Statistik
für
Informatik im Wintersemester 2016/2017

Was?	Wer?	Wann?	Wo?	ab wann?
Vorlesung	Anett Wins	Di., 14:00 - 17:00 Uhr	W3.02	Di., 04. Oktober 2016
Übung 1 Statistik	Eugen Ivanov	Do., 14:00 - 15:30 Uhr	J3.19	Do., 13. Oktober 2016
Übung 2 Statistik	Eugen Ivanov	Do., 15:30 - 17:00 Uhr	J3.19	Do., 13. Oktober 2016

Vorlesung Statistik WS 2016/17 Sessionlist

Session	Datum	Thema
1	Di., 04. Oktober 2016	Einführng, R-Installation, R-Studio, Skalen
2	Di., 11. Oktober 2016	univariate deskriptive Statistik, Plots
3	Di., 18. Oktober 2016	Streuung, Konzentrationsmaße
4	Di., 25. Oktober 2016	Kontingenztabellen, Mosaikplots, Korrelation
	Di., 01. November 2016	Feiertag
5	Di., 08. November 2016	Preisindizes, lineare Regression
6	Di., 15. November 2016	Kominatorik, Wahrscheinlichkeiten
7	Di., 22. November 2016	Wahrscheinlichkeit, diskete Zufallsvariablen
8	Di., 29. November 2016	Binomial-, Hypergeom.-, Poisson-Verteilung
9	Di., 06. Dezember 2016	stetige Zufallsvariablen, Gleichverteilung
10	Di., 13. Dezember 2016	Normalverteilung, Verteilungsparameter
11	Di., 20. Dezember 2016	Schätzfunktionen, Punktschätzer
	Di., 27. Dezember 2016	Weihnachtsferien
	Di., 03. Januar 2017	Weihnachtsferien
12	Di., 10. Januar 2017	Konfidenzintervalle
13	Di., 17. Januar 2017	Tests

Übungen

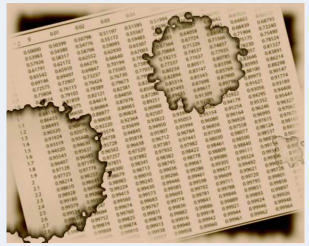
Übung	Datum
1	Do., 13. Oktober 2016
2	Do., 20. Oktober 2016
3	Do., 27. Oktober 2016
4	Do., 03. November 2016
5	Do., 10. November 2016
6	Do., 17. November 2016
7	Do., 24. November 2016
8	Do., 01. Dezember 2016
9	Do., 08. Dezember 2016
10	Do., 15. Dezember 2016
11	Do., 22. Dezember 2016
	Do., 29. Dezember 2016
	Do., 05. Januar 2017
12	Do., 12. Januar 2017
13	Do., 19. Januar 2017

Dipl. Stat. Anett Wins
nach dem Skript von

Prof. Dr. Stefan Etschberger
Hochschule Augsburg

Outline

1. Statistik: Einführung
    - Berühmte Leute zur Statistik
    - Wie lügt man mit Statistik?
    - Gute und schlechte Grafiken
    - Begriff Statistik
    - Grundbegriffe der Datenerhebung
    - R und RStudio
2. Deskriptive Statistik
    - Häufigkeiten
    - Lage und Streuung
    - Konzentration
    - Zwei Merkmale
    - Korrelation
    - Preisindizes
    - Lineare Regression
3. Wahrscheinlichkeitstheorie
    - Kombinatorik
    - Zufall und Wahrscheinlichkeit
    - Zufallsvariablen und Verteilungen
    - Verteilungsparameter
4. Induktive Statistik
    - Grundlagen
    - Punkt-Schätzung
    - Intervall-Schätzung
    - Signifikanztests



Material zur Vorlesung

Kursmaterial:

- Aufgabensatz (beinhaltet Aufgaben zu R)
- Folienskript
- Anmerkungen zur Vorlesung (im Anschluss an die Veranstaltung)
- Beispieldaten
- Alle Auswertungen als -Datei

Literatur:

Bamberg, Günter, Franz Baur und Michael Krapp (2011). **Statistik**. 16. Aufl. München: Oldenbourg Verlag. ISBN: 3486702580.

Dalgaard, Peter (2002). **Introductory Statistics with R**. New York: Springer.

Fahrmeir, Ludwig, Rita Künstler, Iris Pigeot und Gerhard Tutz (2009). **Statistik: Der Weg zur Datenanalyse**. 7. Aufl. Berlin, Heidelberg: Springer. ISBN: 3642019382.

Klausur:

- Klausur am Ende des Semesters
- Bearbeitungszeit: 90 Minuten
- Erreichbare Punktzahl: 90
- R ist prüfungsrelevant: Siehe Anmerkungen in Übungsaufgaben!
- Hilfsmittel:
  - Schreibzeug,
  - Taschenrechner, der nicht 70! berechnen kann,
  - ein Blatt (DIN-A4, vorne und hinten beschrieben) mit handgeschriebenen Notizen (keine Kopien oder Ausdrucke),
- "gemeinsame" Klausur mit der Wiederholungsklausur Statistik für Betriebswirtschaft, Internationales Management, Wirtschaftsinformatik und Informatik vom Sommersemester 2016 (Prof. Dr. Stefan Etschberger)

Datenbasis

Fragebogen

Umfrage Statistik

Bitte beantworten Sie folgende Fragen vollständig und füllen Sie jeweils *nur eine* beliebige Spalte *leserlich* aus.

Ihr Alter (in Jahren)										
Ihre Größe (in cm)										
Ihr Geschlecht (m/w)										
Wie alt ist (bzw. wäre) Ihr Vater heute?										
Wie alt ist (bzw. wäre) Ihre Mutter heute?										
Größe Ihres Vaters (cm)										
Größe Ihrer Mutter (cm)										
Wie viele Geschwister haben Sie?										
Wunschfarbe für Ihr nächstes Smartphone; mögliche Auswahl: (si)lber, (sc)hwarz, (w)eiß, (g)elb, (b)lau, (r)ot										
Ausgaben für Ihre mobile Kommunikation (egal wer bezahlt hat) in den vergangenen 12 Monaten inkl. Hardware (Smartphone, Zubehör), Software (Apps), Vertrag										
Wie viele Paar Schuhe besitzen Sie?										
Ausgaben für Ihre Schuhe (egal wer bezahlt hat) in den letzten 12 Monaten										
Ihre Essgewohnheiten; mögliche Auswahl: (f)ruktarisch, (vegan), (vege)tarisch, (p)escetarisch – also vegetarisch und Fische, (c)arnivor, also auch Fleisch										
Sind Sie Raucher? mögliche Auswahl: (ja), (nein)										
Ihre Note in der Matheklausur (Note oder (n)icht mitgeschrieben)										
Waren Sie mit Ihrer Leistung in der Matheklausur zufrieden? Antworten: (s)ehr zufrieden, (z)ufrieden, (g)eht so, (u)nzufrieden, (n)icht mitgeschrieben										

Statistik: Table of Contents

1 Statistik: Einführung

2 Deskriptive Statistik

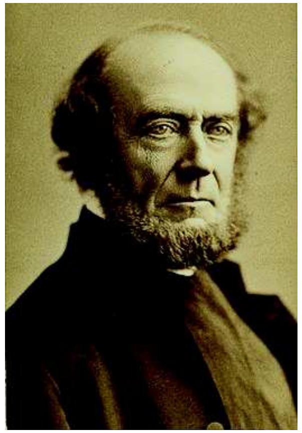
3 Wahrscheinlichkeitstheorie

4 Induktive Statistik

1 Statistik: Einführung
- Berühmte Leute zur Statistik
- Wie lügt man mit Statistik?
- Gute und schlechte Grafiken
- Begriff Statistik
- Grundbegriffe der Datenerhebung
- R und RStudio

- **Leonard Henry Courteney (1832-1918)**:

  „There are three kinds of lies: lies, damned lies and statistics.“



- **Winston Curchill (1874-1965) angeblich**:

  „Ich glaube nur den Statistiken, die ich selbst gefälscht habe.“

- **Andrew Lang (1844-1912)**:

  „Wir benutzen die Statistik wie ein Betrunkener einen Laternenpfahl: Vor allem zur Stütze unseres Standpunktes und weniger zum Beleuchten eines Sachverhalts.“

DRAWN BY BURNE MURDOCK. ENGRAVED BY J. F. JUNGLING.

