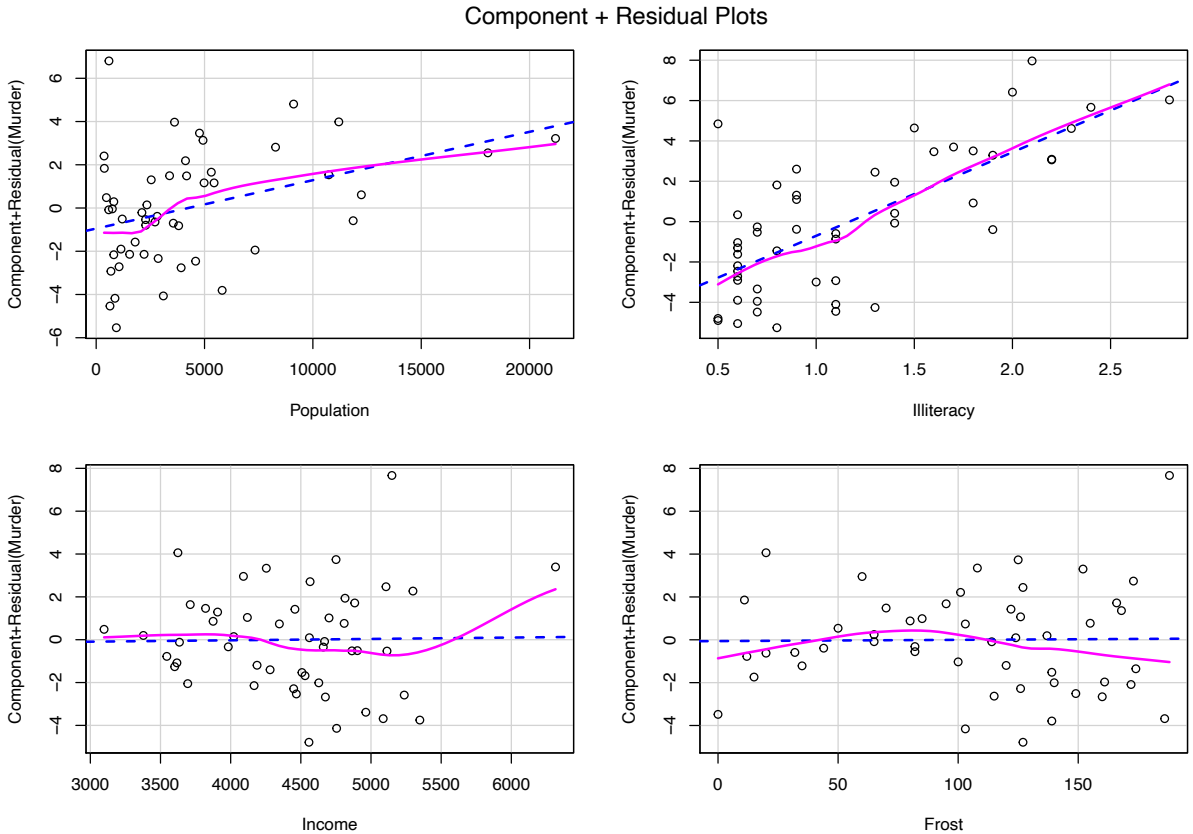


Figure 6: Modelldiagnostik auf Linearität

Homoskedastizität

Wenn Sie die Annahme einer konstanten Varianz erfüllt haben, sollten die Punkte im Diagramm "Skalenposition" (unten links) ein zufälliges Band um eine horizontale Linie sein. Sie scheinen diese Annahme zu erfüllen.

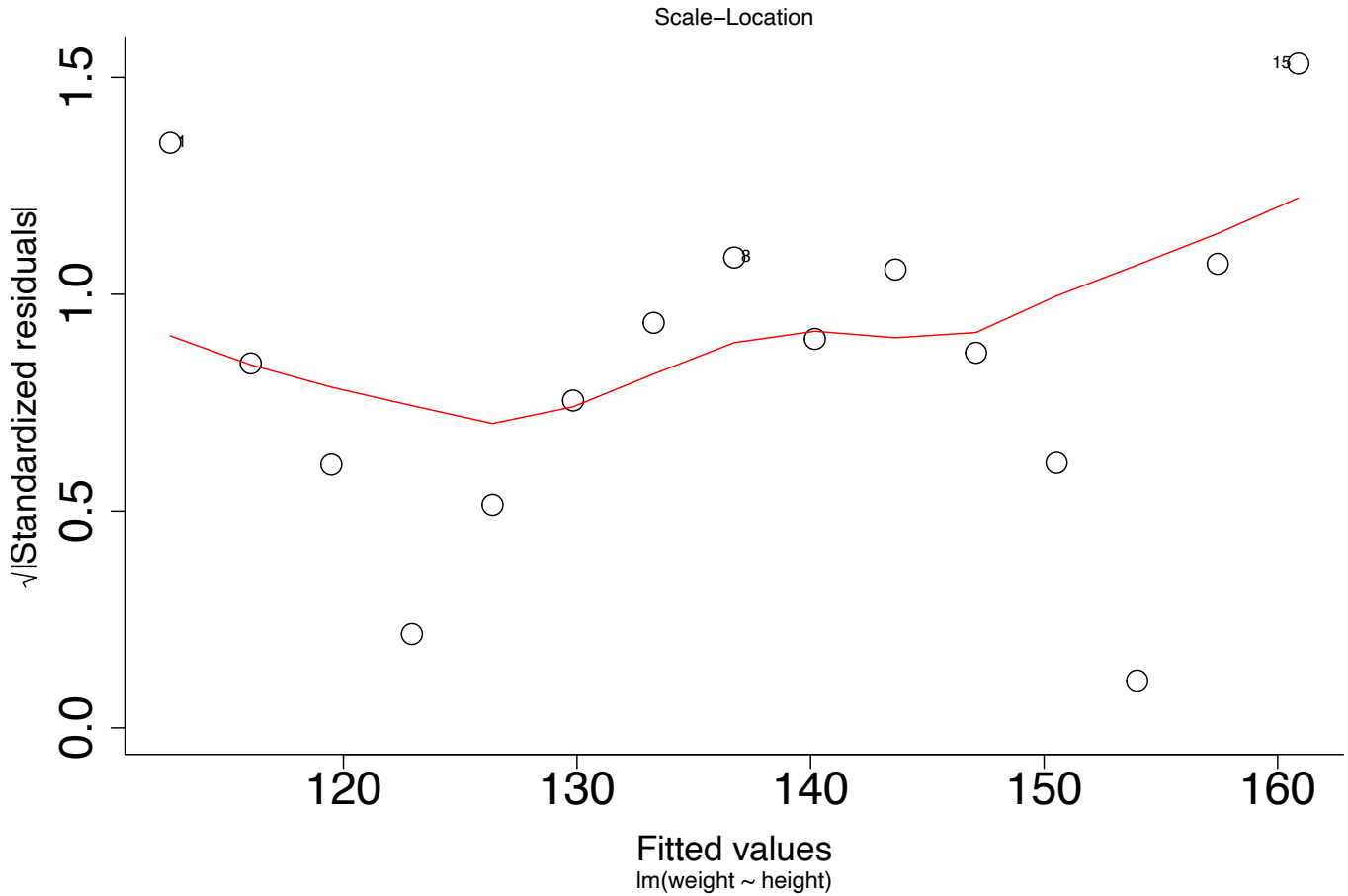


Figure 7: Modelldiagnostik auf Homoskedastizität für Gewicht vs Größe für Frauen

ncvTest()

Die Funktion ncvTest () testet die Hypothese der konstanten Fehlervarianz gegenüber der Alternative, dass sich die Fehlervarianz mit der Höhe der angepassten Werte ändert.

Ein signifikantes Ergebnis deutet auf eine Heteroskedastizität (nicht konstante Fehlervarianz) hin.

```
ncvTest(fit)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 0.8052115, Df = 1, p = 0.36954
```

Gesamtevaluation von den Annahmen des linearen Modells

```
library(gvlma)
gvmodel <- gvlma(fit)
summary(gvmodel)
```

```
##
## Call:
## lm(formula = weight ~ height, data = women)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7333 -1.1333 -0.3833  0.7417  3.1167
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -87.51667    5.93694  -14.74 1.71e-09 ***
## height        3.45000    0.09114   37.85 1.09e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.525 on 13 degrees of freedom
## Multiple R-squared:  0.991,  Adjusted R-squared:  0.9903
## F-statistic:  1433 on 1 and 13 DF,  p-value: 1.091e-14
##
##
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance =  0.05
##
## Call:
##  gvlma(x = fit)
##
##                       Value   p-value                   Decision
## Global Stat        16.5866 0.0023251 Assumptions NOT satisfied!
## Skewness            1.5577 0.2119999    Assumptions acceptable.
## Kurtosis            0.1019 0.7496131    Assumptions acceptable.
## Link Function      14.1218 0.0001713 Assumptions NOT satisfied!
## Heteroscedasticity  0.8052 0.3695398    Assumptions acceptable.
```

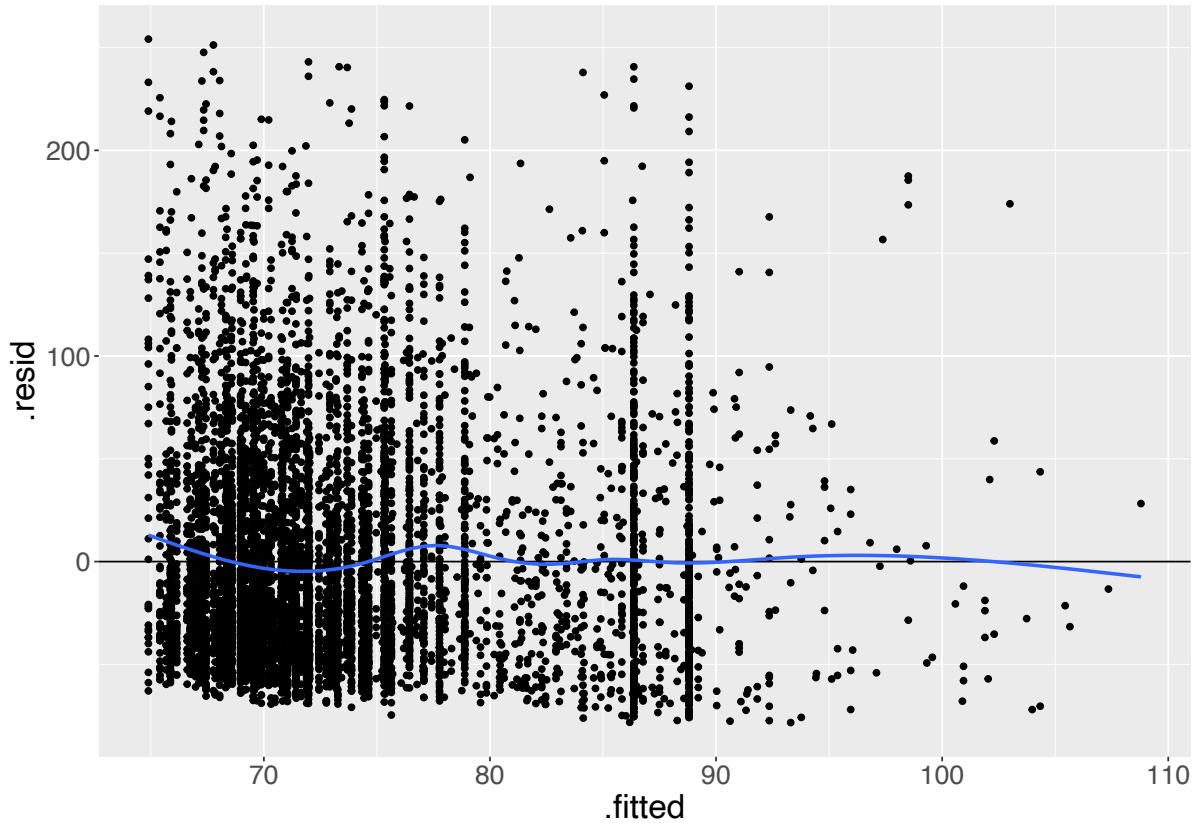


Figure 8: Evaluation of the homoscedasticity assumption for interval from waiting period to conception (wpc) regression.

Homoskedastizität (1)

```
library(car)
ncvTest(model1)
library(lmtest)
bptest(model1)
ggplot(model1, aes(.fitted, .resid)) +
  geom_hline(yintercept = 0) +
  geom_point() +
  geom_smooth(se = FALSE)
```

```
library(car)
ncvTest(model1)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 18.21011, Df = 1, p = 1.9783e-05
```

```
library(lmtest)
bptest(model1)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model1
## BP = 19.601, df = 7, p-value = 0.006499
```

Multikollinearität

Multikollinearität besteht, wenn zwei Variable korreliert sind (z.B. Geburdatum und Alter)

Dies führt zu großen Konfidenzintervallen für Modellparameter und erschwert die Interpretation einzelner Koeffizienten.

Multikollinearität kann mithilfe einer Statistik namens *Varianzinflationsfaktor* (VIF) erfasst werden.

Für jede Prädiktorvariable gibt die Quadratwurzel des VIF den Grad an, in dem das Konfidenzintervall für den Regressionsparameter dieser Variablen relativ zu einem Modell mit unkorrelierten Prädiktoren (daher der Name) erweitert wird.

Multikollinearität - variance inflation factor

```
states <- as.data.frame(state.x77[,c("Murder", "Population",
                        "Illiteracy", "Income", "Frost")])
fit <- lm(Murder ~ Population + Illiteracy + Income + Frost, data=states)
vif(fit)
```

```
## Population Illiteracy     Income      Frost
##   1.245282   2.165848   1.345822   2.082547
```

```
sqrt(vif(fit)) > 2
```

```
## Population Illiteracy     Income      Frost
##      FALSE      FALSE      FALSE      FALSE
```

Für jede Prädiktorvariable gibt die Quadratwurzel des VIF den Grad an, in dem das Konfidenzintervall für den Regressionsparameter dieser Variablen relativ zu einem Modell mit unkorrelierten Prädiktoren (daher der Name) erweitert wird.

$\sqrt{\text{VIF}} > 2$ indicates a multicollinearity problem.

Ausreißer

- Ein *Ausreißer* ist eine Beobachtung, die vom angepassten Regressionsmodell nicht gut vorhergesagt wird (d.h. ein großes positives oder negatives Residuum aufweist). Faustregel: standardisierte Residuen, die größer als 2 oder kleiner als –2 sind.
- Eine Beobachtung mit einem hohen *Hebelwert* (*leverage*) weist eine ungewöhnliche Kombination von Prädikatorwerten auf. Das heißt, es ist ein Ausreißer im Prädiktorbereich. Der Wert der abhängigen Variablen wird nicht zur Berechnung des Hebels einer Beobachtung verwendet.
- Eine *einflussreiche* (*influential*) Beobachtung ist eine Beobachtung, die einen unverhältnismäßigen Einfluss auf die Bestimmung der Modellparameter hat. Einflussreiche Beobachtungen werden anhand einer Statistik identifiziert, die als *Cooks-Distanz* oder *Cooks D* bezeichnet wird.

Ausreißer (1)

Ausreißer (2)

```
fit1 <- lm(weight ~ height, data=women)
outlierTest(fit1)
```

```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##    rstudent unadjusted p-value Bonferroni p
## 15 2.970125           0.011698      0.17548
```

Beispiel: ist polynomiale Regression besser?

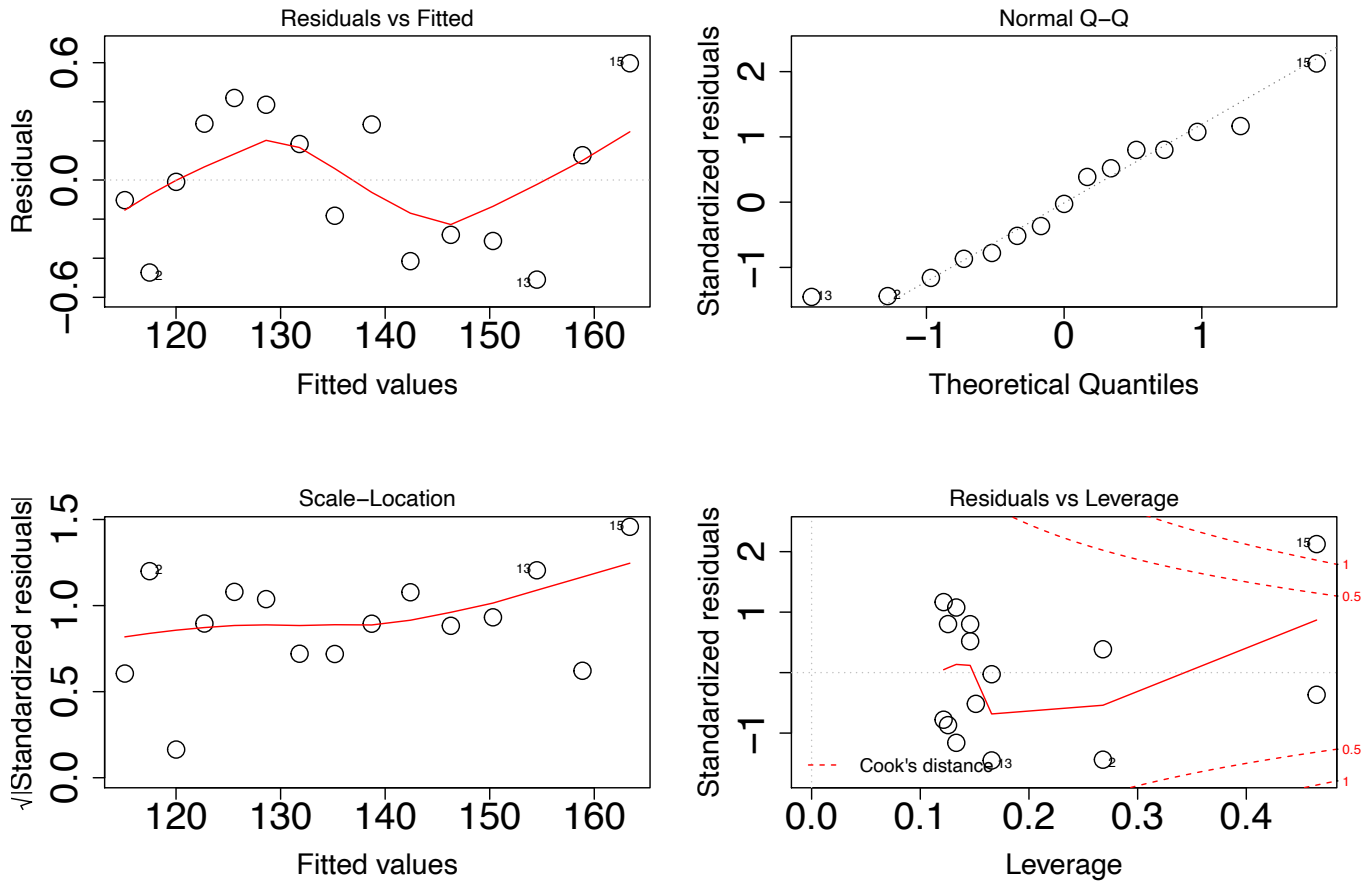


Figure 9: Modelldiagnostik für polynomiale Regression für Gewicht vs Größe für Frauen

Beispiel: Ausreißer rausnehmen.

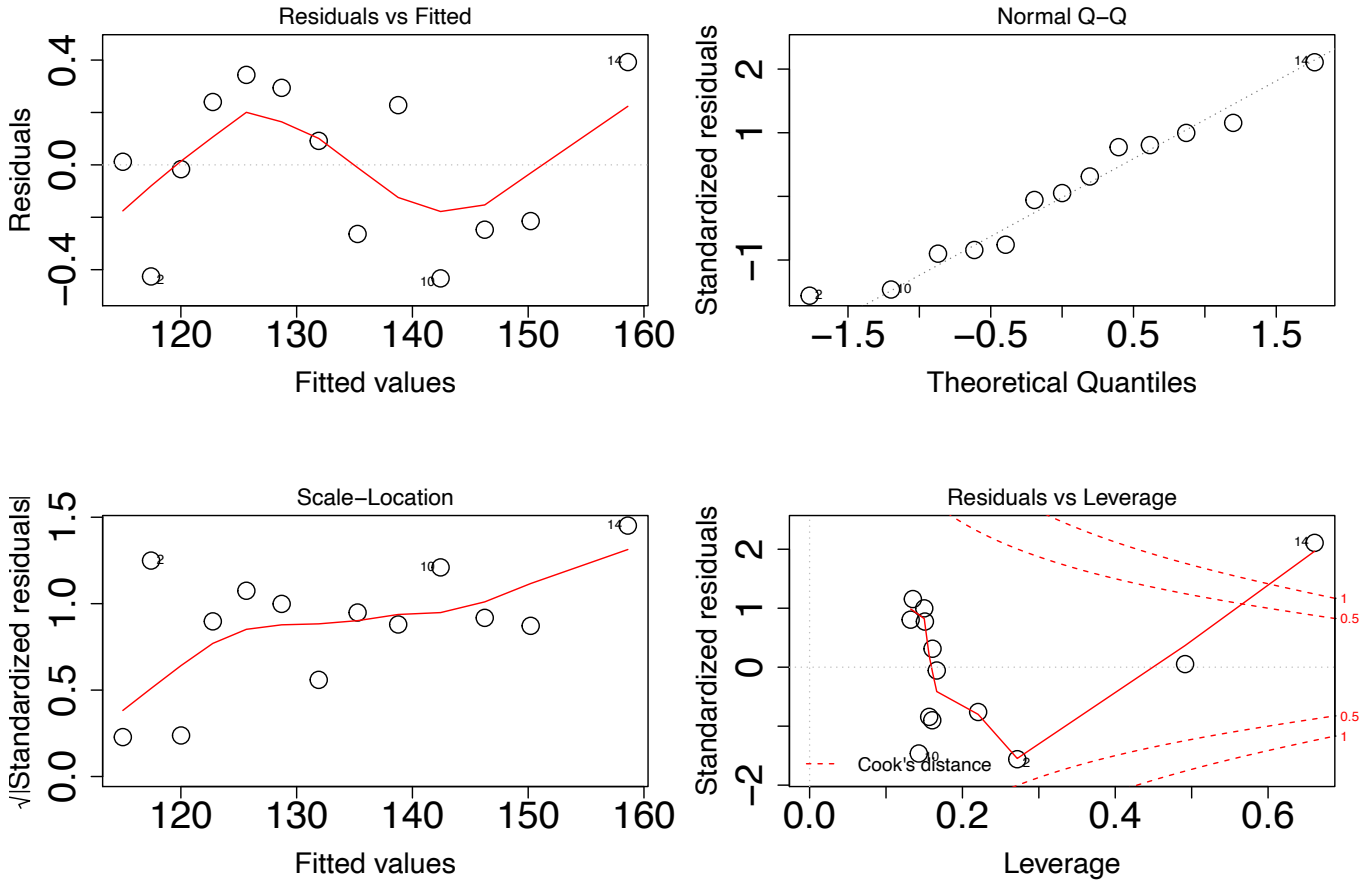


Figure 10: Modelldiagnostik für polynomiale Regression ohne 13. und 15. Messwerte für Gewicht vs Größe für Frauen

Man soll sehr vorsichtig beim Rausnehmen der Messwerte vorgehen. Das Modell soll an die Daten angepasst werden, nicht umgekehrt!

Beispiel 2

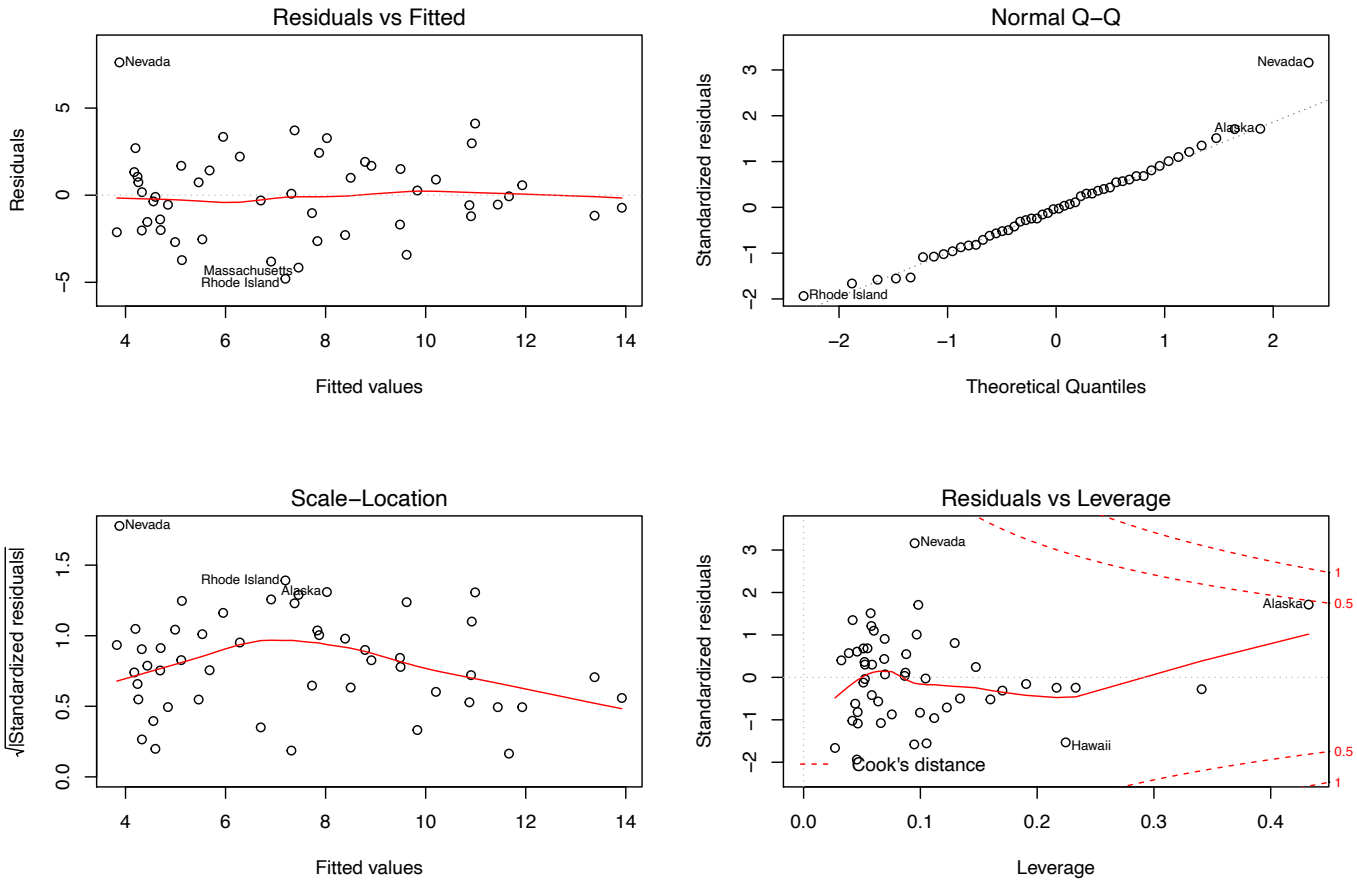


Figure 11: Modelldiagnostik für Tötungsrate

Hebelwert (Leverage)

Beobachtungen mit hoher Hebelwirkung werden durch die *hat-Statistik* identifiziert.

Für einen gegebenen Datensatz beträgt der durchschnittliche *hat*-Wert p/n, wobei p die Anzahl der im Modell geschätzten Parameter (einschließlich des Abschnitts) und n die Stichprobengröße ist.

Grob gesagt sollte eine Beobachtung mit einem Hutwert, der das Zwei- oder Dreifache des durchschnittlichen *hat*-Werts übersteigt, untersucht werden.

Beobachtungen mit hohem Hebel können einflussreiche Beobachtungen sein oder auch nicht. Das hängt davon ab, ob sie auch Ausreißer sind.

Hebelwert (Leverage) (1)

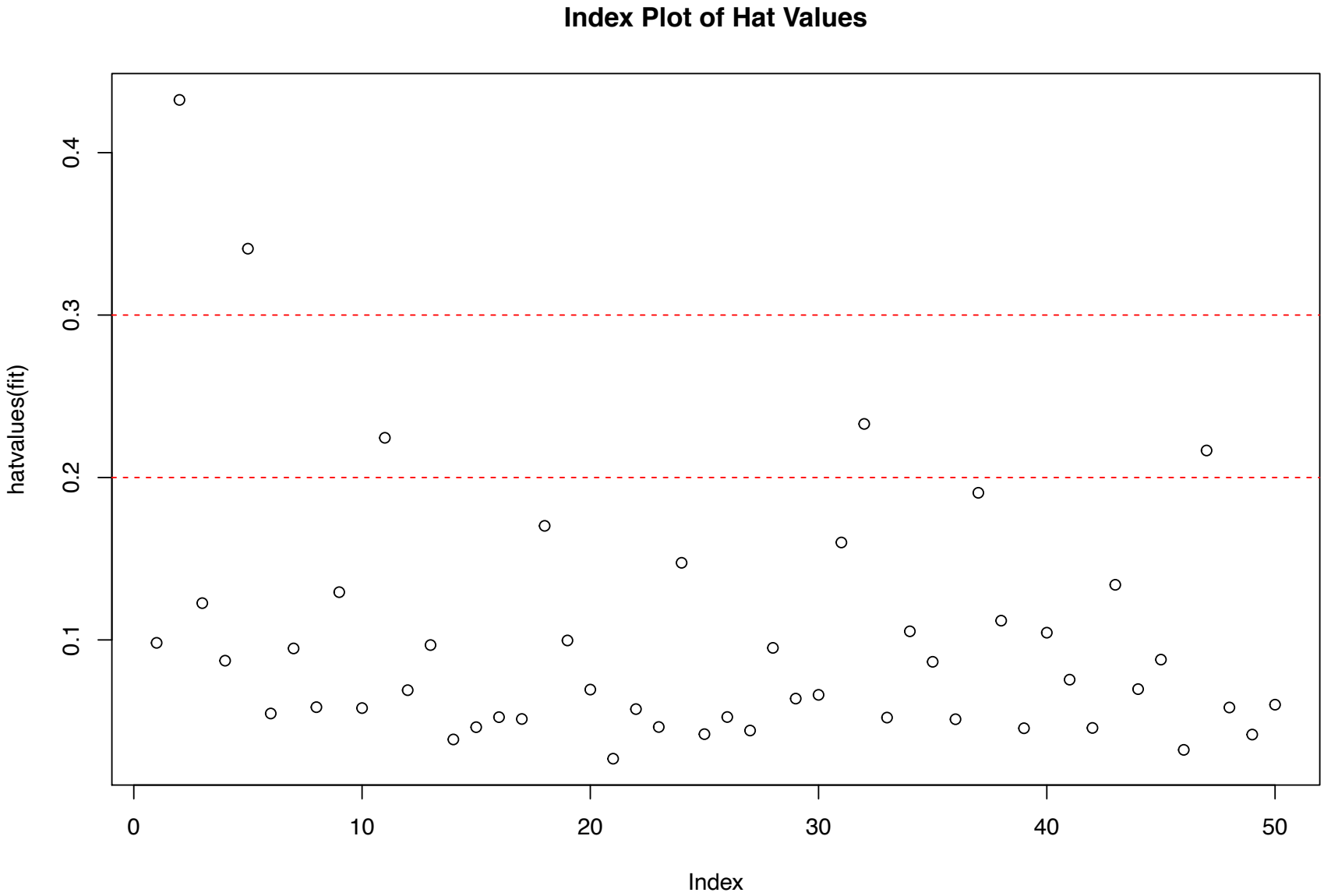


Figure 12: hat-Statistik.

Einflüßreicher Messwerte: Cooks D-Wert

Cooks D-Werte von mehr als $4/(n-k-1)$, wobei n die Stichprobengröße und k die Anzahl der Prädiktorvariablen ist, weisen auf einflussreiche Beobachtungen hin.

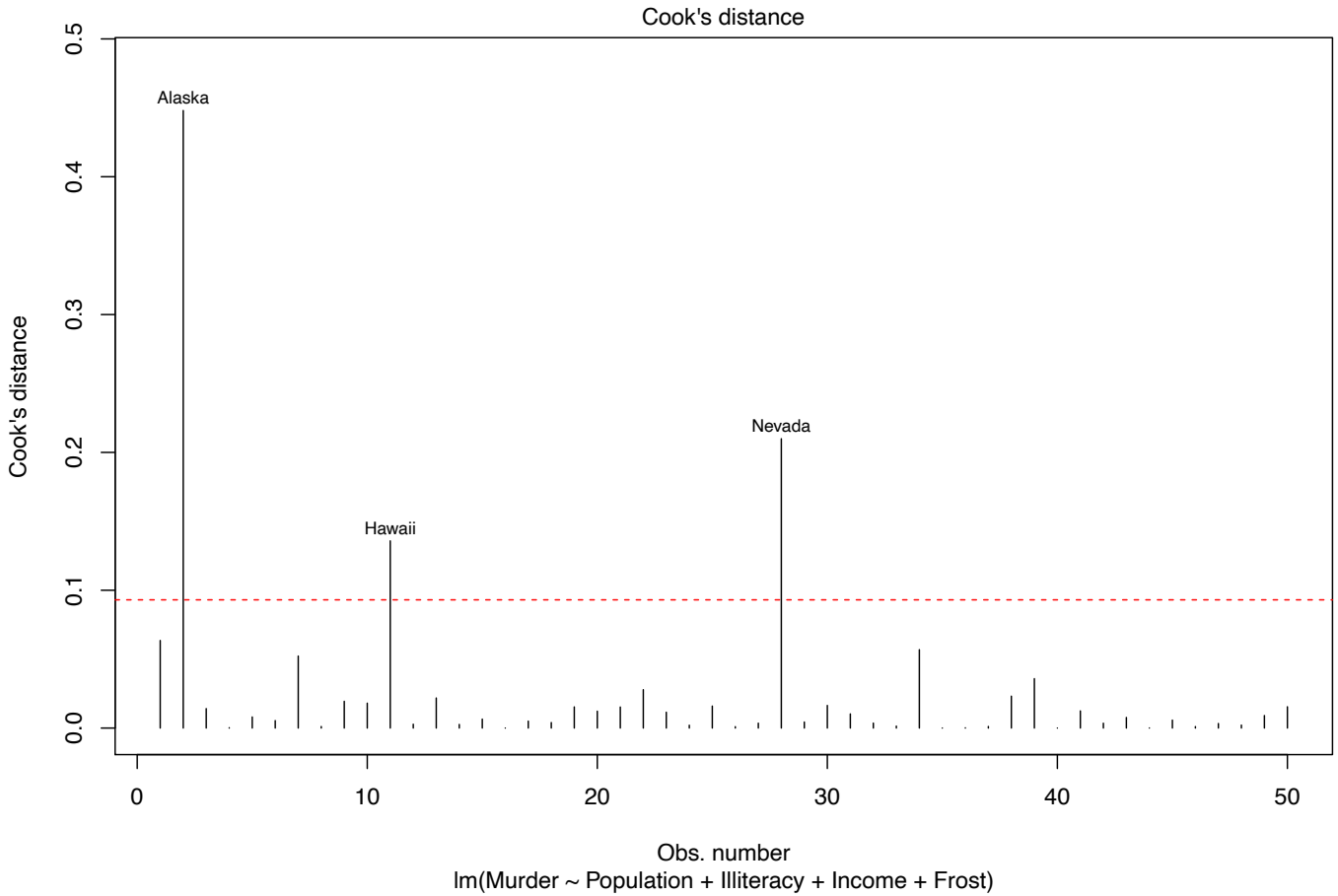


Figure 13: Cooks D-Werte.

Einflüßreicher Messwerte: Added-variable plots

Cooks D-Diagramme können helfen, einflussreiche Beobachtungen zu identifizieren, liefern jedoch keine Informationen darüber, wie sich diese Beobachtungen auf das Modell auswirken.

Für eine Antwortvariable und k Prädiktorvariablen erstellen man k *Added-variable plots* wie folgt.

Für jeden Prädiktor X_k die Residuen aus der Regression der Antwortvariablen auf den anderen $k-1$ Prädiktoren werden gegen die Residuen aus der Regression von X_k auf den anderen $k-1$ Prädiktoren dargestellt.

Einflüßreicher Messwerte: Added-variable plots (1)

```
avPlots(fit, ask=FALSE, id.method="identify")
```

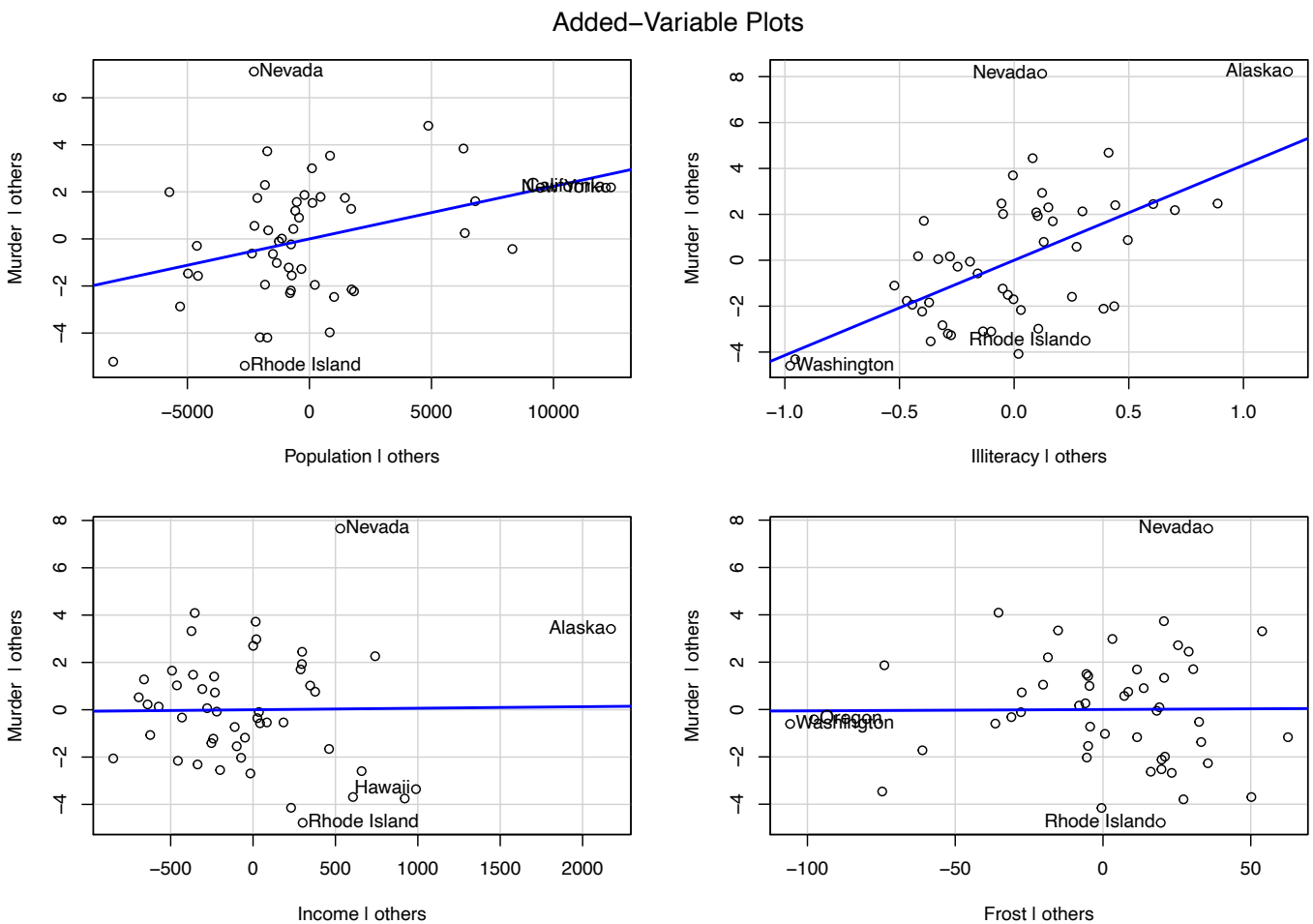


Figure 14: Added-variable plots zur Bewertung der Auswirkung einflussreicher Beobachtungen.

influencePlot()

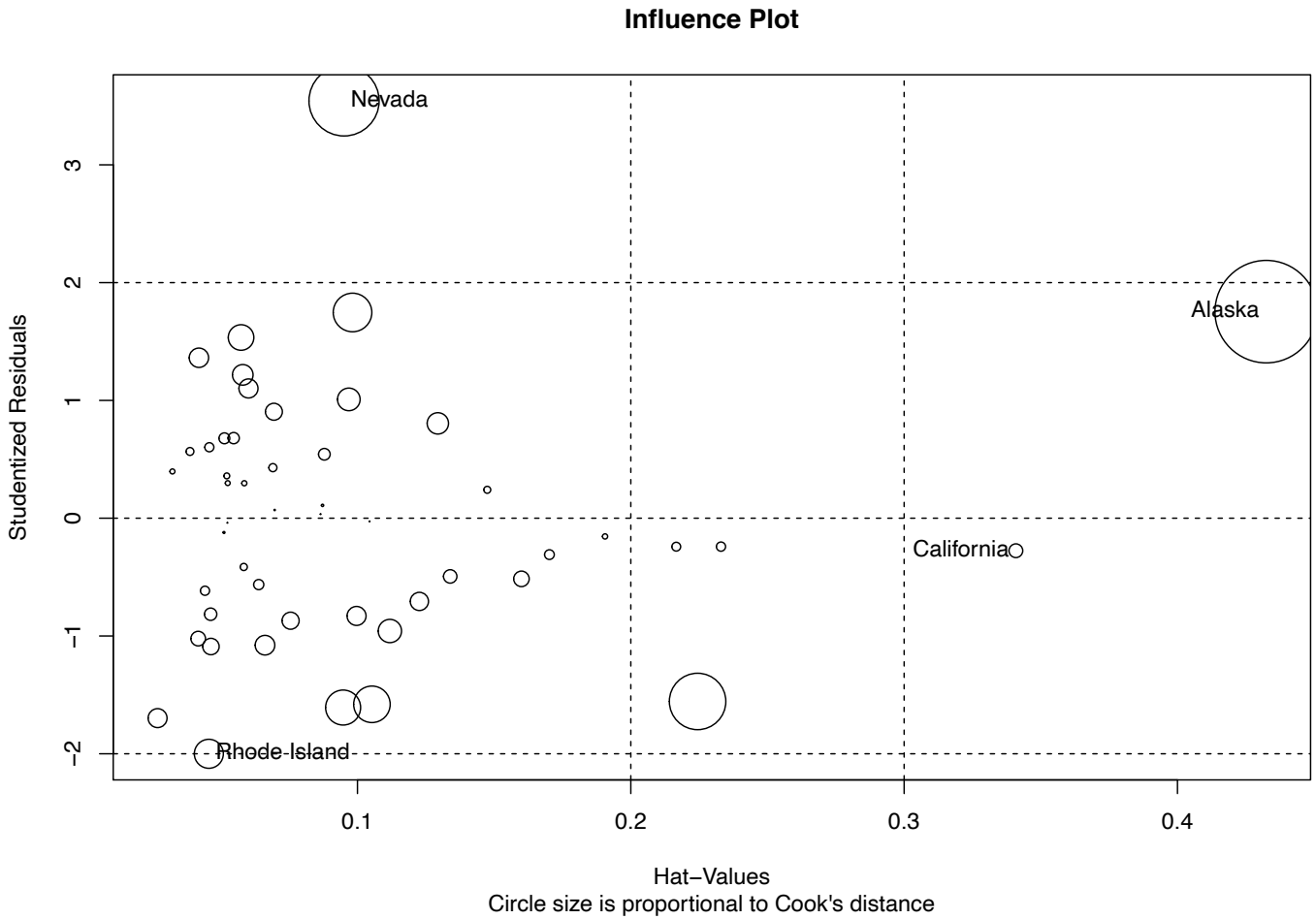


Figure 15: Outliers, leverage and influence zusammen.

influencePlot() (1)

```
library(car)
influencePlot(fit, id.method="identify", main="Influence Plot",
              sub="Circle size is proportional to Cook's distance")
```

```
##                  StudRes        Hat       CookD
## Alaska         1.7536917 0.43247319 0.448050997
## California    -0.2761492 0.34087628 0.008052956
## Nevada         3.5429286 0.09508977 0.209915743
## Rhode Island -2.0001631 0.04562377 0.035858963
#
```

ANOVA (Einfache)

Vitaly Belik

Institut für Veterinär-Epidemiologie und Biometrie, FU Berlin

24/11/2020

Wiederholung

ANOVA

Falls die Antwortvariable (*response*) kontinuierlich ist, und die unabhängigen Variablen (*Faktoren*) kategoriell, macht es Sinn die Unterschiede zwischen den verschieden Gruppen von Faktorenniveaus oder Faktorenabstufugen zu betrachten.

Die statistischen Techniken für den Vergleich zwischen den Mittelwerten mehreren Gruppen bezeichnet man als **ANOVA**.

ANOVA kann und soll als GLM (veralgemeinerte *Regression*) betrachtet werden.

Terminologie

- Einfache ANOVA: eine erklärende (kategorielle) Variable (Faktor). Vergleiche zwischen den Niveaus von diesem Faktor.
- Zweifache (faktorielle) ANOVA: zwei erklärende (kategorielle) Variable (Faktor). Vergleiche zwischen den Niveaus von diesen Faktoren.
- Multiple ANOAVA – **MANOVA**: mehrere unabhägige Variable.
- ANCOVA (Analysis of covariances): zusätzlich zur ANOVA-Situtation kommt noch eine Kovariate für die abhängige Variable hinzu.
- Repeated measures (Messwiederholungen): bei den gleichen Objekten wurden die Messungen mehrmals vorgenommen.

ANOVA Tabelle

$$\hat{Y} = \beta_0 + \beta_1 X$$

Um zu beurteilen, wie viel Information die Variable X über die Variable Y enthält, betrachten wir, wie viel von der *Summe der Quadrate* (SS) der Variablen Y durch die Kenntnis der Variablen X erklärt werden könnte.

Table 1: Zerlegung von Quadratsummen im Regressionsmodell mit k Prädiktorvariablen [Dohoo, p. 327]

Source of variation	Sum of squares	Degrees of freedom	Mean square	F-test
Model	$SSM = \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2$	$dfM = k$	$MSM = \frac{SSM}{dfM}$	$\frac{MSM}{MSE}$
Error	$SSE = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2$	$dE = n - (k+1)$	$MSE = \frac{SSE}{dfE}$	
Total	$SST = \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2$	$dfT = n - 1$	$MST = \frac{SST}{dfT}$	

$MSE = \sigma^2$ and σ is called *standard error of prediction*.

F-Test zur Beurteilung der Modellgüte

H_0: $\beta_1 = \beta_2 ... \beta_k = 0$, $\beta_0 \neq 0$

H_1: zumindest manche von Koeffizienten $\beta_i \neq 0$, $i \neq 0$

- Variablen werden so asugewählt, dass die F-Statistik maximiert wird
- F-Test hat eine einfache Bedeutung, wenn unabhängige Variablen in einem kontrollierten Experiment manipuliert werden
- Vorsicht wird bei Beobachtungsvariablen geboten (beeinflusst von der Anzahl der Variablen, ihren Korrelationen und der Stichprobengröße)

Einfache ANOVA

ANOVA Beispiel

```
df <- read.table('/Users/vitaly/Lehre/WS1920/Beuth/Data files/Viagra.dat', sep = '', header = TRUE)
knitr::kable(df, caption = "Viagra-Datensatz [A. Field]")
```

Table 2: Viagra-Datensatz [A. Field]

person	dose	libido
1	1	3
2	1	2
3	1	1
4	1	1
5	1	4
6	2	5
7	2	2
8	2	4
9	2	2
10	2	3
11	3	7
12	3	4
13	3	5
14	3	3
15	3	6

ANOVA Beispiel (1)

```
library(gplots)
  plotmeans(df$libido ~ df$dose)
```

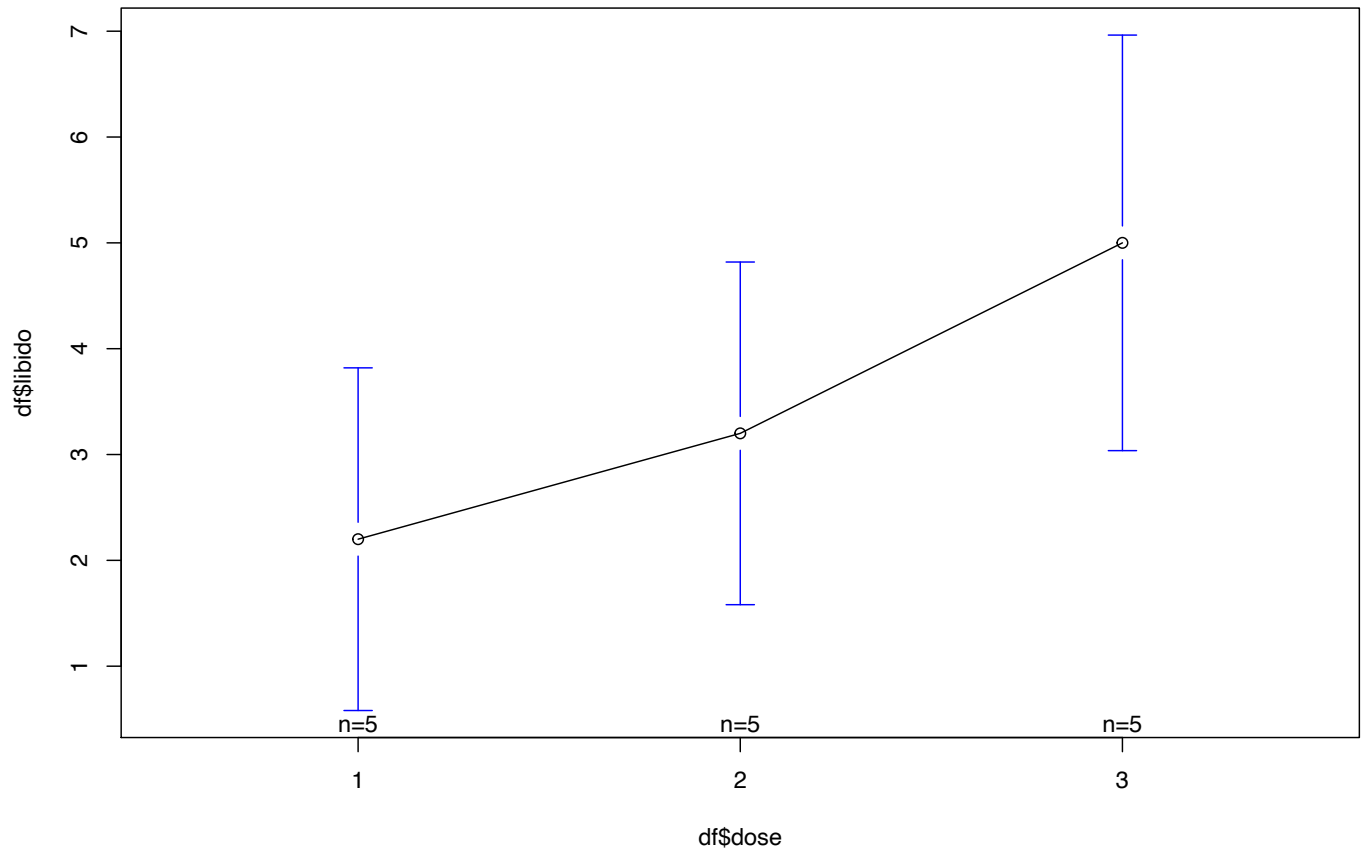


Figure 1: Viagra-Datensatz mit Konfidenzintervallen

ANOVA Beispiel (2)

```
library(pastecs)
by(df$libido, df$dose, stat.desc)
```

```
## df$dose: 1
##      nbr.val     nbr.null       nbr.na          min          max        range
##    5.0000000    0.0000000    0.0000000    1.0000000    4.0000000    3.0000000
##          sum       median         mean      SE.mean CI.mean.0.95          var
##   11.0000000    2.0000000    2.2000000    0.5830952    1.6189318    1.7000000
##      std.dev     coef.var
##    1.3038405    0.5926548
## ------------------------------------------------------------
## df$dose: 2
##      nbr.val     nbr.null       nbr.na          min          max        range
##    5.0000000    0.0000000    0.0000000    2.0000000    5.0000000    3.0000000
##          sum       median         mean      SE.mean CI.mean.0.95          var
##   16.0000000    3.0000000    3.2000000    0.5830952    1.6189318    1.7000000
##      std.dev     coef.var
##    1.3038405    0.4074502
## ------------------------------------------------------------
## df$dose: 3
##      nbr.val     nbr.null       nbr.na          min          max        range
##    5.0000000    0.0000000    0.0000000    3.0000000    7.0000000    4.0000000
##          sum       median         mean      SE.mean CI.mean.0.95          var
##   25.0000000    5.0000000    5.0000000    0.7071068    1.9632432    2.5000000
##      std.dev     coef.var
##    1.5811388    0.3162278
```

ANOVA Beispiel (Levene's test)

```
library(pastecs)
library(car)
leveneTest(df$libido, df$dose, center = median)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  2  0.1176   0.89
##       12
```

If the *Levene's test* is non-significant, the data variances are very similar.

Falls nicht signifikant, kann man *Welch's F* oder robuste ANOVA benutzen.

ANOVA Beispiel (2)

$$\text{libido}_i = \text{dose}_i + \text{error}$$

ANOVA Beispiel (*lm*)

```
df <- read.table('/Users/vitaly/Lehre/WS1920/Beuth/Data files/Viagra.dat', sep = '', header = TRUE)
df$dose <- as.factor(df$dose)
library(ggplot2)
viagraModel <- lm(libido ~ dose, data = df)
summary(viagraModel)
```

```
##
## Call:
## lm(formula = libido ~ dose, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##   -2.0   -1.2   -0.2    0.9    2.0
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.2000     0.6272   3.508  0.00432 **
## dose2         1.0000     0.8869   1.127  0.28158
## dose3         2.8000     0.8869   3.157  0.00827 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.402 on 12 degrees of freedom
## Multiple R-squared:  0.4604, Adjusted R-squared:  0.3704
## F-statistic: 5.119 on 2 and 12 DF,  p-value: 0.02469
```

ANOVA Beispiel (*aov*)

```
ViagraModel <- aov(libido ~ dose, data = df)
summary(ViagraModel)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## dose         2  20.13  10.067   5.119 0.0247 *
## Residuals   12  23.60   1.967
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

F-Test prüft, ob Unterschiede zwischen den Gruppen vorliegen.

F-Test sagt uns nicht zwischen welchen Gruppen die Unterschiede vorliegen.

ANOVA Beispiel (*aov* 1)

```
library(car)
qqPlot(lm(libido ~ dose, data = df), simulate = TRUE, main = 'Q-Q Plot', labels = FALSE)
```

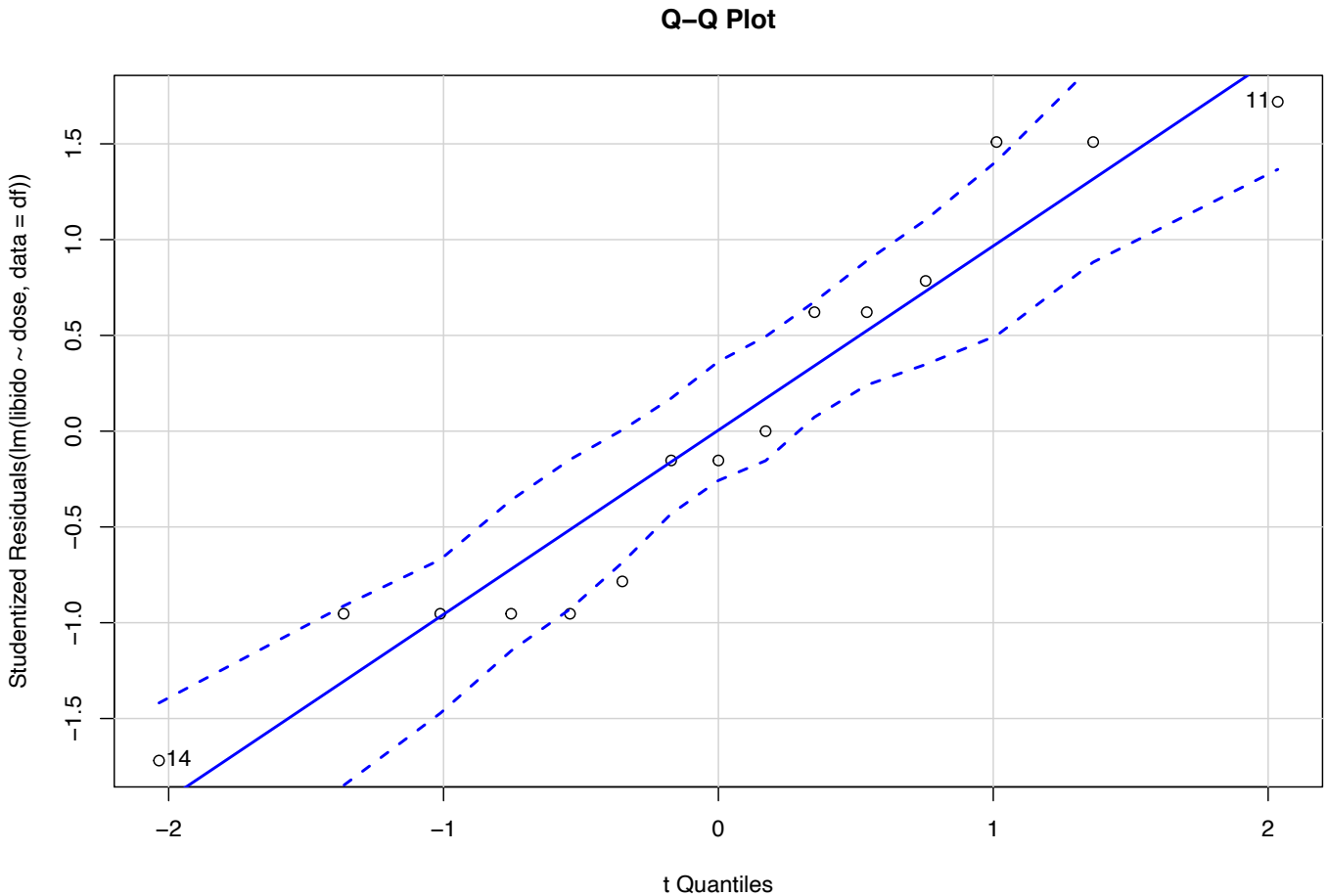


Figure 2: Q-Q Plot.

```
## [1] 11 14
```

ANOVA Beispiel: Bartlett-Test

```
bartlett.test(libido ~ dose, data = df)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  libido by dose
## Bartlett's K-squared = 0.1853, df = 2, p-value = 0.9115
```

ANOVA Beispiel: Aussreißer

```
outlierTest(ViagraModel)
```

```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##    rstudent unadjusted p-value Bonferroni p
## 11  1.71959            0.11348           NA
```

ANOVA Beispiel (*aov* 3)

```
par(mfrow=c(2,2))
plot(ViagraModel)
```

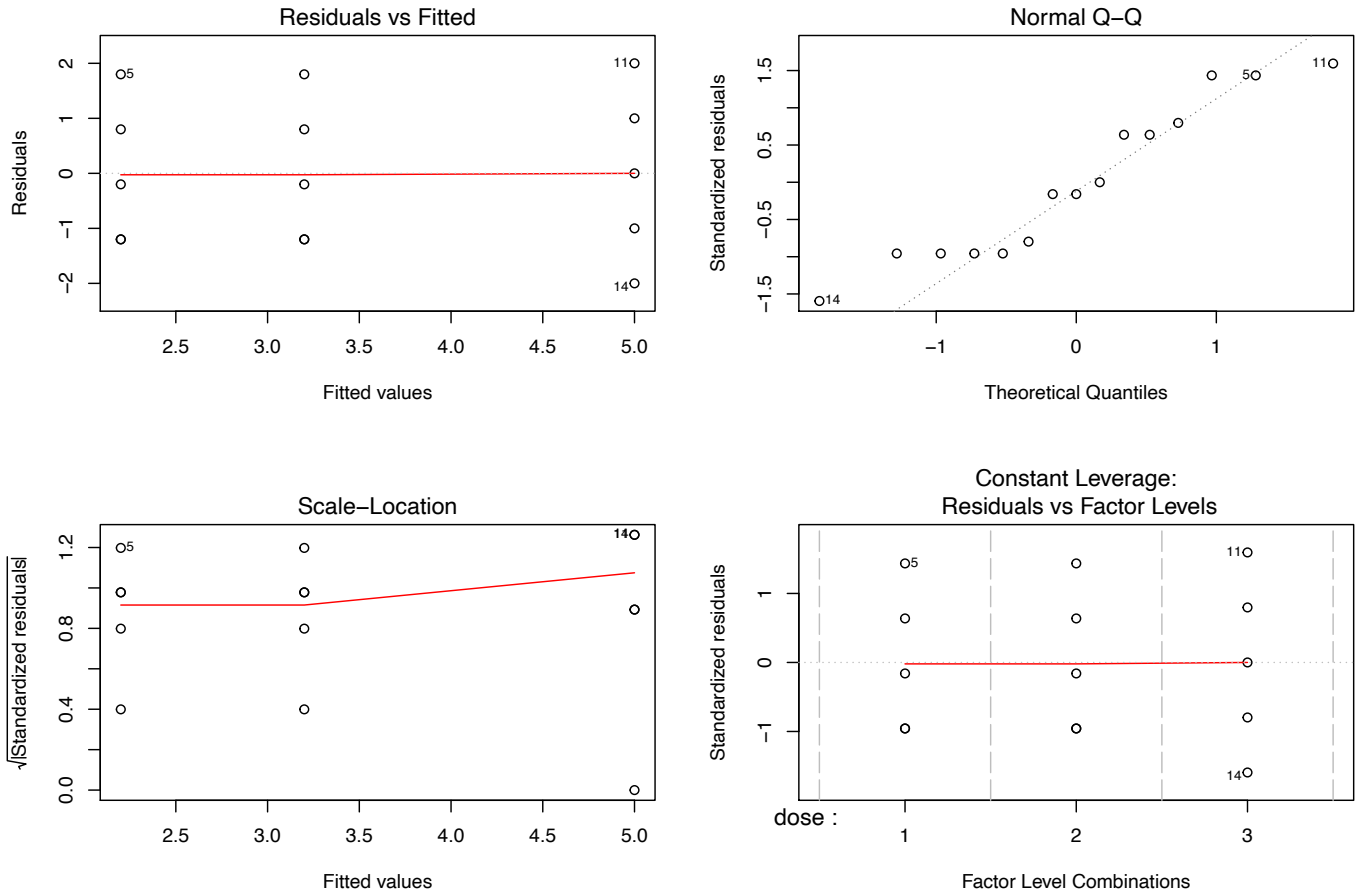


Figure 3: Ausgabe von aov-Funktion

Welch's F

Welch's F-Test berücksichtigt die vorhandenen Unterschiede in den Gruppenvarianzen.

```
oneway.test(libido ~ dose, data = df)
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  libido and dose
## F = 4.3205, num df = 2.0000, denom df = 7.9434, p-value = 0.05374
```

In unserem Beispiel gibt's keine Unterschiede in den Gruppenvarianzen!

Robuste ANOVA

```
viagraWide <- unstack(df, libido ~ dose)
viagraWide1 <- data.frame(data = viagraWide)
colnames(viagraWide1) = c('Placebo','Low.Dose','High.Dose')
knitr::kable(viagraWide1, caption = "Viagra-Datensatz im breiten Format [A. Field]")
```

Table 3: Viagra-Datensatz im breiten Format [A. Field]

Placebo	Low.Dose	High.Dose
3	5	7
2	2	4
1	4	5
1	2	3
4	3	6

Robuste ANOVA (1)

```
library(WRS2)
t1way(libido ~ dose, data = df, tr = 0.1)
```

```
## Call:
## t1way(formula = libido ~ dose, data = df, tr = 0.1)
##
## Test statistic: F = 4.3205
## Degrees of freedom 1: 2
## Degrees of freedom 2: 7.94
## p-value: 0.05374
##
## Explanatory measure of effect size: 0.71
```

```
med1way(libido ~ dose, data = df) # median
```

```
## Call:
## med1way(formula = libido ~ dose, data = df)
##
## Test statistic F: 4.7829
## Critical value: 5.473
## p-value: 0.07
```

```
t1waybt(libido ~ dose, data = df)
```

```
## Call:
## t1waybt(formula = libido ~ dose, data = df)
##
## Effective number of bootstrap samples was 384.
##
## Test statistic: 3
## p-value: 0.09115
## Variance explained 0.623
## Effect size 0.789
```

Alle Tests sind nicht signifikant und somit hat die Dosis keine Aswirkung auf Libido.

Kontraste

Kontraste

```
summary.lm(viagraModel)
```

```
##
## Call:
## lm(formula = libido ~ dose, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##   -2.0   -1.2   -0.2    0.9    2.0
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.2000     0.6272   3.508  0.00432 **
## dose2         1.0000     0.8869   1.127  0.28158
## dose3         2.8000     0.8869   3.157  0.00827 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.402 on 12 degrees of freedom
## Multiple R-squared:  0.4604, Adjusted R-squared:  0.3704
## F-statistic: 5.119 on 2 and 12 DF,  p-value: 0.02469
```

Kontraste

```
contrasts(df$dose)
```

```
##   2 3
## 1 0 0
## 2 1 0
## 3 0 1
```

Geplante Kontraste

Sei gegeben die Variablen oder deren Statistiken $\theta_1, \theta_2, ..., \theta_k$ und die Konstanten $a_1, a_2, ..., a_k$.

Die lineare Kombination $\sum_i a_i \theta_i$ heißt *Kontrast*.

Zwei Kontraste $\sum_i a_i \theta_i$ und $\sum_i b_i \theta_i$ heißen *orthogonal* wenn $\sum_i a_i b_i = 0$.

Geplante Kontraste (1)

Contrast	Description
`contr.helmert`	Contrasts the second level with the first, the third level with the average of the first two, the fourth level with the average of the first three, and so on.
`contr.poly`	Contrasts are used for trend analysis (linear, quadratic, cubic, and so on) based on orthogonal polynomials. Use for ordered factors with equally spaced levels.
`contr.sum`	Contrasts are constrained to sum to zero. Also called *deviation contrasts*, they compare the mean of each level to the overall mean across levels.
`contr.treatment`	Contrasts each level with the baseline level (first level by default). Also called *dummy coding*.
`contr.SAS`	Similar to `contr.treatment`, but the baseline level is the last level. This produces coefficients similar to contrasts used in most SAS procedures.

Figure 4: Verschiedene eingebaute Kontraste [Quelle: A. Kabacoff]

```
contrasts(df$dose) <- contr.helmert(3)
df$dose
```

```
##  [1] 1 1 1 1 1 2 2 2 2 2 3 3 3 3 3
## attr(,"contrasts")
##   [,1] [,2]
## 1   -1   -1
## 2    1   -1
## 3    0    2
## Levels: 1 2 3
```

Geplante Kontraste (3)

```
contrasts(df$dose) <- cbind(c(-2,1,1),c(0,-1,1))
df$dose
```

```
##  [1] 1 1 1 1 1 2 2 2 2 2 3 3 3 3 3
## attr(,"contrasts")
##   [,1] [,2]
## 1   -2    0
## 2    1   -1
## 3    1    1
## Levels: 1 2 3
```

- Bei den geplanten Kontrasten vergleichen wir die Gruppen mit dem positiven Vorzeichen gegen die Gruppen mit dem negativen.

- Im Kontrast 1 vergleichen wir Placebo-Gruppe und die zwei Versuchsgruppen
- Im Kontrast 2 vergleichen wir niedrige Dosis-Gruppe und die höhe Dosis-Gruppe

Geplante Kontraste (4)

```
ViagraModel3 <- aov(libido ~ dose, data = df)
summary.lm(ViagraModel3)
```

```
##
## Call:
## aov(formula = libido ~ dose, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##   -2.0   -1.2   -0.2    0.9    2.0
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.4667     0.3621   9.574 5.72e-07 ***
## dose1         0.6333     0.2560   2.474   0.0293 *
## dose2         0.9000     0.4435   2.029   0.0652 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.402 on 12 degrees of freedom
## Multiple R-squared:  0.4604, Adjusted R-squared:  0.3704
## F-statistic: 5.119 on 2 and 12 DF,  p-value: 0.02469
```

Geplante Kontraste (5)

- Weil wir davon ausgehen, dass Libido durch die Gabe des Medikaments steig, können wir den zweiseitigen Test durch den einseitigen ersetzen und die entsprechenden p-Werte halbieren.
- Ohne die Hypothese, dass die Wirkung mit der Dosis steigt, wäre unser Ergebnis nicht signifikant
- Das zeigt, dass es wichtig ist, die Hypothese zu haben, bevor wir die Daten sammeln.

Post hoc-Tests

Post hoc-Tests

- *F*-Test sagt uns nur, dass es die Unterschiede vorhanden sind, aber nicht zwischen welchen Gruppen.
- Um die paarweisen Vergleiche durchzuführen, brauchen wir *post hoc*-Tests.

Bonferroni und Benjamini-Hochberg *post hoc*-Tests

```
pairwise.t.test(df$libido, df$dose, p.adjust.method = "bonferroni")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  df$libido and df$dose
##
##   1     2
## 2 0.845 -
## 3 0.025 0.196
##
## P value adjustment method: bonferroni
```

```
pairwise.t.test(df$libido, df$dose, p.adjust.method = "BH")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  df$libido and df$dose
##
##   1     2
## 2 0.282 -
## 3 0.025 0.098
##
## P value adjustment method: BH
```

Tukey HSD paarweise Vergleiche

```
TukeyHSD(ViagraModel3)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = libido ~ dose, data = df)
##
## $dose
##     diff         lwr      upr     p adj
## 2-1  1.0 -1.3662412 3.366241 0.5162761
## 3-1  2.8  0.4337588 5.166241 0.0209244
## 3-2  1.8 -0.5662412 4.166241 0.1474576
```

Tukey HSD paarweise Vergleiche (1)

```
par(las=2)
par(mar=c(5,8,4,2))
plot(TukeyHSD(ViagraModel3), cex = 2, cex.main=2, cex.lab=2, cex.axis=2)
```

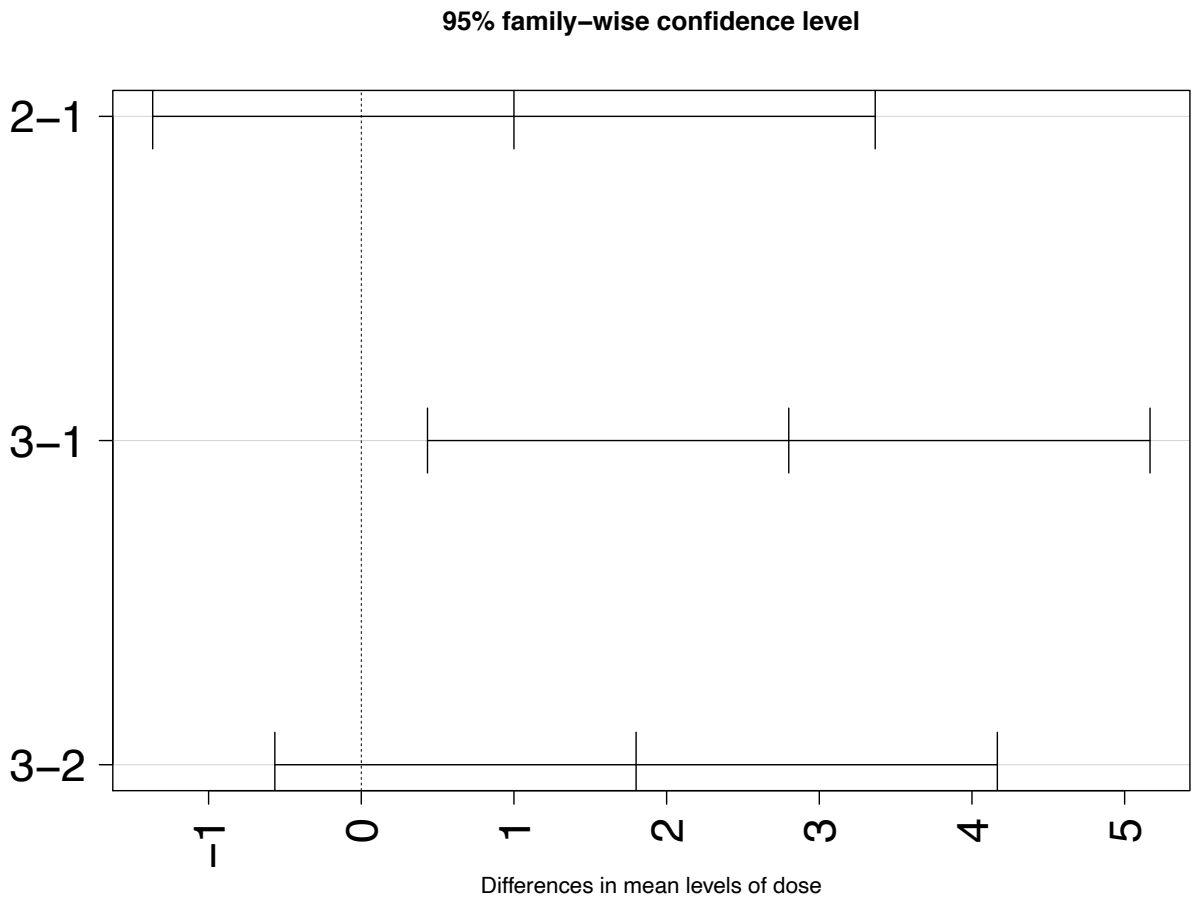


Figure 5: Tukey HSD paarweise Vegleiche.

Tukey HSD paarweise Vergleiche (2)

```
library(multcomp)
par(mar=c(5,4,6,2))
tuk <- glht(ViagraModel3, linfct = mcp(dose = "Tukey"))
summary(tuk)
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: aov(formula = libido ~ dose, data = df)
##
## Linear Hypotheses:
##            Estimate Std. Error t value Pr(>|t|)
## 2 - 1 == 0   1.0000     0.8869   1.127   0.5163
## 3 - 1 == 0   2.8000     0.8869   3.157   0.0209 *
## 3 - 2 == 0   1.8000     0.8869   2.029   0.1475
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tukey HSD paarweise Vergleiche (3)

```
confint(tuk)
```

```
##
##   Simultaneous Confidence Intervals
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: aov(formula = libido ~ dose, data = df)
##
## Quantile = 2.6658
## 95% family-wise confidence level
##
##
## Linear Hypotheses:
##            Estimate lwr     upr
## 2 - 1 == 0  1.0000  -1.3644  3.3644
## 3 - 1 == 0  2.8000   0.4356  5.1644
## 3 - 2 == 0  1.8000  -0.5644  4.1644
```

Für signifikante Unterschiede dürfen die Konfidenzintervalle die Null nicht durchkreuzen!

Tukey HSD paarweise Vergleiche (4)

```
par(mar=c(5,8,4,2))
plot(cld(tuk, level = .05, col = "lightblue"), cex = 2, cex.main=2, cex.lab=2, cex.axis=2)
```

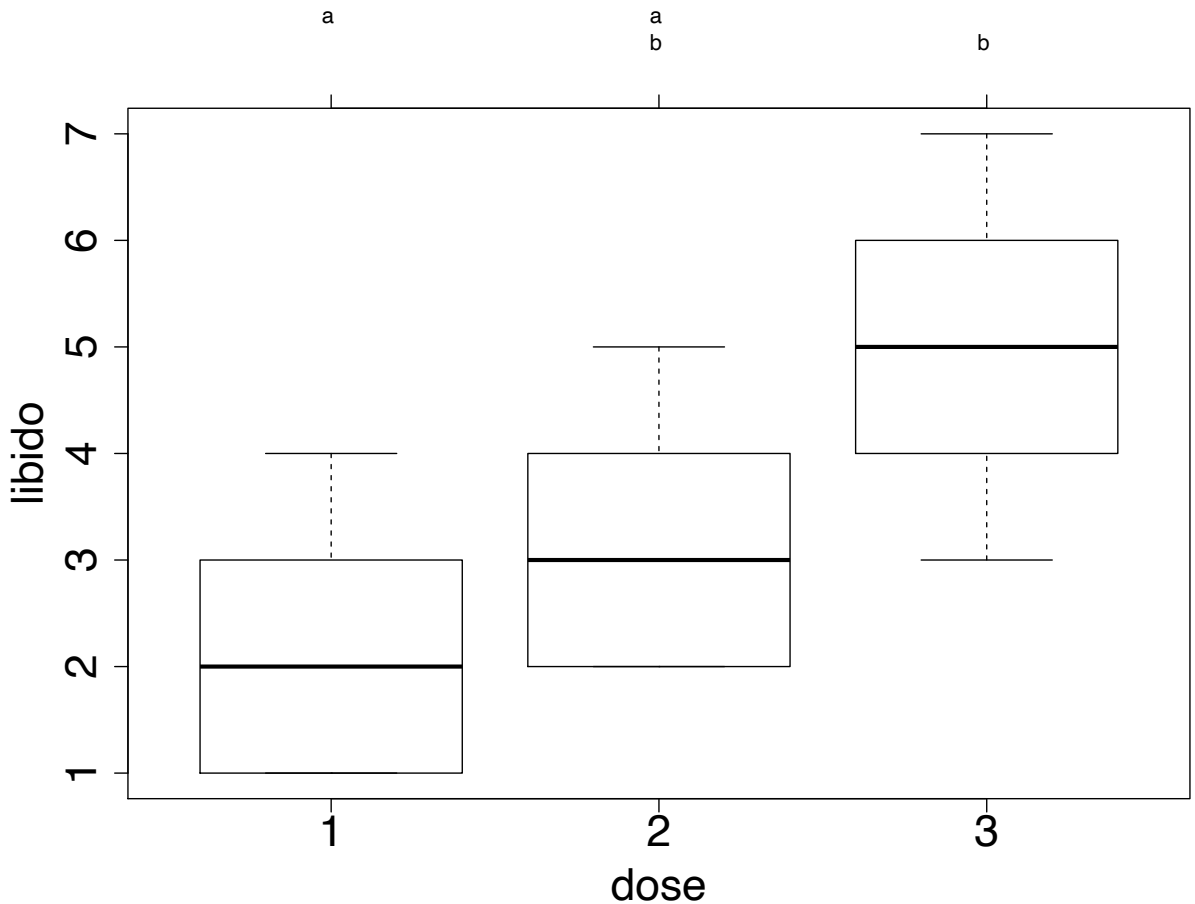


Figure 6: Tukey HSD paarweise Vegleiche. Gruppen, die die gleichen Buchstaben oben zu stehen haben, sind nicht signifikant unterschiedlich.

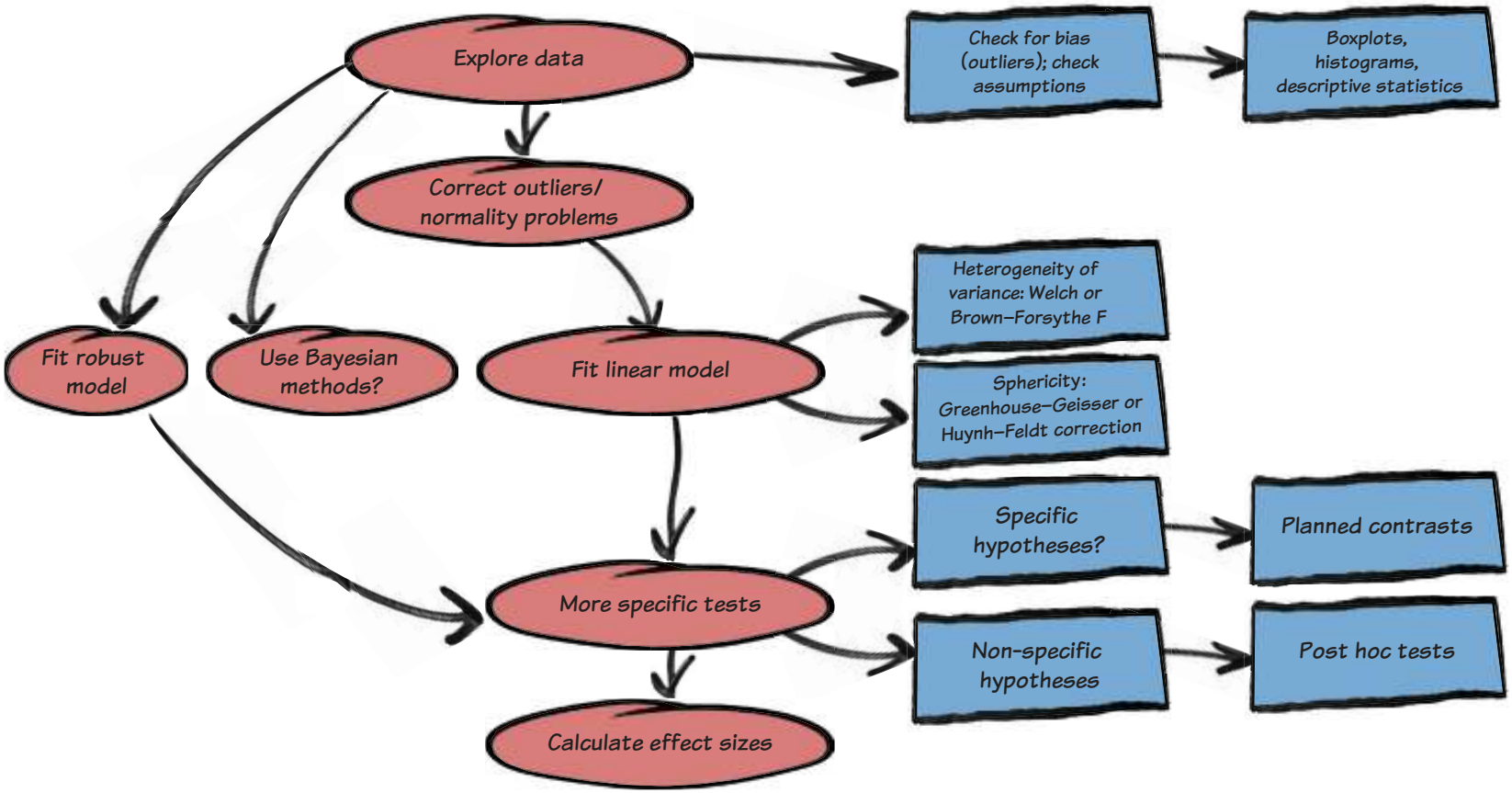


Figure 7: Vergleich mehrerer Mittelwerte [Quelle: A. Field]

Effect size

$$R^2 = \frac{SSM}{SST}$$

Effektgröße ist gegeben durch R (0.68)

Anstatt r^2 benutzt man in ANOVA η^2.

$$\omega^2 = \frac{SSM - (df_M)MSE}{SST + MSE}$$

$\omega = 0.60$

- $\omega^2 = 0.01$ (small effects)
- $\omega^2 = 0.06$ (medium effects)
- $\omega^2 = 0.14$ (large effects)

ANOVA (Fortsetzung)

Vitaly Belik

Institut für Veterinär-Epidemiologie und Biometrie, FU Berlin

1/12/2020

Wiederholung

- ANCOVA
- Factorial ANOVA with repeated measures
- Mixed designes
- Multi-level models

Beispiel (Viagra)

Table 1: Viagra-Datensatz [A. Field]

person	dose	libido
1	1	3
2	1	2
3	1	1
4	1	1
5	1	4
6	2	5
7	2	2
8	2	4
9	2	2
10	2	3
11	3	7
12	3	4
13	3	5
14	3	3
15	3	6

Beispiel (Viagra) (1)

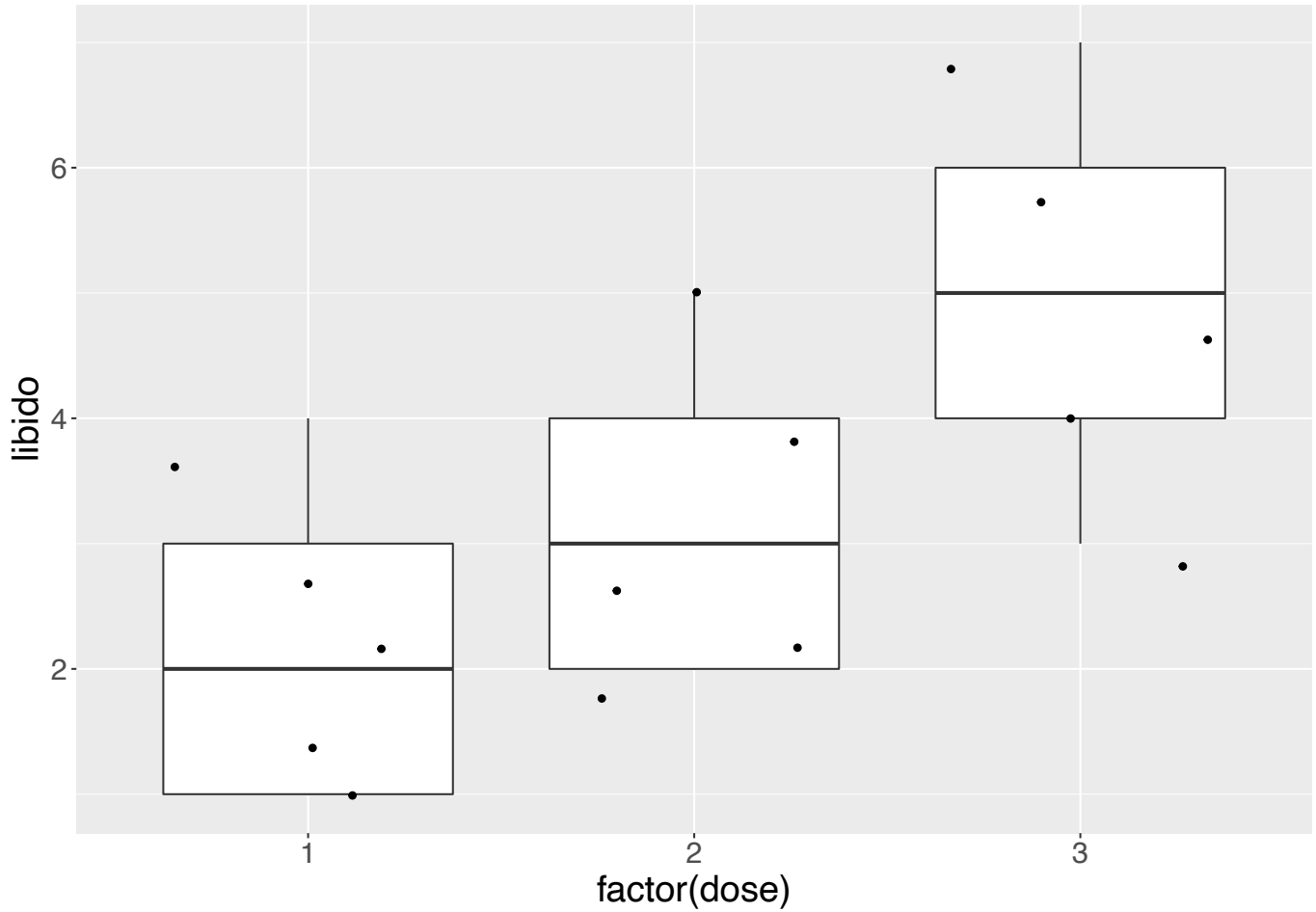


Figure 1: Viagra-Datensatz mit Konfidenzintervallen

ANOVA

Die Antwortvariable (*response*) ist kontinuierlich und die unabhängigen Variablen (*Faktoren*) kategoriell.

- ANOVA kann und soll als GLM (veralgemeinerte *Regression*) betrachtet werden.

Einfache ANOVA: eine erklärende (kategorielle) Variable (Faktor).

- F-Test prüft ob Unterschiede zwischen den Gruppen vorliegen
- Test für Homogenität der Varianz (Levene's Test)
- Kontraste
- post hoc-Tests für paarweise Vergleiche. Bonferroni und Tukey.

Einfache ANOVA mit wiederholten Messungen (repeated measures)

Einfache ANOVA mit wiederholten Messungen (repeated measures)

Bis jetzt haben wir *unabhägige* ANOVA betreachtet. Alle Messungen stammten von verscheidenen Individien (Objekten).

Beispiel

Figure 2: Ich bin ein Star: holt mich hier raus! [Quelle: RTL]

Beispiel (1)

Table 2: Dschungel-Datensatz [A. Field]

participant	stick_insect	kangaroo_testicle	fish_eye	witchetty_grub
P1	8	7	1	6
P2	9	5	2	5
P3	6	2	3	8
P4	5	3	1	9
P5	8	4	5	8
P6	7	5	6	7
P7	10	2	7	2
P8	12	6	8	1

Einfache ANOVA mit wiederholten Messungen (repeated measures) (1)

Falls die Messungen von den gleichen Individuen (Objekten) stammen, braucht man die ANOVA mit *wiederholten Messungen* (*repeated measures*).

Somit ist die Annahme des F-Test über die unabhängige Messungen ist nicht mehr gültig.

Dabei kann man weiter den F-Test benutzen, wenn die Annahme der *Spherizität* erfüllt ist.

Sphärizität

Sphärizität bezieht sich auf die Gleichheit der Varianzen der Differenzen zwischen den Niveaus der Messungen. Daher braucht man mindestens 3 Messniveaus um sich überhaupt mit Sphärizität zu befassen.

Um die Daten auf Sphärizität zu prüfen, nutzt man den *Mauchly-Test*

Der *Mauchly-Test* überprüft die Hypothese, dass die Varianzen gleich sind. Falls der Test signifikant ist, sind die Varianzen nicht gleich.

Sphärizität bereitet auch Probleme bei post hoc-Test. Dabei ist Bonferroni-Methode verlässlicher als Tukey-Test.

Spherizität-Korrekturen

Falls Spherizität nicht erfüllt ist, nimmt man Korrekturen in den Freiheitsgraden für das F-Verhältnis.

- Greenhouse-Geisser-Korrektur
- Huynh-Feldt-Korrektur

Des weiteren kann man hierarchisches Modell benutzen, indem man die Messwerte von gleichen Individuen zu einem Hierarchieebene zusammenfügt.

Repeated measures ANOVA: Theorie

Im Falle einfacher unabhängiger ANOVA

$$SS_{\mathrm{T}} = SS_{\mathrm{M}} + SS_{\mathrm{R}}.$$

Dabei m ist die Anzahl verschiedener Gruppen

$$SS_{\mathrm{M}} = \sum_{k=1}^{m} n_k \left(\overline{x}_k - \overline{x}_{\mathrm{grand}}\right)^2$$

und n_k ist die Größe der Gruppen ($df = m - 1$)

$$SS_{\mathrm{R}} = \sum_{i=1}^{n} n_k \left(\overline{x}_{ik} - \overline{x}_k\right)^2 = \sum_{k=1}^{m} s^2_{\mathrm{group},k}(n_k - 1)$$

und $df_{\mathrm{R}} = df_{\mathrm{total}} - df_{\mathrm{M}}$, $df_{\mathrm{total}} = n - 1$.

Repeated measures ANOVA: Theorie (1)

Im Falle einfacher ANOVA mit wiederholten Messungen

$$SS_{\mathrm{T}} = SS_{\mathrm{B}} + SS_{\mathrm{W}}.$$

wobei

SS_{B} - Quadratensumme für die Unterschiede zwischen (**between**) den Individuen

SS_{W} - Quadratensumme für die Unterschiede innerhalb (**within**) der Individuen

Es folgt

$$SS_{\mathrm{B}} = SS_{\mathrm{T}} - SS_{\mathrm{W}}.$$

Repeated measures ANOVA: Theorie (2)

$$SS_{\text{w}} = \sum_{i=1}^{N} s^2_{\text{person},i}(n_i - 1)$$

wobei n_i - Anzahl der Messungen, die dem Individuum i entsprechen (Anzahl der Niveaus). Insgesamt haben wir N Individuen.
$df = (m-1) \cdot N$.

$$SS_{\text{R}} = SS_{\text{W}} - SS_{\text{M}}$$

und $df_{\text{R}} = df_{\text{W}} - df_{\text{M}}$.

F-Verhältnis ist wie gewöhnt berechnet

$$F = \frac{MS_{\text{M}}}{MS_{\text{R}}}$$

wobei $MS_{\text{M}} = SS_{\text{M}}/df_{\text{M}}$ und $MS_{\text{R}} = SS_{\text{R}}/df_{\text{R}}$.

Zerlegung der Varianzen

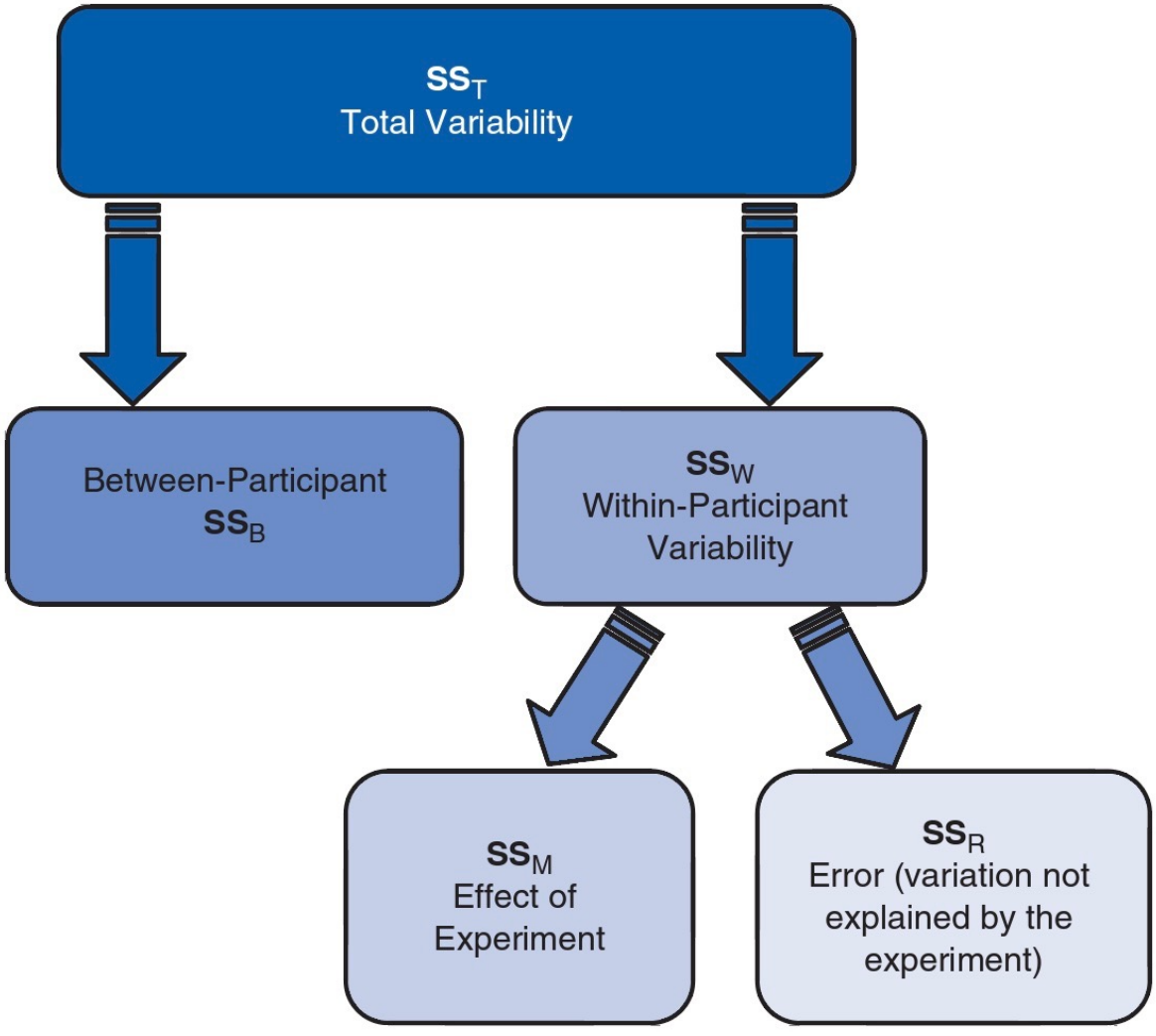


Figure 3: Zerlegung der Varianzen für ANOVA mit wiederholten Messungen [A. Field]

Beispiel

Table 3: Dschungel-Datensatz [A. Field]

Participant	Animal	Time
P1	stick_insect	8
P2	stick_insect	9
P3	stick_insect	6
P4	stick_insect	5
P5	stick_insect	8
P6	stick_insect	7
P7	stick_insect	10
P8	stick_insect	12
P1	kangaroo_testicle	7
P2	kangaroo_testicle	5
P3	kangaroo_testicle	2
P4	kangaroo_testicle	3
P5	kangaroo_testicle	4
P6	kangaroo_testicle	5
P7	kangaroo_testicle	2
P8	kangaroo_testicle	6
P1	fish_eye	1
P2	fish_eye	2
P3	fish_eye	3
P4	fish_eye	1
P5	fish_eye	5
P6	fish_eye	6
P7	fish_eye	7
P8	fish_eye	8
P1	witchetty_grub	6
P2	witchetty_grub	5
P3	witchetty_grub	8
P4	witchetty_grub	9
P5	witchetty_grub	8
P6	witchetty_grub	7
P7	witchetty_grub	2
P8	witchetty_grub	1

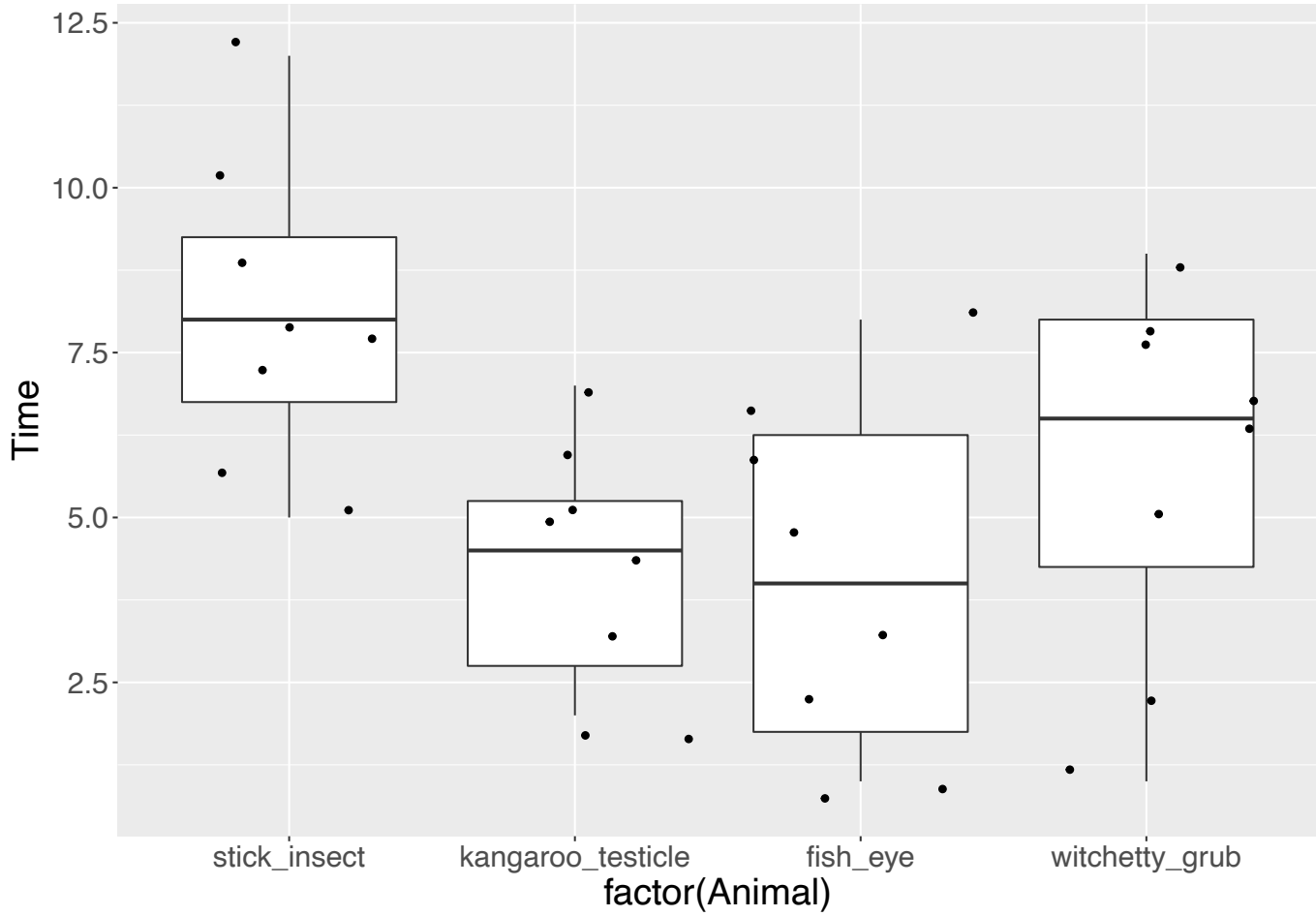


Figure 4: Dschungel-Datensatz [A. Field]

Deskriptive Statistik

```
## longBush$Animal: stick_insect
##      nbr.val     nbr.null       nbr.na          min          max        range
##    8.0000000    0.0000000    0.0000000    5.0000000   12.0000000    7.0000000
##          sum       median         mean      SE.mean CI.mean.0.95          var
##   65.0000000    8.0000000    8.1250000    0.7891564    1.8660584    4.9821429
##      std.dev     coef.var
##    2.2320714    0.2747165
## ------------------------------------------------------------
## longBush$Animal: kangaroo_testicle
##      nbr.val     nbr.null       nbr.na          min          max        range
##    8.0000000    0.0000000    0.0000000    2.0000000    7.0000000    5.0000000
##          sum       median         mean      SE.mean CI.mean.0.95          var
##   34.0000000    4.5000000    4.2500000    0.6477985    1.5318000    3.3571429
##      std.dev     coef.var
##    1.8322508    0.4311178
## ------------------------------------------------------------
## longBush$Animal: fish_eye
##      nbr.val     nbr.null       nbr.na          min          max        range
##    8.0000000    0.0000000    0.0000000    1.0000000    8.0000000    7.0000000
##          sum       median         mean      SE.mean CI.mean.0.95          var
##   33.0000000    4.0000000    4.1250000    0.9716977    2.2977000    7.5535714
##      std.dev     coef.var
##    2.7483761    0.6662730
## ------------------------------------------------------------
## longBush$Animal: witchetty_grub
##      nbr.val     nbr.null       nbr.na          min          max        range
##    8.0000000    0.0000000    0.0000000    1.0000000    9.0000000    8.0000000
##          sum       median         mean      SE.mean CI.mean.0.95          var
##   46.0000000    6.5000000    5.7500000    1.0307764    2.4373989    8.5000000
##      std.dev     coef.var
##    2.9154759    0.5070393
```

Kontraste

Positive Zahlen werden mit negativen verglichen. Null bedeutet, dass die Gruppe im Vergleich nicht vorkommt.

```
knitr::kable(contrasts(longBush$Animal))
```

	PartsvsWhole	KengvsEye	StickvsGrub
stick_insect	1	0	-1
kangaroo_testicle	-1	-1	0
fish_eye	-1	1	0
witchetty_grub	1	0	1

```
## $ANOVA
##        Effect DFn DFd     SSn     SSd          F            p p<.05       ges
## 1 (Intercept)   1   7 990.125  17.375 398.899281 1.973536e-07     * 0.8529127
## 2      Animal   3  21  83.125 153.375   3.793806 2.557030e-02     * 0.3274249
##
## $`Mauchly's Test for Sphericity`
##   Effect        W          p p<.05
## 2 Animal 0.136248 0.04684581     *
##
## $`Sphericity Corrections`
##   Effect       GGe      p[GG] p[GG]<.05       HFe      p[HF] p[HF]<.05
## 2 Animal 0.5328456 0.06258412           0.6657636 0.04833061         *
```

Greenhouse-Geisser-Korrektur (konservativ) und *Huynh-Feldt-Korrektur* sind zu beachten. Spherizität ist nicht erfüllt.