Quellen: Wikimedia Commons

1. Einführung
Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung
R und RStudio
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Morgens in Zeitung: Mehr Statistiken als Goethe und Schiller im ganzen Leben gesehen haben:

- Arbeitslosenzahlen wachsen
- Vogelgrippe breitet sich aus
- 78,643% der Deutschen unzufrieden mit Löw
- Bundesbürger verzehrt 5,8 Liter Speiseeis pro Jahr
- Musiker leben länger als andere Leute
- Tennispieler B hat noch nie gegen einen brilletragenden Linkshänder verloren, der jünger ist als er
- in New York schläft man am sichersten im Central Park

Viele dieser Statistiken: Falsch, bewußt manipuliert oder unpassend ausgesucht.

Fehlerquellen:

- Zahlenmanipulation
- irreführende Darstellung der Zahlen
- ungenügendes Wissen

1. Einführung
Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung
R und RStudio
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Fehlerhafter Sprachgebrauch: „Richtig“ Fragen

1. Frage:

„Finden Sie, dass in einem Betrieb alle Arbeiter in der Gewerkschaft sein sollten?“

Resultat:

- Dafür: 44%
- Dagegen: 20%
- Unentschieden: 36%

2. Frage:

„Finden Sie, dass in einem Betrieb alle Arbeiter in der Gewerkschaft sein sollten oder muss man es jedem einzelnen überlassen, ob er in der Gewerkschaft sein will oder nicht?“

Resultat:

- Dafür: 24%
- Dagegen: 70%
- Unentschieden: 6%

Geänderte Definitionen

Laut einem „Bericht zur Bekämpfung des Analphabetismus in Deutschland“:

- Heute gibt es in Deutschland ca. 7 Millionen Analphabeten
- Zu Kaiser Wilhelms Zeiten gab es weniger als 10 000

Was leiten Sie daraus ab?

BILDUNG

7,5 Millionen Deutsche sind Analphabeten

Ein Siebtel der erwerbsfähigen Bevölkerung kann laut einer Studie kaum lesen und schreiben – doppelt so viel wie bisher gedacht. Bildungsministerin Schavan will reagieren [weiter...]

ANALPHABETISMUS

Ein Land verlernt das Lesen

Studenten verstehen abstrakte Texte nicht mehr, ein Schulbuchverlag kürzt Klassiker, Banker besuchen Lesekurse: Viele Deutsche haben keine Lust mehr zu lesen. [weiter...]

ANALPHABETISMUS

Buchstäblich resigniert

Mehr als sieben Millionen Deutsche können kaum lesen und schreiben. Erst jetzt hat die Politik das Problem erkannt. Aber es gibt zu wenig Geld für Kurse. Von M. Spiewak [weiter...]

Quelle: Zeit.de

Definition

Zu Kaiser Wilhelms Zeiten:

„Analphabet ist, wer seinen Namen nicht schreiben kann.“

Definition heute:

„Ein Analphabet ist eine Person, die sich nicht beteiligen kann an all den zielgerichteten Aktivitäten ihrer Gruppe und ihrer Gemeinschaft, bei denen Lesen, Schreiben und Rechnen erforderlich ist und an der weiteren Nutzung dieser Kulturtechniken für ihre weitere Entwicklung und die der Gesellschaft“.

1. Einführung
Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung
R und RStudio
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Aussage des Vertriebsleiters:

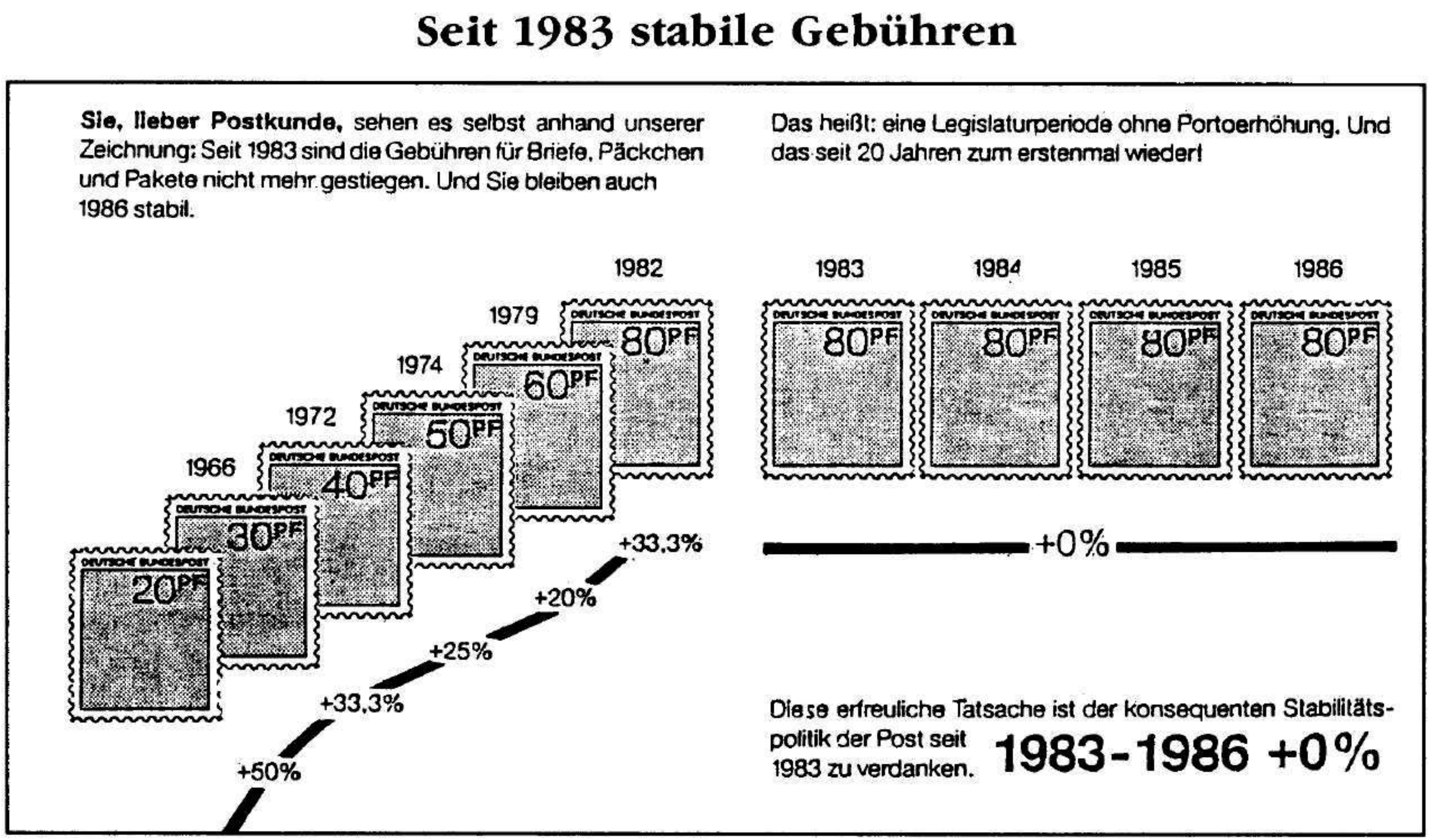
„Unser Umsatz stieg vor einem Jahr um 1%. Dieses Jahr stieg das Umsatzwachstum um 50%!“