Es folgt, dass die Art des Tieres einen wesentlichen Einflüß auf die Zeit hat. Der F-Test (falls überhaupt signifikant) sagt uns nicht welche Tiere sich von einander unterscheiden.

post hoc-Tests

```
##
##  Pairwise comparisons using paired t tests
##
## data:  longBush$Time and longBush$Animal
##
##                   stick_insect kangaroo_testicle fish_eye
## kangaroo_testicle 0.0121       -                 -
## fish_eye          0.0056       1.0000            -
## witchetty_grub    1.0000       1.0000            1.0000
##
## P value adjustment method: bonferroni
```

Repeated Measures ANOVA als hierarchisches Modell

- Wichtige Annahme der Regression ist die Unabhängigkeit der Residuen.
- Wiederholte Messungen beinhalten abhängige Wette, und deshalb führen zu abhängigen Residuen.
- Hierarchisches Modell (multilevel model) ist eine Verallgemeinerung der Regression, dass die Abhängigkeiten in den Daten berücksichtigt.

```
library(nlme)
jungle_ml <- lme(Time~Animal, random = ~1|Participant/Animal, data = longBush, method = "ML")
```

- Der Funktion wird beigebracht, dass die Variable "Animal" aus derselben Teilnehmern besteht.

Repeated Measures ANOVA als hierarchisches Modell (1)

Um das Effekt von "Animal" festzustellen, erzeugen wir ein anderes Modell ohne den Prädiktor.

```
library(nlme)
jungle_ml_0 <- lme(Time ~ 1, random = ~1|Participant/Animal, data = longBush, method = "ML")
```

Jetzt vergleichen wir beide Modelle

```
anova(jungle_ml_0, jungle_ml)
```

```
##             Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## jungle_ml_0     1  4 165.0875 170.9504 -78.54373
## jungle_ml       2  7 158.3949 168.6551 -72.19747 1 vs 2 12.69253  0.0054
```

Repeated Measures ANOVA als hierarchisches Modell (2)

```
summary(jungle_ml)
```

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: longBush
##        AIC      BIC    logLik
##   158.3949 168.6551 -72.19747
##
## Random effects:
##  Formula: ~1 | Participant
##          (Intercept)
## StdDev: 7.756253e-05
##
##  Formula: ~1 | Animal %in% Participant
##         (Intercept)   Residual
## StdDev:    2.309935 0.01176165
##
## Fixed effects: Time ~ Animal
##                       Value Std.Error DF   t-value p-value
## (Intercept)          5.5625 0.4365423 21 12.742178  0.0000
## AnimalPartsvsWhole   1.3750 0.4365423 21  3.149752  0.0048
## AnimalKengvsEye     -0.0625 0.6173641 21 -0.101237  0.9203
## AnimalStickvsGrub   -1.1875 0.6173641 21 -1.923500  0.0681
##  Correlation:
##                     (Intr) AnmlPW AnmlKE
## AnimalPartsvsWhole 0
## AnimalKengvsEye     0      0
## AnimalStickvsGrub   0      0      0
##
## Standardized Within-Group Residuals:
##           Min            Q1           Med            Q3           Max
## -0.0104701088 -0.0046839960  0.0001377646  0.0041329377  0.0085414045
##
## Number of Observations: 32
## Number of Groups:
##             Participant Animal %in% Participant
##                       8                      32
```

Repeated Measures ANOVA als hierarchisches Modell (3)

```
library(multcomp)
postHocs <- glht(jungle_ml, linfct = mcp(Animal = "Tukey") )
summary(postHocs)
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lme.formula(fixed = Time ~ Animal, data = longBush, random = ~1 |
##     Participant/Animal, method = "ML")
##
## Linear Hypotheses:
##                                         Estimate Std. Error z value Pr(>|z|)
## kangaroo_testicle - stick_insect == 0     -3.875      1.155  -3.355  0.00414 **
## fish_eye - stick_insect == 0              -4.000      1.155  -3.463  0.00315 **
## witchetty_grub - stick_insect == 0        -2.375      1.155  -2.056  0.16767
## fish_eye - kangaroo_testicle == 0         -0.125      1.155  -0.108  0.99955
## witchetty_grub - kangaroo_testicle == 0    1.500      1.155   1.299  0.56374
## witchetty_grub - fish_eye == 0             1.625      1.155   1.407  0.49492
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Repeated Measures ANOVA als hierarchisches Modell (4)

```
confint(postHocs)
```

```
##
##   Simultaneous Confidence Intervals
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lme.formula(fixed = Time ~ Animal, data = longBush, random = ~1 |
##     Participant/Animal, method = "ML")
##
## Quantile = 2.5691
## 95% family-wise confidence level
##
##
## Linear Hypotheses:
##                                        Estimate lwr     upr
## kangaroo_testicle - stick_insect == 0   -3.8750  -6.8423 -0.9077
## fish_eye - stick_insect == 0            -4.0000  -6.9673 -1.0327
## witchetty_grub - stick_insect == 0      -2.3750  -5.3423  0.5923
## fish_eye - kangaroo_testicle == 0       -0.1250  -3.0923  2.8423
## witchetty_grub - kangaroo_testicle == 0  1.5000  -1.4673  4.4673
## witchetty_grub - fish_eye == 0           1.6250  -1.3423  4.5923
```

Effektgröße

$$\omega^2 = \frac{\left[\frac{k-1}{nk}\left(MS_{\text{M}} - MS_{\text{R}}\right)\right]}{MS_{\text{R}} + \frac{MS_{\text{B}} - MS_{\text{R}}}{k} + \left[\frac{k-1}{nk}\left(MS_{\text{M}} - MS_{\text{R}}\right)\right]}.$$

Für gegebene Kontraste kann man ausrechen:

$$r = \sqrt{\frac{t^2}{t^2 + df}}.$$

Faktorielle (unabhängige) ANOVA

Beispiel

Wenn im Versuch mehr als zwei unabhägige Variable vorkommen, nennt man das faktorelles Design.

Man will die Wirkungs des Alkohols auf die wahrgenommene Attraktivität des jeweils anderen Geschlechts

Table 5: Attraktivität-Datensatz [A. Field]

gender	alcohol	attractiveness	alcohol1
Female	None	65	None
Female	None	70	None
Female	None	60	None
Female	None	60	None
Female	None	60	None
Female	None	55	None
Female	None	60	None
Female	None	55	None
Female	2 Pints	70	2 Pints
Female	2 Pints	65	2 Pints
Female	2 Pints	60	2 Pints
Female	2 Pints	70	2 Pints
Female	2 Pints	65	2 Pints
Female	2 Pints	60	2 Pints
Female	2 Pints	60	2 Pints

Beispiel

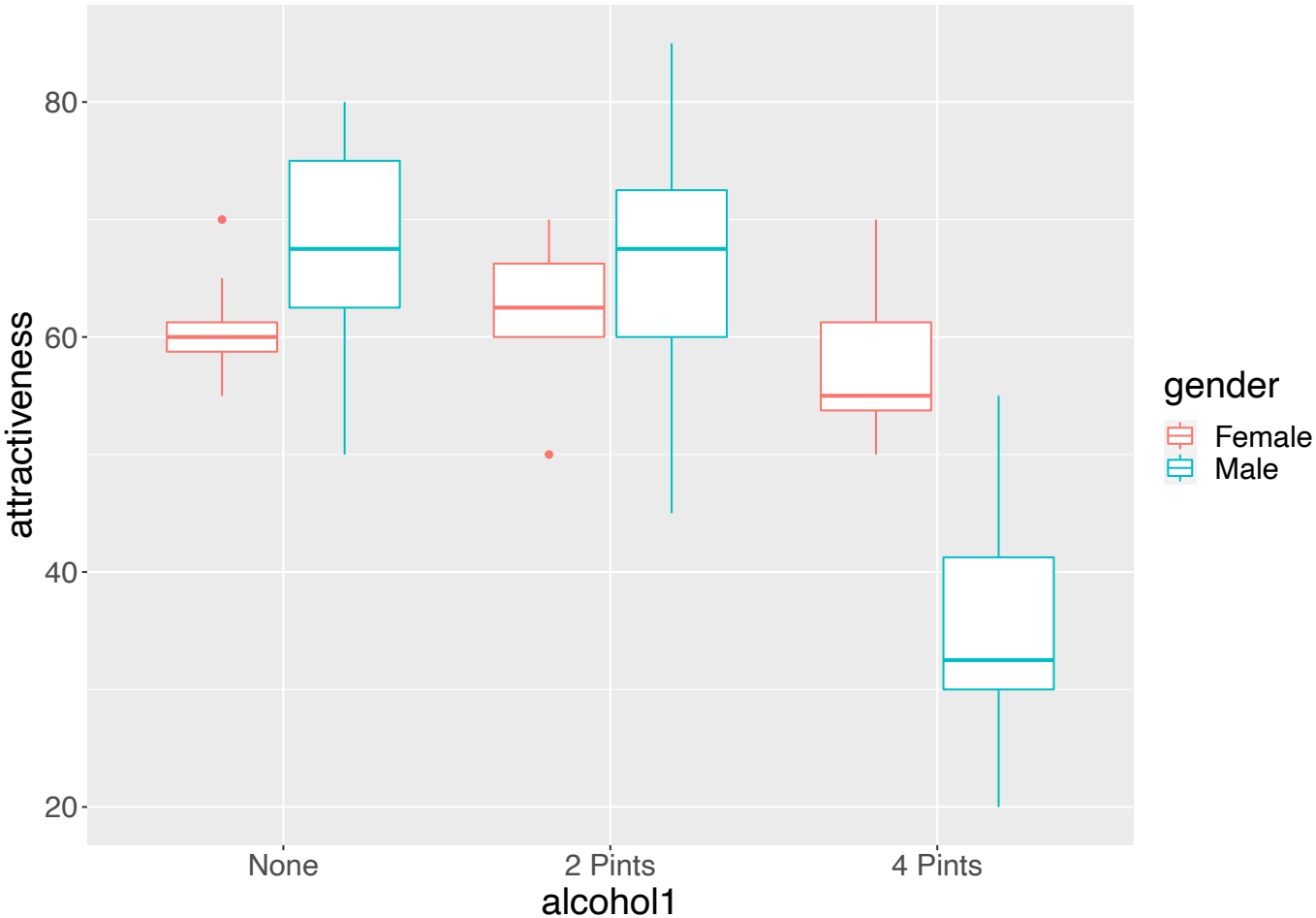
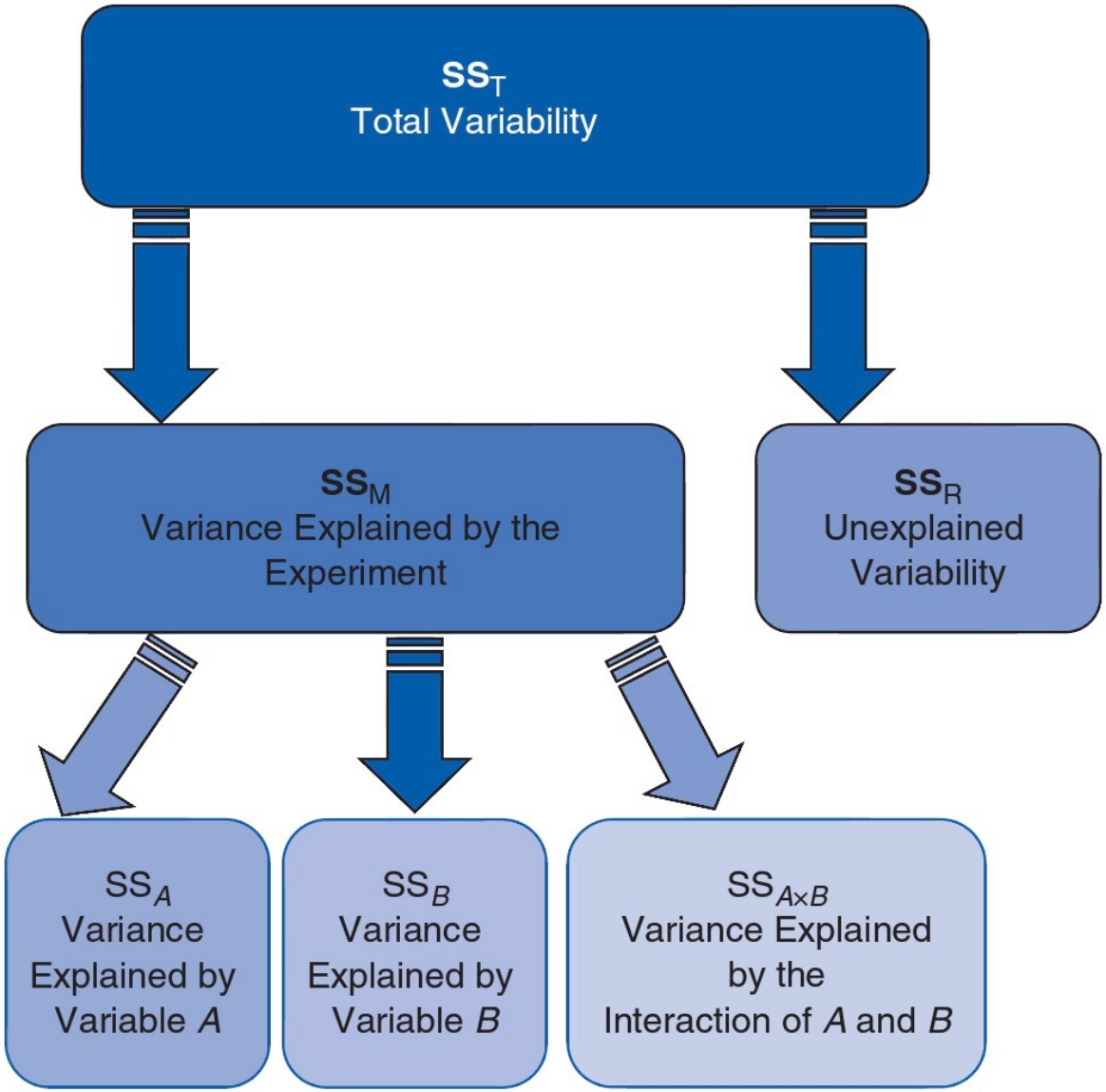


Figure 5: Attraktivität-Data [A. Field]

Zerlegung der Varianzen



Haupteffekt von Geschlecht

```
by(df$attr, df$gender, stat.desc)
```

```
## df$gender: Female
##      nbr.val     nbr.null       nbr.na          min          max        range
##   24.0000000    0.0000000    0.0000000   50.0000000   70.0000000   20.0000000
##          sum       median         mean      SE.mean CI.mean.0.95          var
## 1445.0000000   60.0000000   60.2083333    1.2937687    2.6763646   40.1721014
##      std.dev     coef.var
##    6.3381465    0.1052703
## ------------------------------------------------------------
## df$gender: Male
##      nbr.val     nbr.null       nbr.na          min          max        range
##   24.0000000    0.0000000    0.0000000   20.0000000   85.0000000   65.0000000
##          sum       median         mean      SE.mean CI.mean.0.95          var
## 1355.0000000   60.0000000   56.4583333    3.7768263    7.8129606  342.3460145
##      std.dev     coef.var
##   18.5025948    0.3277212
```

Haupteffekt von Alkohol

```
by(df$attr, df$alcohol, stat.desc)
```

```
## df$alcohol: 2 Pints
##      nbr.val     nbr.null       nbr.na          min          max        range
##   16.0000000    0.0000000    0.0000000   45.0000000   85.0000000   40.0000000
##          sum       median         mean      SE.mean CI.mean.0.95          var
## 1035.0000000   65.0000000   64.6875000    2.4777657    5.2812326   98.2291667
##      std.dev     coef.var
##    9.9110628    0.1532145
## ------------------------------------------------------------
## df$alcohol: 4 Pints
##      nbr.val     nbr.null       nbr.na          min          max        range
##   16.0000000    0.0000000    0.0000000   20.0000000   70.0000000   50.0000000
##          sum       median         mean      SE.mean CI.mean.0.95          var
##  745.0000000   50.0000000   46.5625000    3.5858155    7.6429849  205.7291667
##      std.dev     coef.var
##   14.3432621    0.3080432
## ------------------------------------------------------------
## df$alcohol: None
##      nbr.val     nbr.null       nbr.na          min          max        range
##    16.000000     0.000000     0.000000    50.000000    80.000000    30.000000
##          sum       median         mean      SE.mean CI.mean.0.95          var
##  1020.000000    62.500000    63.750000     2.116404     4.511009    71.666667
##      std.dev     coef.var
##     8.465617     0.132794
```

Haupteffekt von Interaktion

```
by(df$attr, list(df$alcohol, df$gender), stat.desc)
```

```
## : 2 Pints
## : Female
##      nbr.val     nbr.null       nbr.na          min          max        range
##    8.0000000    0.0000000    0.0000000   50.0000000   70.0000000   20.0000000
##          sum       median         mean      SE.mean CI.mean.0.95          var
##  500.0000000   62.5000000   62.5000000    2.3145502    5.4730417   42.8571429
##      std.dev     coef.var
##    6.5465367    0.1047446
## ------------------------------------------------------------
## : 4 Pints
## : Female
##      nbr.val     nbr.null       nbr.na          min          max        range
##    8.0000000    0.0000000    0.0000000   50.0000000   70.0000000   20.0000000
##          sum       median         mean      SE.mean CI.mean.0.95          var
##  460.0000000   55.0000000   57.5000000    2.5000000    5.9115606   50.0000000
##      std.dev     coef.var
##    7.0710678    0.1229751
## ------------------------------------------------------------
## : None
## : Female
##      nbr.val     nbr.null       nbr.na          min          max        range
##   8.00000000   0.00000000   0.00000000  55.00000000  70.00000000  15.00000000
##          sum       median         mean      SE.mean CI.mean.0.95          var
## 485.00000000  60.00000000  60.62500000   1.75191222   4.14261412  24.55357143
##      std.dev     coef.var
##   4.95515604   0.08173453
## ------------------------------------------------------------
## : 2 Pints
## : Male
##      nbr.val     nbr.null       nbr.na          min          max        range
##    8.0000000    0.0000000    0.0000000   45.0000000   85.0000000   40.0000000
##          sum       median         mean      SE.mean CI.mean.0.95          var
##  535.0000000   67.5000000   66.8750000    4.4257263   10.4651798  156.6964286
##      std.dev     coef.var
##   12.5178444    0.1871827
## ------------------------------------------------------------
## : 4 Pints
## : Male
```

Interaktionen

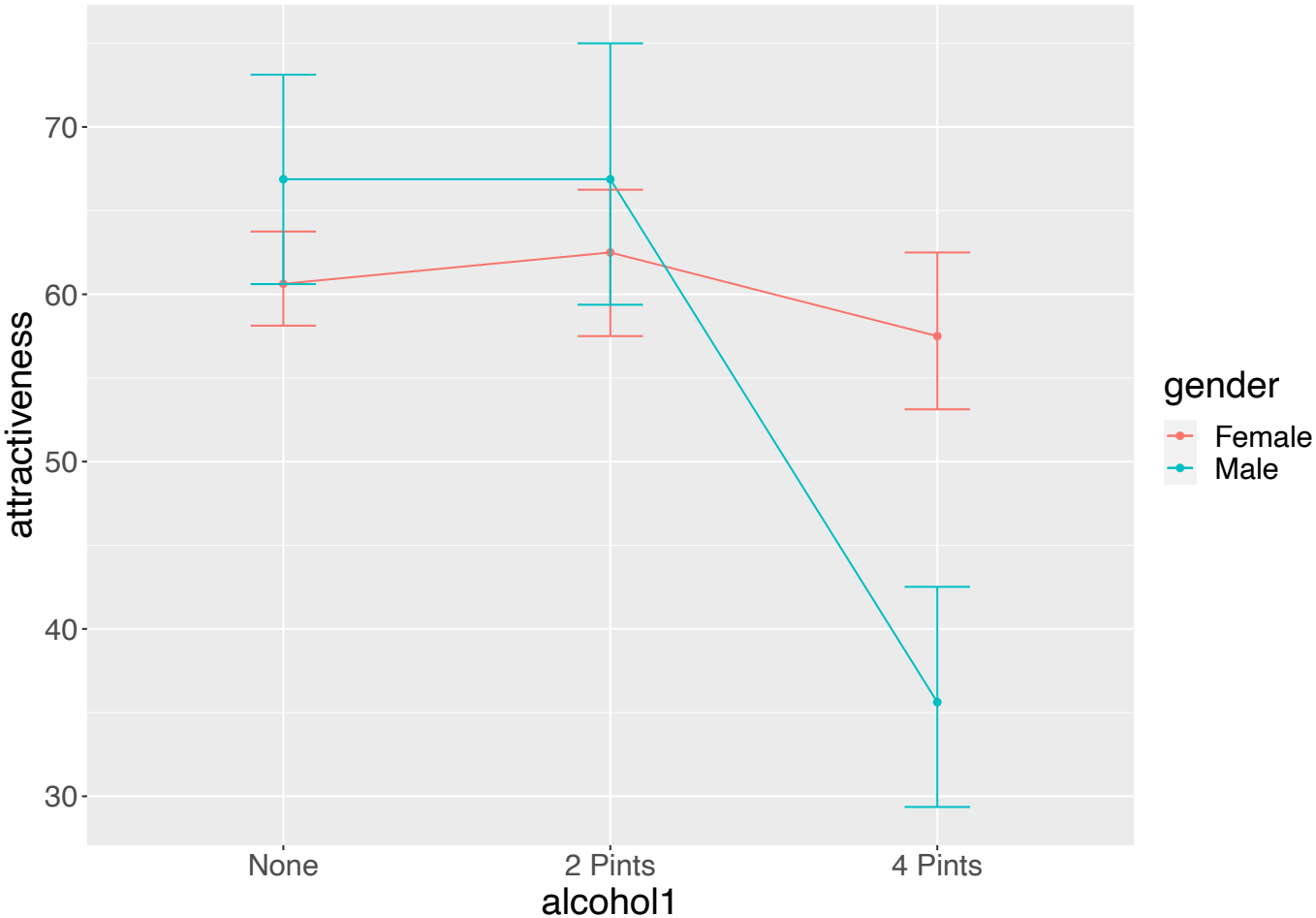


Figure 7: Attraktivität-Data [A. Field]

Leven-Test

Nicht signifikante Test bedeuten die Erfüllung der Anforderungen – Homogenität der Varianz.

```
library(car)
```

```
## Loading required package: carData
```

```
leveneTest(df$attractiveness, df$gender)
```

```
## Warning in leveneTest.default(df$attractiveness, df$gender): df$gender coerced
## to factor.

## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value   Pr(>F)
## group  1  19.979 5.08e-05 ***
##       46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
leveneTest(df$attractiveness, df$alcohol)
```

```
## Warning in leveneTest.default(df$attractiveness, df$alcohol): df$alcohol coerced
## to factor.

## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  2  2.3238 0.1095
##       45
```

```
leveneTest(df$attractiveness, interaction(df$alcohol, df$gender))
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  5  1.4252 0.2351
##       42
```

Kontraste

- Für *alcohol*

```
##           [,1] [,2]
## 2 Pints    -2    0
## 4 Pints     1   -1
## None        1    1
```

- Für *gender*

```
##          [,1]
## Female    -1
## Male       1
```

ANOVA

```r
model <- aov(attractiveness ~ alcohol*gender, data = df)
Anova(model, type = "III")
```

```
## Anova Table (Type III tests)
##
## Response: attractiveness
##                Sum Sq Df   F value    Pr(>F)
## (Intercept)    163333  1 1967.0251 < 2.2e-16 ***
## alcohol          3332  2   20.0654 7.649e-07 ***
## gender            169  1    2.0323    0.1614
## alcohol:gender   1978  2   11.9113 7.987e-05 ***
## Residuals        3487 42
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary.lm(model)
```

```
##
## Call:
## aov(formula = attractiveness ~ alcohol * gender, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -21.875  -5.625  -0.625   5.156  19.375
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        58.333      1.315  44.351  < 2e-16 ***
## alcohol1           -3.177      0.930  -3.416  0.00142 **
## alcohol2            8.594      1.611   5.335 3.57e-06 ***
## gender1            -1.875      1.315  -1.426  0.16138
## alcohol1:gender1   -2.031      0.930  -2.184  0.03459 *
## alcohol2:gender1    7.031      1.611   4.365 8.12e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.112 on 42 degrees of freedom
## Multiple R-squared:  0.6111, Adjusted R-squared:  0.5648
## F-statistic:  13.2 on 5 and 42 DF,  p-value: 9.609e-08
```

ANOVA 2 (Fortsetzung)

Vitaly Belik

Institut für Veterinär-Epidemiologie und Biometrie, FU Berlin

08/12/2020

Wiederholung

Types of sum of squares

Factorial ANOVA with repeated measures

Beispiel (Koma-Saufen)

Zweifache ANOVA mit wiederholten Messungen vergleicht verschiedene Gruppen im Falle von zwei unabhängigen Variablen, wobei alle Teilnehmer an allen Versuchsbedingungen teilgenommen haben.

Beispiel (Koma-Saufen)

Zweifache ANOVA mit wiederholten Messungen vergleicht verschiedene Gruppen im Falle von zwei unabhängigen Variablen, wobei alle Teilnehmer an allen Versuchsbedingungen teilgenommen haben.

Hat die Werbung mit negativen visuellen Inhalten eine Auswirkung auf die Einstellung zum excessiven Alkoholkomsum?

Daten (Koma-Saufen)

Table 1: Datensatz über Einstellung zum Trinken und Werbung [A. Field]

beerpos	beerneg	beerneut	winepos	wineneg	wineneut	waterpos	waterneg	waterneu	participant
1	6	5	38	-5	4	10	-14	-2	P1
43	30	8	20	-12	4	9	-10	-13	P2
15	15	12	20	-15	6	6	-16	1	P3
40	30	19	28	-4	0	20	-10	2	P4
8	12	8	11	-2	6	27	5	-5	P5
17	17	15	17	-6	6	9	-6	-13	P6
30	21	21	15	-2	16	19	-20	3	P7
34	23	28	27	-7	7	12	-12	2	P8
34	20	26	24	-10	12	12	-9	4	P9
26	27	27	23	-15	14	21	-6	0	P10

Beispiel

```
##     participant   groups attitude drink  imagery
## 1            P1  beerpos        1  Beer  Postive
## 21           P1  beerneg        6  Beer Negative
## 41           P1 beerneut        5  Beer  Neutral
## 61           P1  winepos       38  Wine  Postive
## 81           P1  wineneg       -5  Wine Negative
## 101          P1 wineneut        4  Wine  Neutral
## 121          P1 waterpos       10 Water  Postive
## 141          P1 waterneg      -14 Water Negative
## 161          P1 waterneu       -2 Water  Neutral
## 10          P10  beerpos       26  Beer  Postive
## 30          P10  beerneg       27  Beer Negative
## 50          P10 beerneut       27  Beer  Neutral
## 70          P10  winepos       23  Wine  Postive
## 90          P10  wineneg      -15  Wine Negative
## 110         P10 wineneut       14  Wine  Neutral
## 130         P10 waterpos       21 Water  Postive
## 150         P10 waterneg       -6 Water Negative
## 170         P10 waterneu        0 Water  Neutral
## 11          P11  beerpos        1  Beer  Postive
## 31          P11  beerneg      -19  Beer Negative
## 51          P11 beerneut      -10  Beer  Neutral
## 71          P11  winepos       28  Wine  Postive
## 91          P11  wineneg      -13  Wine Negative
## 111         P11 wineneut       13  Wine  Neutral
## 131         P11 waterpos       33 Water  Postive
## 151         P11 waterneg       -2 Water Negative
## 171         P11 waterneu        9 Water  Neutral
## 12          P12  beerpos        7  Beer  Postive
## 32          P12  beerneg      -18  Beer Negative
## 52          P12 beerneut        6  Beer  Neutral
## 72          P12  winepos       26  Wine  Postive
## 92          P12  wineneg      -16  Wine Negative
## 112         P12 wineneut       19  Wine  Neutral
## 132         P12 waterpos       23 Water  Postive
## 152         P12 waterneg      -17 Water Negative
## 172         P12 waterneu        5 Water  Neutral
## 13          P13  beerpos       22  Beer  Postive
## 33          P13  beerneg       -8  Beer Negative
## 53          P13 beerneut        4  Beer  Neutral
```

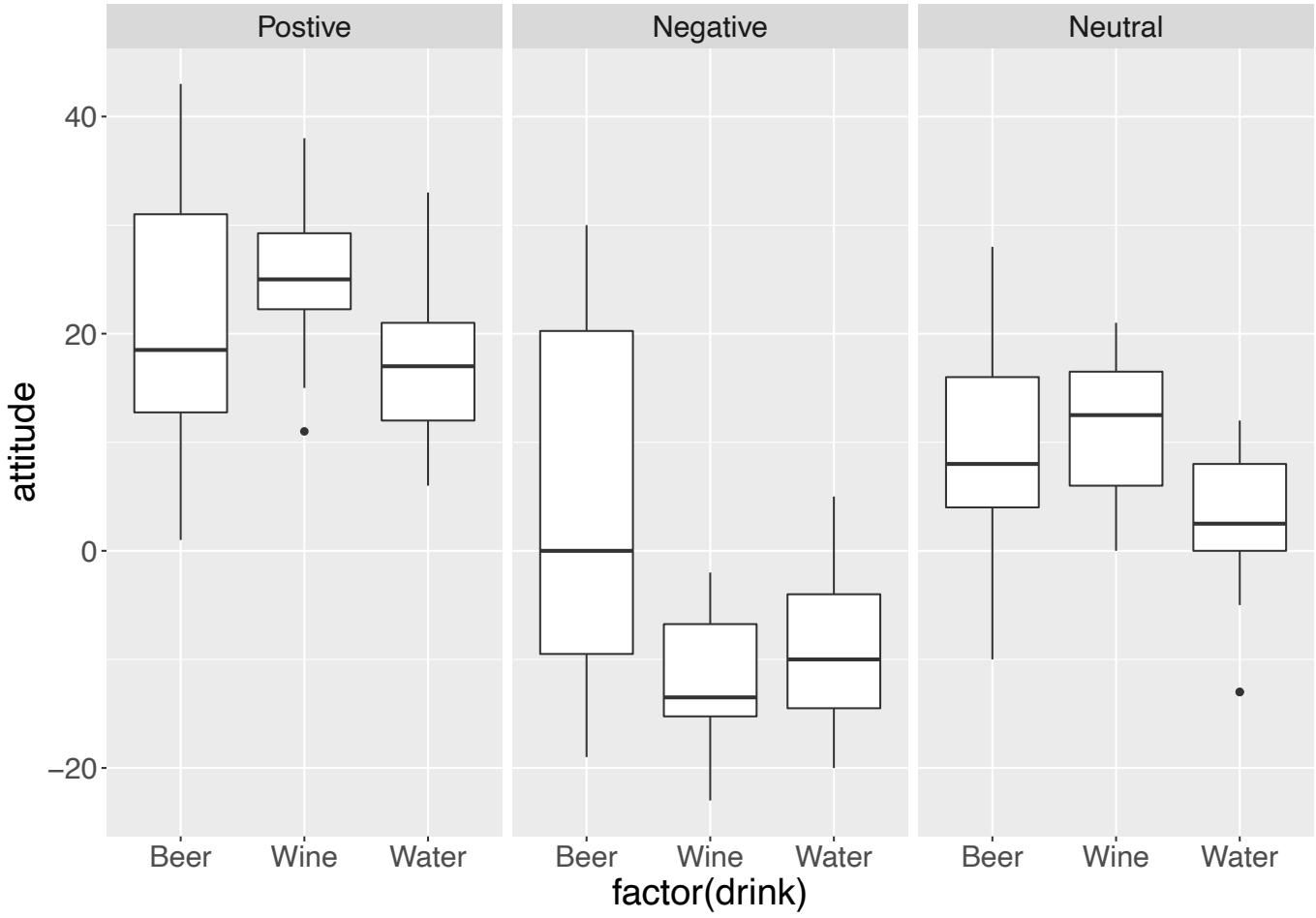


Figure 1: Boxplots von Daten über Alkoholkomsumeinstellung.

Beispiel (Deskriptive Statistik) I

```
## : Beer
## : Postive
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##   18.5000000   21.0500000    2.9086758    6.0879284  169.2078947   13.0079935
##     coef.var
##    0.6179569
## ------------------------------------------------------------
## : Wine
## : Postive
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##   25.0000000   25.3500000    1.5066083    3.1533673   45.3973684    6.7377569
##     coef.var
##    0.2657892
## ------------------------------------------------------------
## : Water
## : Postive
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##   17.0000000   17.4000000    1.5818044    3.3107547   50.0421053    7.0740445
##     coef.var
##    0.4065543
## ------------------------------------------------------------
## : Beer
## : Negative
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##     0.000000     4.450000     3.869227     8.098386   299.418421    17.303711
##     coef.var
##     3.888474
## ------------------------------------------------------------
## : Wine
## : Negative
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##  -13.5000000  -12.0000000    1.3822179    2.8930153   38.2105263    6.1814664
##     coef.var
##   -0.5151222
## ------------------------------------------------------------
## : Water
```

Beispiel (Deskriptive Statistik) II

```
## : Negative
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##  -10.0000000   -9.2000000    1.5210799    3.1836569   46.2736842    6.8024763
##     coef.var
##   -0.7393996
## ------------------------------------------------------------
## : Beer
## : Neutral
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##     8.000000    10.000000     2.302173     4.818503   106.000000    10.295630
##     coef.var
##     1.029563
## ------------------------------------------------------------
## : Wine
## : Neutral
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##   12.5000000   11.6500000    1.3959999    2.9218614   38.9763158    6.2431015
##     coef.var
##    0.5358885
## ------------------------------------------------------------
## : Water
## : Neutral
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##     2.500000     2.350000     1.529147     3.200541    46.765789     6.838552
##     coef.var
##     2.910022
```

Kontraste für Getränke

1. Ist die Wirkung für *alkoholische* und *nicht alkoholische* Getränke unterschiedlich?

Kontraste für Getränke

1. Ist die Wirkung für *alkoholische* und *nicht alkoholische* Getränke unterschiedlich?
2. Ist die Wirkung für *verschiedene alkoholische* Getränke unterschiedlich?

Kontraste für Getränke

1. Ist die Wirkung für *alkoholische* und *nicht alkoholische* Getränke unterschiedlich?
2. Ist die Wirkung für *verschiedene alkoholische* Getränke unterschiedlich?

```
##       AlcoholvsWater BeervsWine
## Beer               1         -1
## Wine               1          1
## Water             -2          0
```

Kontraste für Bilder

1. Haben *negative* Bilder andere Wirkung im Vergleich zu den anderen Formen?

Kontraste für Bilder

1. Haben *negative* Bilder andere Wirkung im Vergleich zu den anderen Formen?
2. Ist die Wirkung unterschiedlich für *positive* und *negative* Bilder?

Kontraste für Bilder

1. Haben *negative* Bilder andere Wirkung im Vergleich zu den anderen Formen?
2. Ist die Wirkung unterschiedlich für *positive* und *negative* Bilder?

```
##          NegativevsOther PositivevsNeutral
## Postive                1                -1
## Negative              -2                 0
## Neutral                1                 1
```

Factorielle ANOVA mit wiederholten Messungen

```
library(ez)
attitudeModel <- ezANOVA(data = longAttitude, dv = .(attitude), wid = .(participant),
    within = .(imagery, drink), type = 3, detailed = TRUE)
```

```
## Warning: Converting "participant" to factor for ANOVA.
```

```
attitudeModel
```

```
## $ANOVA
##          Effect DFn DFd       SSn      SSd          F            p p<.05
## 1   (Intercept)   1  19 11218.006 1920.106 111.005411 2.255322e-09     *
## 2       imagery   2  38 21628.678 3352.878 122.564825 2.680197e-17     *
## 3         drink   2  38  2092.344 7785.878   5.105981 1.086293e-02     *
## 4 imagery:drink   4  76  2624.422 2906.689  17.154922 4.589040e-10     *
##         ges
## 1 0.4126762
## 2 0.5753191
## 3 0.1158687
## 4 0.1411741
##
## $`Mauchly's Test for Sphericity`
##          Effect         W            p p<.05
## 2       imagery 0.6621013 2.445230e-02     *
## 3         drink 0.2672411 6.952302e-06     *
## 4 imagery:drink 0.5950440 4.356587e-01
##
## $`Sphericity Corrections`
##          Effect       GGe        p[GG] p[GG]<.05       HFe        p[HF]
## 2       imagery 0.7474407 1.757286e-13         * 0.7968420 3.142804e-14
## 3         drink 0.5771143 2.976868e-02         * 0.5907442 2.881391e-02
## 4 imagery:drink 0.7983979 1.900249e-08         * 0.9785878 6.809640e-10
##   p[HF]<.05
## 2         *
## 3         *
## 4         *
```

- Signifikanz vom Mauchly's Test für **drink** und **imagery** deutet auf Nichterfüllung der Sphärizität. Daher wird F-Verhältnis angepasst.

Hauptwirkung (main effect) von Getränk

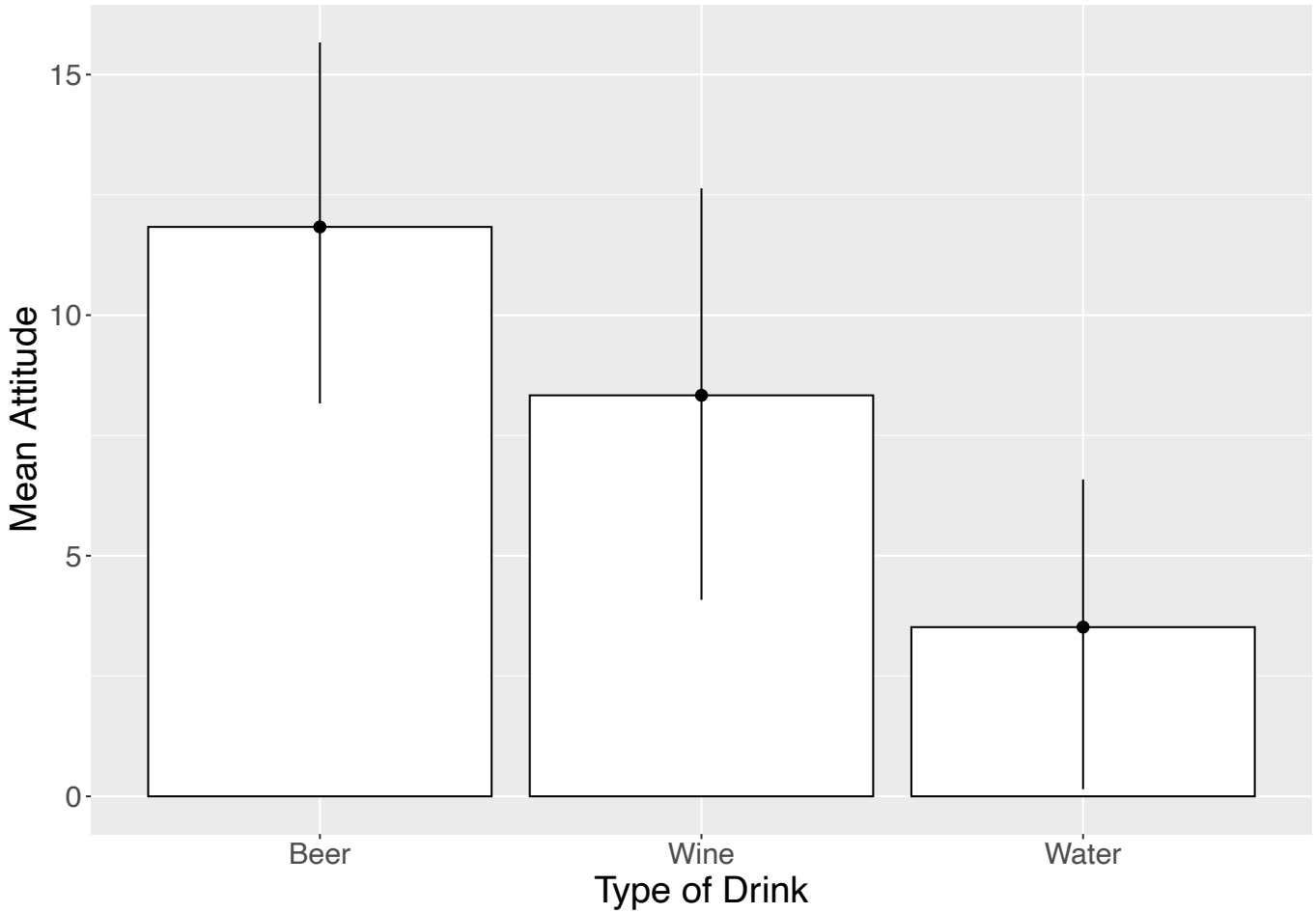


Figure 2: Balkendiagramme für die Hauptwirkung von Getränk.

Getränk: deskriptive Statistik

```
by(longAttitude$attitude, longAttitude$drink, stat.desc, basic = FALSE)
```

```
## longAttitude$drink: Beer
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##    12.500000    11.833333     1.972576     3.947115   233.463277    15.279505
##     coef.var
##     1.291226
## ------------------------------------------------------------
## longAttitude$drink: Wine
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##    12.000000     8.333333     2.166080     4.334316   281.514124    16.778383
##     coef.var
##     2.013406
## ------------------------------------------------------------
## longAttitude$drink: Water
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##     3.500000     3.516667     1.666806     3.335272   166.694633    12.911028
##     coef.var
##     3.671382
```

Hauptwirkung (main effect) von Bild

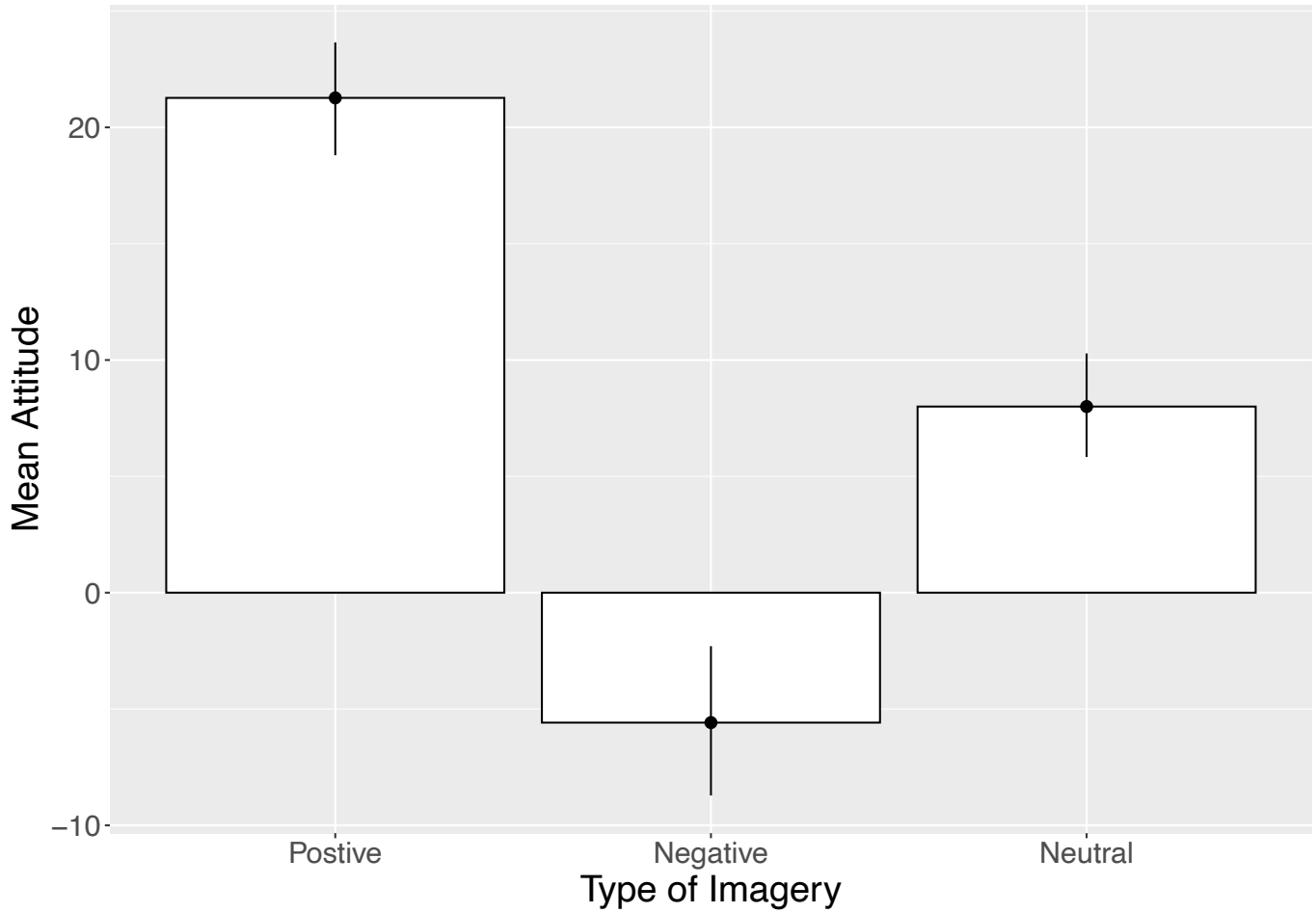


Figure 3: Balkendiagramme für die Hauptwirkung von Bild.

Bild: deskriptive Statistik

```
by(longAttitude$attitude, longAttitude$imagery, stat.desc, basic = FALSE)
```

```
## longAttitude$imagery: Postive
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##   20.5000000   21.2666667    1.2646579    2.5305747   95.9615819    9.7959983
##     coef.var
##    0.4606269
## ------------------------------------------------------------
## longAttitude$imagery: Negative
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##    -9.000000    -5.583333     1.713405     3.428516   176.145480    13.271981
##     coef.var
##    -2.377071
## ------------------------------------------------------------
## longAttitude$imagery: Neutral
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##     7.000000     8.000000     1.143392     2.287922    78.440678     8.856674
##     coef.var
##     1.107084
```

Wechselwirkungen **drink** × **imagery** (interactions)

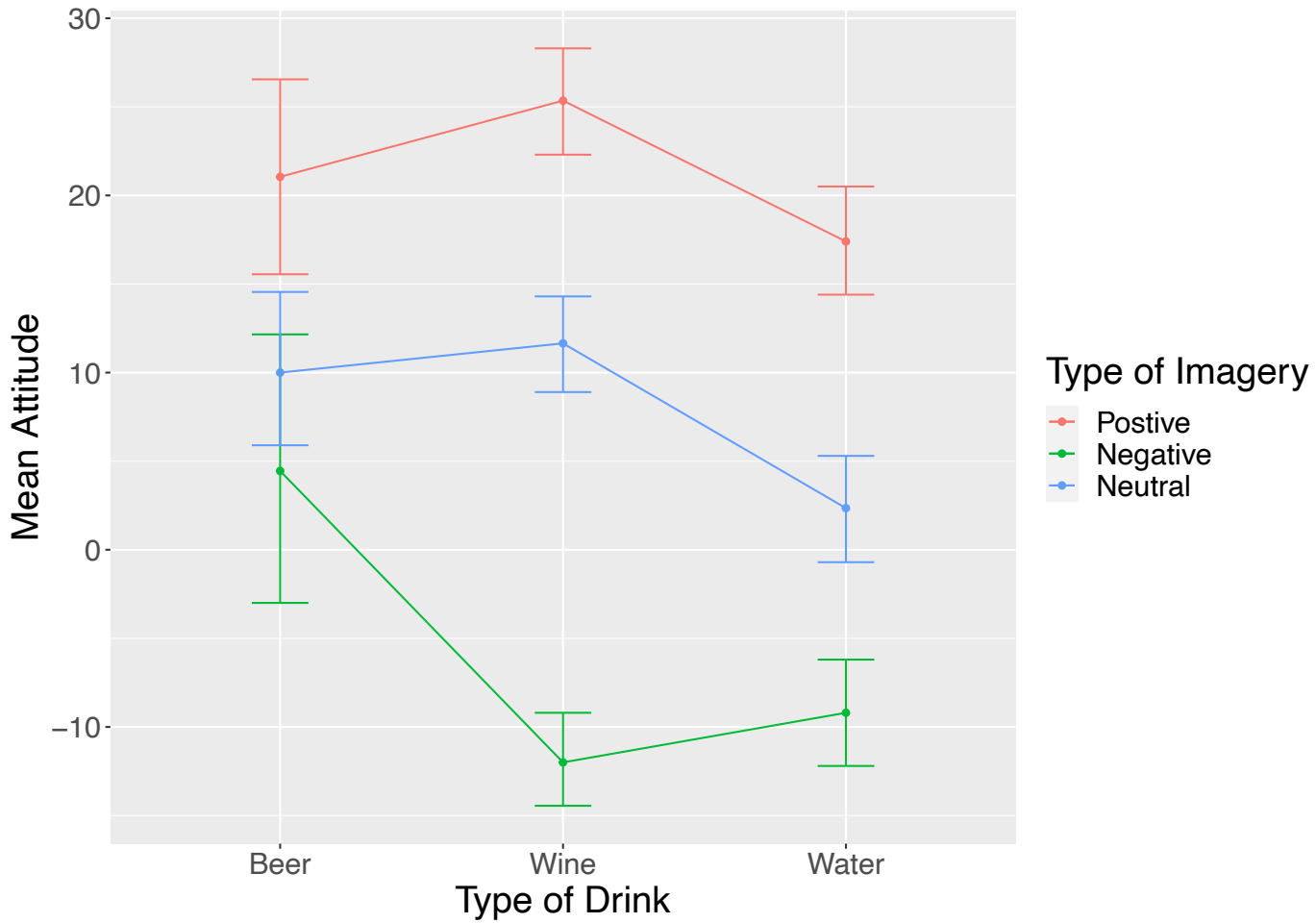


Figure 4: Wechselwirkung zwischen drink und imagery. Man sieht, dass die negativen Bilder andere Wirkung haben als positive oder neutrale Bilder.

Paarweiser *t*-Test (post hoc Test)

```
pairwise.t.test(longAttitude$attitude, longAttitude$groups, paired = TRUE, p.adjust.method = "bonferroni")
```

```
##
##  Pairwise comparisons using paired t tests
##
## data:  longAttitude$attitude and longAttitude$groups
##
##          beerpos beerneg beerneut winepos wineneg wineneut waterpos waterneg
## beerneg  0.00217 -       -        -       -       -        -        -
## beerneut 0.01982 1.00000 -        -       -       -        -        -
## winepos  1.00000 0.01105 0.00310  -       -       -        -        -
## wineneg  5.6e-08 0.00265 2.0e-07  1.9e-10 -       -        -        -
## wineneut 0.39905 1.00000 1.00000  2.2e-05 2.3e-07 -        -        -
## waterpos 1.00000 0.47584 1.00000  0.07300 1.3e-09 0.10547  -        -
## waterneg 2.9e-06 0.18860 0.00010  3.2e-10 1.00000 1.1e-07  4.9e-11  -
## waterneu 0.00212 1.00000 0.74838  4.3e-10 0.00041 8.1e-05  9.0e-07  0.00068
##
## P value adjustment method: bonferroni
```

```
options(digits = 7)
```

Faktorelles Design mit wiederholten Messungen als GLM

```
library(nlme)
baseline <- lme(attitude ~ 1, random = ~1 | participant/drink/imagery, data = longAttitude,
    method = "ML")
```

random $= \sim 1|$*participant/drink/imagery* bedeutet, dass für den zufälligen Teil des Modells die Variablen **drink** und **imagery** s.g. *nested* Variablen innerhalb der Variablen **participant** sind. Die Werte für diese Variablen sind für jeden Teilnehmer vorhanden.

Faktorelles Design mit wiederholten Messungen als GLM

```
library(nlme)
baseline <- lme(attitude ~ 1, random = ~1 | participant/drink/imagery, data = longAttitude,
    method = "ML")
```

random = ~1|participant/drink/imagery bedeutet, dass für den zufälligen Teil des Modells die Variablen **drink** und **imagery** s.g. *nested* Variablen innerhalb der Variablen **participant** sind. Die Werte für diese Variablen sind für jeden Teilnehmer vorhanden.

Falls wir den Haupteffekt von einzelnen Faktoren untersuchen wollen, müssen wir sie dem Modell beifügen.

```
drinkModel <- update(baseline, . ~ . + drink)
imageryModel <- update(drinkModel, . ~ . + imagery)
attitudeModel <- update(imageryModel, . ~ . + drink:imagery)
```

ANOVA-Vergleich von Modellen

```
anova(baseline, drinkModel, imageryModel, attitudeModel)
```

```
##                Model df      AIC      BIC    logLik   Test   L.Ratio p-value
## baseline           1  5 1503.590 1519.555 -746.7950
## drinkModel         2  7 1498.461 1520.812 -742.2306 1 vs 2   9.12891  0.0104
## imageryModel       3  9 1350.529 1379.265 -666.2644 2 vs 3 151.93237  <.0001
## attitudeModel      4 13 1316.512 1358.020 -645.2560 3 vs 4  42.01676  <.0001
```

Modell-Ausgabe I

```
summary(attitudeModel)
```

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: longAttitude
##        AIC     BIC   logLik
##   1316.512 1358.02 -645.256
##
## Random effects:
##  Formula: ~1 | participant
##          (Intercept)
## StdDev: 0.0007437312
##
##  Formula: ~1 | drink %in% participant
##         (Intercept)
## StdDev:    6.044143
##
##  Formula: ~1 | imagery %in% drink %in% participant
##         (Intercept)  Residual
## StdDev:    7.217035 0.2803831
##
## Fixed effects: attitude ~ drink + imagery + drink:imagery
##                                                Value Std.Error  DF   t-value
## (Intercept)                                 7.894444 0.9726049 114  8.116805
## drinkAlcoholvsWater                         2.188889 0.6877355  38  3.182748
## drinkBeervsWine                            -1.750000 1.1911928  38 -1.469116
## imageryNegativevsOther                      6.738889 0.3905470 114 17.255002
## imageryPositivevsNeutral                   -6.633333 0.6764472 114 -9.806136
## drinkAlcoholvsWater:imageryNegativevsOther  0.190278 0.2761584 114  0.689017
## drinkBeervsWine:imageryNegativevsOther      3.237500 0.4783204 114  6.768475
## drinkAlcoholvsWater:imageryPositivevsNeutral 0.445833 0.4783204 114  0.932081
## drinkBeervsWine:imageryPositivevsNeutral   -0.662500 0.8284753 114 -0.799662
##                                               p-value
## (Intercept)                                    0.0000
## drinkAlcoholvsWater                            0.0029
## drinkBeervsWine                                0.1500
## imageryNegativevsOther                         0.0000
```

Modell-Ausgabe II

```
## imageryPositivevsNeutral                       0.0000
## drinkAlcoholvsWater:imageryNegativevsOther     0.4922
## drinkBeervsWine:imageryNegativevsOther         0.0000
## drinkAlcoholvsWater:imageryPositivevsNeutral   0.3533
## drinkBeervsWine:imageryPositivevsNeutral       0.4256
##  Correlation:
##                                               (Intr) drnkAW drnkBW imgrNO imgrPN
## drinkAlcoholvsWater                           0
## drinkBeervsWine                               0      0
## imageryNegativevsOther                        0      0      0
## imageryPositivevsNeutral                      0      0      0      0
## drinkAlcoholvsWater:imageryNegativevsOther    0      0      0      0      0
## drinkBeervsWine:imageryNegativevsOther        0      0      0      0      0
## drinkAlcoholvsWater:imageryPositivevsNeutral 0      0      0      0      0
## drinkBeervsWine:imageryPositivevsNeutral      0      0      0      0      0
##                                               dAW:NO dBW:NO dAW:PN
## drinkAlcoholvsWater
## drinkBeervsWine
## imageryNegativevsOther
## imageryPositivevsNeutral
## drinkAlcoholvsWater:imageryNegativevsOther
## drinkBeervsWine:imageryNegativevsOther        0
## drinkAlcoholvsWater:imageryPositivevsNeutral 0      0
## drinkBeervsWine:imageryPositivevsNeutral      0      0      0
##
## Standardized Within-Group Residuals:
##          Min           Q1          Med           Q3          Max
## -0.086767486 -0.020849266  0.000768403  0.025120590  0.103677229
##
## Number of Observations: 180
## Number of Groups:
##                          participant              drink %in% participant
##                                   20                                  60
## imagery %in% drink %in% participant
##                                  180
```

Post hoc-Tests (**drink**) I

```
library(multcomp)
postHocs <- glht(attitudeModel, linfct = mcp(drink = "Tukey"))
```

```
## Warning in mcp2matrix(model, linfct = linfct): covariate interactions found --
## default contrast might be inappropriate
```

```
summary(postHocs)
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lme.formula(fixed = attitude ~ drink + imagery + drink:imagery,
##     data = longAttitude, random = ~1 | participant/drink/imagery,
##     method = "ML")
##
## Linear Hypotheses:
##                   Estimate Std. Error z value Pr(>|z|)
## Wine - Beer == 0    -3.500      2.322  -1.507  0.28743
## Water - Beer == 0   -8.317      2.322  -3.582  0.00102 **
## Water - Wine == 0   -4.817      2.322  -2.074  0.09522 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

```
confint(postHocs)
```

Post hoc-Tests (**drink**) II

```
##
##   Simultaneous Confidence Intervals
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lme.formula(fixed = attitude ~ drink + imagery + drink:imagery,
##     data = longAttitude, random = ~1 | participant/drink/imagery,
##     method = "ML")
##
## Quantile = 2.3432
## 95% family-wise confidence level
##
##
## Linear Hypotheses:
##                   Estimate lwr      upr
## Wine - Beer == 0   -3.5000  -8.9411   1.9411
## Water - Beer == 0  -8.3167 -13.7577  -2.8756
## Water - Wine == 0  -4.8167 -10.2577   0.6244
```

Post hoc-Tests (**drink**) I

```
postHocs <- glht(attitudeModel, linfct = mcp(imagery = "Tukey"))
```

```
## Warning in mcp2matrix(model, linfct = linfct): covariate interactions found --
## default contrast might be inappropriate
```

```
summary(postHocs)
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lme.formula(fixed = attitude ~ drink + imagery + drink:imagery,
##     data = longAttitude, random = ~1 | participant/drink/imagery,
##     method = "ML")
##
## Linear Hypotheses:
##                          Estimate Std. Error z value Pr(>|z|)
## Negative - Postive == 0  -26.850      1.319  -20.36   <2e-16 ***
## Neutral - Postive == 0   -13.267      1.319  -10.06   <2e-16 ***
## Neutral - Negative == 0   13.583      1.319   10.30   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

```
confint(postHocs)
```

Post hoc-Tests (**drink**) II

```
##
##   Simultaneous Confidence Intervals
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lme.formula(fixed = attitude ~ drink + imagery + drink:imagery,
##     data = longAttitude, random = ~1 | participant/drink/imagery,
##     method = "ML")
##
## Quantile = 2.3437
## 95% family-wise confidence level
##
##
## Linear Hypotheses:
##                          Estimate lwr      upr
## Negative - Postive == 0 -26.8500 -29.9405 -23.7595
## Neutral - Postive == 0  -13.2667 -16.3572 -10.1761
## Neutral - Negative == 0  13.5833  10.4928  16.6739
```

mixed design ANOVA

Mixed design ANOVA

Mischung aus zwischen-Gruppen-Variablen und Variablen mit wiederholten Messungen nennt man *gemischtes Design*.

Mixed design ANOVA

Mischung aus zwischen-Gruppen-Variablen und Variablen mit wiederholten Messungen nennt man *gemischtes Design*.

Es is empfehlenswert mit nicht mehr als *drei* unabhängigen Variablen zu arbeiten. Anderenfalls können die Interaktionen schwer interpretiert werden.

Beispiel (Speed Dating)

Beispiel: Beim Speed Dating werden die Personen beurteilt nach *Attraktivität* (**looks**) und *Charisma* (**personality**). Die beiden Variablen beinhalten wiederholte Messungen. Die Person, die die Beurteilung abgibt, kann männlich oder weiblich sein. Deher ist Geschlecht (**gender**) eine zwischen-Gruppen-Variable.

Beispiel (Speed Dating) (1)

Table 2: Datensatz über Speed Dating [A. Field]

participant	gender	att_high	av_high	ug_high	att_some	av_some	ug_some	att_none	av_none	ug_nor
P01	Male	86	84	67	88	69	50	97	48	4
P02	Male	91	83	53	83	74	48	86	50	4
P03	Male	89	88	48	99	70	48	90	45	4
P04	Male	89	69	58	86	77	40	87	47	5
P05	Male	80	81	57	88	71	50	82	50	4
P06	Male	80	84	51	96	63	42	92	48	4
P07	Male	89	85	61	87	79	44	86	50	4
P08	Male	100	94	56	86	71	54	84	54	4
P09	Male	90	74	54	92	71	58	78	38	4
P10	Male	89	86	63	80	73	49	91	48	3

Beispiel (Speed Dating) (2)

```
##     participant gender   groups dateRating personality      looks
## 1           P01   Male att_high         86 Charismatic Attractive
## 21          P01   Male  av_high         84 Charismatic    Average
## 41          P01   Male  ug_high         67 Charismatic       Ugly
## 61          P01   Male att_some         88     Average Attractive
## 81          P01   Male  av_some         69     Average    Average
## 101         P01   Male  ug_some         50     Average       Ugly
## 121         P01   Male att_none         97     Dullard Attractive
## 141         P01   Male  av_none         48     Dullard    Average
## 161         P01   Male  ug_none         47     Dullard       Ugly
## 2           P02   Male att_high         91 Charismatic Attractive
```

Speed Dating (Boxplots)

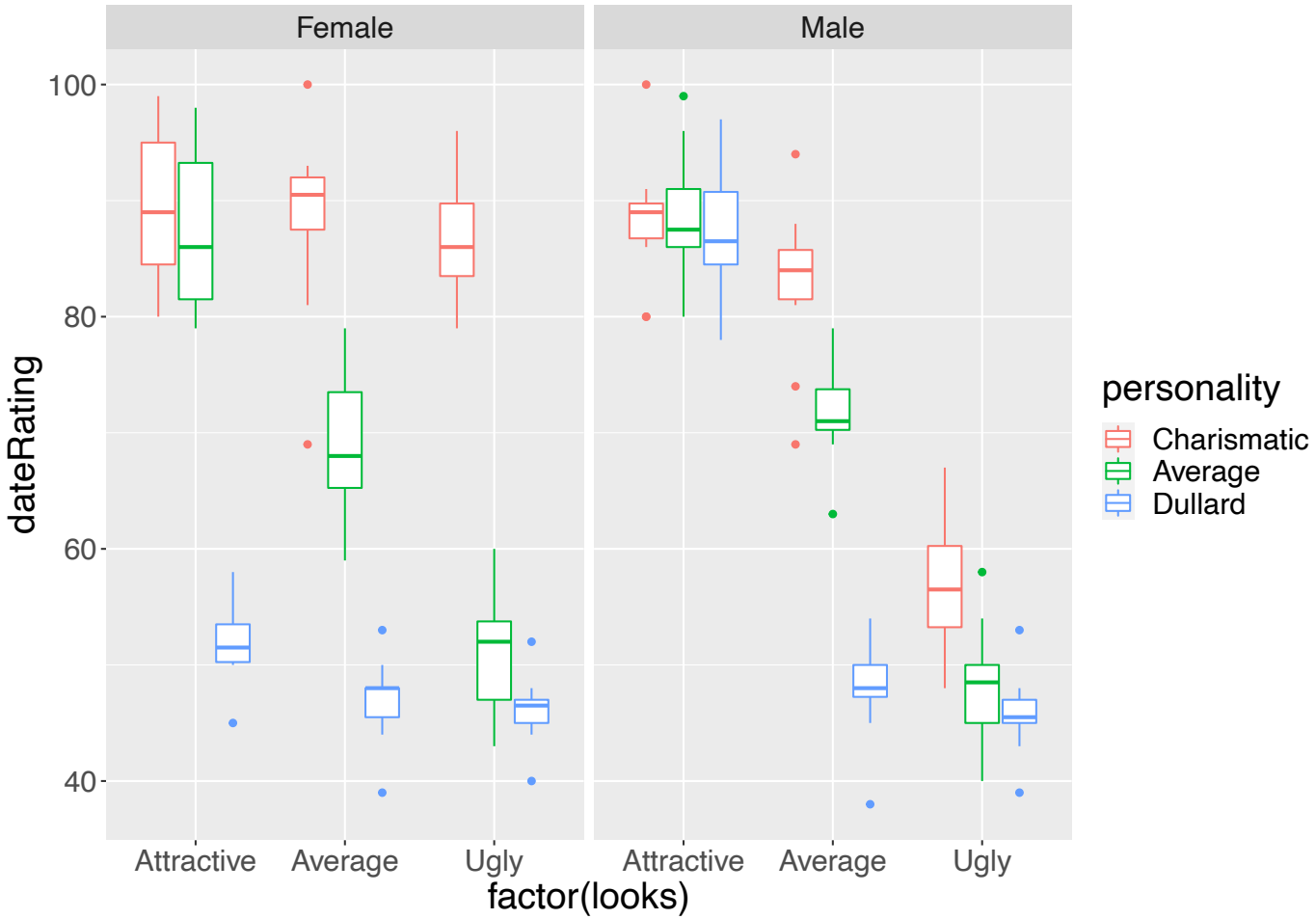


Figure 5: Boxplots von Daten über Speed Dating.

Speed Dating (Deskriptive Statistik) I

```
by(speedData$dateRating, list(speedData$looks, speedData$personality, speedData$gender),
    stat.desc, basic = FALSE)
```

```
## : Attractive
## : Charismatic
## : Female
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##  89.00000000  89.60000000   2.09867683   4.74753683  44.04444444   6.63659886
##     coef.var
##   0.07406918
## ------------------------------------------------------------
## : Average
## : Charismatic
## : Female
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##  90.50000000  88.40000000   2.63396617   5.95844544  69.37777778   8.32933237
##     coef.var
##   0.09422322
## ------------------------------------------------------------
## : Ugly
## : Charismatic
## : Female
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##  86.00000000  86.70000000   1.71949605   3.88977031  29.56666667   5.43752395
##     coef.var
##   0.06271654
## ------------------------------------------------------------
## : Attractive
## : Average
## : Female
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##  86.00000000  87.10000000   2.15225979   4.86874991  46.32222222   6.80604307
##     coef.var
##   0.07814056
## ------------------------------------------------------------
## : Average
```

Speed Dating (Deskriptive Statistik) II

```
## : Average
## : Female
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##  68.00000000  68.90000000   1.88237439   4.25822670  35.43333333   5.95259047
##     coef.var
##   0.08639464
## ------------------------------------------------------------
## : Ugly
## : Average
## : Female
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##   52.0000000   51.2000000    1.7243356    3.9007182   29.7333333    5.4528280
##     coef.var
##    0.1065005
## ------------------------------------------------------------
## : Attractive
## : Dullard
## : Female
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##  51.50000000  51.80000000   1.09341463   2.47347574  11.95555556   3.45768066
##     coef.var
##   0.06675059
## ------------------------------------------------------------
## : Average
## : Dullard
## : Female
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##  48.00000000  47.00000000   1.18321596   2.67662045  14.00000000   3.74165739
##     coef.var
##   0.07960973
## ------------------------------------------------------------
## : Ugly
## : Dullard
## : Female
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##  46.50000000  46.10000000   0.97125349   2.19712803   9.43333333   3.07137320
```

Speed Dating (Deskriptive Statistik) III

```
##     coef.var
##   0.06662415
## ------------------------------------------------------------
## : Attractive
## : Charismatic
## : Male
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##  89.00000000  88.30000000   1.80154255   4.07537238  32.45555556   5.69697776
##     coef.var
##   0.06451843
## ------------------------------------------------------------
## : Average
## : Charismatic
## : Male
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##  84.00000000  82.80000000   2.21509970   5.01090365  49.06666667   7.00476029
##     coef.var
##   0.08459855
## ------------------------------------------------------------
## : Ugly
## : Charismatic
## : Male
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##    56.500000    56.800000     1.812304     4.099716    32.844444     5.731007
##     coef.var
##     0.100898
## ------------------------------------------------------------
## : Attractive
## : Average
## : Male
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##  87.50000000  88.50000000   1.81506045   4.10595200  32.94444444   5.73972512
##     coef.var
##   0.06485565
## ------------------------------------------------------------
## : Average
```

Speed Dating (Deskriptive Statistik) IV

```
## : Average
## : Male
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##  71.00000000  71.80000000   1.39682179   3.15983042  19.51111111   4.41713834
##     coef.var
##   0.06152003
## ------------------------------------------------------------
## : Ugly
## : Average
## : Male
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##   48.5000000   48.3000000    1.7000000    3.8456672   28.9000000    5.3758720
##     coef.var
##    0.1113017
## ------------------------------------------------------------
## : Attractive
## : Dullard
## : Male
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##   86.5000000   87.3000000    1.7194961    3.8897703   29.5666667    5.4375239
##     coef.var
##    0.0622855
## ------------------------------------------------------------
## : Average
## : Dullard
## : Male
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##  48.00000000  47.80000000   1.32329555   2.99350251  17.51111111   4.18462795
##     coef.var
##   0.08754452
## ------------------------------------------------------------
## : Ugly
## : Dullard
## : Male
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##  45.50000000  45.80000000   1.13333333   2.56377812  12.84444444   3.58391468
```

Speed Dating (Deskriptive Statistik) V

```
##     coef.var
##   0.07825141
```

Kontraste

Man betrachte die untersten Kategorien (*Dullard* und *Ugly*) als Kontrolle.

1. Kontrast für **personality** vergleicht *Average* und *Charismatic* mit *Dullard*.
2. Kontrast für **personality** vergleicht *Charismatic* mit *Average*.
3. Kontrast für **looks** vergleicht *Average* und *Attractive* mit *Ugly*.
4. Kontrast für **looks** vergleicht *Attractive* mit *Average*.

(orthogonale) Kontraste

(orthogonale) Kontraste

```
##              SomevsNone HivsAv
## Charismatic           1      1
## Average               1     -1
## Dullard              -2      0
```

(orthogonale) Kontraste

```
##              SomevsNone HivsAv
## Charismatic           1      1
## Average               1     -1
## Dullard              -2      0
```

```
##              AttractivevsUgly AttractivevsAv
## Attractive                  1              1
## Average                     1             -1
## Ugly                       -2              0
```

Factorielle ANOVA mit gemischtem Design I

between bedeutet dass **gender** ist zwischen-Gruppen-Variable und *within* bedeutet, dass **looks** und **personality** are Variablen mit wiederholten Messungen.

```
library(ez)
speedModel <- ezANOVA(data = speedData, dv = .(dateRating), wid = .(participant),
    between = .(gender), within = .(looks, personality), type = 3, detailed = TRUE)
```

```
## Warning: Converting "participant" to factor for ANOVA.
```

```
## Warning: Converting "gender" to factor for ANOVA.
```

```
speedModel
```

```
## $ANOVA
##                       Effect DFn DFd        SSn       SSd            F
## 1                (Intercept)   1  18 846249.800  760.2222 2.003690e+04
## 2                     gender   1  18      0.200  760.2222 4.735457e-03
## 3                      looks   2  36  20779.633  882.7111 4.237325e+02
## 5                personality   2  36  23233.600 1274.0444 3.282498e+02
## 4              gender:looks   2  36   3944.100  882.7111 8.042699e+01
## 6        gender:personality   2  36   4420.133 1274.0444 6.244868e+01
## 7         looks:personality   4  72   4055.267 1992.6222 3.663253e+01
## 8 gender:looks:personality   4  72   2669.667 1992.6222 2.411596e+01
##                p p<.05          ges
## 1 7.013239e-29     * 9.942319e-01
## 2 9.458958e-01       4.073486e-05
## 3 9.594812e-26     * 8.088849e-01
## 5 7.689430e-24     * 8.255493e-01
## 4 5.234187e-14     * 4.454748e-01
## 6 1.974986e-12     * 4.737685e-01
## 7 1.101308e-16     * 4.523510e-01
## 8 1.107801e-12     * 3.522328e-01
```

Factorielle ANOVA mit gemischtem Design II

```
##
## $`Mauchly's Test for Sphericity`
##                     Effect         W         p p<.05
## 3                    looks 0.9602054 0.7081010
## 4              gender:looks 0.9602054 0.7081010
## 5               personality 0.9293298 0.5363446
## 6        gender:personality 0.9293298 0.5363446
## 7         looks:personality 0.6133545 0.5339382
## 8 gender:looks:personality 0.6133545 0.5339382
##
## $`Sphericity Corrections`
##                     Effect       GGe        p[GG] p[GG]<.05       HFe
## 3                    looks 0.9617284 7.624114e-25         * 1.0744125
## 4              gender:looks 0.9617284 1.487026e-13         * 1.0744125
## 5               personality 0.9339944 2.056621e-22         * 1.0380537
## 6        gender:personality 0.9339944 9.442426e-12         * 1.0380537
## 7         looks:personality 0.7993543 9.003598e-14         * 0.9922411
## 8 gender:looks:personality 0.7993543 1.470422e-10         * 0.9922411
##          p[HF] p[HF]<.05
## 3 9.594812e-26         *
## 4 5.234187e-14         *
## 5 7.689430e-24         *
## 6 1.974986e-12         *
## 7 1.426883e-16         *
## 8 1.337876e-12         *
```

Factorielle ANOVA mit gemischtem Design als *glm*

Wir definieren neue (nicht orthogonale) Kontraste um z.B. die Vergleiche zu der Norm zu machen.

1. Kontrast vergleicht für **looks** *Attractive* mit *Average* (baseline = 0)
2. Kontrast vergleicht für **looks** *Ugly* mit *Average* (baseline = 0)

```
AttractivevsAv <- c(1, 0, 0)
UglyvsAv <- c(0, 0, 1)
contrasts(speedData$looks) <- cbind(AttractivevsAv, UglyvsAv)
```

Factorielle ANOVA mit gemischtem Design als *glm* (1)

1. Kontrast vergleicht für **personality** *Charismatic* mit *Average* (baseline = 0)
2. Kontrast vergleicht für **personality** *Dullard* mit *Average* (baseline = 0)

```
HighvsAv <- c(1, 0, 0)
DullvsAv <- c(0, 0, 1)
contrasts(speedData$personality) <- cbind(HighvsAv, DullvsAv)
```

(nicht orthogonale) Kontraste

```
attr(speedData$looks, "contrasts")
```

```
##             AttractivevsAv UglyvsAv
## Attractive               1        0
## Average                  0        0
## Ugly                     0        1
```

```
attr(speedData$personality, "contrasts")
```

```
##              HighvsAv DullvsAv
## Charismatic         1        0
## Average             0        0
## Dullard             0        1
```

Factorielle ANOVA mit gemischtem Design als *glm* (2)

```
baseline <- lme(dateRating ~ 1, random = ~1 | participant/looks/personality, data = speedData,
    method = "ML")
```

Sieht ähnlich aus wie faktorielle ANOVA mit wiederholten Messungen.

Factorielle ANOVA mit gemischtem Design als *glm* (2)

```
baseline <- lme(dateRating ~ 1, random = ~1 | participant/looks/personality, data = speedData,
    method = "ML")
```

Sieht ähnlich aus wie faktorielle ANOVA mit wiederholten Messungen.

Weiter Faktoren werde hinzugefügt.

```
looksM <- update(baseline, . ~ . + looks)
personalityM <- update(looksM, . ~ . + personality)
genderM <- update(personalityM, . ~ . + gender)
looks_gender <- update(genderM, . ~ . + looks:gender)
personality_gender <- update(looks_gender, . ~ . + personality:gender)
looks_personality <- update(personality_gender, . ~ . + looks:personality)
speedDateModel <- update(looks_personality, . ~ . + looks:personality:gender)
```

ANOVA-Tabelle I

```
anova(baseline, looksM, personalityM, genderM, looks_gender, personality_gender,
    looks_personality, speedDateModel)
```

```
##                     Model df      AIC      BIC    logLik   Test   L.Ratio
## baseline                1  5 1575.766 1591.730 -782.8829
## looksM                  2  7 1511.468 1533.819 -748.7343 1 vs 2  68.29719
## personalityM            3  9 1376.704 1405.441 -679.3520 2 vs 3 138.76442
## genderM                 4 10 1378.702 1410.632 -679.3511 3 vs 4   0.00180
## looks_gender            5 12 1343.161 1381.477 -659.5808 4 vs 5  39.54079
## personality_gender      6 14 1289.198 1333.899 -630.5988 5 vs 6  57.96394
## looks_personality       7 18 1220.057 1277.530 -592.0283 6 vs 7  77.14102
## speedDateModel          8 22 1148.462 1218.707 -552.2309 7 vs 8  79.59473
##                     p-value
## baseline
## looksM               <.0001
## personalityM         <.0001
## genderM              0.9662
## looks_gender         <.0001
## personality_gender   <.0001
## looks_personality    <.0001
## speedDateModel       <.0001
```

```
summary(speedDateModel)
```

ANOVA-Tabelle II

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: speedData
##        AIC      BIC    logLik
##   1148.462 1218.707 -552.2309
##
## Random effects:
##  Formula: ~1 | participant
##         (Intercept)
## StdDev:    1.158402
##
##  Formula: ~1 | looks %in% participant
##          (Intercept)
## StdDev: 0.0005251677
##
##  Formula: ~1 | personality %in% looks %in% participant
##         (Intercept)  Residual
## StdDev:    5.090892 0.1283062
##
## Fixed effects: dateRating ~ looks + personality + gender + looks:gender + personality:gender +      looks:personality + look
##                                                 Value Std.Error  DF  t-value
## (Intercept)                                      68.9  1.740866 108 39.57799
## looksAttractivevsAv                              18.2  2.400632  36  7.58134
## looksUglyvsAv                                   -17.7  2.400632  36 -7.37306
## personalityHighvsAv                              19.5  2.400632 108  8.12286
## personalityDullvsAv                             -21.9  2.400632 108 -9.12260
## genderMale                                        2.9  2.461957  18  1.17792
## looksAttractivevsAv:genderMale                   -1.5  3.395006  36 -0.44183
## looksUglyvsAv:genderMale                         -5.8  3.395006  36 -1.70839
## personalityHighvsAv:genderMale                   -8.5  3.395006 108 -2.50368
## personalityDullvsAv:genderMale                   -2.1  3.395006 108 -0.61856
## looksAttractivevsAv:personalityHighvsAv         -17.0  3.395006 108 -5.00736
## looksUglyvsAv:personalityHighvsAv                16.0  3.395006 108  4.71280
## looksAttractivevsAv:personalityDullvsAv         -13.4  3.395006 108 -3.94697
## looksUglyvsAv:personalityDullvsAv                16.8  3.395006 108  4.94845
## looksAttractivevsAv:personalityHighvsAv:genderMale   5.8  4.801263 108  1.20802
## looksUglyvsAv:personalityHighvsAv:genderMale       -18.5  4.801263 108 -3.85315
```

ANOVA-Tabelle III

```
## looksAttractivevsAv:personalityDullvsAv:genderMale  36.2  4.801263 108  7.53968
## looksUglyvsAv:personalityDullvsAv:genderMale         4.7  4.801263 108  0.97891
##                                                    p-value
## (Intercept)                                         0.0000
## looksAttractivevsAv                                 0.0000
## looksUglyvsAv                                       0.0000
## personalityHighvsAv                                 0.0000
## personalityDullvsAv                                 0.0000
## genderMale                                          0.2542
## looksAttractivevsAv:genderMale                      0.6613
## looksUglyvsAv:genderMale                            0.0962
## personalityHighvsAv:genderMale                      0.0138
## personalityDullvsAv:genderMale                      0.5375
## looksAttractivevsAv:personalityHighvsAv             0.0000
## looksUglyvsAv:personalityHighvsAv                   0.0000
## looksAttractivevsAv:personalityDullvsAv             0.0001
## looksUglyvsAv:personalityDullvsAv                   0.0000
## looksAttractivevsAv:personalityHighvsAv:genderMale  0.2297
## looksUglyvsAv:personalityHighvsAv:genderMale        0.0002
## looksAttractivevsAv:personalityDullvsAv:genderMale  0.0000
## looksUglyvsAv:personalityDullvsAv:genderMale        0.3298
##  Correlation:
##                                                    (Intr) lksAtA lksUgA prsnHA
## looksAttractivevsAv                                -0.689
## looksUglyvsAv                                      -0.689  0.500
## personalityHighvsAv                                -0.689  0.500  0.500
## personalityDullvsAv                                -0.689  0.500  0.500  0.500
## genderMale                                         -0.707  0.488  0.488  0.488
## looksAttractivevsAv:genderMale                      0.488 -0.707 -0.354 -0.354
## looksUglyvsAv:genderMale                            0.488 -0.354 -0.707 -0.354
## personalityHighvsAv:genderMale                      0.488 -0.354 -0.354 -0.707
## personalityDullvsAv:genderMale                      0.488 -0.354 -0.354 -0.354
## looksAttractivevsAv:personalityHighvsAv             0.488 -0.707 -0.354 -0.707
## looksUglyvsAv:personalityHighvsAv                   0.488 -0.354 -0.707 -0.707
## looksAttractivevsAv:personalityDullvsAv             0.488 -0.707 -0.354 -0.354
## looksUglyvsAv:personalityDullvsAv                   0.488 -0.354 -0.707 -0.354
```

ANOVA-Tabelle IV

```
## looksAttractivevsAv:personalityHighvsAv:genderMale -0.345  0.500  0.250  0.500
## looksUglyvsAv:personalityHighvsAv:genderMale       -0.345  0.250  0.500  0.500
## looksAttractivevsAv:personalityDullvsAv:genderMale -0.345  0.500  0.250  0.250
## looksUglyvsAv:personalityDullvsAv:genderMale       -0.345  0.250  0.500  0.250
##                                                    prsnDA gndrMl lkAA:M lkUA:M
## looksAttractivevsAv
## looksUglyvsAv
## personalityHighvsAv
## personalityDullvsAv
## genderMale                                          0.488
## looksAttractivevsAv:genderMale                     -0.354 -0.689
## looksUglyvsAv:genderMale                           -0.354 -0.689  0.500
## personalityHighvsAv:genderMale                     -0.354 -0.689  0.500  0.500
## personalityDullvsAv:genderMale                     -0.707 -0.689  0.500  0.500
## looksAttractivevsAv:personalityHighvsAv            -0.354 -0.345  0.500  0.250
## looksUglyvsAv:personalityHighvsAv                  -0.354 -0.345  0.250  0.500
## looksAttractivevsAv:personalityDullvsAv            -0.707 -0.345  0.500  0.250
## looksUglyvsAv:personalityDullvsAv                  -0.707 -0.345  0.250  0.500
## looksAttractivevsAv:personalityHighvsAv:genderMale  0.250  0.488 -0.707 -0.354
## looksUglyvsAv:personalityHighvsAv:genderMale        0.250  0.488 -0.354 -0.707
## looksAttractivevsAv:personalityDullvsAv:genderMale  0.500  0.488 -0.707 -0.354
## looksUglyvsAv:personalityDullvsAv:genderMale        0.500  0.488 -0.354 -0.707
##                                                    prHA:M prDA:M lkAA:HA
## looksAttractivevsAv
## looksUglyvsAv
## personalityHighvsAv
## personalityDullvsAv
## genderMale
## looksAttractivevsAv:genderMale
## looksUglyvsAv:genderMale
## personalityHighvsAv:genderMale
## personalityDullvsAv:genderMale                      0.500
## looksAttractivevsAv:personalityHighvsAv             0.500  0.250
## looksUglyvsAv:personalityHighvsAv                   0.500  0.250  0.500
## looksAttractivevsAv:personalityDullvsAv             0.250  0.500  0.500
## looksUglyvsAv:personalityDullvsAv                   0.250  0.500  0.250
```

ANOVA-Tabelle V

```
## looksAttractivevsAv:personalityHighvsAv:genderMale -0.707 -0.354 -0.707
## looksUglyvsAv:personalityHighvsAv:genderMale       -0.707 -0.354 -0.354
## looksAttractivevsAv:personalityDullvsAv:genderMale -0.354 -0.707 -0.354
## looksUglyvsAv:personalityDullvsAv:genderMale       -0.354 -0.707 -0.177
##                                                    lkUA:HA lkAA:DA lkUA:DA
## looksAttractivevsAv
## looksUglyvsAv
## personalityHighvsAv
## personalityDullvsAv
## genderMale
## looksAttractivevsAv:genderMale
## looksUglyvsAv:genderMale
## personalityHighvsAv:genderMale
## personalityDullvsAv:genderMale
## looksAttractivevsAv:personalityHighvsAv
## looksUglyvsAv:personalityHighvsAv
## looksAttractivevsAv:personalityDullvsAv             0.250
## looksUglyvsAv:personalityDullvsAv                   0.500   0.500
## looksAttractivevsAv:personalityHighvsAv:genderMale -0.354  -0.354  -0.177
## looksUglyvsAv:personalityHighvsAv:genderMale       -0.707  -0.177  -0.354
## looksAttractivevsAv:personalityDullvsAv:genderMale -0.177  -0.707  -0.354
## looksUglyvsAv:personalityDullvsAv:genderMale       -0.354  -0.354  -0.707
##                                                    lAA:HA: lUA:HA: lAA:DA:
## looksAttractivevsAv
## looksUglyvsAv
## personalityHighvsAv
## personalityDullvsAv
## genderMale
## looksAttractivevsAv:genderMale
## looksUglyvsAv:genderMale
## personalityHighvsAv:genderMale
## personalityDullvsAv:genderMale
## looksAttractivevsAv:personalityHighvsAv
## looksUglyvsAv:personalityHighvsAv
## looksAttractivevsAv:personalityDullvsAv
## looksUglyvsAv:personalityDullvsAv
```

ANOVA-Tabelle VI

```
## looksAttractivevsAv:personalityHighvsAv:genderMale
## looksUglyvsAv:personalityHighvsAv:genderMale        0.500
## looksAttractivevsAv:personalityDullvsAv:genderMale  0.500   0.250
## looksUglyvsAv:personalityDullvsAv:genderMale        0.250   0.500   0.500
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -0.09479343 -0.01471239  0.00142862  0.01218635  0.05289021
##
## Number of Observations: 180
## Number of Groups:
##                              participant                  looks %in% participant
##                                       20                                      60
## personality %in% looks %in% participant
##                                      180
```

Haupteffekt von **gender**

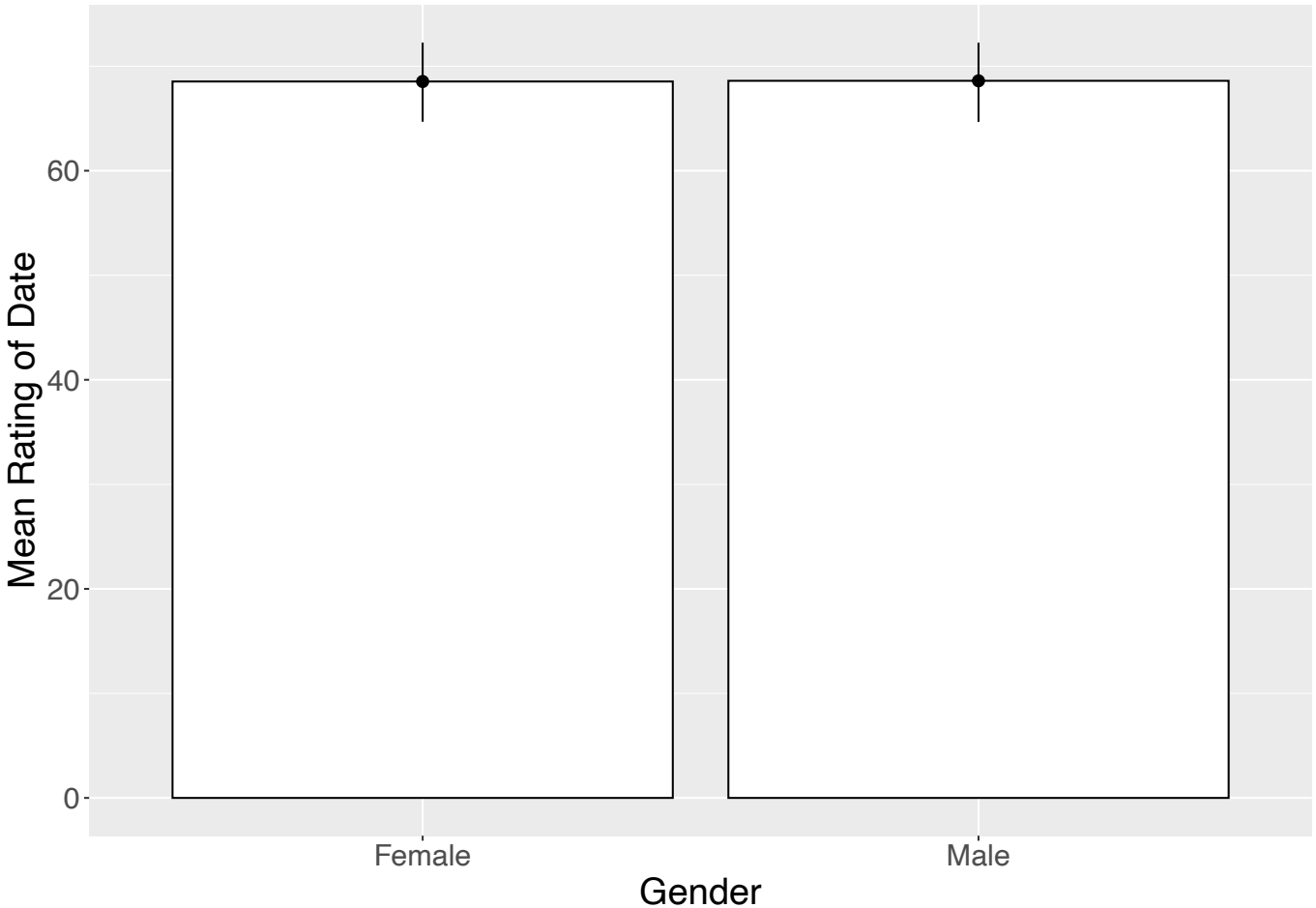


Figure 6: Haupteffekt von gender.

Haupteffekt von **looks**

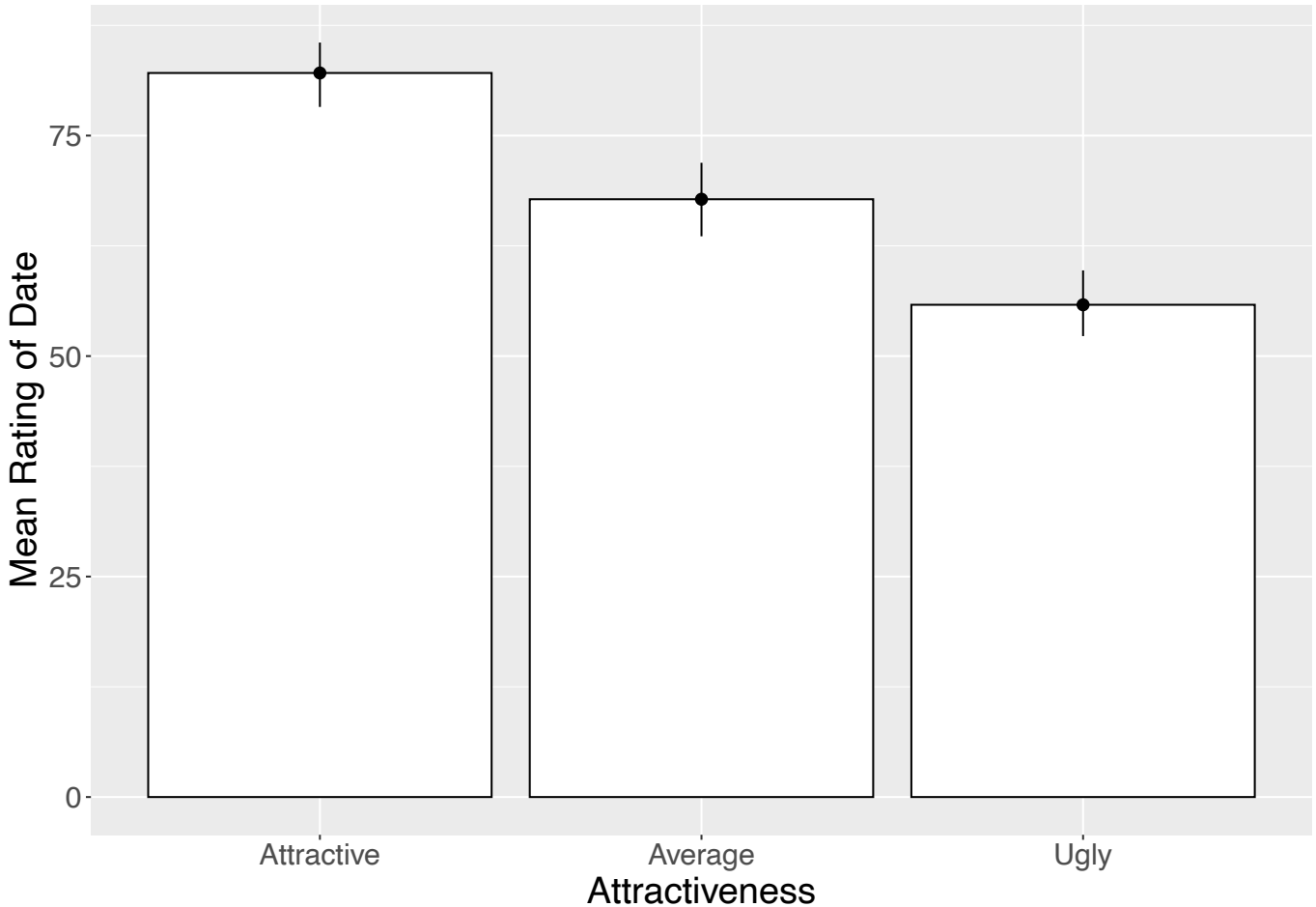


Figure 7: Haupteffekt von looks.

Haupeffekt von **personality**

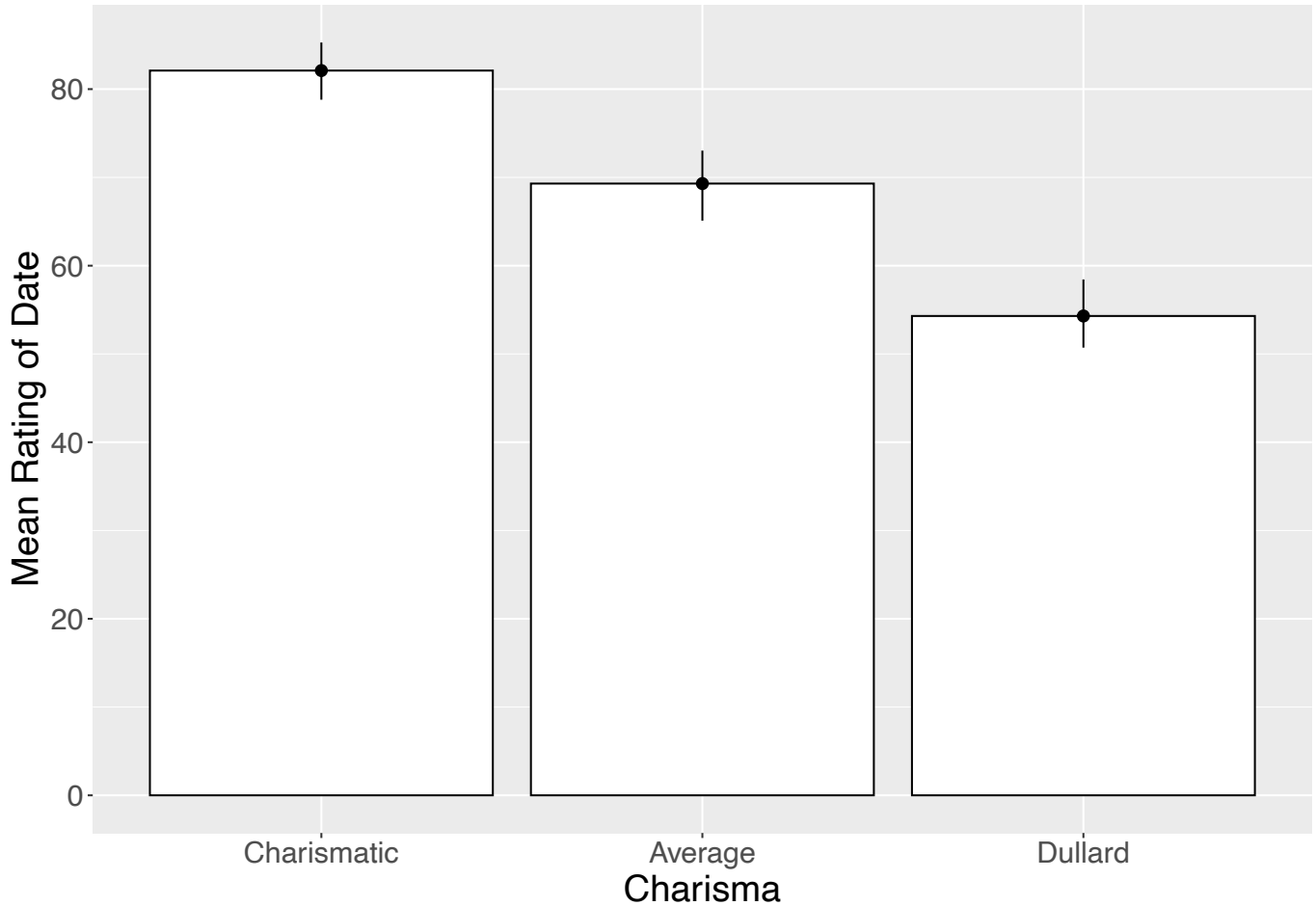


Figure 8: Haupteffekt von personality.

Wechselwirkung **looks**×**gender**

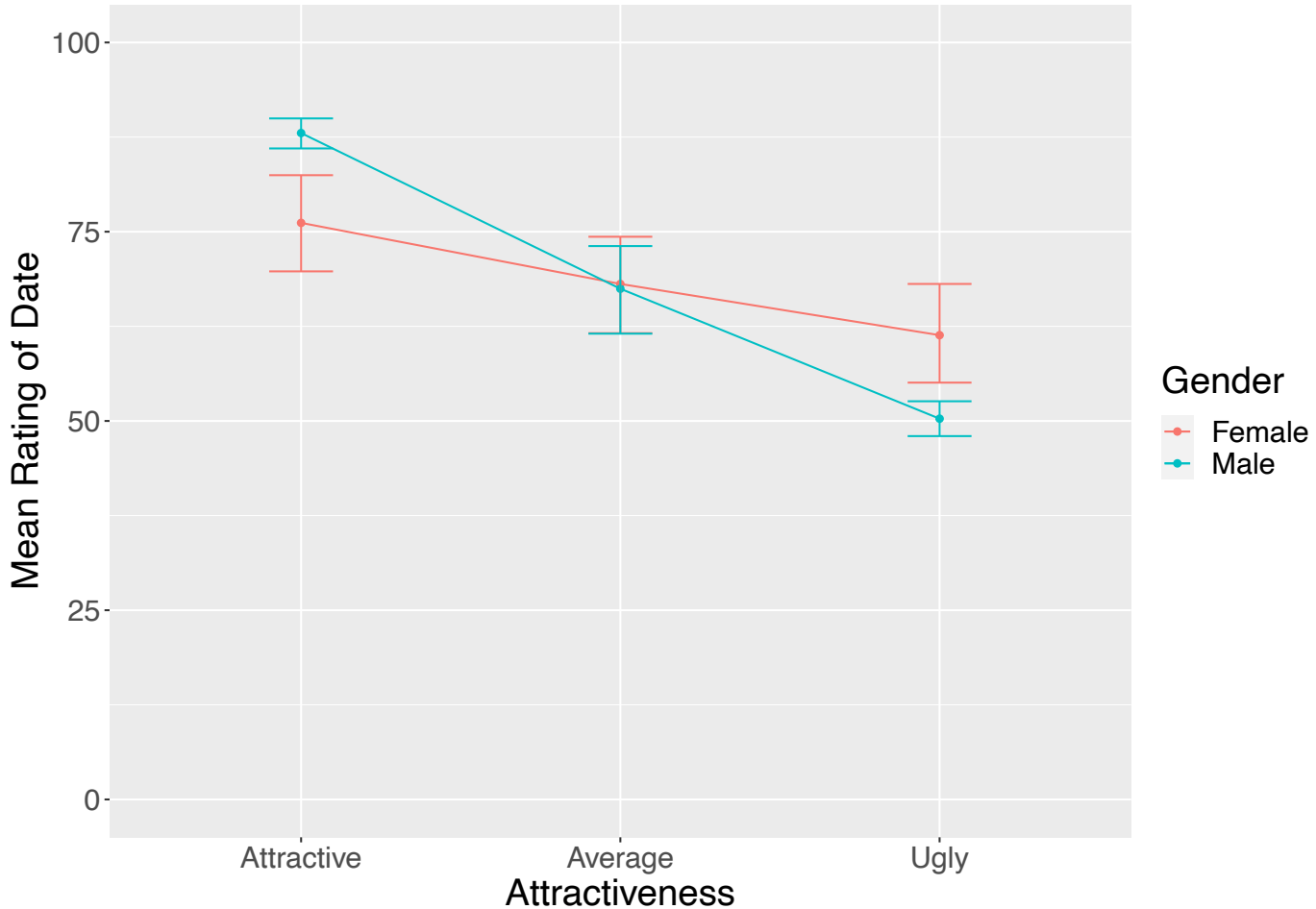


Figure 9: Wechselwirkung zwischen looks und gender.