Im Klartext:

- Basisjahr: Umsatz 100
- Dann: Wachstum auf 101
- Dieses Jahr: Wachstum des Wachstums um 50% bedeutet 1,5% Wachstum. Also Umsatz dann 102,5049

1. Einführung
Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung
R und RStudio
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Verzerrte Zeitachse

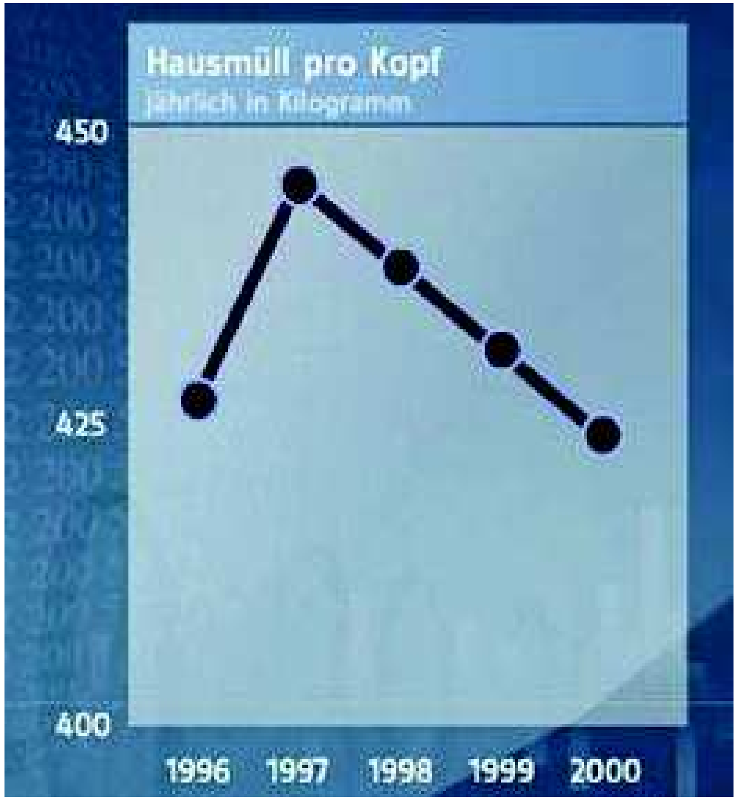


Quelle Kramer (2011)

Statistik

1. Einführung
Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung
R und RStudio
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Grafik aussagekräftig?



Quelle: Bach u. a. (2006)

1. Einführung
Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung
R und RStudio
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

1. Einführung
- Berühmte Leute zur Statistik
- Wie lügt man mit Statistik?
- Gute und schlechte Grafiken
- Begriff Statistik
- Grundbegriffe der Datenerhebung
- R und RStudio