Wechselwirkung **personality**×**gender**

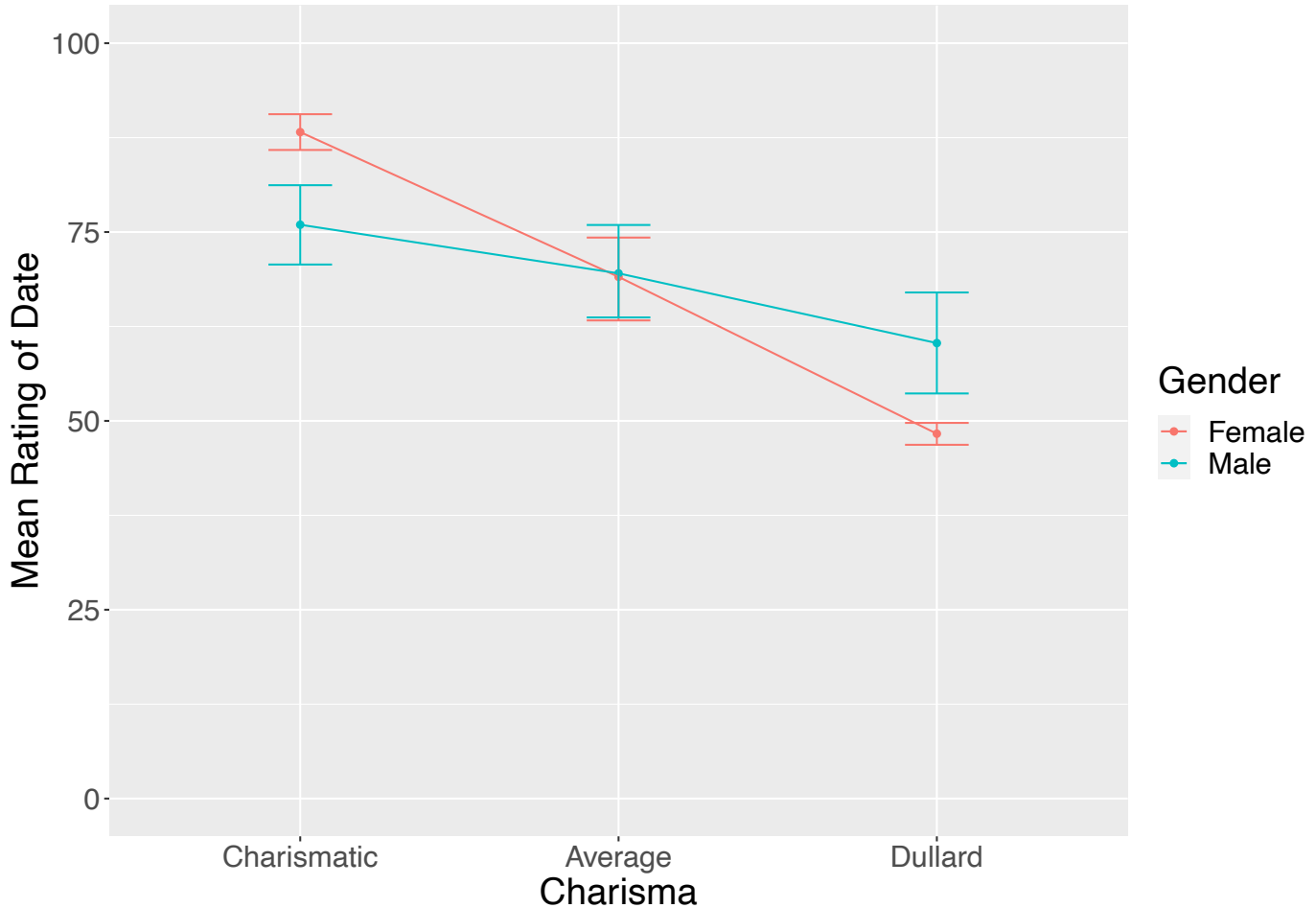


Figure 10: Wechselwirkung looks*imes*gender. Frauen sind mehr von Charisma beeinflusst als Männer.

Wechselwirkung **looks**×**personality**

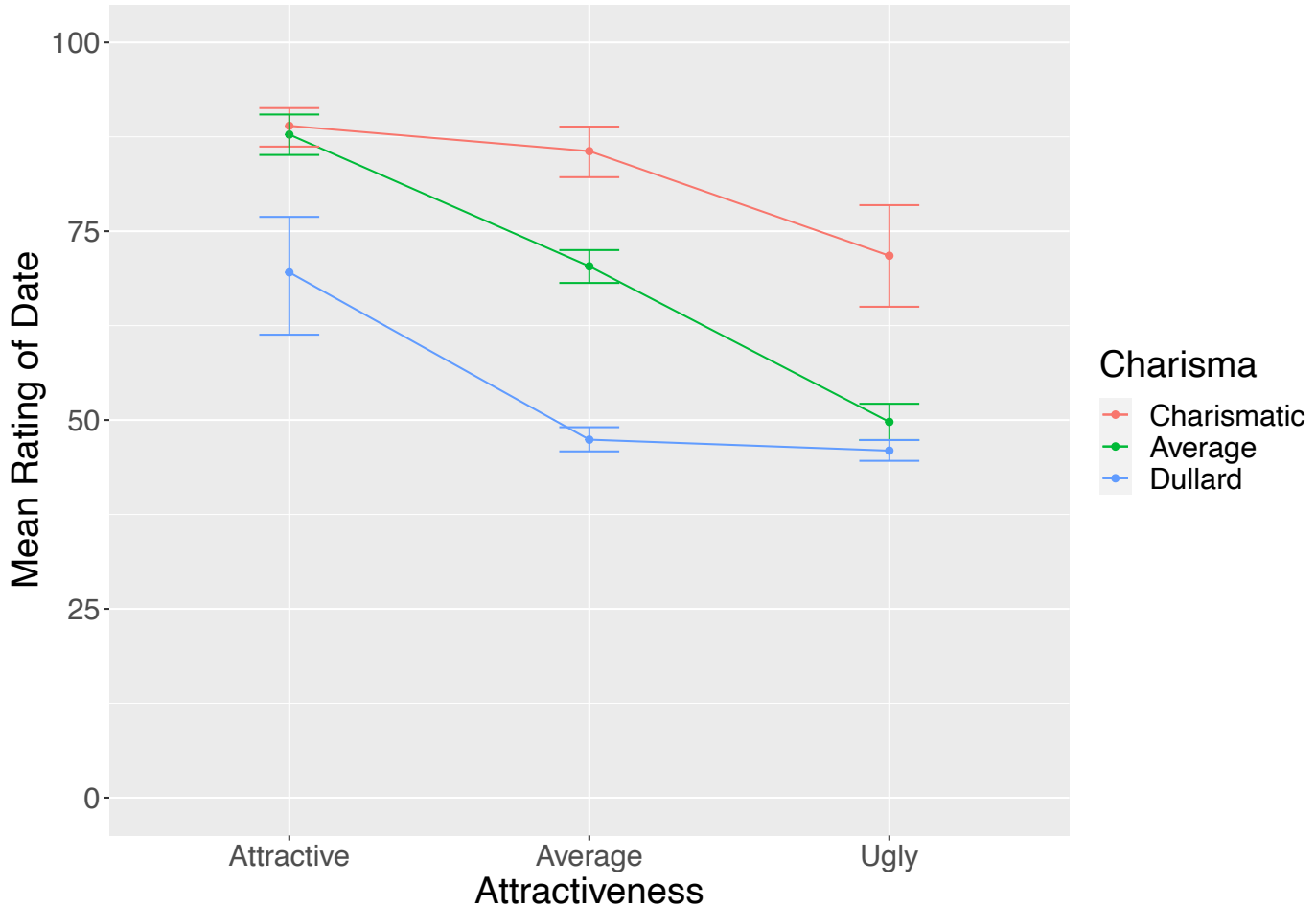


Figure 11: Wechselwirkung zwischen looks und personality.

Wechselwirkung **looks**×**personality**×**gender**

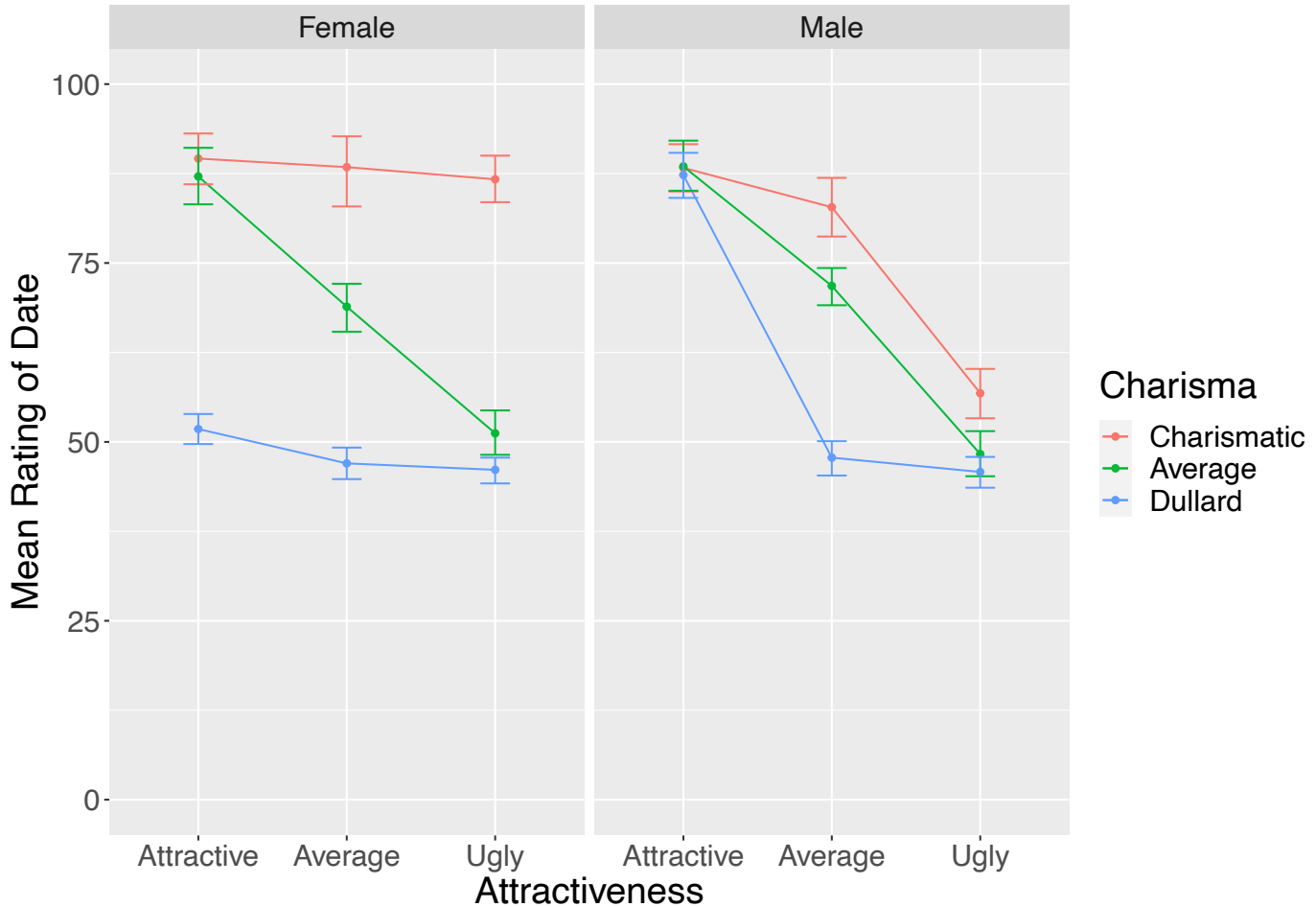


Figure 12: Wechselwirkung zwischen looks, personality und gender

Generalized Linear Models

Vitaly Belik

Institut für Veterinär-Epidemiologie und Biometrie, FU Berlin

15/12/2020

Wiederholung

Generalized Linear Models

Generalized Linear Models

In der Statistik ist das *generalisierte lineare Modell* (GLM) eine flexible Verallgemeinerung der gewöhnlichen linearen Regression, die Antwortvariablen mit anderen (exponenziellen) Fehlerverteilungsmodellen als einer Normalverteilung ermöglicht.

$$y = \beta_0 + \sum_{i=1}^{p} \beta_i X_i,\ y \sim \mathcal{N}(\mu, \sigma)$$

Das GLM verallgemeinert die lineare Regression, indem das lineare Modell über eine *Verknüpfungsfunktion* (*Link function*) mit der Antwortvariablen in Beziehung gesetzt wird und die Größe der Varianz jeder Messung von ihrem vorhergesagten Wert abhängt.

- Linearität nur in den Koeffizienten

$$g(y) = \beta_0 + \sum_{i=1}^{p} \beta_i X_i$$

Generalized Linear Models

- binomiale (dichotomische) logistische Regression
- multinominale logistische Regression
- Poissonsche Regression (count data)

Table 1: Link functions

Datentyp	Transformation	Verteilung
kontinuierlich	$\log(x)$	Log-normal
Anzahl	$\sqrt{x}$ oder $\log(x + 0.5)$	Poisson, Negative binomiale, ...
Verhältnis	$\arcsin\sqrt{x}$ oder logit $= \log\frac{x}{1-x}$	Bernoulli / binomiale, Beta binomiale, ...

Logistische Regression

Es gibt Situationen wann die *Antwortvariable* nicht normal verteilt ist. Z.B. kann sie kategoriell und *binomial* oder *multinomial* sein.

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \sum_{i=1}^{p} \beta_i X_i.$$

Dabei ist $\pi = \mu_Y$ ein bedingter Mittelwert (d.h. die Wahrscheinlichkeit, dass $Y = 1$ vorausgesetzt die vorhandenen X-Werte).

$\frac{\pi}{1-\pi}$ ist das Odds-Ratio, dass $Y = 1$.

$\log\left(\frac{\pi}{1-\pi}\right)$ ist *log odds* oder *logit*.

Logistische Regression: Beispiel

Welche persönliche, demographische, und Beziehungsvariablen können Untreue vorhersagen?

Table 2: Auszug aus dem Datensatz über Untreueverhalten [nach Kabacoff / Green&Fair]

	affairs	gender	age	yearsmarried	children	religiousness	education	occupation	rating
4	0	male	37	10.00	no	3	18	7	4
5	0	female	27	4.00	no	4	14	6	4
11	0	female	32	15.00	yes	1	12	1	4
16	0	male	57	15.00	yes	5	18	6	5
23	0	male	22	0.75	no	2	17	6	3
29	0	female	32	1.50	no	2	17	5	5
44	0	female	22	0.75	no	2	12	1	3
45	0	male	57	15.00	yes	2	14	4	4
47	0	female	32	15.00	yes	4	16	1	2
49	0	male	22	1.50	no	4	14	4	5
50	0	male	37	15.00	yes	2	20	7	2
55	0	male	27	4.00	yes	4	18	6	4
64	0	male	47	15.00	yes	5	17	6	4
80	0	female	22	1.50	no	2	17	5	4
86	0	female	27	4.00	no	4	14	5	4

Logistische Regression: Beispiel(1)

```
summary(Affairs)
```

```
##     affairs          gender         age         yearsmarried    children
##  Min.   : 0.000   female:315   Min.   :17.50   Min.   : 0.125   no :171
##  1st Qu.: 0.000   male  :286   1st Qu.:27.00   1st Qu.: 4.000   yes:430
##  Median : 0.000                Median :32.00   Median : 7.000
##  Mean   : 1.456                Mean   :32.49   Mean   : 8.178
##  3rd Qu.: 0.000                3rd Qu.:37.00   3rd Qu.:15.000
##  Max.   :12.000                Max.   :57.00   Max.   :15.000
##  religiousness     education       occupation       rating
##  Min.   :1.000   Min.   : 9.00   Min.   :1.000   Min.   :1.000
##  1st Qu.:2.000   1st Qu.:14.00   1st Qu.:3.000   1st Qu.:3.000
##  Median :3.000   Median :16.00   Median :5.000   Median :4.000
##  Mean   :3.116   Mean   :16.17   Mean   :4.195   Mean   :3.932
##  3rd Qu.:4.000   3rd Qu.:18.00   3rd Qu.:6.000   3rd Qu.:5.000
##  Max.   :5.000   Max.   :20.00   Max.   :7.000   Max.   :5.000
```

Logistische Regression: Beispiel(1)

```
knitr::kable(table(Affairs$affairs))
```

Var1	Freq
0	451
1	34
2	17
3	19
7	42
12	38

Logistische Regression: Beispiel(2)

- Transformation zu binären Variablen

```
Affairs$ynaffair[Affairs$affair > 0] <- 1
Affairs$ynaffair[Affairs$affair == 0] <- 0
Affairs$ynaffair <- factor(Affairs$ynaffair, levels = c(0, 1), labels = c("No", "yes"))
knitr::kable(table(Affairs$ynaffair))
```

Var1	Freq
No	451
yes	150

Logistische Regression: Beispiel(3)

```
fit.full <- glm(ynaffair ~ gender + age + yearsmarried + children + religiousness +
    education + occupation + rating, data = Affairs, family = binomial())
summary(fit.full)
```

```
##
## Call:
## glm(formula = ynaffair ~ gender + age + yearsmarried + children +
##     religiousness + education + occupation + rating, family = binomial(),
##     data = Affairs)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5713  -0.7499  -0.5690  -0.2539   2.5191
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.37726    0.88776   1.551 0.120807
## gendermale     0.28029    0.23909   1.172 0.241083
## age           -0.04426    0.01825  -2.425 0.015301 *
## yearsmarried   0.09477    0.03221   2.942 0.003262 **
## childrenyes    0.39767    0.29151   1.364 0.172508
## religiousness -0.32472    0.08975  -3.618 0.000297 ***
## education      0.02105    0.05051   0.417 0.676851
## occupation     0.03092    0.07178   0.431 0.666630
## rating        -0.46845    0.09091  -5.153 2.56e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 675.38  on 600  degrees of freedom
## Residual deviance: 609.51  on 592  degrees of freedom
## AIC: 627.51
##
## Number of Fisher Scoring iterations: 4
```

Beispiel: reduziertes Modell

```
fit.reduced <- glm(ynaffair ~ age + yearsmarried + religiousness + rating, data = Affairs,
    family = binomial())
summary(fit.reduced)
```

```
##
## Call:
## glm(formula = ynaffair ~ age + yearsmarried + religiousness +
##     rating, family = binomial(), data = Affairs)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6278  -0.7550  -0.5701  -0.2624   2.3998
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.93083    0.61032   3.164 0.001558 **
## age           -0.03527    0.01736  -2.032 0.042127 *
## yearsmarried   0.10062    0.02921   3.445 0.000571 ***
## religiousness -0.32902    0.08945  -3.678 0.000235 ***
## rating        -0.46136    0.08884  -5.193 2.06e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 675.38  on 600  degrees of freedom
## Residual deviance: 615.36  on 596  degrees of freedom
## AIC: 625.36
##
## Number of Fisher Scoring iterations: 4
```

Beispiel: Modellvergleich (χ^2)

```r
anova(fit.reduced, fit.full, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: ynaffair ~ age + yearsmarried + religiousness + rating
## Model 2: ynaffair ~ gender + age + yearsmarried + children + religiousness +
##     education + occupation + rating
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       596     615.36
## 2       592     609.51  4   5.8474   0.2108
```

Es wird getestet ob die Reduzierung der Restsumme der Quadrate statistisch signifikant ist oder nicht.

Beispiel: Interpretation der Koeffezienten

Regressionskoeffizienten geben die Veränderung (in *log(odds)*) in der Antwortvariable, wenn alle weiteren Variablen konstant bleiben.

```
coef(fit.reduced)
```

```
##   (Intercept)           age  yearsmarried religiousness        rating
##    1.93083017   -0.03527112    0.10062274   -0.32902386   -0.46136144
```

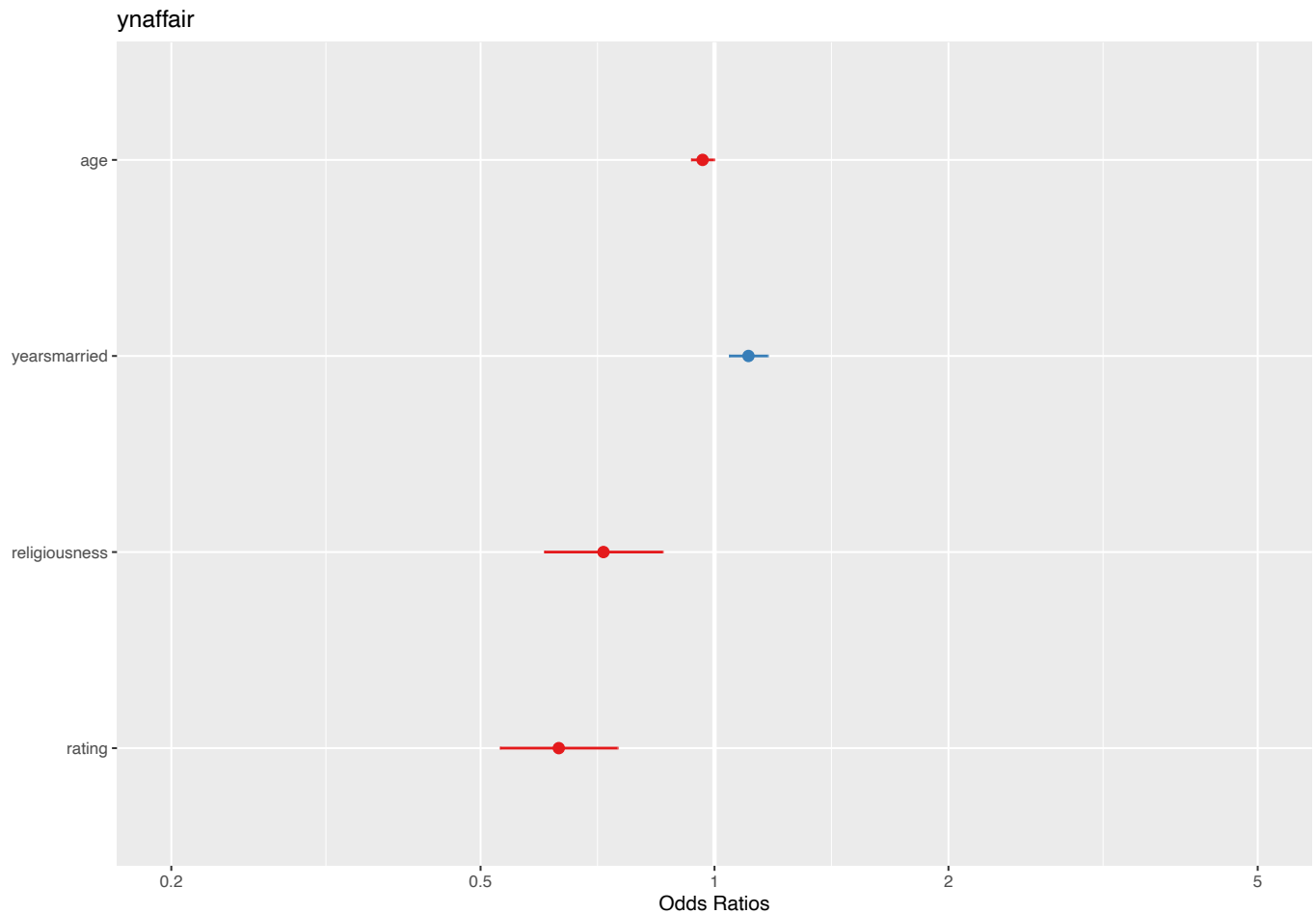
```
exp(coef(fit.reduced))
```

```
##   (Intercept)           age  yearsmarried religiousness        rating
##     6.8952321     0.9653437     1.1058594     0.7196258     0.6304248
```

```
exp(confint(fit.reduced))
```

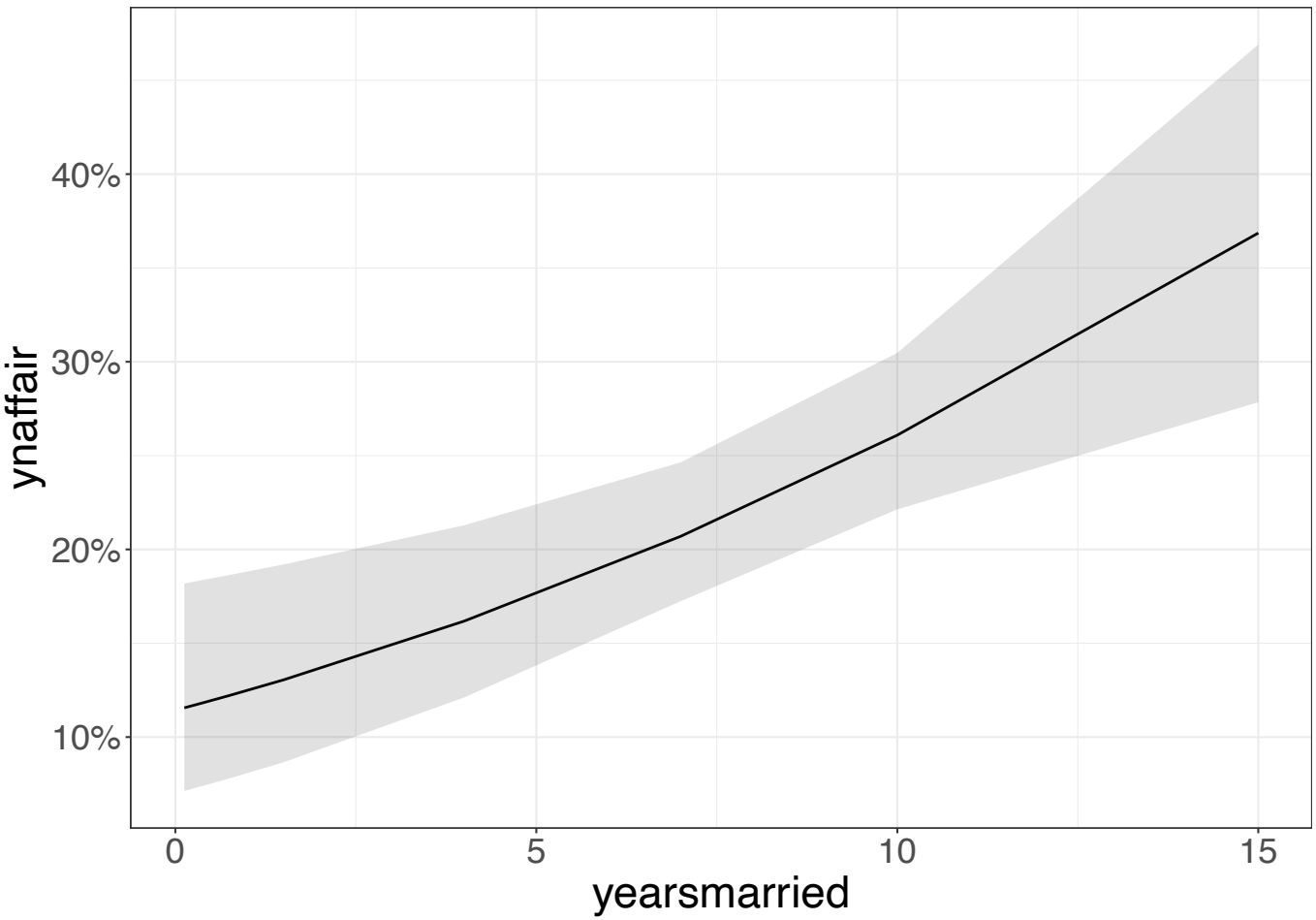
```
##                    2.5 %     97.5 %
## (Intercept)   2.1255764 23.3506030
## age           0.9323342  0.9981470
## yearsmarried  1.0448584  1.1718250
## religiousness 0.6026782  0.8562807
## rating        0.5286586  0.7493370
```

```
library(sjPlot)
library(sjlabelled)
library(sjmisc)
plot_model(fit.reduced, axis.lim = c(0.5, 2))
```



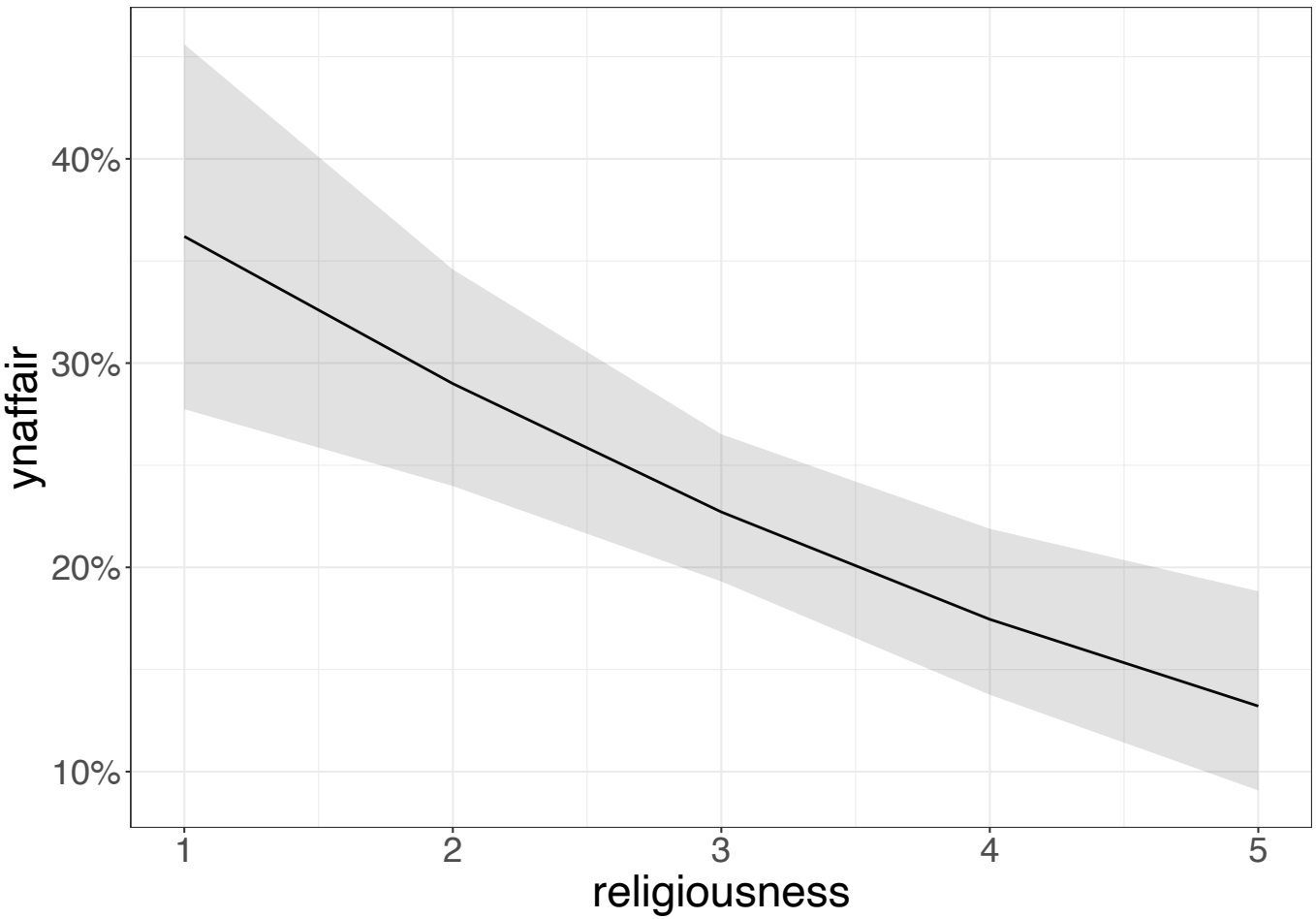
Visualization of the model

```
library(ggeffects)
library(ggthemes)
library(ggplot2)
plot(ggpredict(fit.reduced, "yearsmarried")) + theme_bw() + theme(text = element_text(size = 24)) +
    labs(title = NULL)
```



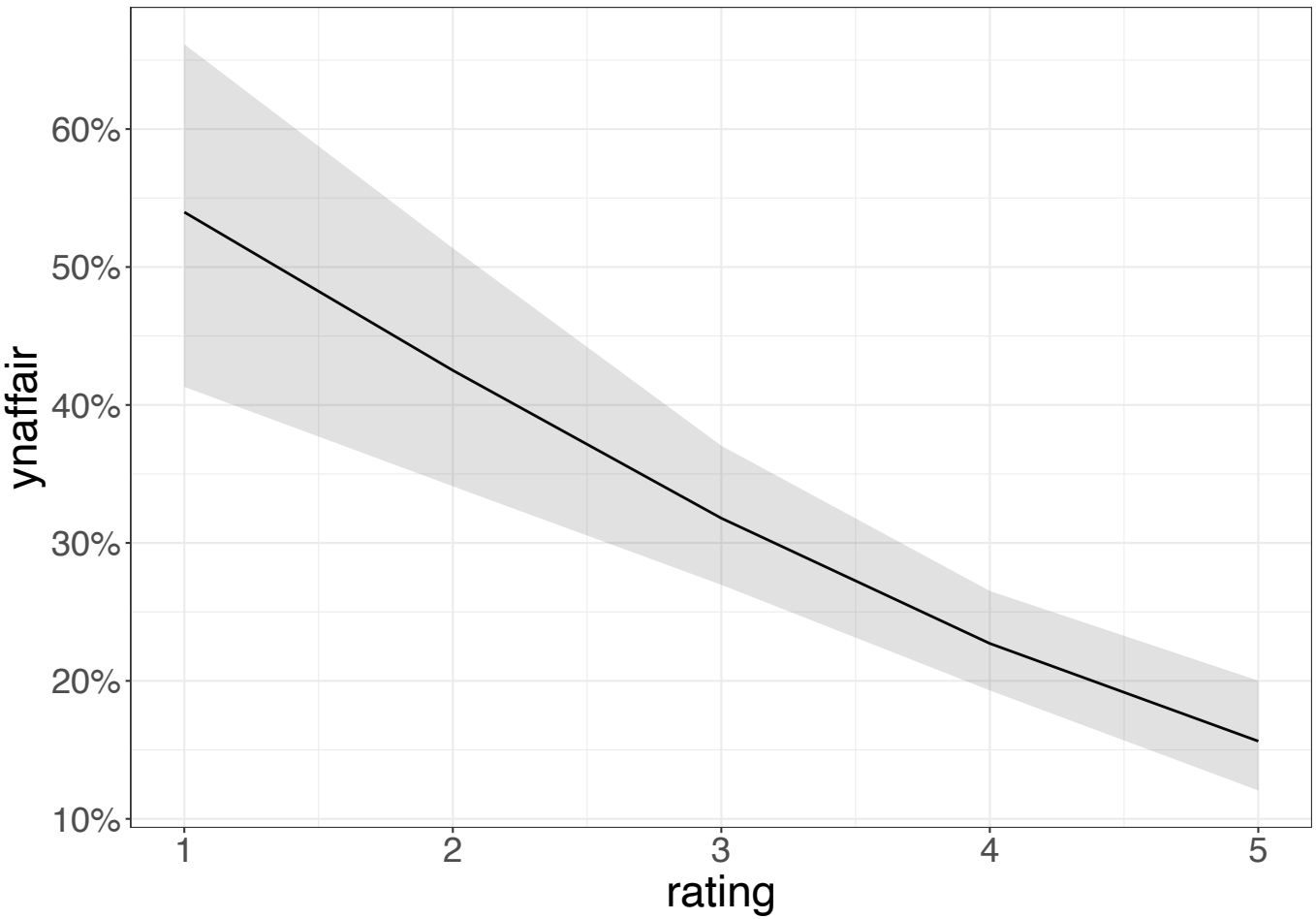
Visualization of the model

```
plot(ggpredict(fit.reduced, "religiousness")) + theme_bw() + theme(text = element_text(size = 24)) +
    labs(title = NULL)
```



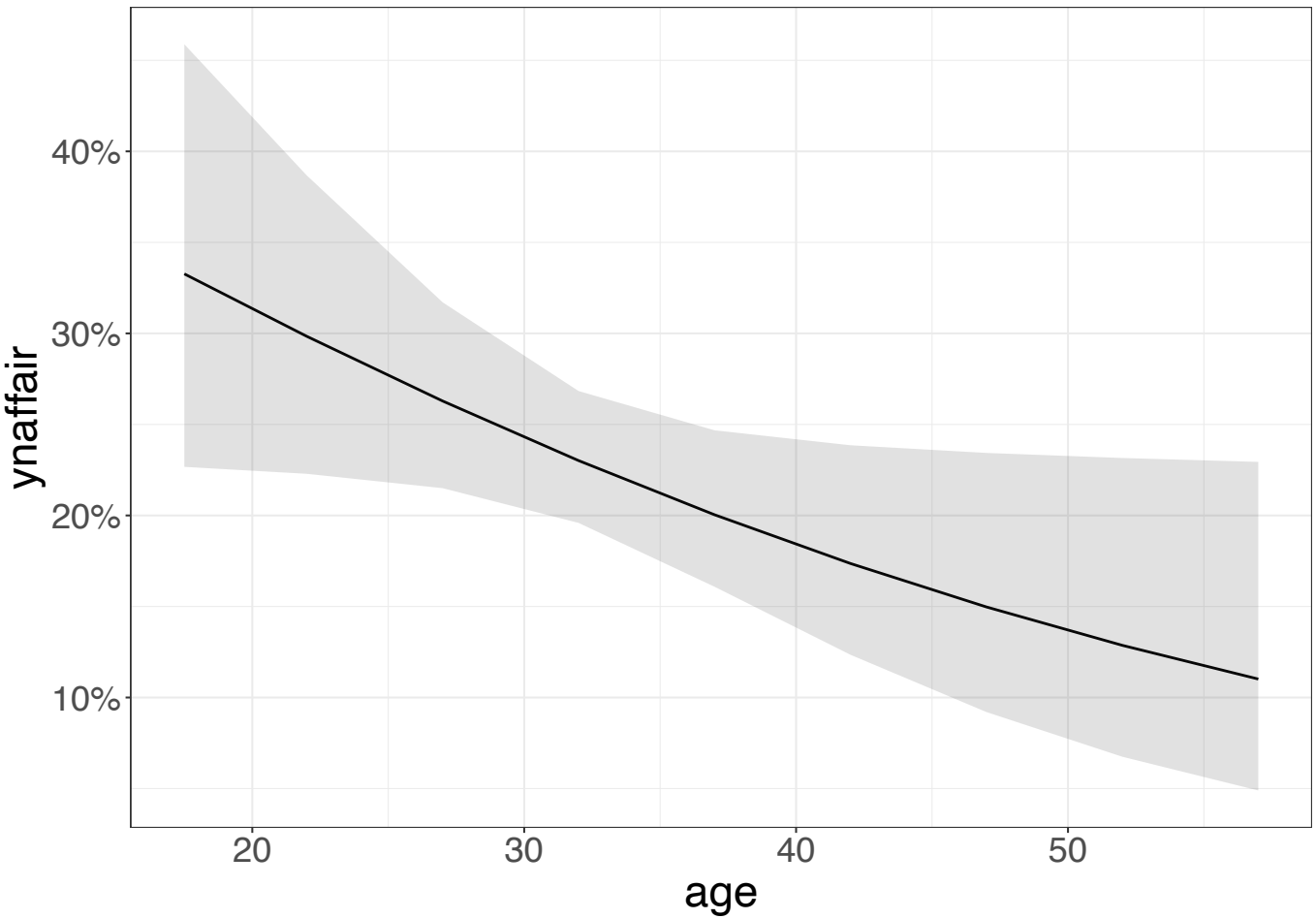
Visualization of the model

```
plot(ggpredict(fit.reduced, "rating")) + theme_bw() + theme(text = element_text(size = 24)) +
    labs(title = NULL)
```



Visualization of the model

```
plot(ggpredict(fit.reduced, "age")) + theme_bw() + theme(text = element_text(size = 24)) +
    labs(title = NULL)
```



Overdispersion

$$\sigma^2 = n\pi(1-\pi).$$

Overdispersion findet dann statt, wenn die beobachtete Varianz von der Zielvariablen größer ist als die nach der Binonialverteilung zu erwartende Varianz.

```
fit <- fit.reduced
fit.od <- glm(ynaffair ~ age + yearsmarried + religiousness + rating, data = Affairs,
    family = quasibinomial())
pchisq(summary(fit.od)$dispersion * fit$df.residual, fit$df.residual, lower = F)
```

```
## [1] 0.340122
```

Der Test ist nicht signifikant, also sind unsere Daten nicht “overdispersed”.

–>

Gemischte Modelle

Vitaly Belik

Institut für Veterinär-Epidemiologie und Biometrie, FU Berlin

05/01/2020

Gemischte Modelle

- Beispiel. Wann braucht man solche Modelle?

Beispiel Kosmetische Chirurgie (nach A. Field)

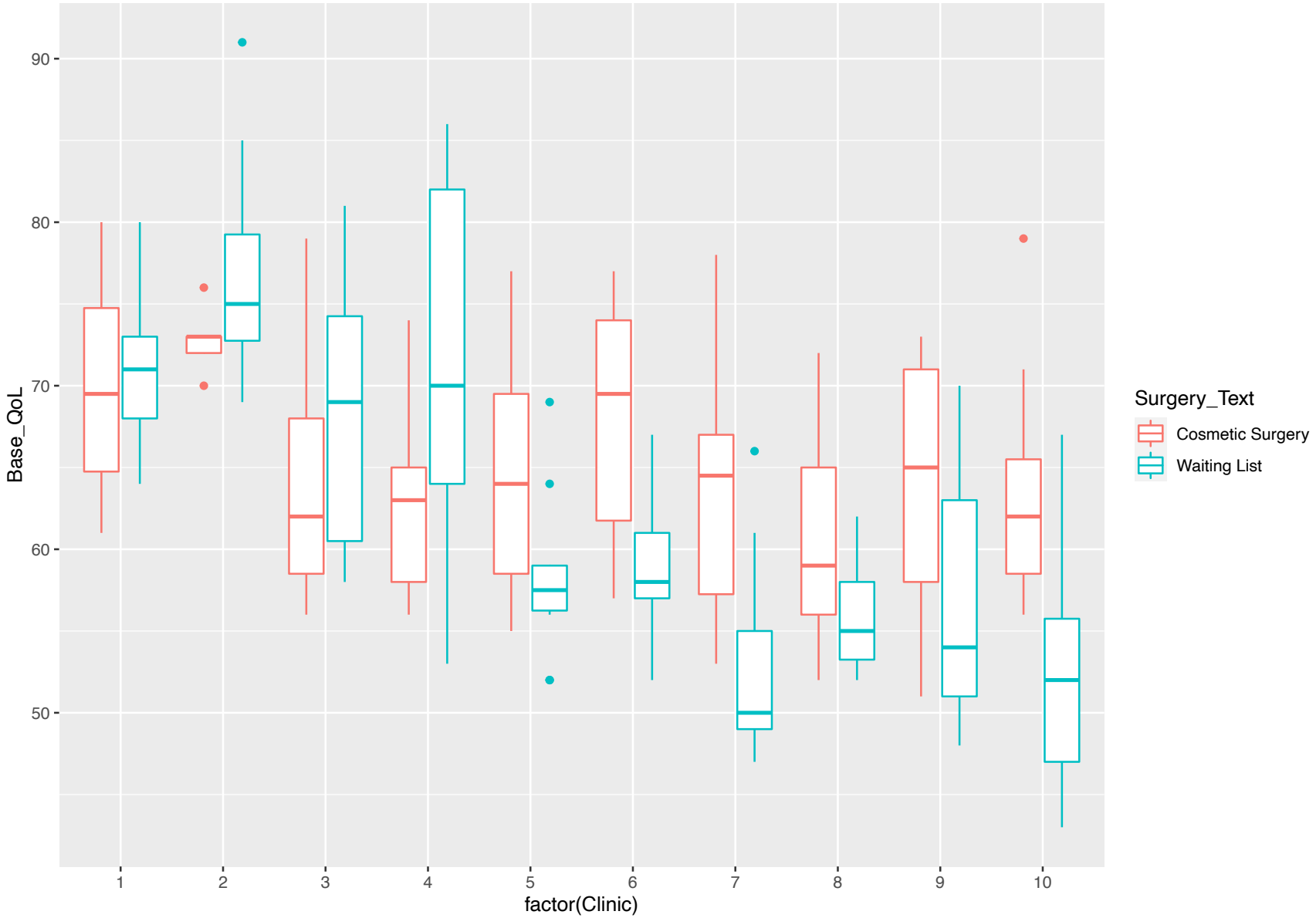
```
d <- read.table("Cosmetic Surgery.dat", sep = "\t", header = TRUE)
head(d)
```

```
##   particnu Post_QoL Base_QoL Clinic Surgery Reason Age Gender BDI Surgery_Text
## 1        1     71.3       73      1       0      0  31      0  12 Waiting List
## 2        2     77.0       74      1       0      0  32      0  16 Waiting List
## 3        3     73.0       80      1       0      0  33      0  13 Waiting List
## 4        4     68.9       76      1       0      0  59      1  11 Waiting List
## 5        5     69.0       71      1       0      0  61      1  11 Waiting List
## 6        6     68.5       72      1       0      1  32      0  10 Waiting List
##          Reason_Text Gender_Text
## 1 Change Appearance      Female
## 2 Change Appearance      Female
## 3 Change Appearance      Female
## 4 Change Appearance        Male
## 5 Change Appearance        Male
## 6   Physical reason      Female
```

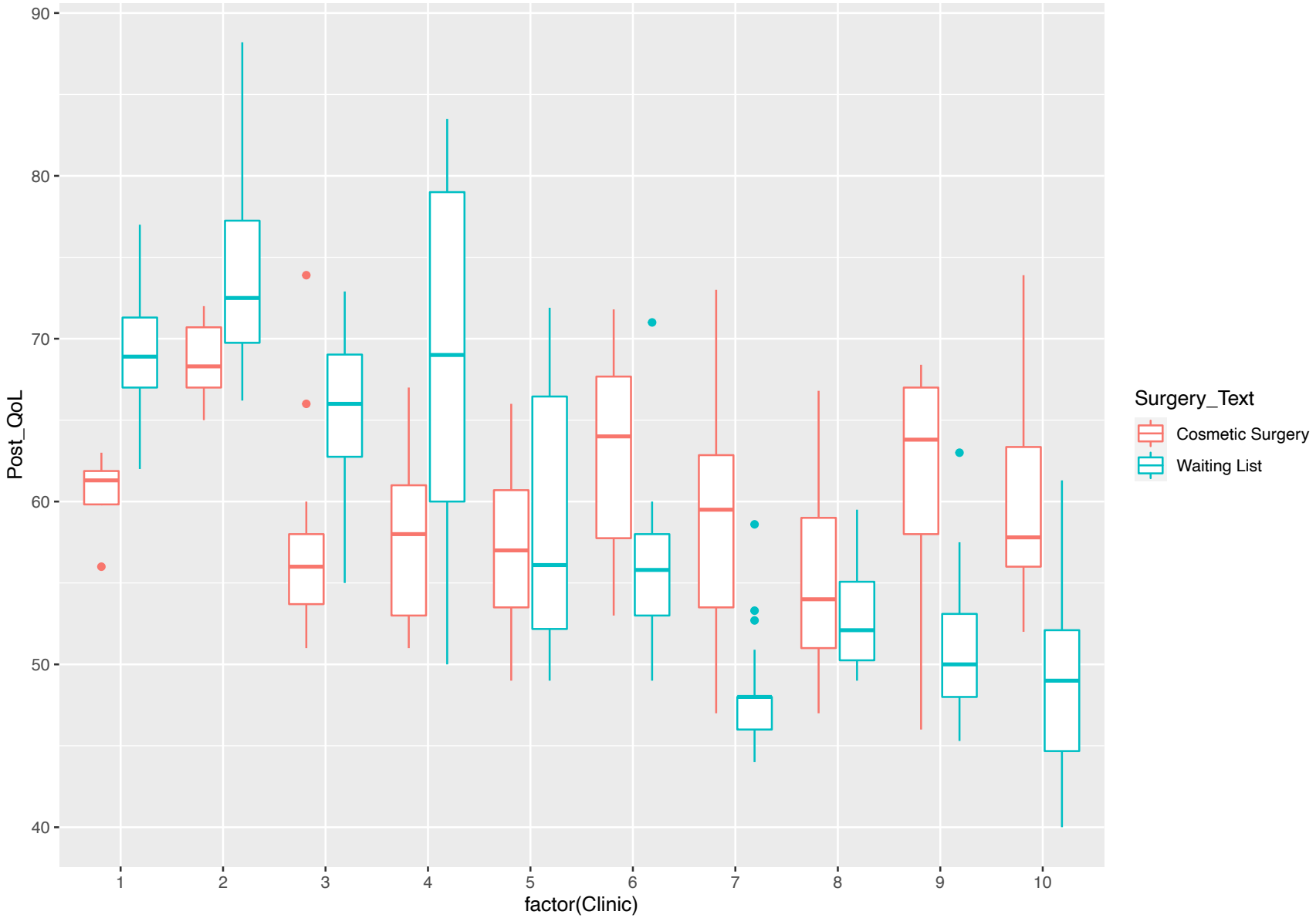
Beschreibung des Datensatzes

- Post_QoL: This is a measure of quality of life after the cosmetic surgery. This is our outcome variable.
- Base_QoL: We need to adjust our outcome for quality of life before the surgery.
- Surgery: This variable is a dummy variable that specifies whether the person has undergone cosmetic surgery (1) or whether they are on the waiting list (0), which acts as our control group.
- Surgery_Text: This variable is the same as above but specifies group membership as text (we will use this variable when we create graphs but not for the main analysis).
- Clinic: This variable specifies which of 10 clinics the person attended to have their surgery.
- Age: This variable tells us the person's age in years.
- BDI: It is becoming increasingly apparent that people volunteering for cosmetic sur- gery (especially when the surgery is purely for vanity) might have very different per- sonality profiles than the general public (Cook, Rossera, Toone, James, & Salmon, 2006). In particular, these people might have low self-esteem or be depressed. When looking at quality of life it is important to assess natural levels of depression, and this variable used the Beck Depression Inventory (BDI) to do just that.
- Reason: This dummy variable specifies whether the person had/is waiting to have surgery purely to change their appearance (0), or because of a physical reason (1).
- Reason_Text: This variable is the same as above but contains text to define each group rather than a number.
- Gender: This variable simply specifies whether the person was a man (1) or a woman (0).

```
library(ggplot2)
ggplot(data = d, aes(y = Base_QoL, x = factor(Clinic), color = Surgery_Text)) +
    geom_boxplot()
```

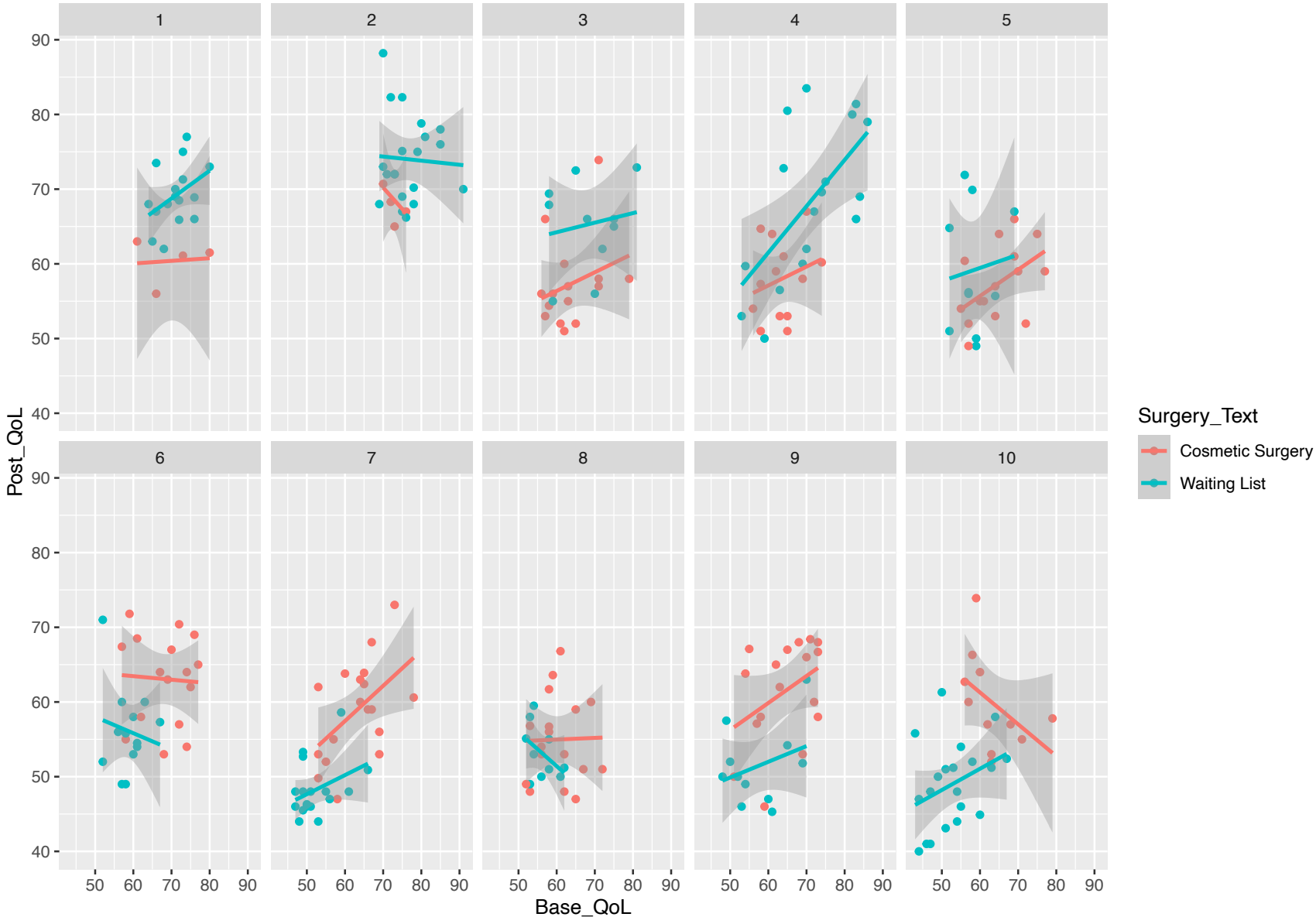


```
library(ggplot2)
ggplot(data = d, aes(y = Post_QoL, x = factor(Clinic), color = Surgery_Text)) +
    geom_boxplot()
```



```
library(ggplot2)
ggplot(data = d, aes(x = Base_QoL, y = Post_QoL, color = Surgery_Text)) +
    geom_point() + facet_wrap(~Clinic, ncol = 5) + geom_smooth(method = lm)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Ein gemischtes Modell ist hier eine gute Wahl:

- Es ermöglicht uns, alle uns vorliegenden Daten zu verwenden (höhere Stichprobengröße) und

Ein gemischtes Modell ist hier eine gute Wahl:

- Es ermöglicht uns, alle uns vorliegenden Daten zu verwenden (höhere Stichprobengröße) und
- die Korrelationen zwischen Daten von den gleichen Kliniken zu berücksichtigen.

Ein gemischtes Modell ist hier eine gute Wahl:

- Es ermöglicht uns, alle uns vorliegenden Daten zu verwenden (höhere Stichprobengröße) und
- die Korrelationen zwischen Daten von den gleichen Kliniken zu berücksichtigen.
- Wir werden auch weniger Parameter schätzen und Probleme mit mehreren Vergleichen vermeiden, die bei Verwendung separater Regressionen auftreten würden.

Feste und zufällige Effekte

- Gewöhnliche Regression

$$Y_i = \beta_0 + \beta_i X_i + \varepsilon_i$$

Feste und zufällige Effekte

- Gewöhnliche Regression

$$Y_i = \beta_0 + \beta_i X_i + \varepsilon_i$$

Es ist wichtig, dass dieser Unterschied wenig mit den Variablen selbst und viel mit der Forschungsfrage zu tun hat! In vielen Fällen kann dieselbe Variable entweder als zufälliger oder als fester Effekt betrachtet werden (und manchmal sogar als beide gleichzeitig!). Beziehen Sie sich daher immer auf Ihre Fragen und Hypothesen, um Ihre Modelle entsprechend zu konstruieren.

$$Y_{ij} = \beta_{0j} + \beta_{ij} X_{ij} + \varepsilon_{ij}$$

$$b_{0j} = b_0 + u_{0j} \text{ (Random intercepts)}$$

$$b_{ij} = b_i + u_{ij} \text{ (Random slopes)}$$

Sollten meine Variablen feste oder zufällige Effekte sein?

Im Großen und Ganzen sind **feste Effekte** Variablen, von denen wir erwarten, dass sie sich auf die abhängige Variable auswirken: Sie werden in einer linearen Standardregression als erklärende Variablen bezeichnet. In unserem Fall sind wir daran interessiert, Schlussfolgerungen darüber zu ziehen, wie sich die schon vorhandene Lebensqualität auf die Lebensqualität nach der OP auswirkt. Die schon vorhandene Lebensqualität ist also ein fester Effekt.

Sollten meine Variablen feste oder zufällige Effekte sein?

Im Großen und Ganzen sind **feste Effekte** Variablen, von denen wir erwarten, dass sie sich auf die abhängige Variable auswirken: Sie werden in einer linearen Standardregression als erklärende Variablen bezeichnet. In unserem Fall sind wir daran interessiert, Schlussfolgerungen darüber zu ziehen, wie sich die schon vorhandene Lebensqualität auf die Lebensqualität nach der OP auswirkt. Die schon vorhandene Lebensqualität ist also ein fester Effekt.

Auf der anderen Seite sind **zufällige Effekte** normalerweise *Gruppierungsfaktoren*, für die wir versuchen zu *kontrollieren*. Sie sind immer kategorisch, da Sie R nicht zwingen können, eine kontinuierliche Variable als zufälligen Effekt zu behandeln. In den meisten Fällen sind wir nicht speziell an ihren Auswirkungen auf die Antwortvariable interessiert, aber wir wissen, dass sie möglicherweise die Muster beeinflussen, die wir sehen.

Sollten meine Variablen feste oder zufällige Effekte sein?

Im Großen und Ganzen sind **feste Effekte** Variablen, von denen wir erwarten, dass sie sich auf die abhängige Variable auswirken: Sie werden in einer linearen Standardregression als erklärende Variablen bezeichnet. In unserem Fall sind wir daran interessiert, Schlussfolgerungen darüber zu ziehen, wie sich die schon vorhandene Lebensqualität auf die Lebensqualität nach der OP auswirkt. Die schon vorhandene Lebensqualität ist also ein fester Effekt.

Auf der anderen Seite sind **zufällige Effekte** normalerweise *Gruppierungsfaktoren*, für die wir versuchen zu *kontrollieren*. Sie sind immer kategorisch, da Sie R nicht zwingen können, eine kontinuierliche Variable als zufälligen Effekt zu behandeln. In den meisten Fällen sind wir nicht speziell an ihren Auswirkungen auf die Antwortvariable interessiert, aber wir wissen, dass sie möglicherweise die Muster beeinflussen, die wir sehen.

Darüber hinaus sind die Daten für unseren Zufallseffekt nur ein Beispiel für alle Möglichkeiten. Und wir verallgemeinern die Ergebnisse anhand repräsentativer Stichproben auf eine ganze Population. Es ist uns egal, wie viel besser Patienten in Klinik A im Vergleich zu Patienten in Klinik B abgeschnitten haben, aber wir wissen, dass ihre jeweiligen Chirurgen ein Grund sein könnten, warum ihre Lebensqualität unterschiedlich wären, und wir möchten gerne wissen, wie viel Abweichungen sind darauf zurückzuführen, wenn wir die Lebensqualität für Patienten der Klinik Z vorhersagen.

Mehr über zufällige Effekte

Beachten Sie, dass die goldene Regel lautet, dass Ihr zufälliger Effekt im Allgemeinen mindestens fünf Ebenen haben soll. Z.B.~würden wir Sex (ein zweistufiger Faktor: männlich oder weiblich) als festen, nicht zufälligen Effekt betrachten.

Mehr über zufällige Effekte

Beachten Sie, dass die goldene Regel lautet, dass Ihr zufälliger Effekt im Allgemeinen mindestens fünf Ebenen haben soll. Z.B.~würden wir Sex (ein zweistufiger Faktor: männlich oder weiblich) als festen, nicht zufälligen Effekt betrachten.

Dies liegt einfach daran, dass die Schätzung der Varianz an wenigen Datenpunkten sehr ungenau ist. Mathematisch könnten Sie, aber Sie würden nicht viel Vertrauen in sie haben. Wenn Sie nur zwei oder drei Ebenen haben, wird das Modell Schwierigkeiten haben, die Varianz zu partitionieren - es gibt Ihnen eine Ausgabe, aber nicht unbedingt eine, der Sie vertrauen können.

Mehr über zufällige Effekte

Beachten Sie, dass die goldene Regel lautet, dass Ihr zufälliger Effekt im Allgemeinen mindestens fünf Ebenen haben soll. Z.B.~würden wir Sex (ein zweistufiger Faktor: männlich oder weiblich) als festen, nicht zufälligen Effekt betrachten.

Dies liegt einfach daran, dass die Schätzung der Varianz an wenigen Datenpunkten sehr ungenau ist. Mathematisch könnten Sie, aber Sie würden nicht viel Vertrauen in sie haben. Wenn Sie nur zwei oder drei Ebenen haben, wird das Modell Schwierigkeiten haben, die Varianz zu partitionieren - es gibt Ihnen eine Ausgabe, aber nicht unbedingt eine, der Sie vertrauen können.

Der Name zufällig hat nicht viel mit mathematischer Zufälligkeit zu tun obwohl das verwirrend ist. Stellen Sie sie sich vorerst als *Gruppierungsvariablen* vor. Genau genommen geht es darum, unsere Modelle für unsere Fragen repräsentativ zu machen und bessere Schätzungen zu erhalten.

Crossed and Nested random effects

Wiederholen wir dies mit einem anderen Beispiel: Ein Effekt wird (vollständig) gekreuzt, wenn alle Probanden alle Stufen dieses Effekts erfahren haben. Wenn Sie beispielsweise ein Düngungsexperiment an Sämlingen durchgeführt haben, die in einem saisonalen Wald wachsen, und in jeder Jahreszeit wiederholte Messungen über die Zeit (z. B. 3 Jahre) durchgeführt haben, möchten Sie möglicherweise einen gekreuzten Faktor namens Jahreszeit (Sommer1, Herbst1, Winter1, Frühling1, Summer2,. . . , Spring3), dh ein Faktor für jede Jahreszeit eines jeden Jahres. Dieser Gruppierungsfaktor würde die Tatsache erklären, dass alle Pflanzen im Experiment, unabhängig vom festen (Behandlungs-) Effekt (dh gedüngt oder nicht gedüngt), im zweiten Jahr einen sehr heißen Sommer oder im dritten einen sehr regnerischen Frühling erlebt haben könnten Jahr, und diese Bedingungen könnten Störungen in den erwarteten Mustern verursachen. Sie müssen nicht einmal Klimadaten zuordnen, um dies zu berücksichtigen! Sie wissen nur, dass alle Beobachtungen ab Frühjahr 3 einander ähnlicher sein können, weil sie dieselben Umwelteigenschaften hatten und nicht, weil sie auf Ihre Behandlung ansprechen.

Nested random effects

Pseudoreplikation ist eine massive Erhöhung Ihrer Stichprobengröße durch Verwendung nicht unabhängiger Daten. Bei einer Stichprobengröße von 60.000 würden Sie mit ziemlicher Sicherheit einen „signifikanten“ Behandlungseffekt erzielen, der möglicherweise überhaupt keine ökologische Bedeutung hat. Und es verstößt gegen die Annahme der Unabhängigkeit von Beobachtungen, die für die lineare Regression von zentraler Bedeutung ist.

Feste Effekte

- Sie sollten die ML verwenden, wenn Sie Modelle mit verschiedenen festen Effekten vergleichen, da ML nicht auf den Koeffizienten der festen Effekte beruht

Feste Effekte

- Sie sollten die ML verwenden, wenn Sie Modelle mit verschiedenen festen Effekten vergleichen, da ML nicht auf den Koeffizienten der festen Effekte beruht
- Obwohl Sie ML zum Vergleichen von Modellen verwenden, sollten Sie Parameterschätzungen aus Ihrem endgültigen „besten“ REML-Modell melden, da ML die Varianz der zufälligen Effekte möglicherweise unterschätzt.

Zufällige Effekte

- Über experimentelles Design, Ihr System und die gesammelten Daten sowie Ihre Fragen nachdenken.

Zufällige Effekte

- Über experimentelles Design, Ihr System und die gesammelten Daten sowie Ihre Fragen nachdenken.
- Wenn Ihre zufälligen Effekte dazu dienen, sich mit Pseudoreplikation zu befassen, spielt es keine Rolle, ob sie "signifikant" sind oder nicht: Sie sind Teil Ihres Designs und müssen einbezogen werden.

Zufällige Effekte

- Über experimentelles Design, Ihr System und die gesammelten Daten sowie Ihre Fragen nachdenken.
- Wenn Ihre zufälligen Effekte dazu dienen, sich mit Pseudoreplikation zu befassen, spielt es keine Rolle, ob sie "signifikant" sind oder nicht: Sie sind Teil Ihres Designs und müssen einbezogen werden.
- Es wird empfohlen REML-Schätzer zum Vergleich von Modellen mit verschiedenen zufälligen Effekten (wir halten feste Effekte konstant) zu nutzen.

Zufällige Effekte

- Über experimentelles Design, Ihr System und die gesammelten Daten sowie Ihre Fragen nachdenken.
- Wenn Ihre zufälligen Effekte dazu dienen, sich mit Pseudoreplikation zu befassen, spielt es keine Rolle, ob sie "signifikant" sind oder nicht: Sie sind Teil Ihres Designs und müssen einbezogen werden.
- Es wird empfohlen REML-Schätzer zum Vergleich von Modellen mit verschiedenen zufälligen Effekten (wir halten feste Effekte konstant) zu nutzen.

HINWEIS: Variieren Sie NICHT zufällige und feste Effekte gleichzeitig - entweder mit Ihrer Zufallseffektstruktur oder mit Ihrer festen Effektstruktur an einem bestimmten Punkt.

HINWEIS 2: Vergleichen Sie keine lmer-Modelle mit lm-Modellen (oder glmer mit glm).

- Ein vollständiges Modell anpassen (sogar komplexer als erwartet oder gewünscht)
- Sortieren Sie die Zufallseffektstruktur (verwenden Sie REML-Likelihood oder REML AIC oder BIC).
- Struktur mit festen Effekten sortieren (entweder REML, die F-Statistik oder die t-Statistik verwenden oder verschachtelte ML-Modelle vergleichen – halten Sie Ihre zufälligen Effekte konstant)
- Sobald Sie das endgültige Modell erreicht haben, präsentieren Sie es mithilfe der REML-Schätzung

HINWEIS: Es besteht die Gefahr, dass es wie ein gebrochener Rekord klingt: Ich denke, es ist am besten, zu entscheiden, was Ihr Modell auf Biologie / Ökologie / Datenstruktur usw. basiert, als blind der Modellauswahl zu folgen. Nur weil etwas nicht signifikant ist, heißt das nicht unbedingt, dass Sie es immer loswerden sollten.

Kovarianzenstruktur

- Bei den zufälligen Effekten brauchen wir Annahmen über Korrelationen (variance–covariance matrix)
- Simple Struktur – Fehler I. Art (parameter überwiegend signifikant). Komplizierte Struktur – Fehler II. Art (parameter überwiegend nicht signifikant).
- Variance components

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- Unstructured

$$\begin{pmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{pmatrix}$$

Beispiel. Standard-Regression (ANOVA)

```r
m <- lm(Post_QoL ~ factor(Surgery), data = d)
summary(m)
```

```
##
## Call:
## lm(formula = Post_QoL ~ factor(Surgery), data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.916  -7.271  -1.271   7.084  28.284
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        59.9159     0.7731  77.498   <2e-16 ***
## factor(Surgery)1   -0.6449     1.1222  -0.575    0.566
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.31 on 274 degrees of freedom
## Multiple R-squared:  0.001204,   Adjusted R-squared:  -0.002442
## F-statistic: 0.3302 on 1 and 274 DF,  p-value: 0.566
```

```r
anova(m)
```

```
## Analysis of Variance Table
##
## Response: Post_QoL
##                   Df  Sum Sq Mean Sq F value Pr(>F)
## factor(Surgery)    1    28.6  28.620  0.3302  0.566
## Residuals        274 23747.9  86.671
```

```r
m1 <- lm(Post_QoL ~ factor(Surgery) + Base_QoL, data = d)
summary(m1)
```

```
##
## Call:
## lm(formula = Post_QoL ~ factor(Surgery) + Base_QoL, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.4142  -5.1326  -0.6495   4.0540  23.5005
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       18.14702    2.90767   6.241 1.65e-09 ***
## factor(Surgery)1 -1.69723    0.84404  -2.011   0.0453 *
## Base_QoL           0.66504    0.04537  14.659  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.977 on 273 degrees of freedom
## Multiple R-squared:  0.4411, Adjusted R-squared:  0.437
## F-statistic: 107.7 on 2 and 273 DF,  p-value: < 2.2e-16
```

```r
anova(m1)
```

```
## Analysis of Variance Table
##
## Response: Post_QoL
##                   Df  Sum Sq Mean Sq F value Pr(>F)
## factor(Surgery)    1    28.6    28.6   0.588 0.4439
## Base_QoL           1 10459.6 10459.6 214.888 <2e-16 ***
## Residuals        273 13288.3    48.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Nur intercept

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
library(nlme)
```

```
##
## Attaching package: 'nlme'

## The following object is masked from 'package:lme4':
##
##     lmList
```

```
m_intercept <- gls(Post_QoL ~ 1, data = d, method = "ML")  # uses ML-Estimation
summary(m_intercept)
```

```
## Generalized least squares fit by maximum likelihood
##   Model: Post_QoL ~ 1
##   Data: d
##        AIC      BIC    logLik
##   2017.124 2024.365 -1006.562
##
## Coefficients:
##                Value Std.Error  t-value p-value
## (Intercept) 59.60978 0.5596972 106.5036       0
##
## Standardized residuals:
##        Min         Q1        Med         Q3        Max
## -2.1127754 -0.7875625 -0.1734394  0.7962286  3.0803354
##
## Residual standard error: 9.281527
## Degrees of freedom: 276 total; 275 residual
```

Zufallseffekt Clinic

```
# m_rand_intercept <- lmer(Post_QoL ~ 1|Clinic, data = d,
# REML = FALSE) # uses ML-Estimation
randomInterceptOnly <- lme(Post_QoL ~ 1, random = ~1 | Clinic,
    data = d, method = "ML")  # uses ML-Estimation
anova(m_intercept, randomInterceptOnly)
```

```
##                     Model df      AIC      BIC     logLik   Test  L.Ratio
## m_intercept             1  2 2017.124 2024.365 -1006.5622
## randomInterceptOnly     2  3 1911.473 1922.334  -952.7364 1 vs 2 107.6517
##                     p-value
## m_intercept
## randomInterceptOnly  <.0001
```

Zufällige Effekte mit *Surgery* als fester Effekt

```r
# Add surgery as a predictor
randomInterceptSurgery <- lme(Post_QoL ~ Surgery, data = d, random = ~1 |
    Clinic, method = "ML")
summary(randomInterceptSurgery)
```

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: d
##        AIC      BIC    logLik
##   1910.137 1924.619 -951.0686
##
## Random effects:
##  Formula: ~1 | Clinic
##         (Intercept) Residual
## StdDev:    6.099513  7.18542
##
## Fixed effects: Post_QoL ~ Surgery
##                Value Std.Error  DF  t-value p-value
## (Intercept) 59.30517 2.0299632 265 29.21490   0.000
## Surgery      1.66583 0.9091314 265  1.83233   0.068
##  Correlation:
##         (Intr)
## Surgery -0.21
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -1.8904290 -0.7191399 -0.1420998  0.7177762  2.8644538
##
## Number of Observations: 276
## Number of Groups: 10
```

Zufällige Effekte mit *Surgery* und *Base_QoL* als feste Effekte

```r
## Fit effect of surgery and baseline QoL- random intercepts
## across clinics
randomInterceptSurgeryQoL <- lme(Post_QoL ~ Surgery + Base_QoL,
    data = d, random = ~1 | Clinic, method = "ML")
summary(randomInterceptSurgeryQoL)
```

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: d
##        AIC      BIC    logLik
##   1847.49 1865.592 -918.745
##
## Random effects:
##  Formula: ~1 | Clinic
##         (Intercept) Residual
## StdDev:    3.039264 6.518986
##
## Fixed effects: Post_QoL ~ Surgery + Base_QoL
##                 Value Std.Error  DF   t-value p-value
## (Intercept) 29.563601  3.471879 264  8.515160  0.0000
## Surgery     -0.312999  0.843145 264 -0.371228  0.7108
## Base_QoL     0.478630  0.052774 264  9.069465  0.0000
##  Correlation:
##          (Intr) Surgry
## Surgery   0.102
## Base_QoL -0.947 -0.222
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -1.8872666 -0.7537675 -0.0954987  0.5657241  3.0020852
##
## Number of Observations: 276
## Number of Groups: 10
```

Modellenvergleich

```
anova(randomInterceptOnly, randomInterceptSurgery, randomInterceptSurgeryQoL)
```

```
##                            Model df      AIC      BIC    logLik   Test  L.Ratio
## randomInterceptOnly            1  3 1911.473 1922.334 -952.7364
## randomInterceptSurgery         2  4 1910.137 1924.619 -951.0686 1 vs 2  3.33564
## randomInterceptSurgeryQoL      3  5 1847.490 1865.592 -918.7450 2 vs 3 64.64721
##                            p-value
## randomInterceptOnly
## randomInterceptSurgery      0.0678
## randomInterceptSurgeryQoL  <.0001
```

Random slopes I

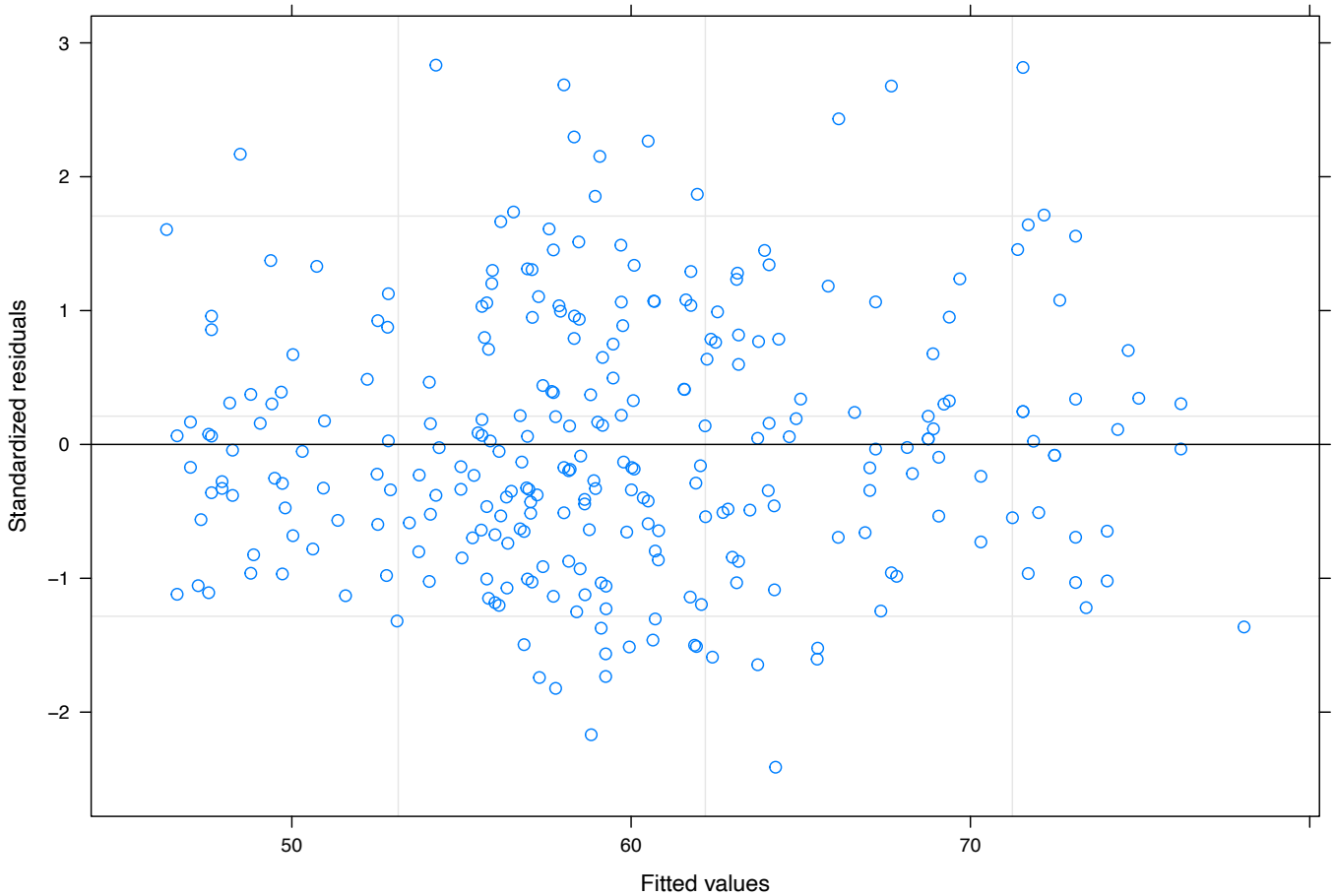
```r
## Fit effect of surgery and baseline QoL- random slopes and
## intercepts across clinics
addRandomSlope <- lme(Post_QoL ~ Surgery + Base_QoL, data = d,
    random = ~Surgery | Clinic, method = "ML")
summary(addRandomSlope)
```

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: d
##        AIC      BIC    logLik
##   1812.624 1837.967 -899.3119
##
## Random effects:
##  Formula: ~Surgery | Clinic
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev   Corr
## (Intercept) 6.132655 (Intr)
## Surgery     6.197489 -0.965
## Residual    5.912335
##
## Fixed effects: Post_QoL ~ Surgery + Base_QoL
##                Value Std.Error  DF   t-value p-value
## (Intercept) 40.10253  3.892945 264 10.301334  0.0000
## Surgery     -0.65453  2.110917 264 -0.310069  0.7568
## Base_QoL     0.31022  0.053506 264  5.797812  0.0000
##  Correlation:
##          (Intr) Surgry
## Surgery  -0.430
## Base_QoL -0.855 -0.063
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -2.4114778 -0.6628574 -0.1138411  0.6833110  2.8334730
##
## Number of Observations: 276
## Number of Groups: 10
```

Random slopes II

```
anova(randomInterceptSurgeryQoL, addRandomSlope)
```

```
##                            Model df      AIC      BIC    logLik   Test  L.Ratio
## randomInterceptSurgeryQoL      1  5 1847.490 1865.592 -918.7450
## addRandomSlope                 2  7 1812.624 1837.966 -899.3119 1 vs 2 38.86626
##                            p-value
## randomInterceptSurgeryQoL
## addRandomSlope              <.0001
```


Standardized residuals
3
2
1
0
−1
−2
50
60
70
Fitted values

```r
## Fit effect of surgery and baseline QoL, Reason and
## Reason*Surgery Interaction- random slopes and intercepts
## across clinics

addReason <- lme(Post_QoL ~ Surgery + Base_QoL + Reason, data = d,
    random = ~Surgery | Clinic, method = "ML")
addReason <- update(addRandomSlope, . ~ . + Reason)
summary(addReason)
```

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: d
##        AIC      BIC    logLik
##   1810.825 1839.788 -897.4124
##
## Random effects:
##  Formula: ~Surgery | Clinic
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev   Corr
## (Intercept) 5.838006 (Intr)
## Surgery     6.171602 -0.969
## Residual    5.886964
##
## Fixed effects: Post_QoL ~ Surgery + Base_QoL + Reason
##                Value Std.Error  DF   t-value p-value
## (Intercept) 41.43182  3.929923 263 10.542653  0.0000
## Surgery     -0.56817  2.106684 263 -0.269697  0.7876
## Base_QoL     0.30535  0.053417 263  5.716335  0.0000
## Reason      -1.69137  0.854411 263 -1.979571  0.0488
##  Correlation:
##          (Intr) Surgry Bas_QL
## Surgery  -0.400
## Base_QoL -0.862 -0.064
## Reason   -0.231 -0.042  0.118
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -2.2006350 -0.6830482 -0.1256909  0.6730347  2.9817687
##
## Number of Observations: 276
## Number of Groups: 10
```

```r
finalModel <- lme(Post_QoL ~ Surgery + Base_QoL + Reason + Reason:Surgery,
    data = d, random = ~Surgery | Clinic, method = "ML")
summary(finalModel)
```

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: d
##        AIC      BIC    logLik
##   1807.045 1839.629 -894.5226
##
## Random effects:
##  Formula: ~Surgery | Clinic
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev   Corr
## (Intercept) 5.482359 (Intr)
## Surgery     5.417495 -0.946
## Residual    5.818911
##
## Fixed effects: Post_QoL ~ Surgery + Base_QoL + Reason + Reason:Surgery
##                    Value Std.Error  DF   t-value p-value
## (Intercept)    42.51781  3.875317 262 10.971441  0.0000
## Surgery        -3.18768  2.185367 262 -1.458646  0.1459
## Base_QoL        0.30536  0.053125 262  5.747835  0.0000
## Reason         -3.51515  1.140934 262 -3.080939  0.0023
## Surgery:Reason  4.22129  1.700269 262  2.482718  0.0137
##  Correlation:
##                (Intr) Surgry Bas_QL Reason
## Surgery        -0.356
## Base_QoL       -0.865 -0.078
## Reason         -0.233  0.306  0.065
## Surgery:Reason  0.096 -0.505  0.024 -0.661
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -2.2331483 -0.6972193 -0.1541073  0.6326387  3.1641799
##
## Number of Observations: 276
## Number of Groups: 10
```

```r
intervals(finalModel, 0.95)
```

```
## Approximate 95% confidence intervals
##
##  Fixed effects:
##                     lower       est.      upper
## (Intercept)    34.9565174 42.5178133 50.0791092
## Surgery        -7.4516406 -3.1876776  1.0762854
## Base_QoL        0.2017009  0.3053562  0.4090115
## Reason         -5.7412741 -3.5151488 -1.2890236
## Surgery:Reason  0.9038215  4.2212894  7.5387574
## attr(,"label")
## [1] "Fixed effects:"
##
##  Random Effects:
##   Level: Clinic
##                                lower       est.     upper
## sd((Intercept))           3.3130516  5.4823585  9.072076
## sd(Surgery)               3.1323585  5.4174947  9.369697
## cor((Intercept),Surgery) -0.9937882 -0.9455544 -0.598258
##
##  Within-group standard error:
##    lower     est.    upper
## 5.331220 5.818911 6.351214
```

```
anova(addRandomSlope, addReason, finalModel)
```

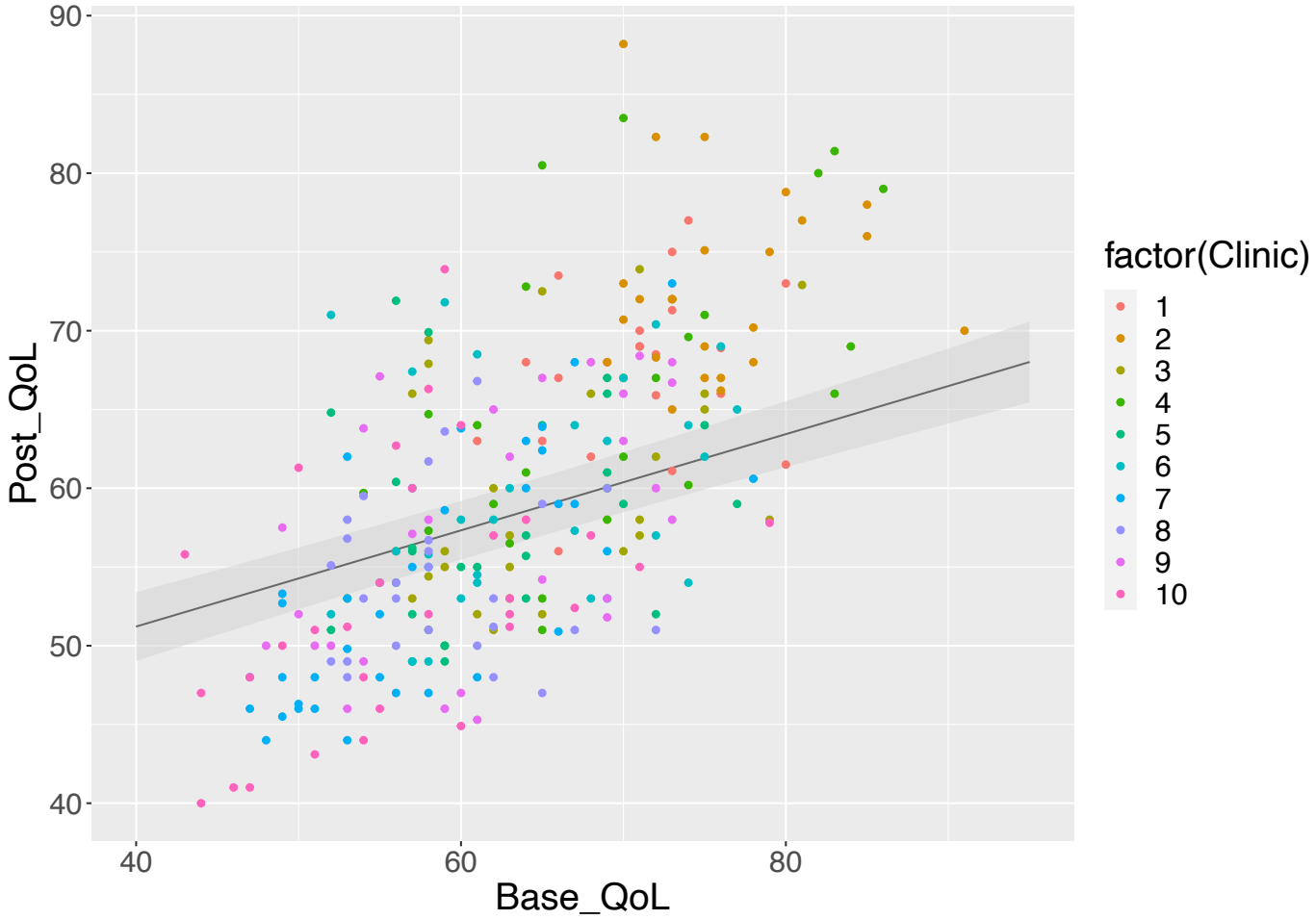
```
##                Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## addRandomSlope     1  7 1812.624 1837.966 -899.3119
## addReason          2  8 1810.825 1839.788 -897.4124 1 vs 2 3.798961  0.0513
## finalModel         3  9 1807.045 1839.629 -894.5226 2 vs 3 5.779555  0.0162
```

Visualisierung der Ergebnisse I

```
library(ggeffects)
# Extract the prediction data frame
pred.mm <- ggpredict(finalModel, terms = c("Base_QoL"))

ggplot(pred.mm) +
   geom_line(aes(x = x, y = predicted)) +          # slope
   geom_ribbon(aes(x = x, ymin = predicted - std.error, ymax = predicted + std.error),
               fill = "lightgrey", alpha = 0.5) +  # error band
   geom_point(data = d,                      # adding the raw data (scaled values)
              aes(x = Base_QoL, y = Post_QoL, colour = factor(Clinic))) +
   labs(x = "Base_QoL", y = "Post_QoL") +
  theme(text = element_text(size = 20))
```

Visualisierung der Ergebnisse II



References

https://ourcodingclub.github.io/tutorials/mixed-models/

A. Field

Hinweise

- Gemischte Modelle konvergieren manchmal nicht
- Da kann Reskalierung helfen

```
d$BDI_Centred <- scale(d$BDI, scale = F)
head(d$BDI)
```

```
## [1] 12 16 13 11 11 10
```

```
head(d$BDI_Centred)
```

```
##              [,1]
## [1,] -11.054348
## [2,]  -7.054348
## [3,] -10.054348
## [4,] -12.054348
## [5,] -12.054348
## [6,] -13.054348
```

Diskriminanzanalyse und Klassifikation

Vitaly Belik

Institut für Veterinär-Epidemiologie und Biometrie, FU Berlin

17/12/2019

Wiederholung

Was ist der Unterschied zwischen linearen *Diskriminanzanalyse* (als eine Art Klassifikationsverfahren) und *Clustering*?

LDA ist eng mit der Varianzanalyse (ANOVA) und der Regressionsanalyse verbunden, die auch versuchen, eine abhängige Variable als lineare Kombination anderer Merkmale oder Messungen auszudrücken. ANOVA verwendet jedoch qualitative erklärende Variablen und eine kontinuierliche abhängige Variable, während eine Diskriminanzanalyse kontinuierliche erklärende Variablen und eine qualitative abhängige Variable (dh Klassenbezeichnung) enthält.

Die logistische Regression und die Probit-Regression sind der LDA ähnlicher als die Varianzanalyse, da sie die qualitative Variable auch anhand kontinuierlicher erklärender Variablen erklären. Diese anderen Methoden sind in Anwendungen vorzuziehen, bei denen kein Grund zu der Annahme besteht, dass die unabhängigen Variablen *normal verteilt* sind (obwohl es auch für die leichte Abweichung davon auch funktioniert). Dies ist die Grundannahme der LDA-Methode.

Diskriminanzanalyse

Diskriminanzanalyse

- Diskiminanzanalyse benutzt man um die Wahrscheinlichkeit zu bestimmen, dass eine Beobachtung zu einer *bestimmten Klasse* oder Kategorie aus mehreren zu gehören.
- Die Prädiktoren können sowohl kontinuierlich als auch kategoriell sein

Lineare Diskriminanzanalyse

- Der LDA-Algorithmus beginnt mit der Suche nach Richtungen, die die Trennung zwischen Klassen maximieren, und verwendet diese Richtungen, um die Klassen vorherzusagen.
- Diese Richtungen, lineare Diskriminanten genannt, sind eine lineare Kombination von Prädiktorvariablen.
- Annahmen:
  1) Normalverteilung der Prädiktoren
  2) Die Klassen haben die gleichen Varianzen oder Kovarianzmatrizen.

Funktionsweise

Es wird nach einer linearen Kombination (2-dimensionaler Fall) $w_x x + w_z z$ gesucht dass das Verhältnis maximiert:

$$\frac{SS_{\text{between}}}{SS_{\text{within}}}$$

Lineare Diskriminanzanalyse: Vorarbeit

- Überprüfen Sie die univariaten Verteilungen jeder Variablen und stellen Sie sicher, dass sie normal verteilt sind. Andernfalls können Sie sie mit log und Wurzel für Exponentialverteilungen und mit Box-Cox für schräge Verteilungen transformieren.
- Entfernen Sie Ausreißer aus Ihren Daten und standardisieren Sie die Variablen, um deren Maßstab vergleichbar zu machen.

Box-Cox-Transformationen

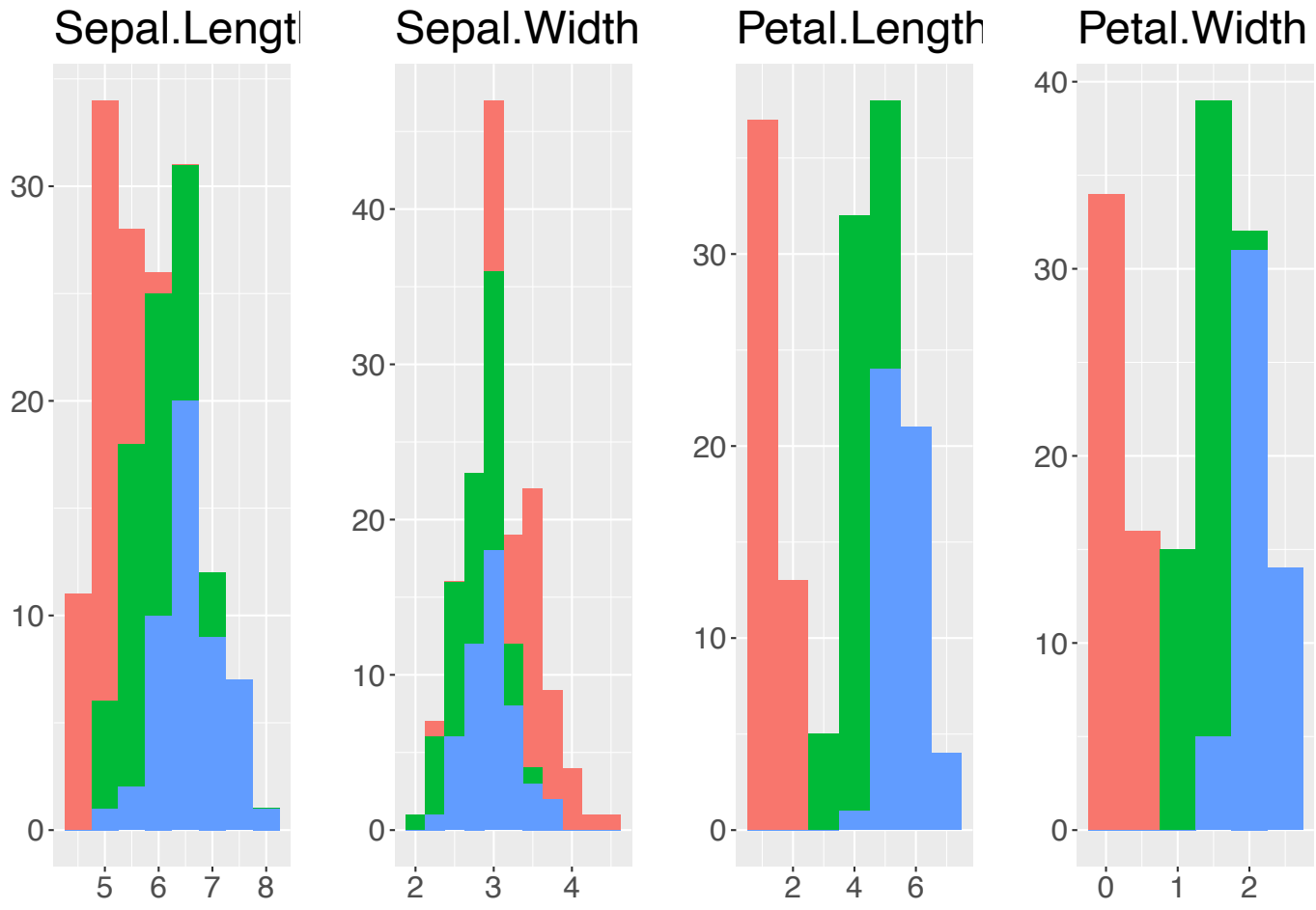
$$y(\lambda) = \begin{cases} \dfrac{y^\lambda - 1}{\log y} & \text{if } \lambda \neq 0 \\ \log y & \text{if } \lambda = 0 \end{cases}$$

Fisher's Iris-Datensatz

Figure 1: R. A. Fisher (1936). The use of multiple measurements in taxonomic problems. Annals of Eugenics. 7 (2): 179–188

```
library(tidyverse)
library(caret)
data("iris")
knitr::kable(head(iris))
```

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa


Sepal.Lengt
Sepal.Width
Petal.Length
Petal.Width
30
20
10
0
40
30
20
10
0
30
20
10
0
40
30
20
10
0
5
6
7
8
2
3
4
2
4
6
0
1
2

Lineare Diskriminanzanalyse(1)

Der Datensatz wird in Test-Datensatz und Training-Datensatz aufgeteilt.

```
set.seed(123)
training.samples <- iris$Species %>% createDataPartition(p = 0.8, list = FALSE)
train.data <- iris[training.samples, ]
test.data <- iris[-training.samples, ]
```

Die Daten werden normalisiert. In diesem Fall, können die Gewichte von Diskriminanten als Maß für die Wichtigkeit der Merkmale verwendet werden.

```
preproc.param <- train.data %>% preProcess(method = c("center", "scale"))
train.transformed <- preproc.param %>% predict(train.data)
test.transformed <- preproc.param %>% predict(test.data)
```

Lineare Diskriminanzanalyse(3)

```
library(MASS)
model <- lda(Species ~ ., data = train.transformed)
predictions <- model %>% predict(test.transformed)
mean(predictions$class == test.transformed$Species)
```

```
## [1] 0.9666667
```

```
model
```

```
## Call:
## lda(Species ~ ., data = train.transformed)
##
## Prior probabilities of groups:
##     setosa versicolor  virginica
##  0.3333333  0.3333333  0.3333333
##
## Group means:
##            Sepal.Length Sepal.Width Petal.Length Petal.Width
## setosa       -1.0112835  0.78048647   -1.2900001  -1.2453195
## versicolor    0.1014181 -0.68674658    0.2566029   0.1472614
## virginica     0.9098654 -0.09373989    1.0333972   1.0980581
##
## Coefficients of linear discriminants:
##                     LD1         LD2
## Sepal.Length  0.6794973  0.04463786
## Sepal.Width   0.6565085 -1.00330120
## Petal.Length -3.8365047  1.44176147
## Petal.Width  -2.2722313 -1.96516251
##
## Proportion of trace:
##    LD1    LD2
## 0.9902 0.0098
```

Visualisierung von Ergebnissen I

```
predictions <- model %>% predict(test.transformed)
head(predictions)
```

```
## $class
##  [1] setosa     setosa     setosa     setosa     setosa     setosa
##  [7] setosa     setosa     setosa     setosa     versicolor versicolor
## [13] versicolor versicolor versicolor versicolor versicolor versicolor
## [19] versicolor versicolor virginica  virginica  virginica  virginica
## [25] virginica  virginica  virginica  versicolor virginica  virginica
## Levels: setosa versicolor virginica
##
## $posterior
##           setosa   versicolor    virginica
## 1   1.000000e+00 3.978425e-22 1.319337e-43
## 2   1.000000e+00 1.038098e-17 3.967605e-38
## 6   1.000000e+00 2.882148e-21 2.041612e-41
## 16  1.000000e+00 8.381782e-28 7.486309e-50
## 23  1.000000e+00 6.531615e-25 3.414098e-47
## 34  1.000000e+00 1.089899e-28 7.865614e-52
## 35  1.000000e+00 1.449641e-17 8.617776e-38
## 38  1.000000e+00 2.569637e-23 2.490647e-45
## 44  1.000000e+00 9.532279e-16 1.078894e-33
## 47  1.000000e+00 1.357831e-22 7.647431e-44
## 51  3.896613e-18 9.999518e-01 4.821738e-05
## 60  2.156329e-20 9.997875e-01 2.124713e-04
## 64  2.984097e-23 9.970966e-01 2.903446e-03
## 73  1.426307e-28 8.761315e-01 1.238685e-01
## 74  7.061573e-22 9.998138e-01 1.862221e-04
## 87  6.615078e-21 9.991080e-01 8.919736e-04
## 91  1.396073e-22 9.997365e-01 2.634948e-04
## 94  2.215203e-14 1.000000e+00 2.929967e-08
## 95  9.053615e-21 9.998731e-01 1.268678e-04
## 97  6.118802e-19 9.999554e-01 4.460634e-05
## 104 5.684420e-38 1.055968e-03 9.989440e-01
## 109 2.940937e-42 2.018347e-04 9.997982e-01
## 111 1.464269e-31 1.419173e-02 9.858083e-01
```

Visualisierung von Ergebnissen II

```
## 113 1.023351e-38 1.717615e-04 9.998282e-01
## 116 9.953348e-40 1.990495e-05 9.999801e-01
## 120 3.529426e-33 2.862129e-01 7.137871e-01
## 127 1.757931e-29 2.351857e-01 7.648143e-01
## 134 3.846340e-28 8.063778e-01 1.936222e-01
## 138 1.920885e-34 6.617982e-03 9.933820e-01
## 147 7.176867e-36 6.395225e-03 9.936048e-01
##
## $x
##            LD1         LD2
## 1    8.1629386 -0.50527678
## 2    7.2027133  0.71110616
## 6    7.8162430 -1.73271506
## 16   9.2840285 -3.10506837
## 23   8.7872943 -1.09878034
## 34   9.6017637 -2.20200431
## 35   7.1488541  0.54633214
## 38   8.4569312 -0.50932474
## 44   6.4915104 -1.35588952
## 47   8.2157619 -1.08017093
## 51  -1.3891224 -0.03180619
## 60  -1.8942686  0.45721896
## 64  -2.5795707  0.65766782
## 73  -3.7839909  1.55889831
## 74  -2.1581786  1.40630549
## 87  -2.0769055 -0.05362730
## 91  -2.3086075  1.62423854
## 94  -0.2436776  1.95140633
## 95  -1.9314725  0.97127628
## 97  -1.5316307  0.48575787
## 104 -5.5155322  0.38658118
## 109 -6.2714692  1.54980508
## 111 -4.3863191 -1.24526156
## 113 -5.6196997 -0.66840529
## 116 -5.7667803 -1.84422852
## 120 -4.7162917  2.35998730
```