2. Deskriptive Statistik

3. W-Theorie

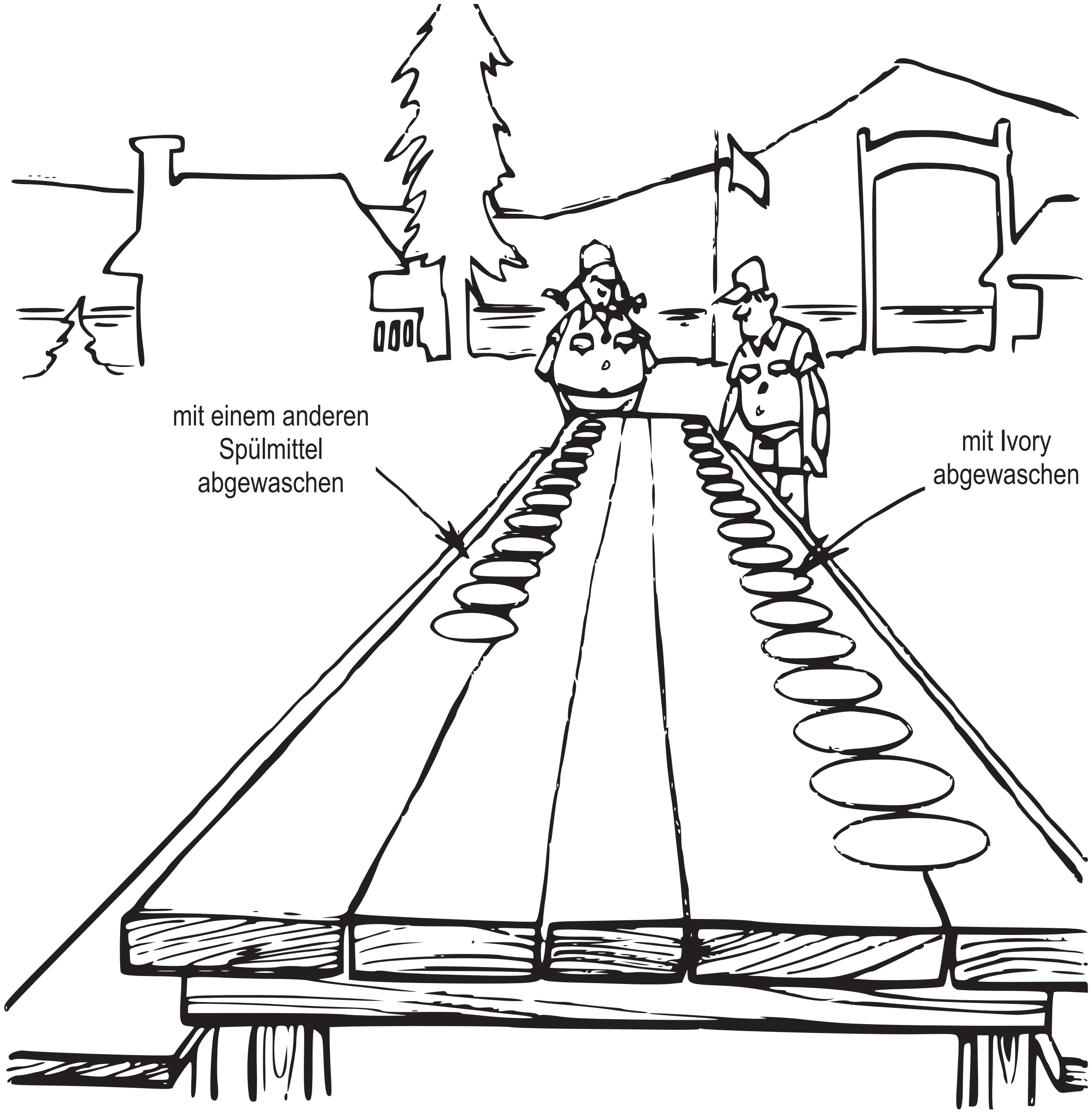
4. Induktive Statistik

Quellen

Tabellen

Größenproportionen

11 zu 15



Falsche Berechnungsbasis

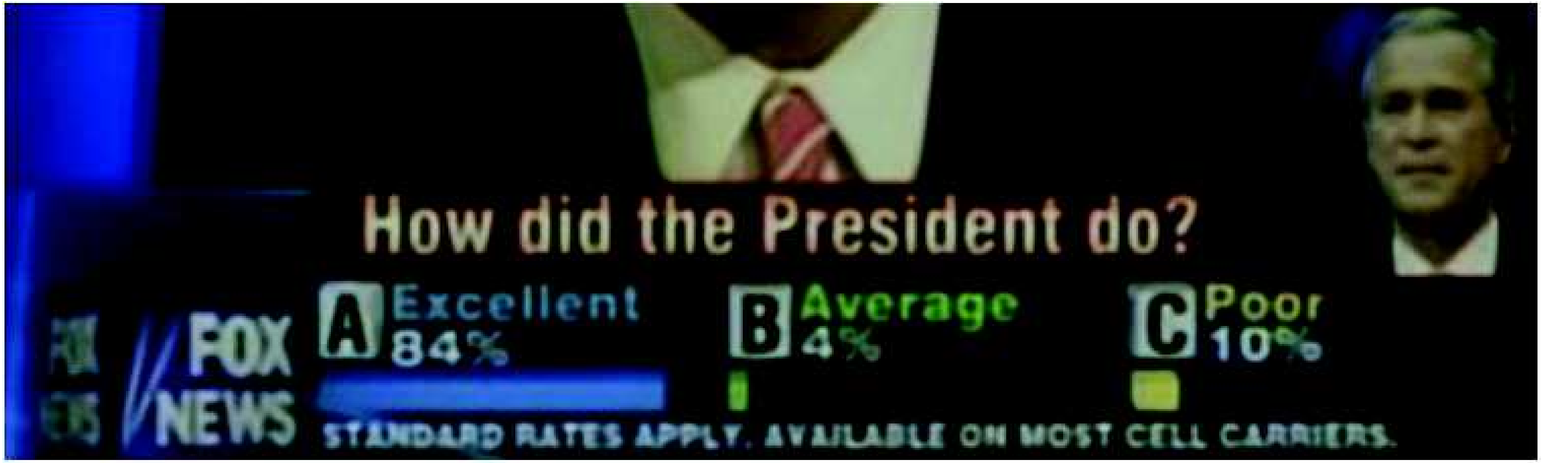
- Ein Einzelhändler bezieht ein Produkt zu 100 € und verkauft es für 200 €. Hat er eine Gewinnspanne von 50% oder 100%?
- Bahn: 9 Tote pro 10 Mio Passagieren je Kilometer
  Flugzeug: 3 Tote pro 10 Mio Passagieren je Kilometer
  Bahn: 7 pro 10 Mio Passagiere je Stunde
  Flugzeug: 24 pro 10 Mio Passagiere je Stunde
- Nur 40 % aller durch Autounfälle Gestorbenen hatten keinen Sicherheitsgut angelegt
  Also: Keinen Gurt anlegen ist sicherer
- Die Hälfte der Todesfälle ereignen sich in Krankenhäusern
  Also: Krankenhäuser sind lebenssgefährlich
- Zwei Drittel aller alkoholabhängigen Personen sind verheiratet
  Also: die Ehe führt zum Alkohol

1. Einführung
Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung
R und RStudio
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Nicht repräsentative Stichproben

1. Einführung
Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung
R und RStudio
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Fernsehumfragen



- Kostenpflichtige Telefonabstimmung nach regierungsfreundlichem Bericht im Fernsehen
- In den meisten Umfragen erreichte Bush zu diesem Zeitpunkt nur 30 % Zustimmung

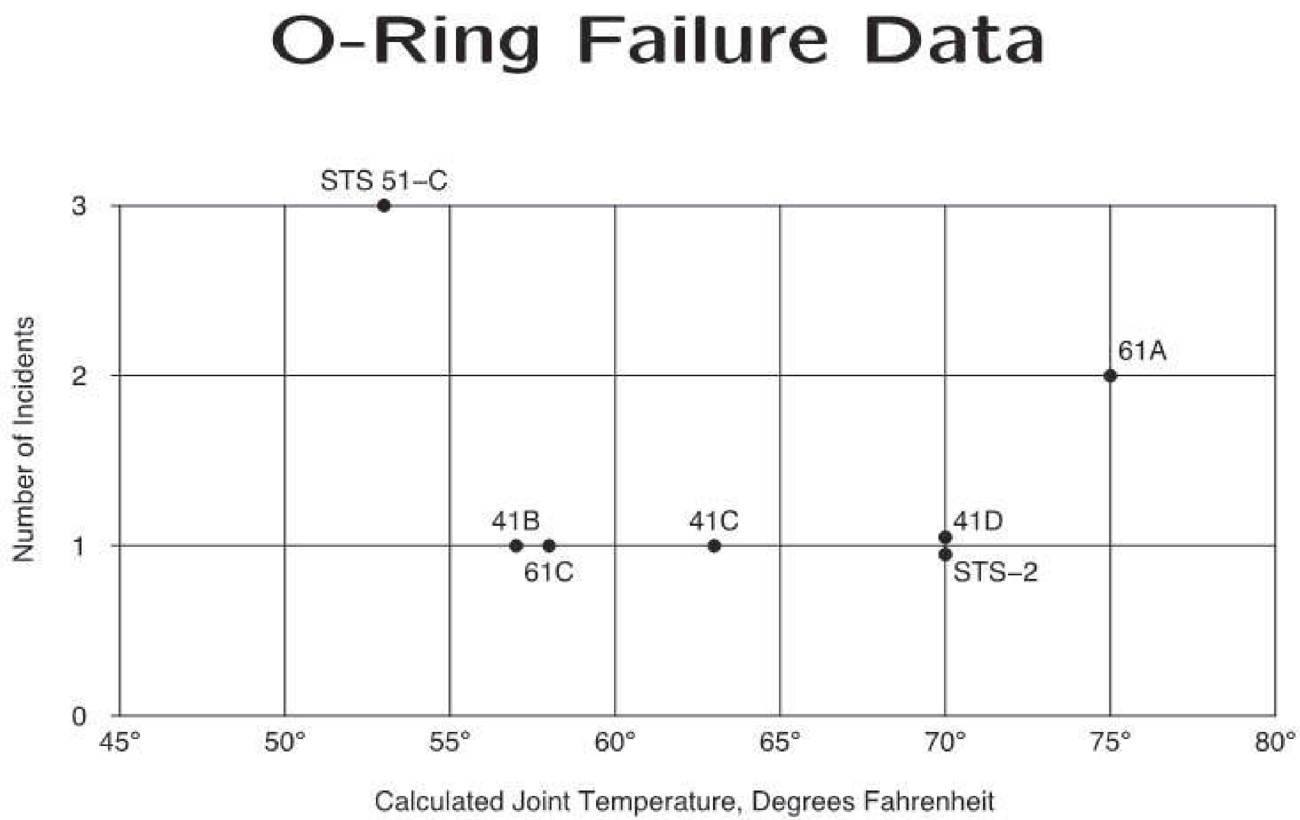
1. Einführung
Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung
R und RStudio
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Challenger-Katastrophe

Am 28. Januar 1986, 73 Sekunden nach dem Start der Mission STS-51-L, brach die Raumfähre in etwa 15 Kilometer Höhe auseinander. Dabei starben alle sieben Astronauten. Es war der bis dahin schwerste Unfall in der Raumfahrtgeschichte der USA.

Challenger Unfall 1986

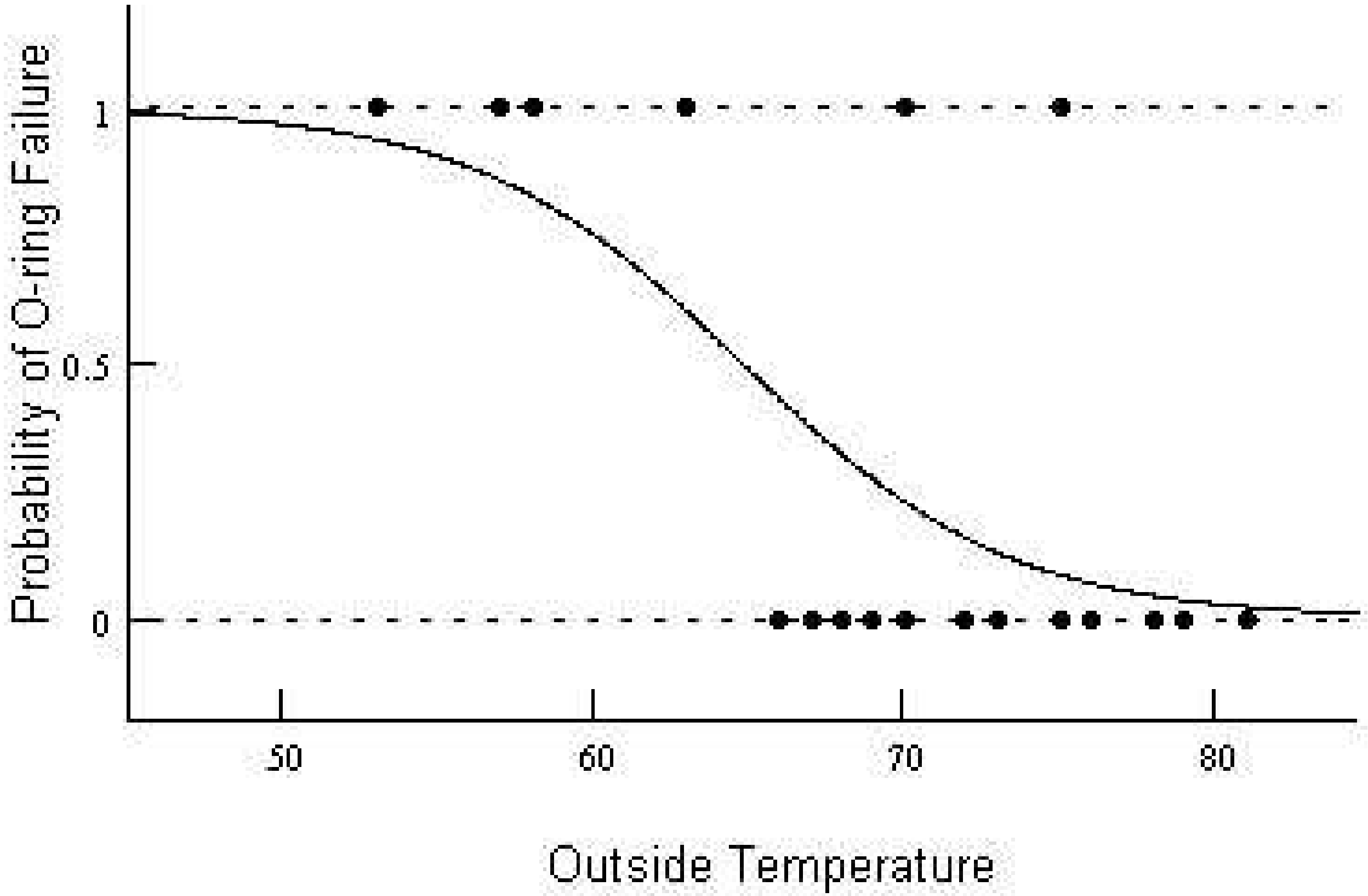
- Grund für Explosion: 2 Gummidichtungsringe waren undicht
- Die Temperatur der Dichtungsringe: Unter 20° F (ca. -6,7° C).
- Probleme mit Dichtungsringen bei Start der vorigen Fähre: Umgebungstemperatur 53° F (ca. 11,7° C).
- Frage: Ist der Dichtungsfehler durch die Umgebungstemperatur zu prognostizieren?



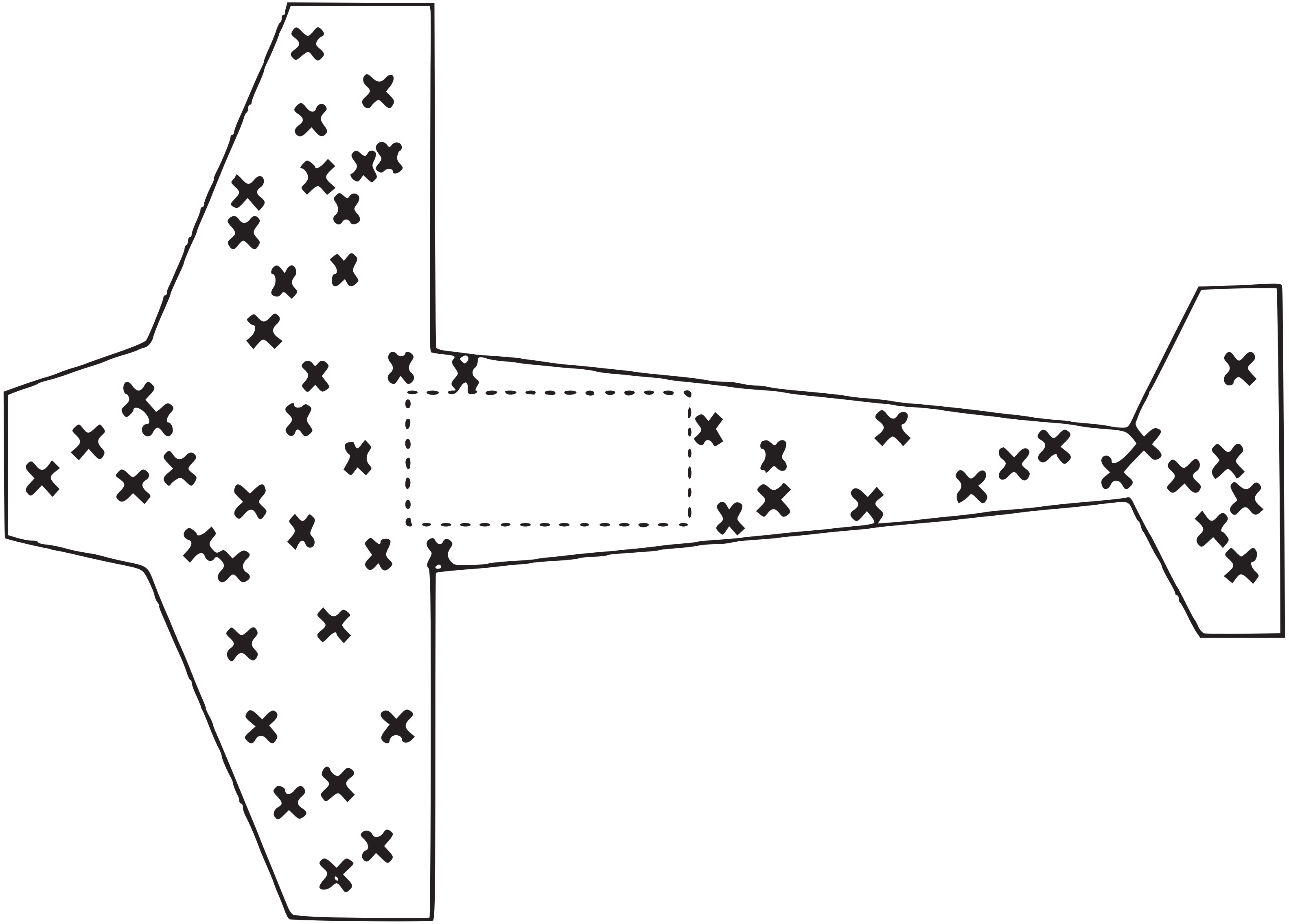
Statistik

1. Einführung
Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung
R und RStudio
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