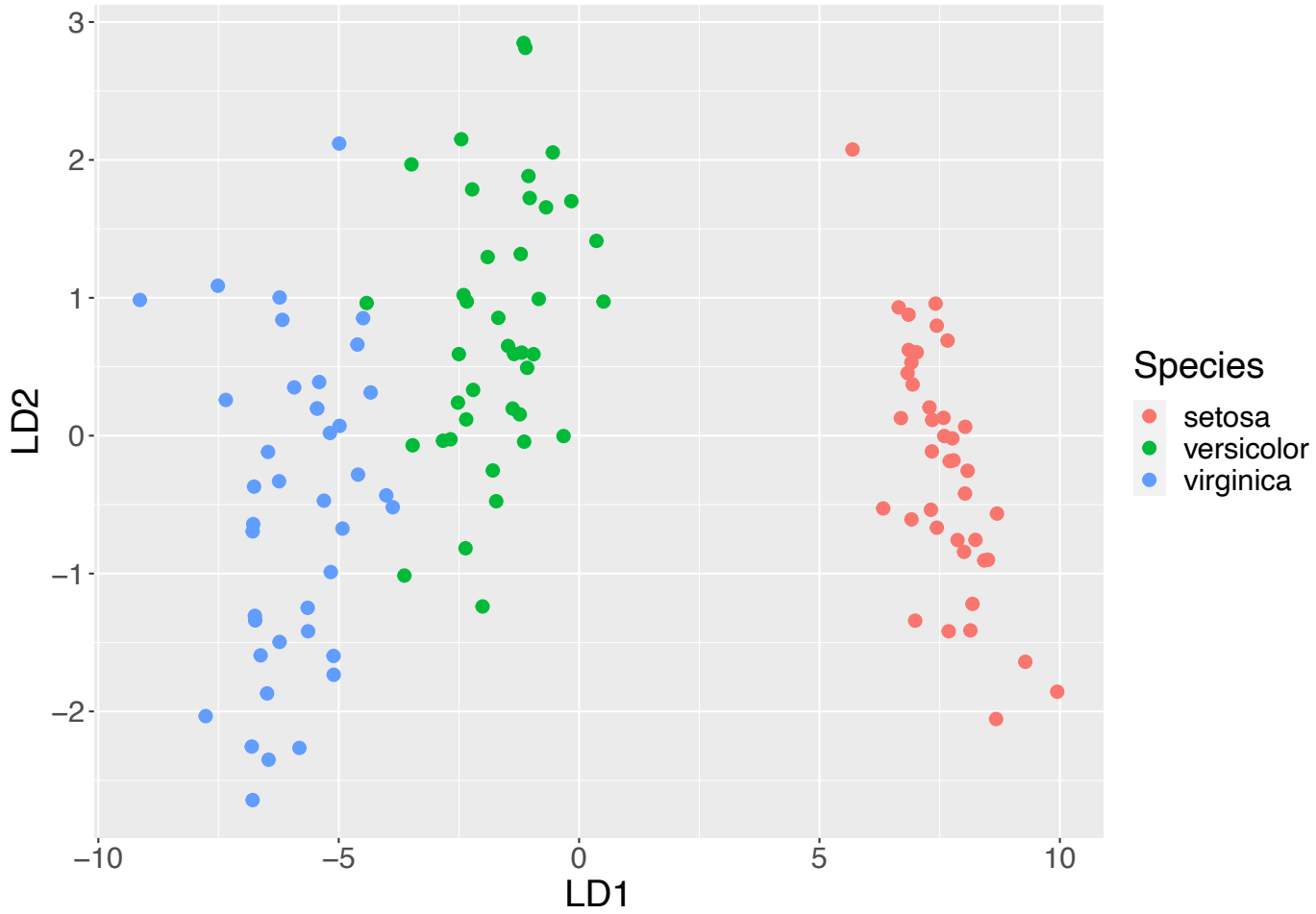
Visualisierung von Ergebnissen III

```
## 127 -4.0396234 -0.01782327
## 134 -3.7311676  0.98400416
## 138 -4.9012893 -0.17950933
## 147 -5.1622599  0.63288729
```

```
# plot(model)
lda.data <- cbind(train.transformed, predict(model)$x)
```

Visualisierung von Ergebnissen

```
# plot(model)
lda.data <- cbind(train.transformed, predict(model)$x)
ggplot(lda.data, aes(LD1, LD2)) + geom_point(aes(color = Species), size = 3) + theme(text = element_text(size = 20))
```

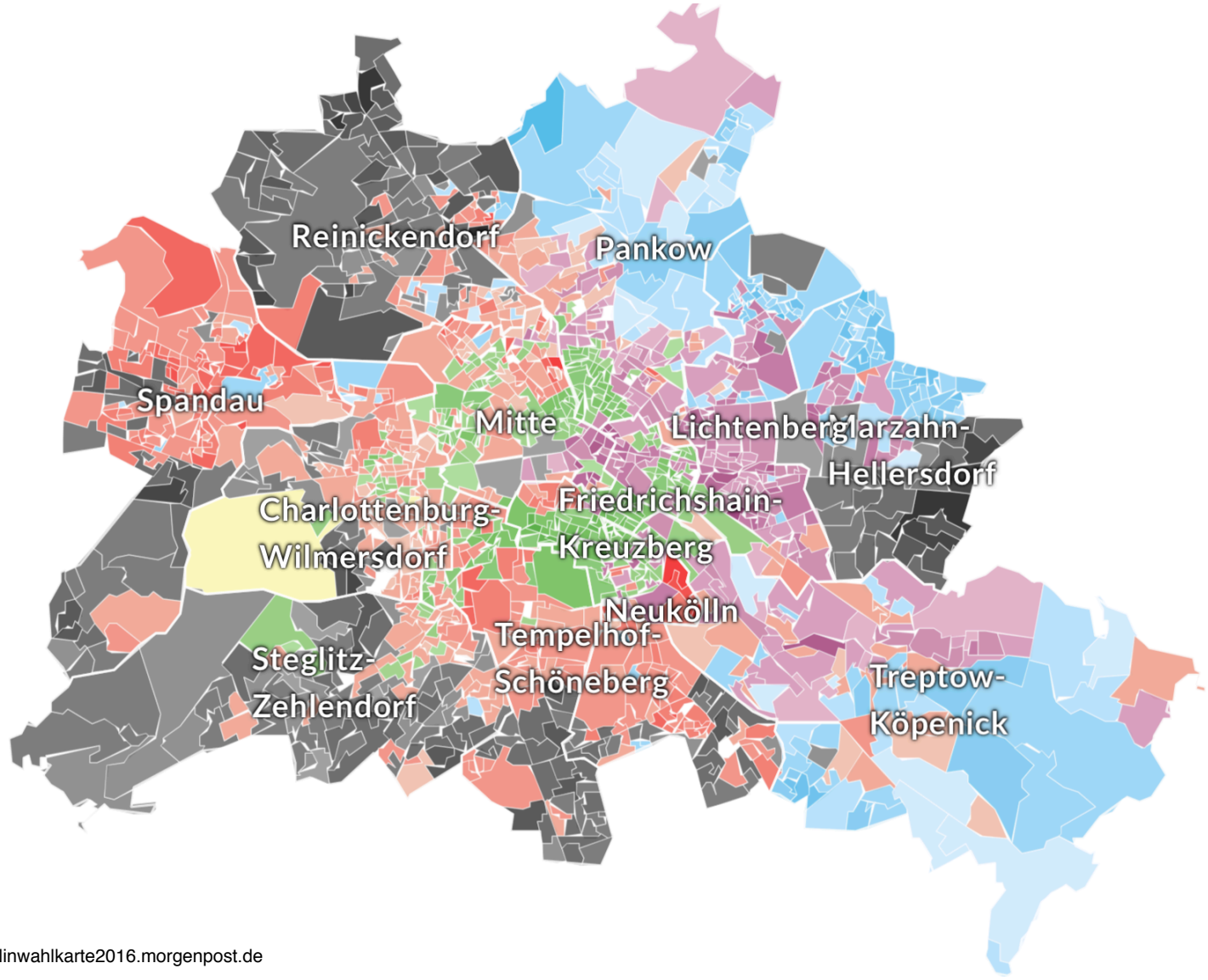


Klassifikationsbäume

Entscheidungsbäume

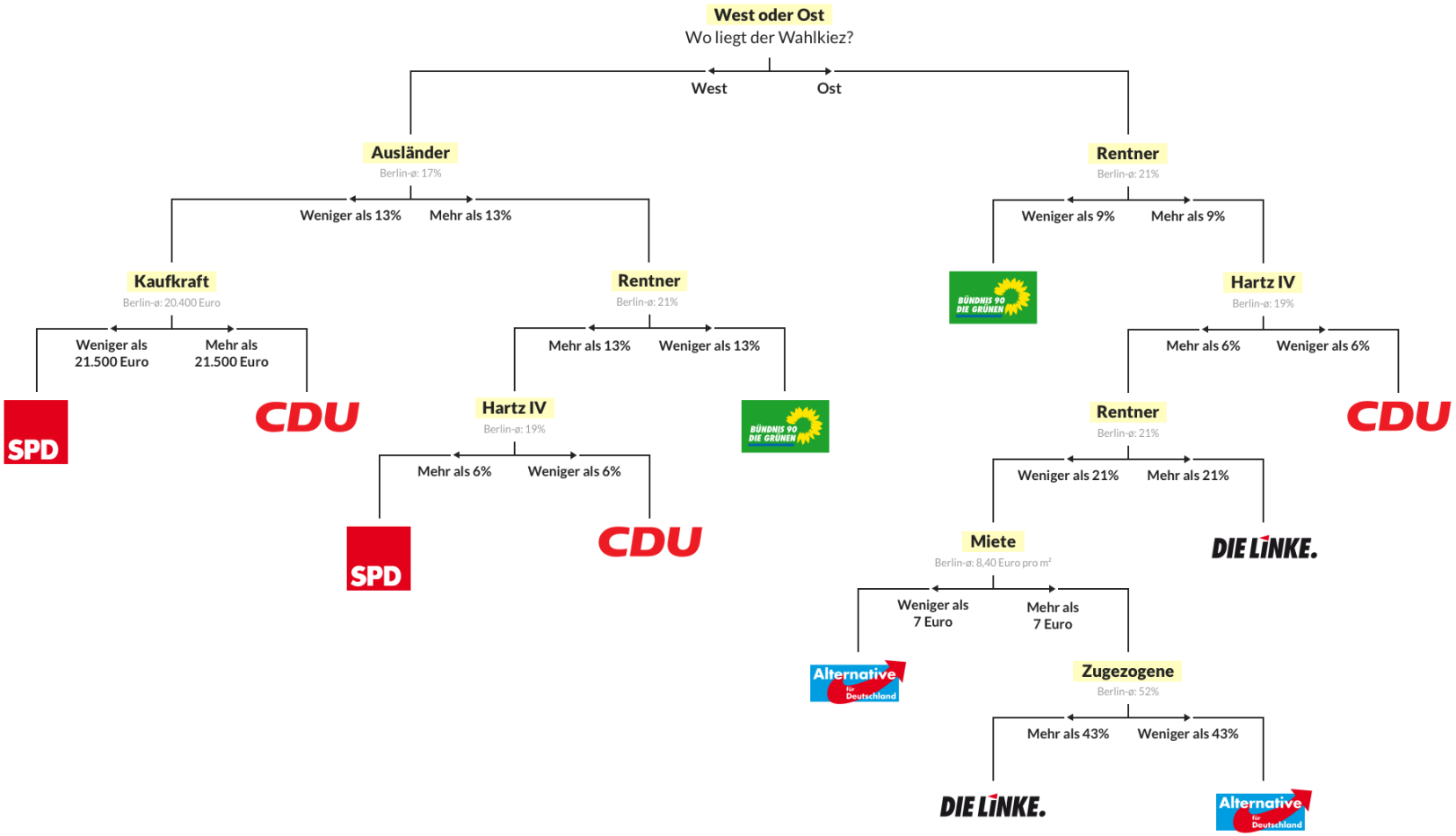
Klassifikationsbäume umfassen das Erstellen einer Reihe von binären Teilungen für die Prädiktorvariablen, um einen Baum zu erstellen, mit dem neue Beobachtungen in eine von zwei oder mehreren Gruppen eingeteilt werden können.

Abgeordnetenhauswahl Berlin 2016



http://berlinwahlkarte2016.morgenpost.de

Abgeordnetenhauswahl Berlin 2016: Entscheidungsbaum



Verfahren

1. Wählen Sie die Prädiktorvariable aus, die die Daten am besten in zwei Gruppen aufteilt, sodass die Reinheit (Homogenität) des Ergebnisses in den beiden Gruppen maximiert wird. Wenn der Prädiktor kontinuierlich ist, wählen Sie einen Schnittpunkt, der die Reinheit für die beiden erstellten Gruppen maximiert. Wenn die Prädiktorvariable kategorisch ist, kombinieren Sie die Kategorien, um zwei Gruppen mit maximaler Reinheit zu erhalten.
2. Trennen Sie die Daten in diese beiden Gruppen und setzen Sie den Vorgang für jede Untergruppe fort.
3. Wiederholen Sie die Schritte 1 und 2, bis eine Untergruppe weniger als eine Mindestanzahl von Beobachtungen enthält oder keine Teilungen die Verunreinigung über einen *festgelegten Schwellenwert* hinaus verringern. Die Untergruppen im endgültigen Satz werden als Endknoten bezeichnet. Jeder Endknoten wird basierend auf dem häufigsten Wert des Ergebnisses für die Stichprobe in diesem Knoten als die eine oder andere Kategorie des Ergebnisses klassifiziert.
4. Um einen Fall zu klassifizieren, führen Sie ihn in der Baumstruktur zu einem Endknoten aus und weisen Sie ihm den in Schritt 3 zugewiesenen modalen Ergebniswert zu.

Classification **A**nd **R**egression **T**ree

Verlustfunktion

$$J(k, t_k) = \frac{m_{\text{left}}}{m} G_{\text{left}} + \frac{m_{\text{right}}}{m} G_{\text{right}}$$

t_k Schwellenwert für das Merkmal k

$m_{\text{left/right}}$ Anzahl Datenpunkte

$G_{\text{left/right}}$ Gini Koeffizient oder Unreinheit (impurity)

$G_i = 1 - \sum_{k=1}^{n} p_{i,k}^2$ für den Knoten i und das Merkmal k

Nachteile von Entscheidungsbäumen

algorithmische Komplexität $O(n \times m \log(m))$

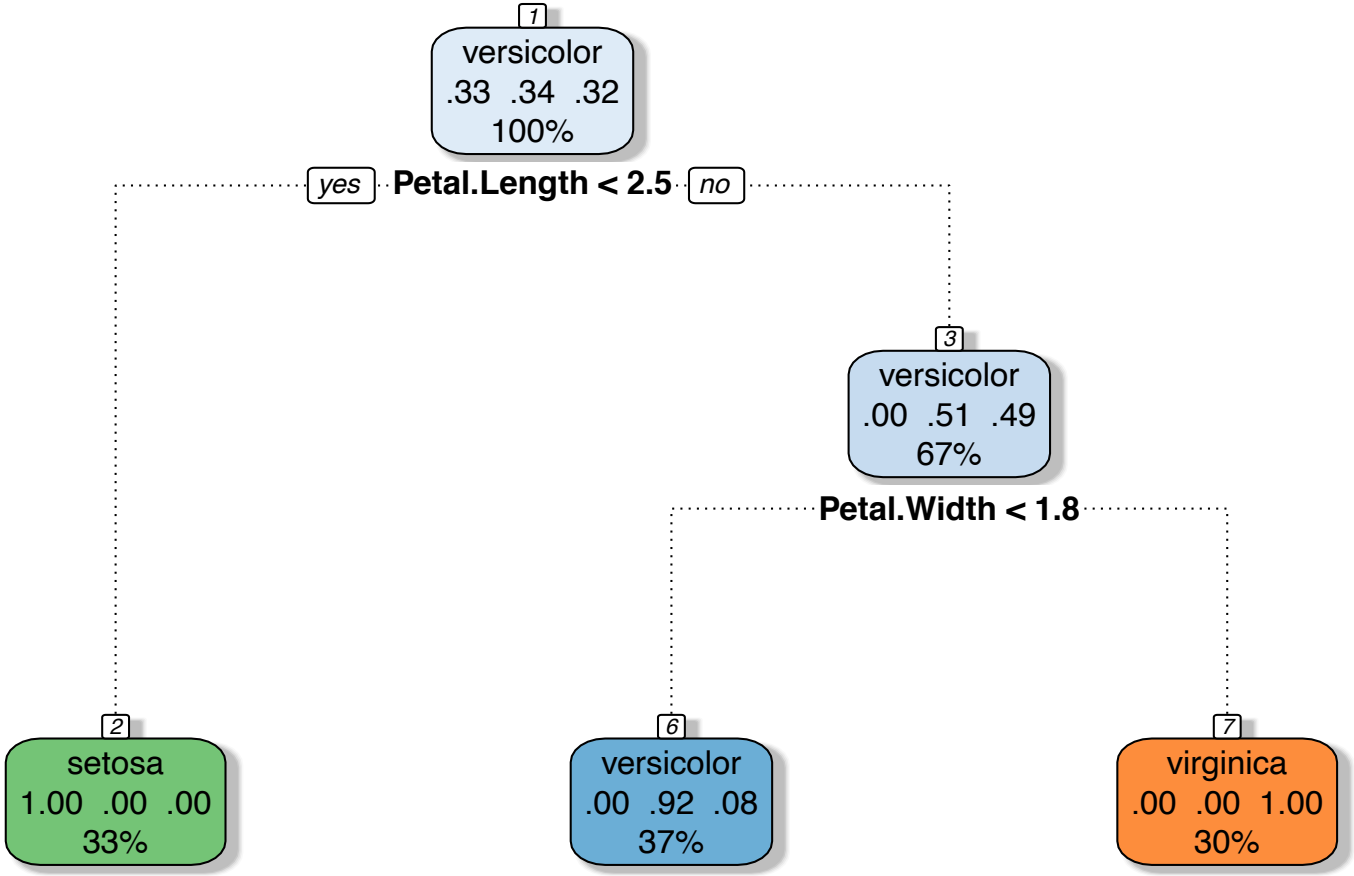
m Anzahl Datenpunkte

n Anzahl Merkmale

- Overfitting (kann z.B. durch Tiefenbegrenzung reduziert werden)
- Sensibel zu kleinen Veränderungen, z.B. Drehungen
- *Random Forest* besteht aus Bäumen, die auf zufällig ausgesuchten Teilmenge der Daten trainiert werden. Am Ende wird die unter allen Bäumen häufigste Vorhersage ausgewählt.

```
library(rpart)
library(gmodels)
library(rpart.plot)
library(rattle)
train <- sample(nrow(iris), 0.7 * nrow(iris))
df.train <- iris[train, ]
df.test <- iris[-train, ]
dtree <- rpart(Species ~ ., data = df.train, method = "class", parms = list(split = "gini"))
```

```
fancyRpartPlot(dtree)
```



Rattle 2021–Jan–12 08:02:01 vitaly

```
(df.test$Species == predict(dtree, df.test, type = "class"))
```

```
##  [1]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [13]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE
## [25]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [37]  TRUE  TRUE  TRUE FALSE  TRUE FALSE  TRUE  TRUE  TRUE
```

```
CrossTable(df.test$Species, predict(dtree, df.test, type = "class"))
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  45
##
##
##                 | predict(dtree, df.test, type = "class")
## df.test$Species |     setosa | versicolor |  virginica |  Row Total |
## ----------------|------------|------------|------------|------------|
##          setosa |         15 |          0 |          0 |         15 |
##                 |     20.000 |      5.000 |      5.000 |            |
##                 |      1.000 |      0.000 |      0.000 |      0.333 |
##                 |      1.000 |      0.000 |      0.000 |            |
##                 |      0.333 |      0.000 |      0.000 |            |
## ----------------|------------|------------|------------|------------|
##      versicolor |          0 |         13 |          1 |         14 |
##                 |      4.667 |     14.881 |      2.881 |            |
##                 |      0.000 |      0.929 |      0.071 |      0.311 |
##                 |      0.000 |      0.867 |      0.067 |            |
##                 |      0.000 |      0.289 |      0.022 |            |
## ----------------|------------|------------|------------|------------|
##       virginica |          0 |          2 |         14 |         16 |
##                 |      5.333 |      2.083 |     14.083 |            |
##                 |      0.000 |      0.125 |      0.875 |      0.356 |
##                 |      0.000 |      0.133 |      0.933 |            |
##                 |      0.000 |      0.044 |      0.311 |            |
## ----------------|------------|------------|------------|------------|
##    Column Total |         15 |         15 |         15 |         45 |
##                 |      0.333 |      0.333 |      0.333 |            |
## ----------------|------------|------------|------------|------------|
##
##
```

```
CrossTable(df.test$Species == predict(dtree, df.test, type = "class"))
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  45
##
##
##            |     FALSE |      TRUE |
##            |-----------|-----------|
##            |         3 |        42 |
##            |     0.067 |     0.933 |
##            |-----------|-----------|
##
##
##
##
```

Confusion matrix

```
library(caret)
confusionMatrix(predict(dtree, df.test, type = "class"), df.test$Species)
```

```
## Confusion Matrix and Statistics
##
##             Reference
## Prediction   setosa versicolor virginica
##   setosa         15          0         0
##   versicolor      0         13         2
##   virginica       0          1        14
##
## Overall Statistics
##
##                Accuracy : 0.9333
##                  95% CI : (0.8173, 0.986)
##     No Information Rate : 0.3556
##     P-Value [Acc > NIR] : 5.426e-16
##
##                   Kappa : 0.9
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: setosa Class: versicolor Class: virginica
## Sensitivity                 1.0000            0.9286           0.8750
## Specificity                 1.0000            0.9355           0.9655
## Pos Pred Value              1.0000            0.8667           0.9333
## Neg Pred Value              1.0000            0.9667           0.9333
## Prevalence                  0.3333            0.3111           0.3556
## Detection Rate              0.3333            0.2889           0.3111
## Detection Prevalence        0.3333            0.3333           0.3333
## Balanced Accuracy           1.0000            0.9320           0.9203
```

Qualitätsmaße für Klassifikationsverfahren

- Accuracy

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN}$$

- Spezifität

$$\text{Sp} = \frac{TN}{TN + FP}$$

- Sensitivität (recall)

$$\text{Sn} = \frac{TP}{TP + FN}$$

Die Entscheidungsbäume sind in der Lage nicht offentsichliche Muster in den Daten erkennen.

Klassifikation II

Vitaly Belik

Institut für Veterinär-Epidemiologie und Biometrie, FU Berlin

2021 January 19

Wiederholung

Wahrheitsmatrix (Confusion matrix)

Zwei Möglichkeiten für ein Klassifikationsverfahren (z.B. Ergebnis des Tests auf Sars-Cov-2[1]): *positive* ("1") und *negative* ("0").

		Vorhersage $\hat{y}=1$	Vorhersage $\hat{y}=0$	
Wirklichkeit	$y=1$	True Positive	False Negative	Sensitivity (Recall) $\mathrm{TP}/(y=1)$
	$y=0$	False Positive	True Negative	Specificity $\mathrm{TN}/(y=0)$
	Prevalence $(y=1)/\mathrm{total}$	Precision $\mathrm{TP}/(\hat{y}=1)$		Accuracy $(\mathrm{TP}+\mathrm{TN})/\mathrm{total}$

[1] PCR-Test auf Sars-Cov-2 hat Sensitivität zwischen 71% und 98%
https://www.aerzteblatt.de/archiv/214370/PCR-Tests-auf-SARS-CoV-2-Ergebnisse-richtig-interpretieren

Qualitätsmaße für Klassifikationsverfahren

- Accuracy

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN}$$

- Spezifität

$$\text{Sp} = \frac{TN}{TN + FP}$$

- Sensitivität (recall)

$$\text{Sn} = \frac{TP}{TP + FN}$$

- Precision

$$\text{Sn} = \frac{TP}{TP + FP}$$

- $F1$

$$F1 = \frac{2}{\text{recall}^{-1} + \text{precision}^{-1}}$$

Beispiel: Lahmheiten bei den Milchkühen I

Ergebnisse I

```
modelcont5 <- glmer(lameness ~ meansteps + meanlayingfreq + lactation + DIM + gemelk +
    Season + meansteps * Season + meanlayingfreq * DIM + meanlayingfreq * Season +
    lactation * DIM + lactation * Season + gemelk * Season + (1 + lactation | cow),
    data = datascaled, family = binomial, nAGQ = 0)
# control=glmerControl(optimizer = 'Nelder_Mead',optCtrl=list(maxfun=100000)))
# library(stargazer) stargazer(as.data.frame(summary(modelcont5)))
summary(modelcont5)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
##  Family: binomial  ( logit )
## Formula: lameness ~ meansteps + meanlayingfreq + lactation + DIM + gemelk +
##     Season + meansteps * Season + meanlayingfreq * DIM + meanlayingfreq *
##     Season + lactation * DIM + lactation * Season + gemelk *
##     Season + (1 + lactation | cow)
##    Data: datascaled
##
##      AIC      BIC   logLik deviance df.resid
##  20098.3  20304.2 -10023.2  20046.3    20274
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.0204 -0.4290 -0.1972  0.4130  4.3659
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  cow    (Intercept) 5.23582  2.2882
##         lactation   0.08462  0.2909   -0.03
## Number of obs: 20300, groups:  cow, 2758
##
## Fixed effects:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 -1.647916   0.114237 -14.425  < 2e-16 ***
## meansteps                   -0.046065   0.034337  -1.342 0.179736
## meanlayingfreq              -0.083298   0.038163  -2.183 0.029059 *
## lactation                    0.581329   0.044328  13.114  < 2e-16 ***
```

Ergebnisse II

```
## DIM                          -0.016491   0.074653  -0.221 0.825173
## gemelk                       -0.301042   0.040054  -7.516 5.65e-14 ***
## SeasonSpring                 -0.227372   0.106417  -2.137 0.032629 *
## SeasonSummer                  0.187096   0.583392   0.321 0.748435
## SeasonFall                    0.048797   0.120928   0.404 0.686567
## meansteps:SeasonSpring       -0.007842   0.050161  -0.156 0.875774
## meansteps:SeasonSummer       -0.338154   0.167204  -2.022 0.043134 *
## meansteps:SeasonFall          0.120655   0.061482   1.962 0.049710 *
## meanlayingfreq:DIM            0.055140   0.025778   2.139 0.032431 *
## meanlayingfreq:SeasonSpring   0.057737   0.048756   1.184 0.236335
## meanlayingfreq:SeasonSummer   0.174144   0.253733   0.686 0.492506
## meanlayingfreq:SeasonFall     0.119313   0.064009   1.864 0.062321 .
## lactation:DIM                -0.114232   0.029710  -3.845 0.000121 ***
## lactation:SeasonSpring        0.083318   0.039586   2.105 0.035316 *
## lactation:SeasonSummer        0.094387   0.157511   0.599 0.549011
## lactation:SeasonFall          0.003039   0.046563   0.065 0.947958
## gemelk:SeasonSpring           0.141362   0.050372   2.806 0.005010 **
## gemelk:SeasonSummer           1.026044   0.313452   3.273 0.001063 **
## gemelk:SeasonFall             0.165618   0.073943   2.240 0.025104 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# library(pander) pander(summary(modelcont5), round=3 library(sjstats)
library(performance)
r2(modelcont5)
```

```
## # R2 for Mixed Models
##
##   Conditional R2: 0.667
##      Marginal R2: 0.083
```

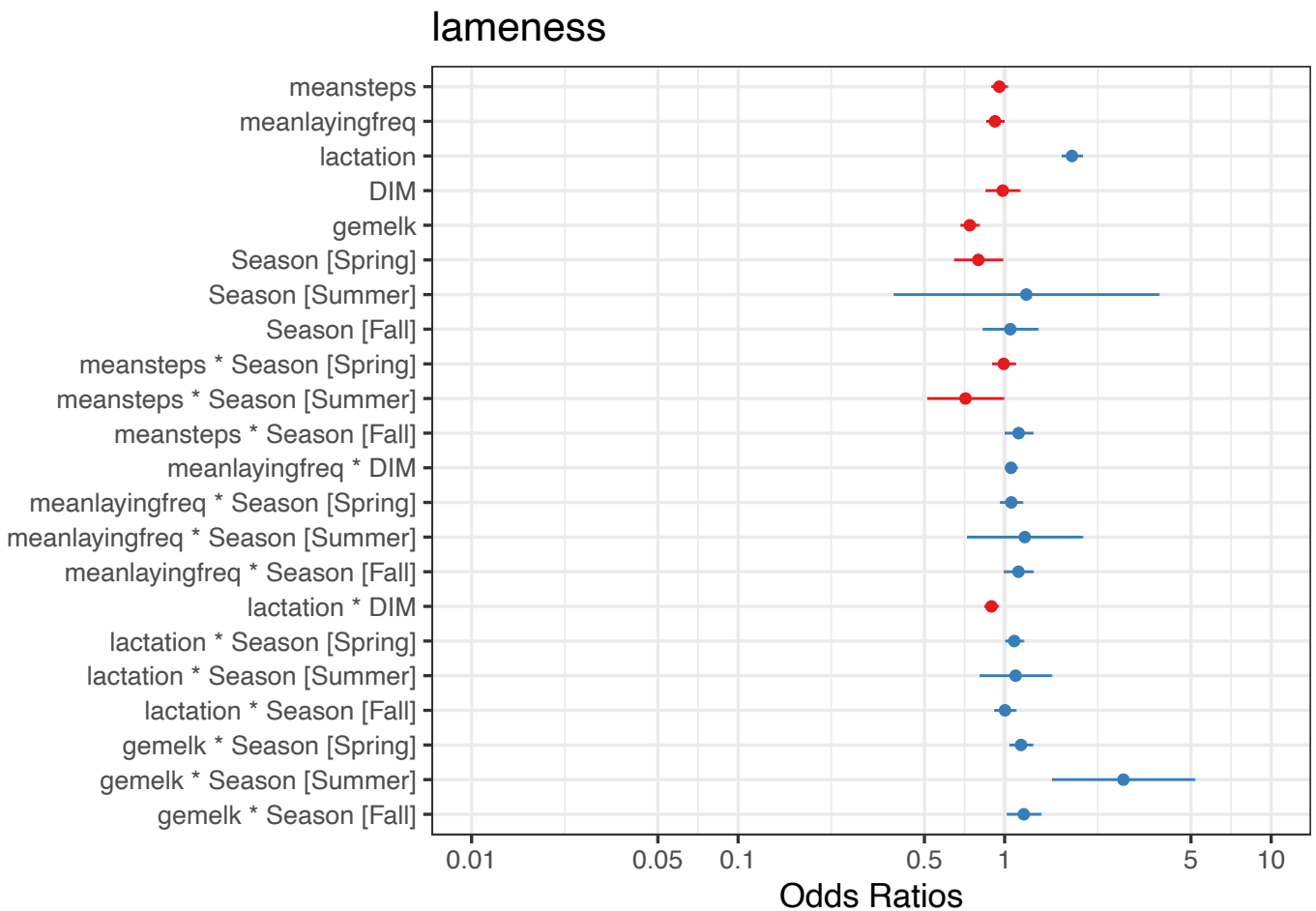
Odds Ratios I

```
library(gtsummary)
library(tibble)
modelcont5 %>% tbl_regression(exponentiate = TRUE, pvalue_fun = function(x) style_pvalue(x,
    digits = 2), estimate_fun = function(x) style_ratio(x, digits = 2)) %>% bold_p(t = 0.1) %>%
    bold_labels() %>% italicize_levels()
```

Characteristic	OR	95% CI	p-value
meansteps	0.95	0.89, 1.02	0.18
meanlayingfreq	0.92	0.85, 0.99	**0.029**
lactation	1.79	1.64, 1.95	**<0.001**
DIM	0.98	0.85, 1.14	0.83
gemelk	0.74	0.68, 0.80	**<0.001**
Season			
Winter			
Spring	0.80	0.65, 0.98	**0.033**
Summer	1.21	0.38, 3.78	0.75
Fall	1.05	0.83, 1.33	0.69
meansteps * Season			
*meansteps * Spring*	0.99	0.90, 1.09	0.88
*meansteps * Summer*	0.71	0.51, 0.99	**0.043**
*meansteps * Fall*	1.13	1.00, 1.27	**0.050**
meanlayingfreq * DIM	1.06	1.00, 1.11	**0.032**
meanlayingfreq * Season			
*meanlayingfreq * Spring*	1.06	0.96, 1.17	0.24
*meanlayingfreq * Summer*	1.19	0.72, 1.96	0.49
*meanlayingfreq * Fall*	1.13	0.99, 1.28	**0.062**
lactation * DIM	0.89	0.84, 0.95	**<0.001**
lactation * Season			
*lactation * Spring*	1.09	1.01, 1.17	**0.035**
*lactation * Summer*	1.10	0.81, 1.50	0.55
*lactation * Fall*	1.00	0.92, 1.10	0.95
gemelk * Season			
*gemelk * Spring*	1.15	1.04, 1.27	**0.005**
*gemelk * Summer*	2.79	1.51, 5.16	**0.001**
*gemelk * Fall*	1.18	1.02, 1.36	**0.025**

Visualisierung von Odds Ratios I

```
library(sjPlot)
library(sjlabelled)
library(sjmisc)
# theme_set(theme_sjplot())
library(ggplot2)
theme_set(theme_bw(base_size = 18))
plot_model(modelcont5, axis.lim = c(0.01, 10))
```



Logistische Regression

Es gibt Situationen wann die *Antwortvariable* nicht normal verteilt ist. Z.B. kann sie kategoriell und *binomial* oder *multinomial* sein.

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \sum_{i=1}^{p} \beta_i X_i.$$

Dabei ist $\pi = \mu_Y$ ein bedingter Mittelwert (d.h. die Wahrscheinlichkeit, dass $Y = 1$ vorausgesetzt die vorhandenen X-Werte).

$\frac{\pi}{1-\pi}$ ist das Odds-Ratio, dass $Y = 1$.

$\log\left(\frac{\pi}{1-\pi}\right)$ ist *log odds* oder *logit*.

Vorhersage gegen Wirklichkeit

```
library(pROC)
datascaled$pred <- predict(modelcont5, datascaled, type = "response", allow.new.levels = TRUE)  # otherwise errors
roc <- roc(datascaled$lameness, datascaled$pred, na.rm = TRUE)
auc(roc)

u <- data.frame(datascaled$lameness, datascaled$pred)
knitr::kable(head(na.omit(u), 15))
```

	datascaled.lameness	datascaled.pred
3	1	0.8702749
4	1	0.8016579
5	1	0.7981376
6	1	0.7784939
7	1	0.7527559
8	0	0.7433584
9	1	0.7410689
10	1	0.7457488
11	1	0.7925206
12	0	0.7679485
13	1	0.1917976
14	0	0.2298744
15	0	0.2186604
17	1	0.5916771
18	1	0.5401616

Calculating Confusion matrix by hand

```
pred <- predict(modelcont5, na.omit(datascaled), type = "response", allow.new.levels = TRUE)
pred_y <- as.numeric(pred > 0.5)
true_y <- as.numeric(na.omit(datascaled)$lameness == 1)
true_pos <- (true_y == 1) & (pred_y == 1)
true_neg <- (true_y == 0) & (pred_y == 0)
false_pos <- (true_y == 0) & (pred_y == 1)
false_neg <- (true_y == 1) & (pred_y == 0)

conf_mat <- matrix(c(sum(true_pos), sum(false_pos), sum(false_neg), sum(true_neg)),
    2, 2)
colnames(conf_mat) <- c("Yhat = 1", "Yhat = 0")
rownames(conf_mat) <- c("Y = 1", "Y = 0")
conf_mat
```

```
##       Yhat = 1 Yhat = 0
## Y = 1     8007     1719
## Y = 0     1437     9137
```

```
# accuracy
(conf_mat[1, 1] + conf_mat[2, 2])/sum(conf_mat[, ])
```

```
## [1] 0.844532
```

```
# precision
conf_mat[1, 1]/sum(conf_mat[, 1])
```

```
## [1] 0.8478399
```

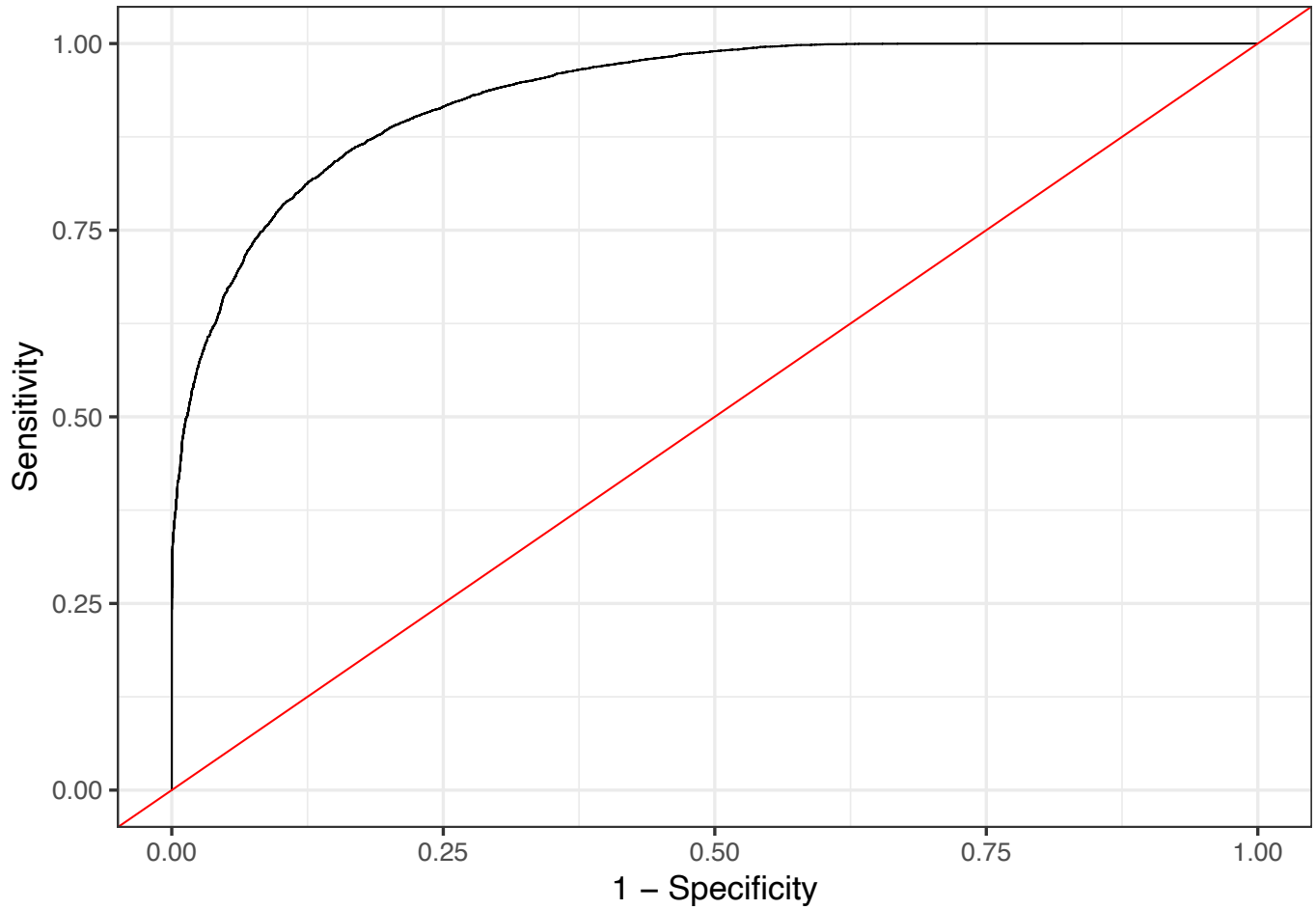
```
# sensitivity (recall)
conf_mat[1, 1]/sum(conf_mat[1, ])
```

```
## [1] 0.8232572
```

```
# specificity
conf_mat[2, 2]/sum(conf_mat[2, ])
```

```
## [1] 0.8641006
```

```r
library(ggthemes)
new_df <- data.frame(roc$specificities, roc$sensitivities)
colnames(new_df) <- c("Specificity", "Sensitivity")
# ggplot(new_df, aes(x = 1- spec, y = sens)) + geom_line()
ggplot() + geom_line(data = new_df, aes(x = 1 - Specificity, y = Sensitivity)) +
    geom_abline(intercept = 0, slope = 1, color = "red", size = 0.5)
```



Source

Peter Bruce, Andrew Bruce & Peter Gedeck. Practical Statistics for Data Scientists

Poisson Regression

Vitaly Belik

Institut für Veterinär-Epidemiologie und Biometrie, FU Berlin

2021 January 26

Wiederholung

Poissonregression

Die Poisson-Regression ist nützlich, wenn Sie eine Ergebnisvariable vorhersagen, die die Anzahl aus einer Reihe kontinuierlicher und / oder kategorialer Prädiktorvariablen darstellt.

Wir interessieren uns für den Einfluss einer Behandlung mit Antiepileptika auf die Anzahl der Anfälle, die über einen Zeitraum von acht Wochen nach Beginn der Therapie auftreten [Breslow seizure data (Breslow, 1993)].

Es wurden Daten zum Alter und zur Anzahl der Anfälle gesammelt, die von Patienten gemeldet wurden, die während eines Zeitraums von acht Wochen vor und acht Wochen nach der randomisierten Aufteilung in eine Arzneimittel- oder Placebo-Gruppe an einfachen oder komplexen partiellen Anfällen litten.

SumY (die Anzahl der Anfälle im Zeitraum von acht Wochen nach der Randomisierung) ist die Antwortvariable. Der Behandlungszustand (*Trt*), das Alter in Jahren (*Age*) und die Anzahl der Anfälle, die im Basiszeitraum von acht Wochen (*Basis*) gemeldet wurden, sind die Prädiktorvariablen.

Die Baseline-Anzahl der Anfälle und das Alter sind aufgrund ihrer möglichen Auswirkung auf die Antwortvariable enthalten.

Fragestellung: verringert die medikamentöse Behandlung die Anzahl der Anfälle nach Berücksichtigung der Kovariaten oder nicht?

Beispiel: Breslow seizure data

```
data(breslow.dat, package = "robust")
names(breslow.dat)
```

```
##  [1] "ID"    "Y1"    "Y2"    "Y3"    "Y4"    "Base"  "Age"   "Trt"   "Ysum"
## [10] "sumY"  "Age10" "Base4"
```

```
summary(breslow.dat[c(6, 7, 8, 10)])
```

```
##       Base             Age               Trt          sumY       
##  Min.   :  6.00   Min.   :18.00   placebo  :28   Min.   :  0.00  
##  1st Qu.: 12.00   1st Qu.:23.00   progabide:31   1st Qu.: 11.50  
##  Median : 22.00   Median :28.00                  Median : 16.00  
##  Mean   : 31.22   Mean   :28.34                  Mean   : 33.05  
##  3rd Qu.: 41.00   3rd Qu.:32.00                  3rd Qu.: 36.00  
##  Max.   :151.00   Max.   :42.00                  Max.   :302.00
```

Beispiel: Breslow seizure data

```
ggplot(breslow.dat, aes(x = sumY)) + geom_histogram(fill = "transparent", color = "black")
```

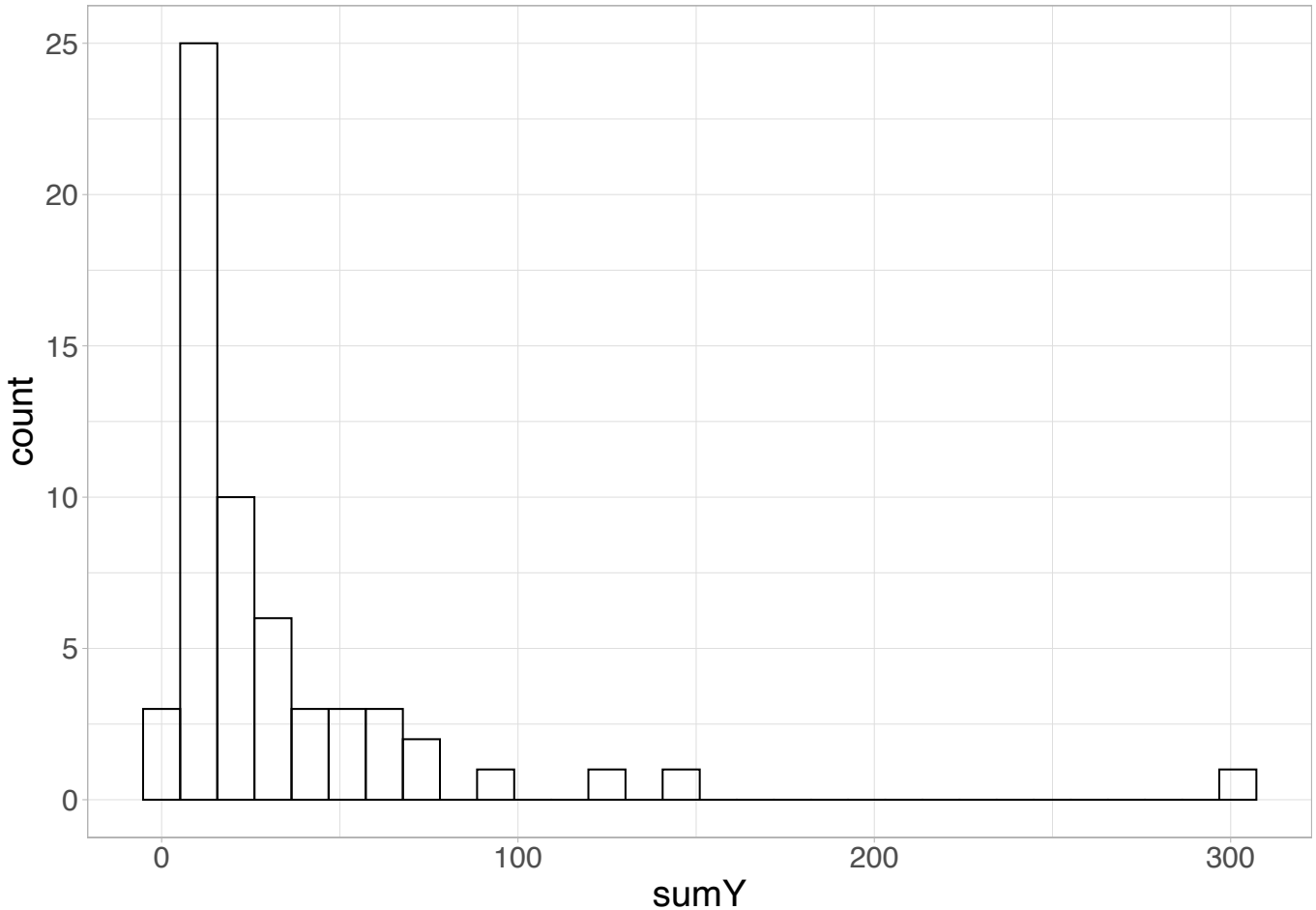
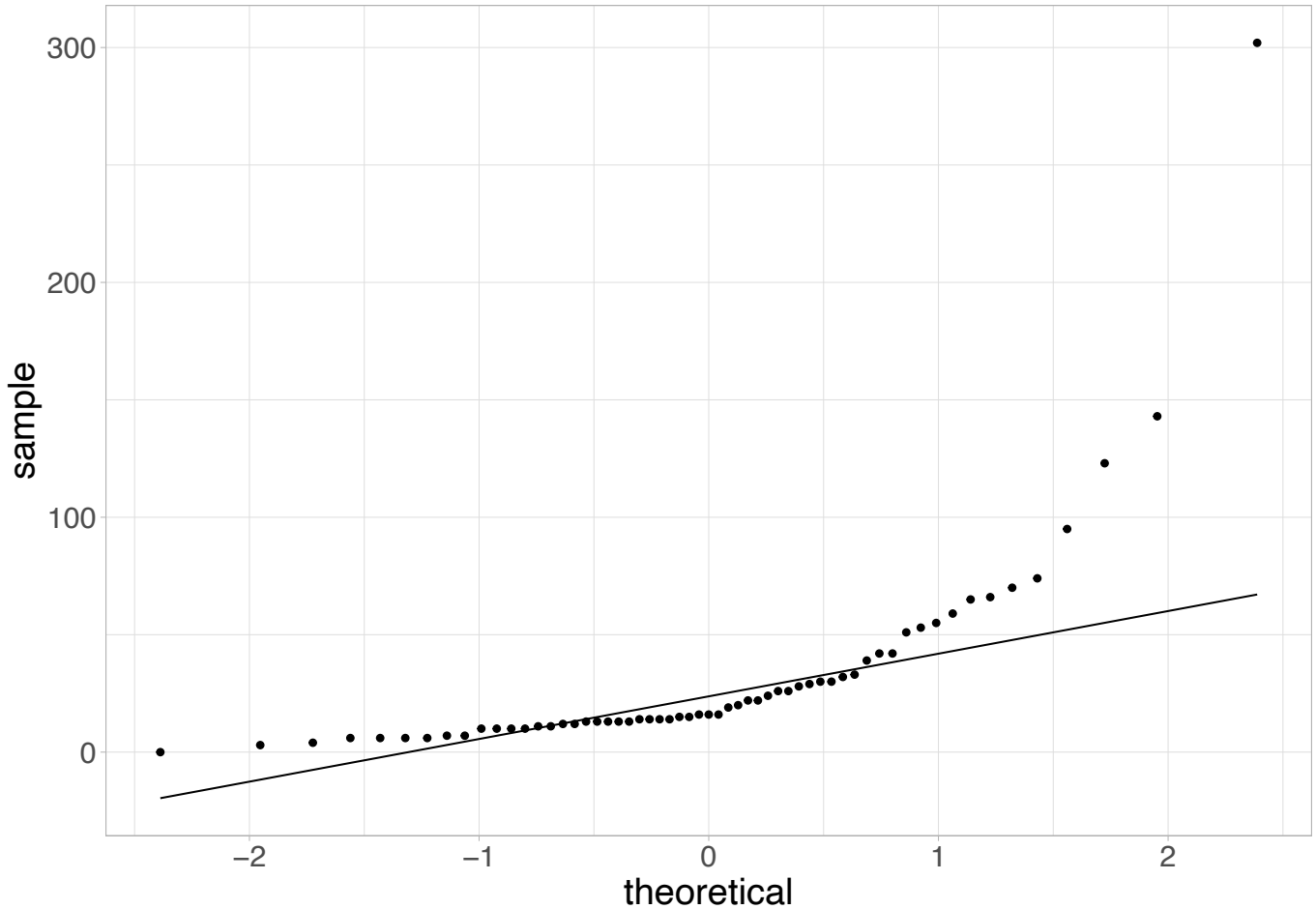


Figure 1: Verteilung von Anzahl der Anfälle nach der Behandlung.

```
ggplot(breslow.dat, aes(sample = sumY)) + stat_qq() + stat_qq_line()
```



```
ggplot(breslow.dat, aes(y = sumY, x = Trt)) + geom_boxplot(fill = "transparent",
    color = "black")
```

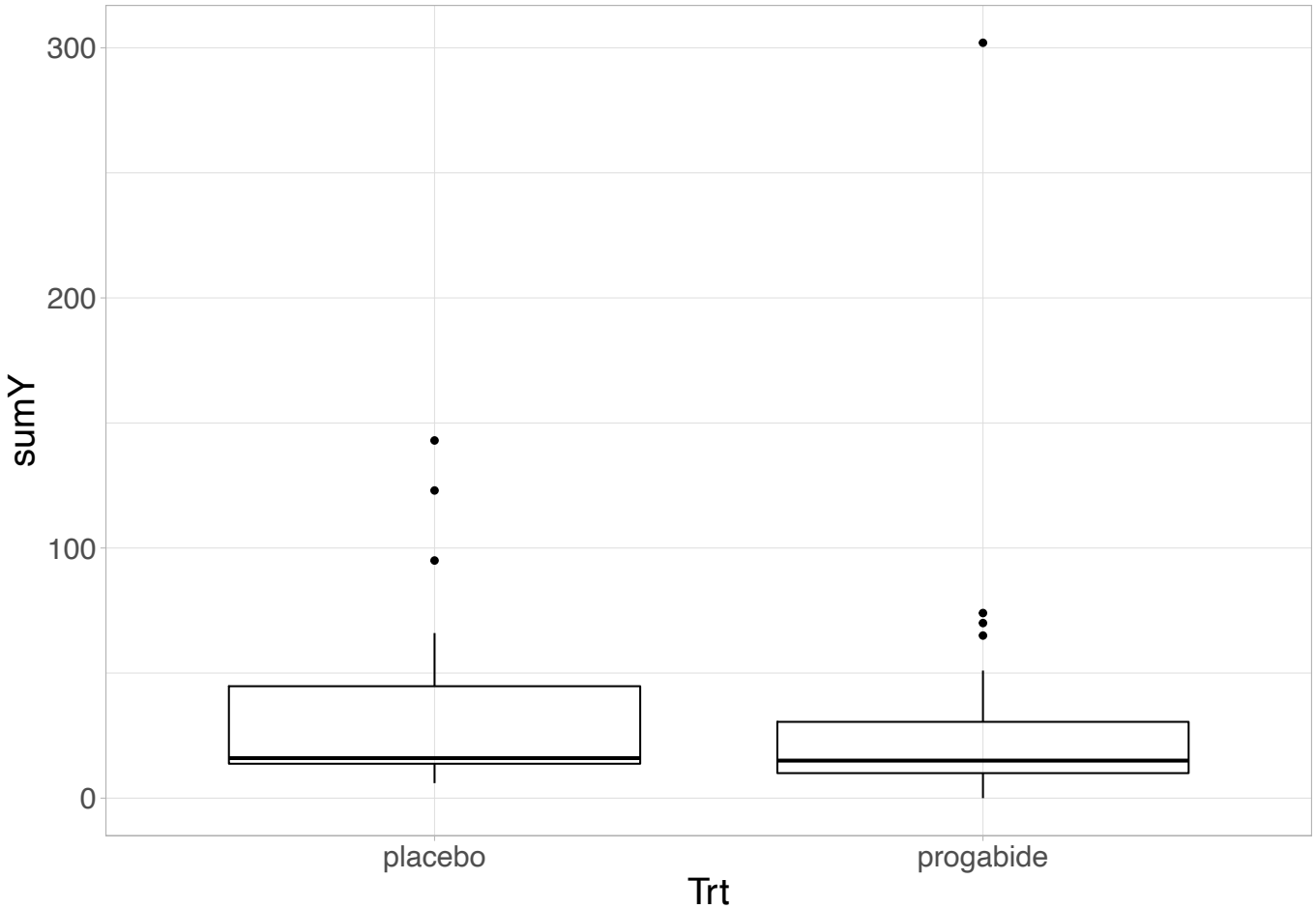


Figure 2: Anzahl der Anfälle vor und nach der Behandlung.

```
ggplot(breslow.dat, aes(y = log10(sumY), x = Trt)) + geom_boxplot(fill = "transparent",
    color = "black")
```

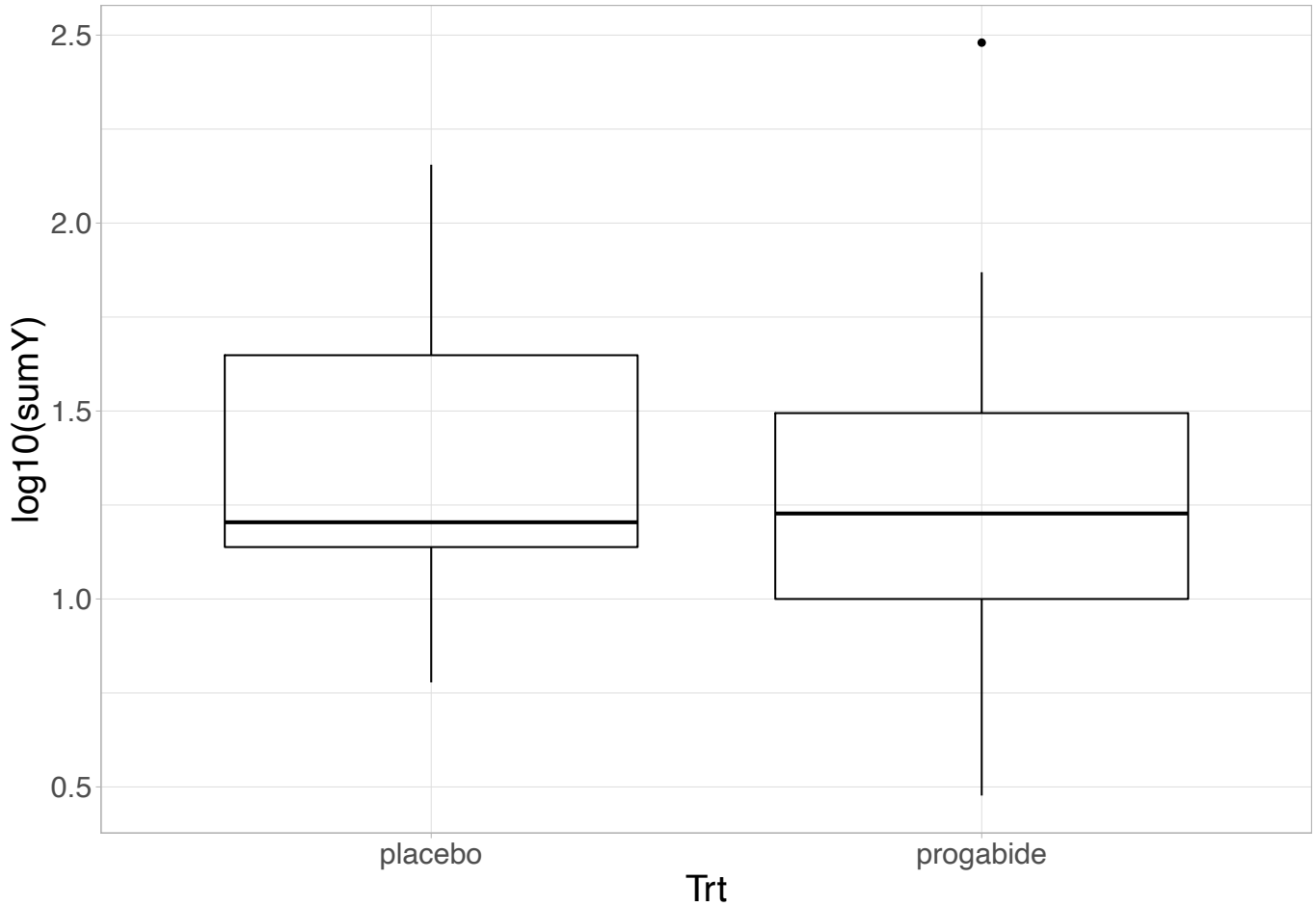


Figure 3: Anzahl der Anfälle vor und nach der Behandlung.