- Fehler in Analyse: Starts ohne Fehler wurden nicht berücksichtigt
- Korrekte Modellierung mittels logistischer Regression liefert:

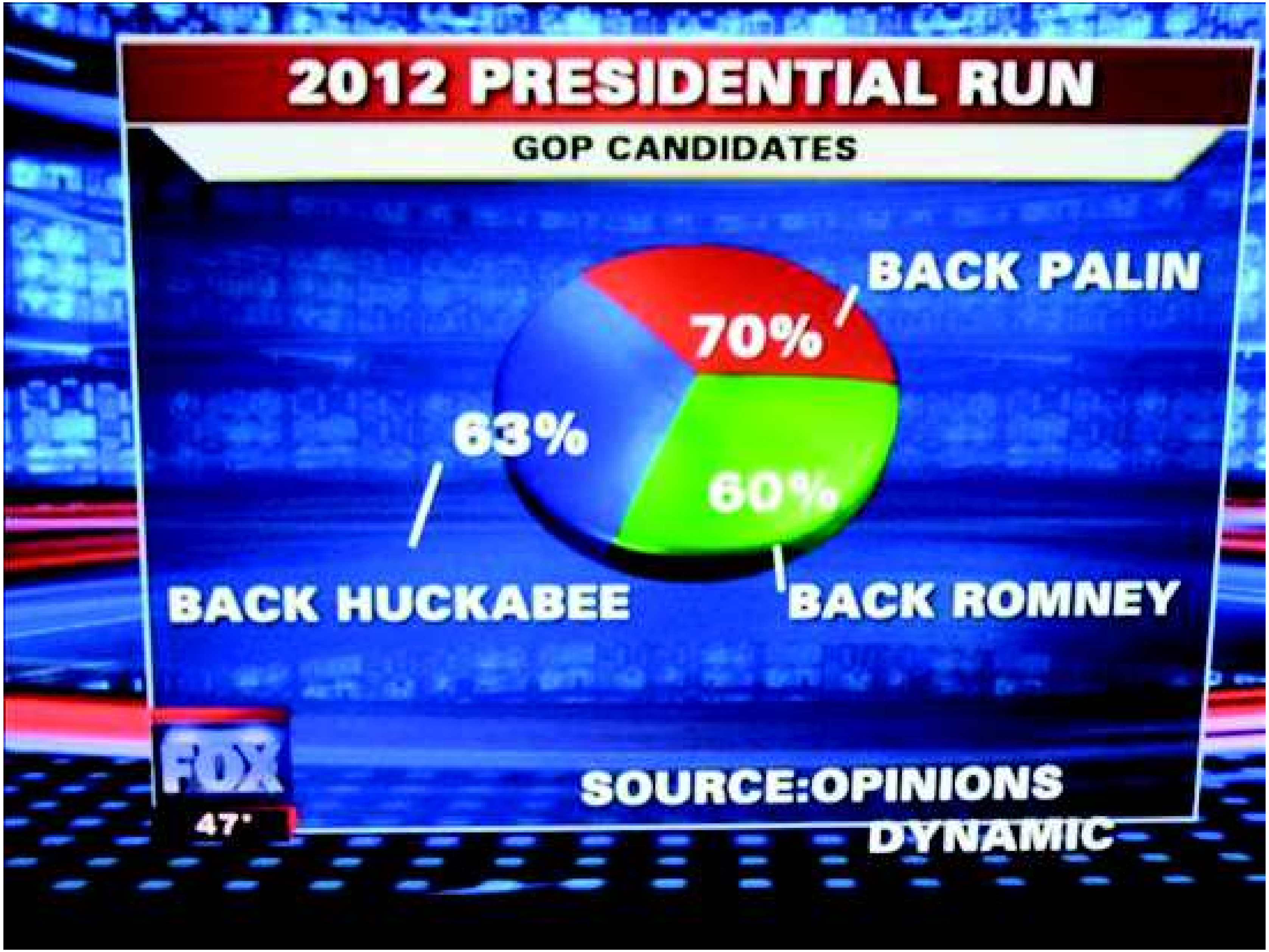


Auswahl der Stichprobe



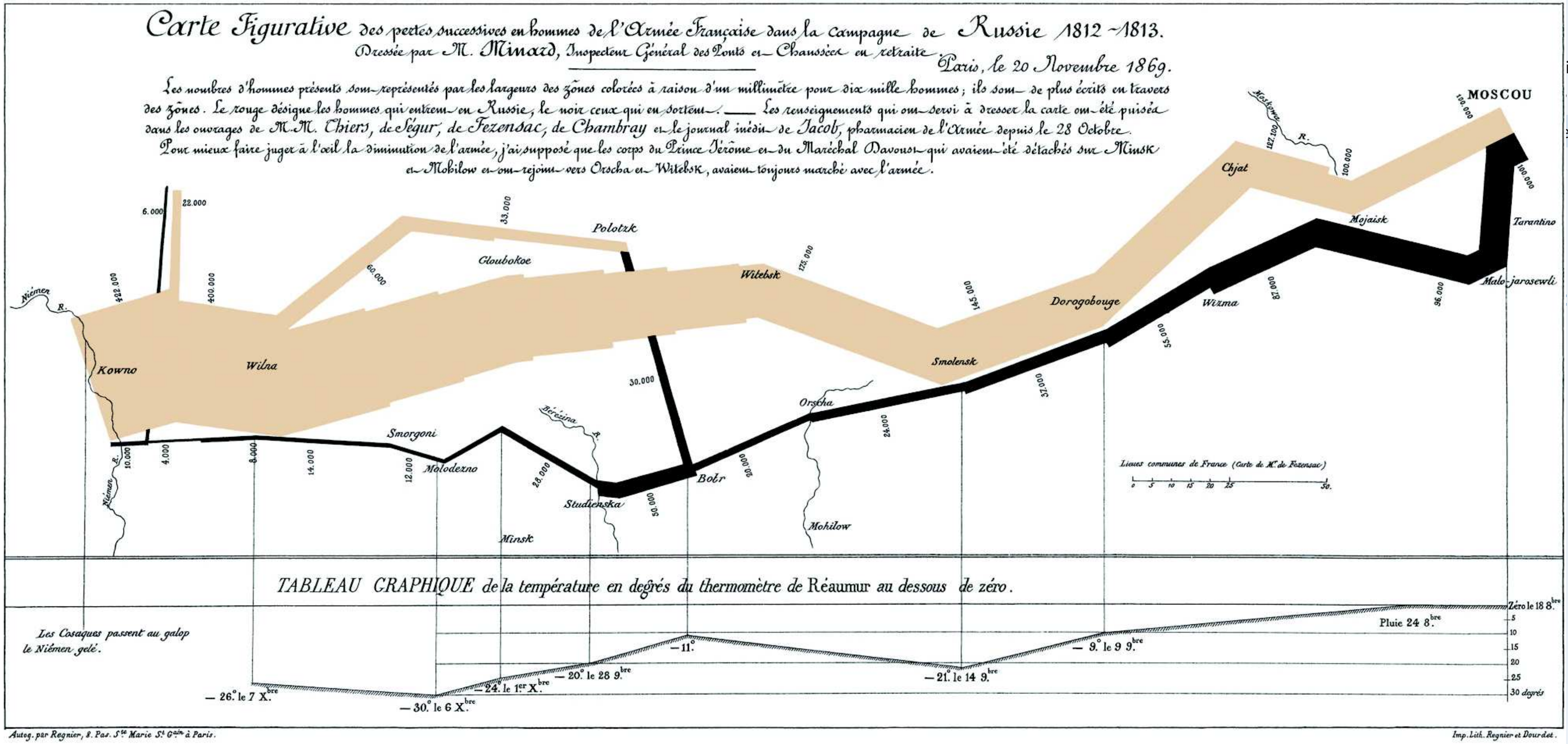
1. Einführung
Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung
R und RStudio
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Aussage?



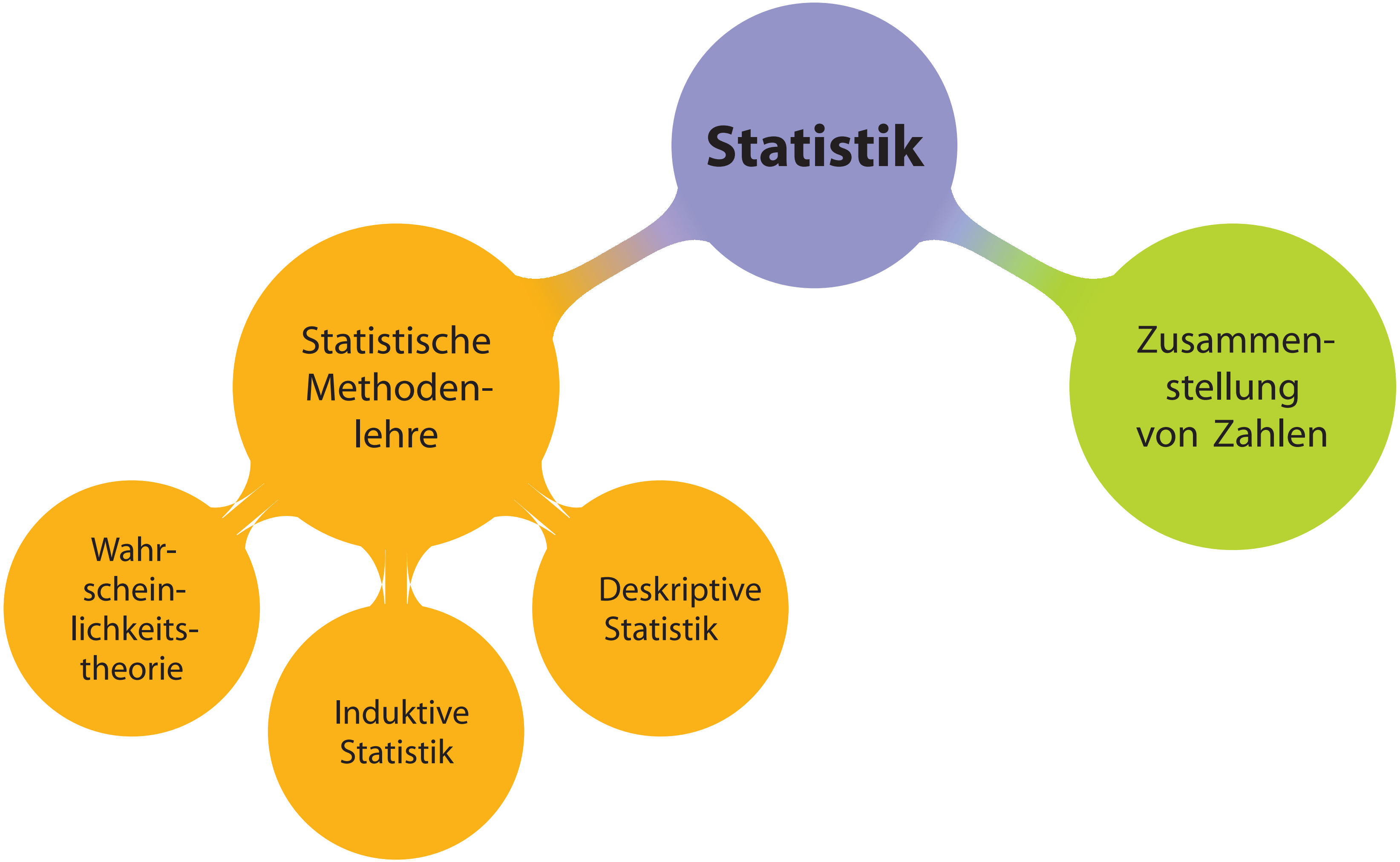
Quelle: Friendly, Michael (2001), Gallery of Data Visualization, `http://www.datavis.ca/gallery/`, Stand November 2014

Minards Grafik von 1869 über Napoleons Rußlandfeldzug



Quelle: Wikimedia Commons, `http://goo.gl/T7ZNme`, Stand November 2014

Bedeutungen des Begriffs Statistik



Einfaches Beispiel

Beispiel

12 Beschäftigte werden nach der Entfernung zum Arbeitsplatz (in km) befragt.

Antworten: 4, 11, 1, 3, 5, 4, 20, 4, 6, 16, 10, 6

- deskriptiv:
  - Durchschnittliche Entfernung: $7{,}5$
  - Klassenbildung:

Klasse	$[0;5)$	$[5;15)$	$[15;30)$
Häufigkeit	5	5	2

- induktiv:
  - Schätze die mittlere Entfernung **aller** Beschäftigten.
  - Prüfe, ob die mittlere Entfernung geringer als 10 km ist.

1. Einführung
Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung
R und RStudio
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Merkmale

- Merkmalsträger: Untersuchte statistische Einheit
- Merkmal: Interessierende Eigenschaft des Merkmalträgers
- (Merkmals-)Ausprägung: Konkret beobachteter Wert des Merkmals
- Grundgesamtheit: Menge aller relevanten Merkmalsträger
- Typen von Merkmalen:

  a) qualitativ – quantitativ
     - qualitativ: z.B. Geschlecht
     - quantitativ: z.B. Schuhgröße
     - Qualitative Merkmale sind quantifizierbar (z.B.: weiblich 1, männlich 0)

  b) diskret – stetig
     - diskret: Abzählbar viele unterschiedliche Ausprägungen
     - stetig: Alle Zwischenwerte realisierbar

1. Einführung
Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung
R und RStudio
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Nominalskala:

- Zahlen haben nur Bezeichnungsfunktion
- z.B. Artikelnummern

Ordinalskala:

- zusätzlich Rangbildung möglich
- z.B. Schulnoten
- Differenzen sind aber **nicht** interpretierbar!
  ➠ Addition usw. ist unzulässig.

Kardinalskala:

- zusätzlich Differenzbildung sinnvoll
- z.B. Gewinn
- Noch feinere Unterscheidung in: Absolutskala, Verhältnisskala, Intervallskala

1. Einführung
Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung
R und RStudio
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Skalendegression und Skalenprogression

Ziel der Skalierung: Gegebene Information angemessen abbilden, möglichst ohne Über- bzw. Unterschätzungen

Es gilt:

- Grundsätzlich können alle Merkmale nominal skaliert werden.
- Grundsätzlich kann jedes metrische Merkmal ordinal skaliert werden.