Sie können die Verzerrung der abhängigen Variablen und das mögliche Vorhandensein von Ausreißern deutlich erkennen.

Auf den ersten Blick scheint die Anzahl der Anfälle im Arzneimittel-Gruppe geringer zu sein und weist eine geringere Varianz auf.

Bei Poisson-verteilten Daten geht eine geringere Varianz mit einen kleineren Mittelwert einher.

Poisson regression I

```r
fit <- glm(sumY ~ Base + Age + Trt, data = breslow.dat, family = poisson())
summary(fit)
```

```
##
## Call:
## glm(formula = sumY ~ Base + Age + Trt, family = poisson(), data = breslow.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -6.0569  -2.0433  -0.9397   0.7929  11.0061
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.9488259  0.1356191  14.370  < 2e-16 ***
## Base          0.0226517  0.0005093  44.476  < 2e-16 ***
## Age           0.0227401  0.0040240   5.651 1.59e-08 ***
## Trtprogabide -0.1527009  0.0478051  -3.194   0.0014 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 2122.73  on 58  degrees of freedom
## Residual deviance:  559.44  on 55  degrees of freedom
## AIC: 850.71
##
## Number of Fisher Scoring iterations: 5
```

```r
anova(fit)
```

Poisson regression II

```
## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: sumY
##
## Terms added sequentially (first to last)
##
##
##      Df Deviance Resid. Df Resid. Dev
## NULL                    58    2122.73
## Base  1  1508.11        57     614.62
## Age   1    44.96        56     569.66
## Trt   1    10.21        55     559.44
```

Poisson regression: theory

- Poisson-Regression

$$\log\left(\mathbb{E}(Y|x)\right) = \beta_0 + \sum_i \beta_i x_i$$

- Poissonsche Wahrscheinlichkeitsdiche:

$$P_{\lambda t}(k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

Wahrscheinlichkeit k Ereignisse in der Zeit t zu sehen. Der Erwartungswert (Mittelwert) und die Varianz von der Poissonverteilung ist durch $\lambda = \mathbb{E}(Y|x)$ gegeben.

```
coef(fit)
```

```
##  (Intercept)         Base          Age Trtprogabide
##   1.94882593   0.02265174   0.02274013  -0.15270095
```

```
confint(fit)
```

```
##                     2.5 %      97.5 %
## (Intercept)   1.68188804  2.21355355
## Base          0.02165019  0.02364694
## Age           0.01483716  0.03061253
## Trtprogabide -0.24646125 -0.05904066
```

Der Regressionsparameter 0,0227 für Alter zeigt an, dass ein Anstieg des Alters um ein Jahr mit einem Anstieg der logarithmischen mittleren Anzahl von Anfällen um 0,02 verbunden ist, wobei die Baseline-Anzahl der Anfälle und der Behandlungsart konstant gehalten werden.

```
exp(coef(fit))
```

```
##  (Intercept)         Base          Age Trtprogabide
##    7.0204403    1.0229102    1.0230007    0.8583864
```

```
exp(confint(fit))
```

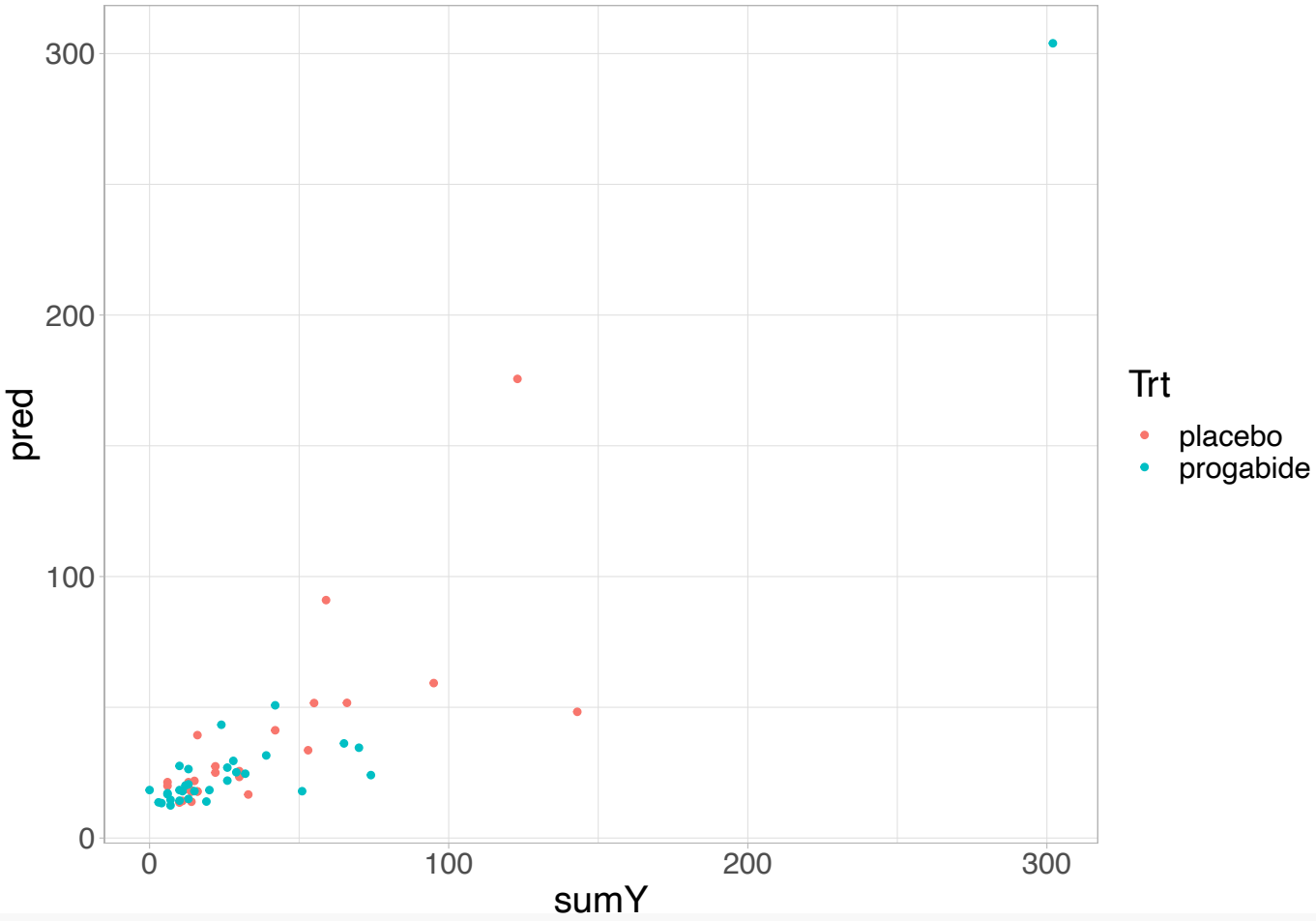
```
##                   2.5 %    97.5 %
## (Intercept)  5.3756959 9.1481671
## Base         1.0218863 1.0239287
## Age          1.0149478 1.0310859
## Trtprogabide 0.7815616 0.9426684
```

Ein Anstieg des Alters um ein Jahr multipliziert die erwartete Anzahl von Anfällen mit 1,023, wobei die anderen Variablen konstant bleiben.

Noch wichtiger ist, dass eine Änderung von *Trt* um eine Einheit (d.h. der Wechsel von Placebo zu Progabid) die erwartete Anzahl von Anfällen mit 0,86 multipliziert. Es ist somit einen Rückgang der Anzahl der Anfälle in der Arzneimittelgruppe um 20% im Vergleich zur Placebogruppe zu erwarten, wobei die Anzahl der Anfälle und das Alter konstant bleiben.

Die potenzierten Parameter im Poisson-Modell wirken sich multiplikativ (wie in der logistischen Regression) und nicht additiv auf die Antwortvariable aus.

Visualisierung




```
# ggplot(breslow.dat, aes(x = Age, y = sumY, colour = Trt)) + geom_point() +
# geom_smooth(method='glm', family='poisson', se=TRUE)
```

Overdispersion

In einer Poisson-Verteilung sind Varianz und Mittelwert gleich. Eine Überdispersion (Overdispersion) tritt bei der Poisson-Regression auf, wenn die beobachtete Varianz der Antwortvariablen größer ist als durch die Poisson-Verteilung vorhergesagt. Beim Umgang mit Zähldaten eine Überdispersion tritt häufig auf und kanns sich negativ auf die Interpretation der Ergebnisse auswirken.

Gründe für Überdispersion

- Das Weglassen einer wichtigen Prädiktorvariablen kann zu einer Überdispersion führen.
- Überdispersion kann auch durch ein Phänomen verursacht werden, das als Zustandsabhängigkeit bezeichnet wird. Innerhalb von Beobachtungen wird angenommen, dass jedes Ereignis in einer Zählung unabhängig ist. Für die Anfallsdaten würde dies bedeuten, dass für jeden Patienten die Wahrscheinlichkeit eines Anfalls unabhängig voneinander ist. Diese Annahme ist jedoch oft unhaltbar. Für eine bestimmte Person ist es unwahrscheinlich, dass die Wahrscheinlichkeit eines ersten Anfalls mit der Wahrscheinlichkeit eines 40. Anfalls übereinstimmt, vorausgesetzt es gab schon 39 Anfälle.
- In Längsschnittstudien kann eine Überdispersion durch die Clusterbildung verursacht werden die in wiederholten Messungen vorkommt.
- Wenn eine Überdispersion vorliegt und diese im Modell nicht berücksichtigt wird, führt es zu sehr kleinen Standardfehlern und Konfidenzintervallen sowie zu sehr liberalen Signifikanztests (d.h. man findet Effekte, die nicht vorhanden sind).

Overdispersion: Beispiel und Test

```
deviance(fit)/df.residual(fit)
```

```
## [1] 10.1717
```

```
library(qcc)
qcc.overdispersion.test(breslow.dat$sumY, type = "poisson")
```

```
##
## Overdispersion test Obs.Var/Theor.Var Statistic p-value
##        poisson data          62.87013  3646.468       0
```

Der Signifikanztest hat einen p-Wert von weniger als 0,05, was stark auf das Vorhandensein einer Überdispersion hinweist.

"Residual deviance" größer als "residual degress of freedom" deutet auf Überdispersion.

Quasi-Poissonische Regression I

Um die Overdispersion zu berücksichtigen kann man quasipoissonische Regression durchführen.

```
fit.od <- glm(sumY ~ Base + Age + Trt, data = breslow.dat, family = quasipoisson())
summary(fit.od)
```

```
##
## Call:
## glm(formula = sumY ~ Base + Age + Trt, family = quasipoisson(),
##     data = breslow.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -6.0569  -2.0433  -0.9397   0.7929  11.0061
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.948826   0.465091   4.190 0.000102 ***
## Base          0.022652   0.001747  12.969  < 2e-16 ***
## Age           0.022740   0.013800   1.648 0.105085
## Trtprogabide -0.152701   0.163943  -0.931 0.355702
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 11.76075)
##
##     Null deviance: 2122.73  on 58  degrees of freedom
## Residual deviance:  559.44  on 55  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

Quasi-Poissonische Regression II

Beachten Sie, dass die Parameterschätzungen im Quasi-Poisson-Ansatz mit denen des Poisson-Ansatzes identisch sind.

Die Standardfehler sind jedoch viel größer. In diesem Fall haben die größeren Standardfehler zu p-Werten für Trt (und Alter) geführt, die größer als 0,05 sind.

Wenn Sie die Überdispersion berücksichtigen, gibt es keine ausreichenden Beweise dafür, dass das Medikamentenschema die Anzahl der Anfälle stärker reduziert als die Einnahme eines Placebos, nachdem die Anfallsrate und das Alter zu Studienbeginn kontrolliert wurden.

Poissonregression mit variablen Zeitintervallen

Unsere Diskussion über die Poisson-Regression beschränkte sich auf Antwortvariablen, die eine Zählung über einen festgelegten Zeitraum messen (z. B. Anzahl der Anfälle in einem Zeitraum von acht Wochen, Anzahl der Verkehrsunfälle im vergangenen Jahr oder Anzahl des sozialen Verhaltens an einem Tag). Die Zeitdauer ist über die Beobachtungen hinweg konstant. Sie können jedoch Poisson-Regressionsmodelle anpassen, bei denen der Zeitraum für jede Beobachtung variieren kann. In diesem Fall ist die Antwortvariable eine Rate.

$$\log\left(\frac{\mathbb{E}(Y|x)}{time}\right) = \beta_0 + \sum_i \beta_i x_i$$

$$\log\left(\mathbb{E}(Y|x)\right) = \log(time) + \beta_0 + \sum_i \beta_i x_i$$

```
glm(sumY ~ Base + Age + Trt, data = breslow.dat, offset = log(time), family = poisson)
```

Zero inflated models

Es gibt Fälle, in denen die Anzahl der Nullen in einem Datensatz größer ist als vom Poisson-Modell vorhergesagt. Dies kann auftreten, wenn es eine Untergruppe der Grundgesamtheit gibt, die sich niemals auf das gezählte Verhalten (z.B. Rauchen) einlassen würde.

In solchen Fällen können Sie die Daten mithilfe eines Ansatzes analysieren, der als *zero-inflated Poisson regression* bezeichnet wird. Stellen Sie sich dies als ein Modell vor, das eine logistische Regression (zur Vorhersage struktureller Nullen) und ein Poisson-Regressionsmodell (zur Vorhersage von Zählungen für Beobachtungen, die keine strukturellen Nullen sind) kombiniert. Die zero inflated Poisson-Regression kann z.B. mithilfe der Funktion *zeroinfl()* im *pscl*-Paket angepasst werden.

Literatur

1. Peter Bruce, Andrew Bruce & Peter Gedeck. Practical Statistics for Data Scientists.
2. Kabakoff, R in Action (2d Edition).

Grundlagen der Versuchsplanung (Design of Experiments)

Vitaly Belik

Institut für Veterinär-Epidemiologie und Biometrie, FU Berlin

02/02/2021

Wiederholung

Einführung in die Versuchsplanung

Versuchsplanung (Einführung)

Figure 1: Bestimmung der Gewichte mit der Waage [nach H. Hotelling]

Versuchsplanung (Einführung)(1)

$$\boldsymbol{X} = M\boldsymbol{\theta} + \boldsymbol{\xi}$$

$\boldsymbol{X} = (X_1, ..., X_8)^T$, $\boldsymbol{\theta} = (\theta_1, ..., \theta_8)^T$, $\boldsymbol{\xi} = (\xi_1, ..., \xi_8)^T$ is iid with $\mu = 0$ and σ.

Versuchsplanung (Einführung)(1)

$$\boldsymbol{X} = M\boldsymbol{\theta} + \boldsymbol{\xi}$$

$\boldsymbol{X} = (X_1, ..., X_8)^T$, $\boldsymbol{\theta} = (\theta_1, ..., \theta_8)^T$, $\boldsymbol{\xi} = (\xi_1, ..., \xi_8)^T$ is iid with $\mu = 0$ and σ.

$$M = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 \end{pmatrix}$$

M ist eine **Hadamar-Matrix** – bestehend aus -1 und 1 und alle Spalten sind orthogonal zueinander, ebenso alle Zeilen.

Versuchsplanung (Einführung)(2)

$$\hat{\boldsymbol{\theta}} = M^{-1}\boldsymbol{X}$$

$$M^{-1} = \frac{1}{8}\begin{pmatrix} 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$\hat{\theta}_1 = \boldsymbol{w}\boldsymbol{X}/8$$

$$\boldsymbol{w} = (1, 1, 1, 1, -1, -1, -1, -1, -1)$$

$$Var(\theta_1) = Var(\boldsymbol{\xi}/8) = \sigma^2/8$$

Treffgenaugkeit und Präzision

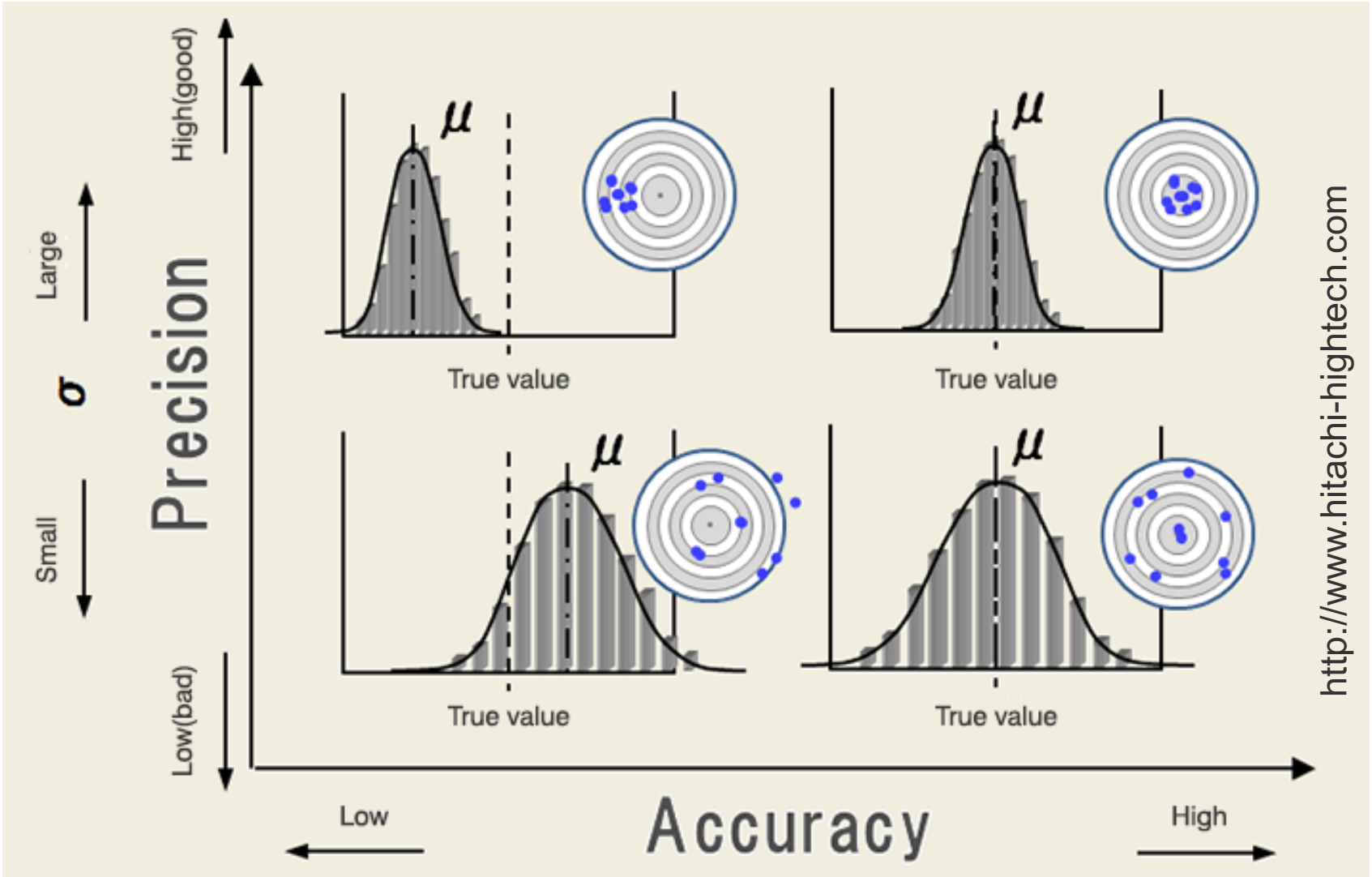


Figure 2: Accuracy vs. precision

Statistische Versuchsplanung: allgemeine Aspekte

Vor dem Versuch muss man sich Gedanken um die folgenden Aspekte machen und ggf. Inanspruchnahme der statistischen Beratung in Erwägung ziehen.

- Das vorher formulierte Modell muss mathematisch formalisiert und die entsprechenden Hypothesen gewählt werden
- Es müss geklärt werden welche Parameter mit welchen statistischen Verfahren und auf welchem Signifikanzniveau ermittelt werden sollen. Dabei soll man beachten ob die gewonnenen Daten die entsprechenden Vorausetzungen für die geplante statistische Auswertung erfüllen.
- Die Wahl der Faktorstufen (möglichst äquidistant) und die Anzahl der Wiederholungen (möglichst balansiert) soll bewusst erfolgen.
- Die Aufzeichnung der Daten muss geplant werden.

Treffgenaugkeit und Präzision

Treffgenaugkeit

Systematische Abweichung (Bias, Verzerrung) des gemessenen Mittelwertes von wahrem Mittelwert der Grundgesamtheit (z.B. Körpergewicht gemessen durch eine alte Waage) ist auf mangelhafte Treffgenaugkeit zurückzuführen.

Präzision

Streuung um den experimentellen Mittelwert gibt Aufschluss über die Präzision der Messungen.

Gründsuatze der Versuchsplanung

- Ceteris-paribus-Prinzip
- Widerholungen
- Randomisieren
- Blockbildung
- Faktorielle Experimente
- Symmetrischer Aufbau
- Wirtschaflichkeit

Ceteris-paribus-Prinzip

Man will den Einfluss ganz bestimmter bekannter Faktoren untersuchen. Daher versucht man alle unbekannten oder nicht berücksichtigten Faktoren, die Einfluss auf die Messergebnisse haben könnten, möglichst konstant zu halten.

Widerholungen

- Durch Wiederholungen lässt sich der Schätzwert für den Versuchsfehler bestimmen.
- Der Versuchsfehler verringert sich mit der Anzahl der Wiederholungen ($s = \sigma/\sqrt{n}$).
- Daher soll man im Experiment stets Wiederholungen einplanen, um die Präzision und ihre Größe bestimmen zu können.

Randomisieren

- Wenn die verschidenen Versuchseinheiten zufällig den jeweiligen Behandlungen zugeordnet werde, spricht man von *Randomisieren*

Beispiel: Drei Präparate sollen an 12 Mäusen getestet werden. Man wird jedes Präparat jeweils 4 "Versuchseinheiten" (Mäusen) verabreichen. Die Zuordung erfolgt zufällig.

- Die "störende" Wirkung von unbekannten und im Versuch nicht berücksichtigten Faktoren soll durch Zufallszuteilung miminisiert werden und somit Treffgenauigkeit des Experiments zu erhöhen.
- Das Randomisieren kann eine verbesserte Normalität der Daten bewirken und Unabhängigkeit erreichen — die wichtigen Voraussätzungen für die vorgestellten statistischen Verfahren.
- Das Randomisieren wird mit Hilfe von Zufallsgeneratoren (im einfachsten Fall – Tabellen) durgeführt.

Blockbildung

In vielen Experimenten gibt es bekannte “Störfaktoren”, die uns im vorgesehenen Versuch nicht interessieren.

Um die Enfüsse solcher Störfaktoren zu reduzieren, kann man die Versuchseinheiten in Gruppen (Blöcke) einteilen, wobei die Störfaktoren innerhalb dieser Blöcke möglichst homogen sind.

Dadurch werden die Störfaktoren “ausgeschaltet” und die Zufallsstreuung reduziert.

Beispiele

- Tiere aus einem Wurf

Blockbildung (1)

Die Sorten A, B, C und D werden auf Ertragsunterschiede untersucht. Falls es in West-Ost-Richtung große Bodenunterschiede gibt, würde der Sorteneffekt vom störenden Bodeneffekt überlagert. Randomisierte Anordnung kann wie folgt aussehen:

Sorte	Sorte	Sorte	Sorte	Sorte	Sorte	Sorte	Sorte	Sorte	Sorte	Sorte	Sorte
D	D	B	A	C	B	C	B	A	D	A	C
1. Wdh	2. Wdh	1. Wdh	1. Wdh	1. Wdh	2. Wdh	2. Wdh	3. Wdh	2. Wdh	3. Wdh	3. Wdh	3. Wdh

Westen $\leftarrow$ $\rightarrow$ Osten

Figure 3: [Köhler et al.]

Diese Anordnung hilft den Bodeneffekt auszuschalten, allerding kann die Reststreuung in der einfaktoriellen ANOVA sich vergrössen.

Blockbildung (2)

Der Einfluss vom störenden Bodenfaktor lässt sich reduzieren bei gleichzeitiger Verbesserung der Versuchsgenaugkeit, indem man das Versuchsfeld in 3 homogenere Blöcke einteilt wobei in jedem Block alle Sorten zifällig verteilt sind.

Block I				**Block II**				**Block III**			
Sorte	Sorte	Sorte	Sorte	Sorte	Sorte	Sorte	Sorte	Sorte	Sorte	Sorte	Sorte
A	*B*	*C*	*D*	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*
1. Wdh	1. Wdh	1. Wdh	1. Wdh	2. Wdh	2. Wdh	2. Wdh	2. Wdh	3. Wdh	3. Wdh	3. Wdh	3. Wdh

Westen ← → Osten

Figure 4: [Köhler et al.]

Bei der Auswertung der Ergebnisse muss man den Faktor "Blöcke" oder in dem Fall "Boden Unterschiede in Ost-West-Richtung" berücksichtigen.

Sind die *Unterschiede zwischen den Blöcken gering*, so ist unter Umständen durch die Blockbildung keine Erhöhung der Empfindlichkeit zu erreichen.

Es dürfen keine Wechselwirkungen zwischen Faktoren und Blöcken auftreten.

Faktorielle Experimente

Im Unterschied zum einfaktoriellen Versuch (z.B. Wirkung vom Sortenfaktor auf Ertrag) wird beim mehrfaktoriellen Versuch gleichzeitig die Wirkung von mehreren Faktoren untersucht. Die Vorteile dieser Vergehensweise:

- Bei einfaktoriellen Versuchen hält man die nicht untersuchten auf wirkürlich konstantem Stufenniveau. Mehrfaktorielle Versuche variieren alle interessierenden Faktoren.
- Im Vergleich zu einfaktoriellen Experimenten wird im mehrfaktoriellen Experimenten beim gleichem Aufwand eine größere Präzision erreicht.
- Die mehrfaktoriellen Versuche ermöglichen die Bestimmung der Wechselwirkungen (z.B. Nebenwirkungen von Medikamenten beim Alkoholkomsum).

Symmetrischer Aufbau

- Mehrfache ANOVA erfordet gleiche Anzahl an Wiederholungen (Balanciertheit). Auch einfache ANOVA führt zu fehlerhaften Entscheidungen im F- und t-Test falls die Daten nicht normalverteilt oder heteroskedastisch sind.
- Bei der Regression ist die gleiche Schrittweite (Äquidistanz) empfohlen was sowohl Vereinfachung als auch Informationsgewinn bringt.

Wirtschaftlichkeit

Es wird die Stichprobe und nicht die Grundgesamtheit untersucht, unter anderem aus Kostengründen.

So muss die Frage nach einem geigneten Stichprobenumfang geklärt werden.

Blöcke mit zufälliger Anordnung

Blockbildung und Randomisieren stellen zwei gegensätzliche Grundsätze der Versuchsplanung dar.

Deswegen buldet man Blöcke und randomisiert innerhalb der Blöcke.

C	D	B	E	A	E	A	D	C	B	E	C	A	B	D	A	D	E	B	C
Block I					**Block II**					**Block III**					**Block IV**				

Figure 5: Jede der fünf Sorten, A, B, C, D und E soll auf vier Parzellen (n=4 Wiederholungen) angebaut werden. Das Versuchsfeld wird in 4 Blöcke von jeweils 5 benachbarten Parzellen aufgeteilt. [Köhler et al.]

ANOVA:

- Unterschiede zwischen den Sorten mit 4 Freiheitsgraden
- Unterschiede zwischen den Blöcken mit 3 Freiheitsgraden
- Reststreung mit 12 Freiheitsgraden

Lateinische Quadrate

Beispiel. (nach A.Linder) Zu prüfen sei die Leistung von drei Sägen A, B, und C, die verschiedene Zahnformen haben. Je zwei Sägen jeder Art sollen überprüft werden, also A1; A2; B1; B2 und C1; C2. Man bildet 6 Gruppen mit je zwei Arbeitern. An 6 verschiedenen Holzarten soll die Schnittzeit jeweils gemessen werden.

		Arbeitergruppen					
		1	2	3	4	5	6
Holzarten	1	C_1	B_2	A_2	B_1	A_1	C_2
	2	B_1	C_2	B_2	A_1	A_2	C_1
	3	B_2	A_1	B_1	C_2	C_1	A_2
	4	C_2	A_2	A_1	C_1	B_2	B_1
	5	A_2	B_1	C_1	B_2	C_2	A_1
	6	A_1	C_1	C_2	A_2	B_1	B_2

Figure 6: Störfaktoren Arbeitsgruppenunterschiede und Holzunterschiede werden reduziert. [Köhler et al.]

Eigenschaft des Lateinischen Quadrates: *jede Zeile und jede Spalte enthält alle 6 verschiedene Symbole genau einmal*.

Lateinische Quadrate (1)

Zu jedem $n > 1$ gibt es jeweils mehrere $n \times n$ Lateinische Quadrate, die man zufällig auswählen soll. Bzw. man kann ein neues Lateinisches Quadrat bilden, indem man in einem bestehenden LQ die Zeilen und Spalten vertauscht.

ANOVA:

- Zwischen den Sägen mit 5 FG
- Zwischen den Arbeitergruppen mit 5 FG
- Zwischen den Holzarten mit 5 FG
- Reststreuung mit 20 FG

Mehrfaktorielle Versuche

Beispiel In einem Düngungsversuch mit Gerste sollen die drei Faktoren K (Kaliumsulfat), N (Ammoniumsulfat), P (Superphosphat) in je zwei Stufen (Beigabe, keine Beigabe) untersucht werden. Wir bezeichnen keine Beigabe mit 0, Beigabe von K und N mit KN, nur Beigabe von K mit K, Beigabe von K, N und P mit KNP.

Um diese 8 Beigaben in einem Feldversuch mit einem Lateinischen Quadrat (Ausschaltung von zwei Störfaktoren) zu untersuchen, wären 64 Parzellen (8×8) notwendig. Wir wählen hier stattdessen eine weniger aufwändige Blockanlage (Ausschaltung von einem Störfaktor) und brauchen nur 32 Parzellen.

Block I								**Block II**							
N	KN	NP	0	KNP	K	KP	P	0	KNP	KP	K	NP	N	P	KN
KN	0	N	KNP	P	NP	KP	K	0	K	NP	KNP	KN	P	KP	N
Block III								**Block IV**							

Figure 7: . [Köhler et al.]

Mehrfaktorielle Versuche (1)

ANOVA:

- Zwischen den jeweils K,N,P-Beigaben mit 1 FG
- Zwischen den Blöcken mit 3 FG

Wechselwirkungen: - Zwischen den K- und N-Beigaben mit 1 FG - Zwischen den K- und P-Beigaben mit 1 FG - Zwischen den N- und P-Beigaben mit 1 FG - Zwischen den K-, N- und P-Beigaben mit 1 FG - Reststreuung mit 21 FG

Kontraste für Erhöhungen oder Vermiderungen der verschiedenen Beigabekombinationen:

$L_N = (N + KN + NP + KNP) - (0 + K + P + KP)$, analog L_K, L_P, L_{KP}, L_{NP}, L_{KN} und L_{KNP}.

Bei mehr als zwei Stufen pro Faktor können multiple Vergleiche durchgeführt werden z.B. auch unter der Anwendung der Kontraste.

Allgemein gibt es mit m Faktoren und n Stufen n^m mögliche Kombinationen. Man kann durch Weglassen bestimmter Kombinationen den Aufwand reduzieren.

(Weitere) Non-parametric Tests

Non-parametric Tests

- Wilcoxson rank-sum-Test der Mann-Whitney-Test
- Wilcoxon signed-rank-Test
- Friedman-Test
- Kruskal-Wallis-Test

Kruskal-Wallis-Test

Falls die Bedingungen der einfachen ANOVA nicht erfüllt sind, kann man den Kruskal-Wallis-Test benutzen um die Unterschiede zwischen unabhängigen Gruppen zu untersuchen.

Ähnlich zum Mann-Witney-Test (und Wilcoxon-Test) wird die folgende Rangstatistik betrachtet (wird mit χ^2-Statistik verglichen)

$$H = \frac{12}{N(N+1)} \sum_{i}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

R_i ist die Rangsumme für jede Gruppe

N ist der Stichprobenumfang

n_i ist die Größe jeder der k Gruppen

$k-1$ sind Freiheitsgrade

```r
soyaData <- read.delim("~/Lehre/WS1920/Beuth/Data files/Soya.dat", header = TRUE)
soyaData$Soya <- factor(soyaData$Soya, levels = c("No Soya Meals", "1 Soya Meal Per Week",
    "4 Soyal Meals Per Week", "7 Soya Meals Per Week"))
knitr::kable(soyaData[1:15, ], caption = "Auszug aus einem ausgedachten Datensatz über Auswirkungen von Soya auf Spermienanzahl
```

Table 1: Auszug aus einem ausgedachten Datensatz über Auswirkungen von Soya auf Spermienanzahl [A. Field]

Soya	Sperm
No Soya Meals	0.35
No Soya Meals	0.58
No Soya Meals	0.88
No Soya Meals	0.92
No Soya Meals	1.22
No Soya Meals	1.51
No Soya Meals	1.52
No Soya Meals	1.57
No Soya Meals	2.43
No Soya Meals	2.79
No Soya Meals	3.40
No Soya Meals	4.52
No Soya Meals	4.72
No Soya Meals	6.90
No Soya Meals	7.58

```r
# soyaData$Soya<-factor(soyaData$Soya, levels = levels(soyaData$Soya)[c(4, 1, 2,
# 3)])
```

Kruskal-Wallis-Test (1)

```
library(car)
library(pastecs)

by(soyaData$Sperm, soyaData$Soya, stat.desc, basic = FALSE)
```

```
## soyaData$Soya: No Soya Meals
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##     3.095000     4.987000     1.136926     2.379614    25.852022     5.084488
##     coef.var
##     1.019549
## ------------------------------------------------------------
## soyaData$Soya: 1 Soya Meal Per Week
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##     2.595000     4.606000     1.044822     2.186837    21.833057     4.672586
##     coef.var
##     1.014456
## ------------------------------------------------------------
## soyaData$Soya: 4 Soyal Meals Per Week
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##    2.9450000    4.1105000    0.9861233    2.0639798   19.4487839    4.4100775
##     coef.var
##    1.0728810
## ------------------------------------------------------------
## soyaData$Soya: 7 Soya Meals Per Week
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
##    1.3350000    1.6535000    0.2479774    0.5190226    1.2298555    1.1089885
##     coef.var
##    0.6706916
```

Kruskal-Wallis-Test (2)

```
by(soyaData$Sperm, soyaData$Soya, stat.desc, desc = FALSE, basic = FALSE, norm = TRUE)
```

```
## soyaData$Soya: No Soya Meals
##    skewness    skew.2SE    kurtosis    kurt.2SE  normtest.W  normtest.p
## 1.546140856 1.509598499 2.328051363 1.172959394 0.805255802 0.001035917
## ------------------------------------------------------------
## soyaData$Soya: 1 Soya Meal Per Week
##    skewness    skew.2SE    kurtosis    kurt.2SE  normtest.W  normtest.p
## 1.350565932 1.318645901 1.422731699 0.716825470 0.825831600 0.002153894
## ------------------------------------------------------------
## soyaData$Soya: 4 Soyal Meals Per Week
##     skewness     skew.2SE     kurtosis     kurt.2SE   normtest.W   normtest.p
## 1.8222369167 1.7791691502 2.7926151830 1.4070240317 0.7427432543 0.0001359072
## ------------------------------------------------------------
## soyaData$Soya: 7 Soya Meals Per Week
##   skewness   skew.2SE   kurtosis   kurt.2SE normtest.W normtest.p
##  0.6086712  0.5942855 -0.9161653 -0.4615984  0.9122606  0.0703908
```

Kruskal-Wallis-Test (3)

```
leveneTest(soyaData$Sperm, soyaData$Soya)
```

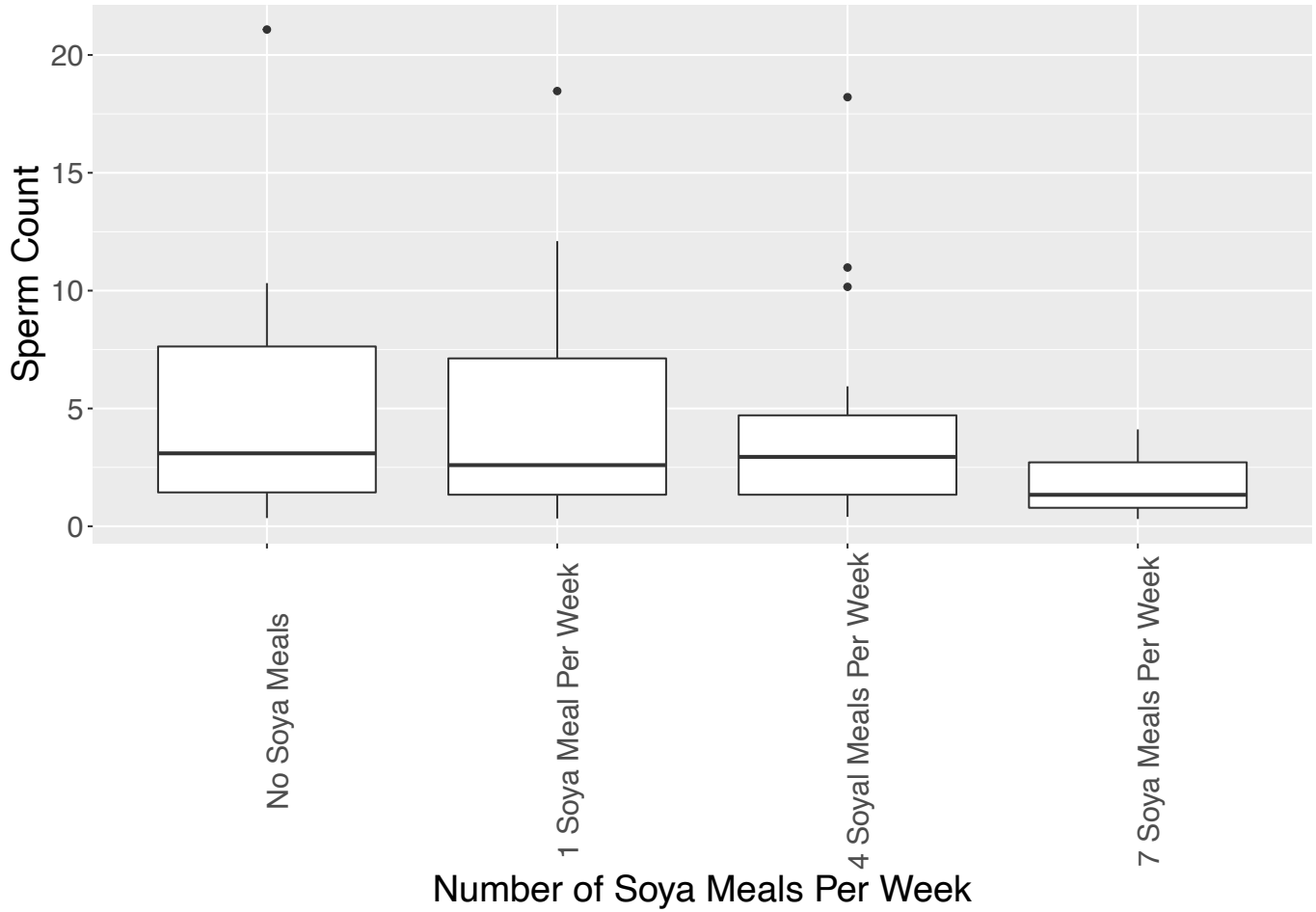
```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value  Pr(>F)
## group  3  2.8606 0.04237 *
##       76
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
kruskal.test(Sperm ~ Soya, data = soyaData)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  Sperm by Soya
## Kruskal-Wallis chi-squared = 8.6589, df = 3, p-value = 0.03419
```

Kruskal-Wallis-Test (4)

```
library(ggplot2)
ggplot(soyaData, aes(Soya, Sperm)) + geom_boxplot() + labs(y = "Sperm Count", x = "Number of Soya Meals Per Week") +
    theme(text = element_text(size = 20)) + theme(axis.text.x = element_text(angle = 90))
```



Post hoc-Tests für den Kruskal-Wallis-Test I

Paarwise Vergleiche:

```
library(pgirmess)
kruskalmc(Sperm ~ Soya, data = soyaData)
```

```
## Multiple comparison test after Kruskal-Wallis
## p.value: 0.05
## Comparisons
##                                                obs.dif critical.dif difference
## No Soya Meals-1 Soya Meal Per Week                 2.2     19.38715      FALSE
## No Soya Meals-4 Soyal Meals Per Week               2.2     19.38715      FALSE
## No Soya Meals-7 Soya Meals Per Week               19.0     19.38715      FALSE
## 1 Soya Meal Per Week-4 Soyal Meals Per Week        0.0     19.38715      FALSE
## 1 Soya Meal Per Week-7 Soya Meals Per Week        16.8     19.38715      FALSE
## 4 Soyal Meals Per Week-7 Soya Meals Per Week      16.8     19.38715      FALSE
```

Die post-hoc-Tests sind sehr konservation wenn wir alle Gruppen untereinander vergleichen. Deswegen werden weniger Vergleiche angestrebt, z.B. mit dem niedrigsten Level (No Soya Meals)

Post hoc-Tests für den Kruskal-Wallis-Test (1) I

Vergleiche *treatment* vs *control* (erster Level)

```
soyaData$Soya <- factor(soyaData$Soya, levels = c("No Soya Meals", "1 Soya Meal Per Week",
    "4 Soyal Meals Per Week", "7 Soya Meals Per Week"))
kruskalmc(Sperm ~ Soya, data = soyaData, cont = "two-tailed")
```

```
## Multiple comparison test after Kruskal-Wallis, treatments vs control (two-tailed)
## p.value: 0.05
## Comparisons
##                                       obs.dif critical.dif difference
## No Soya Meals-1 Soya Meal Per Week        2.2     17.59209      FALSE
## No Soya Meals-4 Soyal Meals Per Week      2.2     17.59209      FALSE
## No Soya Meals-7 Soya Meals Per Week       19.0     17.59209       TRUE
```

Friedman-Test

Falls die Bedingungen der ANOVA mit Wiederholungen nicht erfüllt sind, kann man den Friedman-Test benutzen.

Test-Statistik (wird mit χ^2-Statistik verglichen):

$$F_r = \frac{12}{Nk(k+1)} \sum_i^k \frac{R_i^2}{n_i} - 3N(k+1)$$

R_i ist die Rangsumme für jede Gruppe (aber zuerst werden die Werte für jede Person oder jedes Objekt rangiert)

N ist der Stichprobenumfang

n_i ist die Größe jeder der k Gruppen

$k-1$ sind Freiheitsgrade

Friedman-Test (1)

```
dietData <- read.delim("~/Lehre/WS1920/Beuth/Data files/Diet.dat", header = TRUE)
stat.desc(dietData)
```

```
##                    Start      Month1     Month2
## nbr.val       10.0000000  10.0000000  10.000000
## nbr.null       0.0000000   0.0000000   0.000000
## nbr.na         0.0000000   0.0000000   0.000000
## min           62.0110884  65.3836941  55.431306
## max          120.4644487 119.9637568 114.258942
## range         58.4533603  54.5800626  58.827636
## sum          780.5434341 774.6350144 770.668400
## median        69.2288216  68.5727731  72.249343
## mean          78.0543434  77.4635014  77.066840
## SE.mean        6.3973145   5.8865852   5.093204
## CI.mean.0.95  14.4717308  13.3163808  11.521627
## var          409.2563282 346.5188507 259.407224
## std.dev       20.2300847  18.6150168  16.106124
## coef.var       0.2591795   0.2403069   0.208989
```

Friedman-Test (2)

verlangt *matrix* nicht *DataFrame*

```
dietData <- na.omit(dietData)  # falls es fehlende Werte gibt
friedman.test(as.matrix(dietData))
```

```
##
##  Friedman rank sum test
##
## data:  as.matrix(dietData)
## Friedman chi-squared = 0.2, df = 2, p-value = 0.9048
```

Das Ergebnis ist nicht signifikant (p-value > 0.05). Es gibt keinen Effekt.

Post hoc-Tests für Friedman-Test

```
friedmanmc(as.matrix(dietData))
```

```
## Multiple comparisons between groups after Friedman test
## p.value: 0.05
## Comparisons
##     obs.dif critical.dif difference
## 1-2       1      10.7062      FALSE
## 1-3       2      10.7062      FALSE
## 2-3       1      10.7062      FALSE
```