Das nennt man Skalendegression. Dabei: Informationsverlust

Aber:

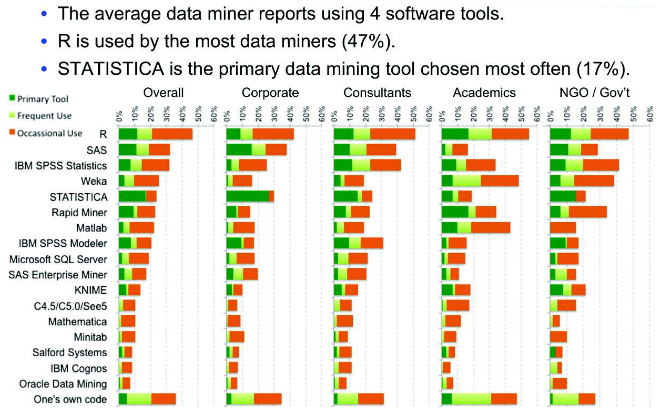
- Nominale Merkmale dürfen **nicht** ordinal- oder metrisch skaliert werden.
- Ordinale Merkmale dürfen **nicht** metrisch skaliert werden.

Das nennt nennt man Skalenprogression. Dabei: Interpretation von **mehr Informationen** in die Merkmale, als inhaltlich vertretbar.
(Gefahr der **Fehlinterpretation**)

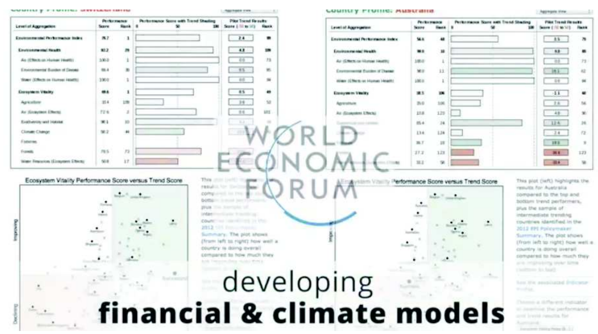
1. Einführung
Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung
R und RStudio
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Was ist R und warum soll man es benutzen?

- R ist ein freies Softwarepaket zu Statistik und Datenanalyse
- R ist sehr mächtig und weit verbreitet in Wissenschaft und Industrie (sogar von mehr Leuten benutzt als z.B. SPSS)



source: `http://goo.gl/axhGhh`



graphics source: `http://goo.gl/W70kms`

Was ist R und warum soll man es benutzen?

- R ist ein freies Softwarepaket zu Statistik und Datenanalyse
- R ist sehr mächtig und weit verbreitet in Wissenschaft und Industrie (sogar von mehr Leuten benutzt als z.B. SPSS)
- Ursprung von R: 1993 an der Universität Auckland von Ross Ihaka and Robert Gentleman entwickelt
- Seitdem: Viele Leute haben R verbessert mit tausenden von Paketen für viele Anwendungen

graphics source: `http://goo.gl/W70kms`

1. Einführung
Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung
R und RStudio
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Was ist R und warum soll man es benutzen?

- R ist ein freies Softwarepaket zu Statistik und Datenanalyse
- R ist sehr mächtig und weit verbreitet in Wissenschaft und Industrie (sogar von mehr Leuten benutzt als z.B. SPSS)
- Ursprung von R: 1993 an der Universität Auckland von Ross Ihaka and Robert Gentleman entwickelt
- Seitdem: Viele Leute haben R verbessert mit tausenden von Paketen für viele Anwendungen
- Nachteil (auf den ersten Blick): Kein point und click tool

```
> summary(diamonds$price)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    326     950    2401    3933    5324   18820
> aveSize <- round(mean(diamonds$carat), 4)
> clarity <- levels(diamonds$clarity)
> p <- qplot(carat, price,
+            data=diamonds, color=clarity,
+            xlab="Carat", ylab="Price",
+            main="Diamond Pricing")
>
> format.plot(p, size=24)
> |
```

graphics source: http://goo.gl/W70kms

1. Einführung
Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung
R und RStudio
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Was ist R und warum soll man es benutzen?

- R ist ein freies Softwarepaket zu Statistik und Datenanalyse
- R ist sehr mächtig und weit verbreitet in Wissenschaft und Industrie (sogar von mehr Leuten benutzt als z.B. SPSS)
- Ursprung von R: 1993 an der Universität Auckland von Ross Ihaka and Robert Gentleman entwickelt
- Seitdem: Viele Leute haben R verbessert mit tausenden von Paketen für viele Anwendungen
- Nachteil (auf den ersten Blick): Kein point und click tool
- Großer Vorteil (auf den zweiten Blick): Kein point und click tool

```
> summary(diamonds$price)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    326     950    2401    3933    5324   18820
> aveSize <- round(mean(diamonds$carat), 4)
> clarity <- levels(diamonds$clarity)
> p <- qplot(carat, price,
+            data=diamonds, color=clarity,
+            xlab="Carat", ylab="Price",
+            main="Diamond Pricing")
>
> format.plot(p, size=24)
> |
```

Download: `R-project.org`

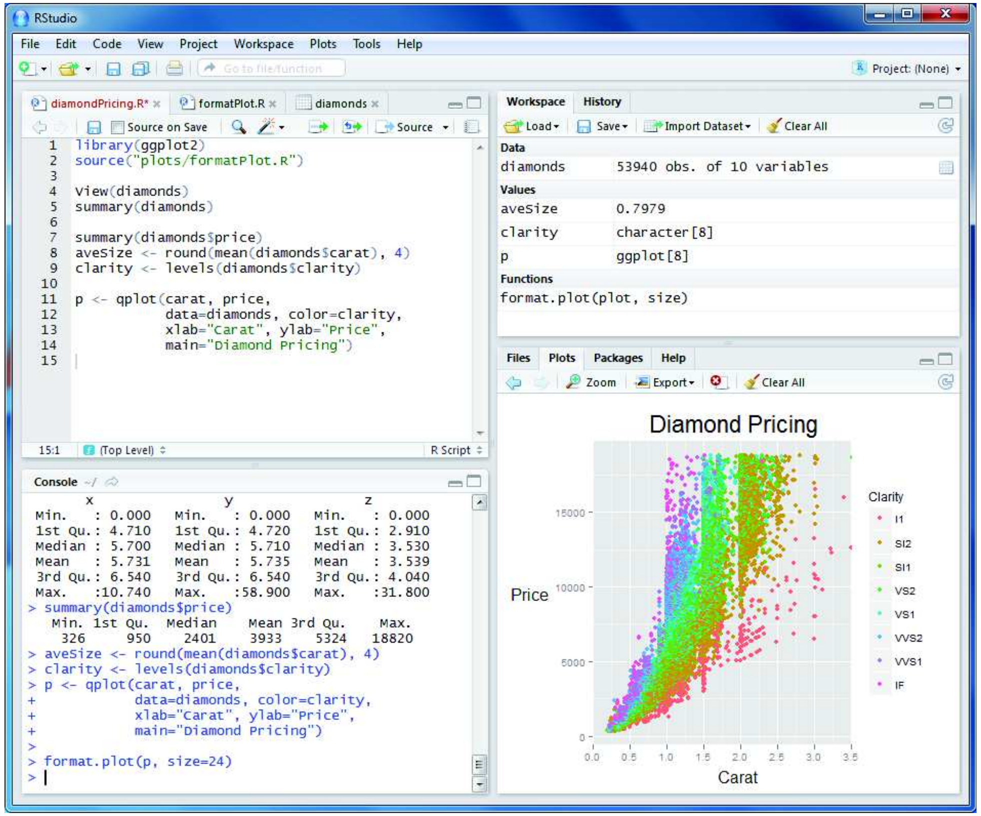
graphics source: `http://goo.gl/W70kms`

Was ist RStudio?

- RStudio ist ein Integrated Development Environment (IDE) um R leichter benutzen zu können.
- Gibt's für OSX, Linux und Windows
- Ist auch frei
- Trotzdem: Sie müssen Kommandos schreiben
- Aber: RStudio unterstützt Sie dabei
- **Download:** `RStudio.com`

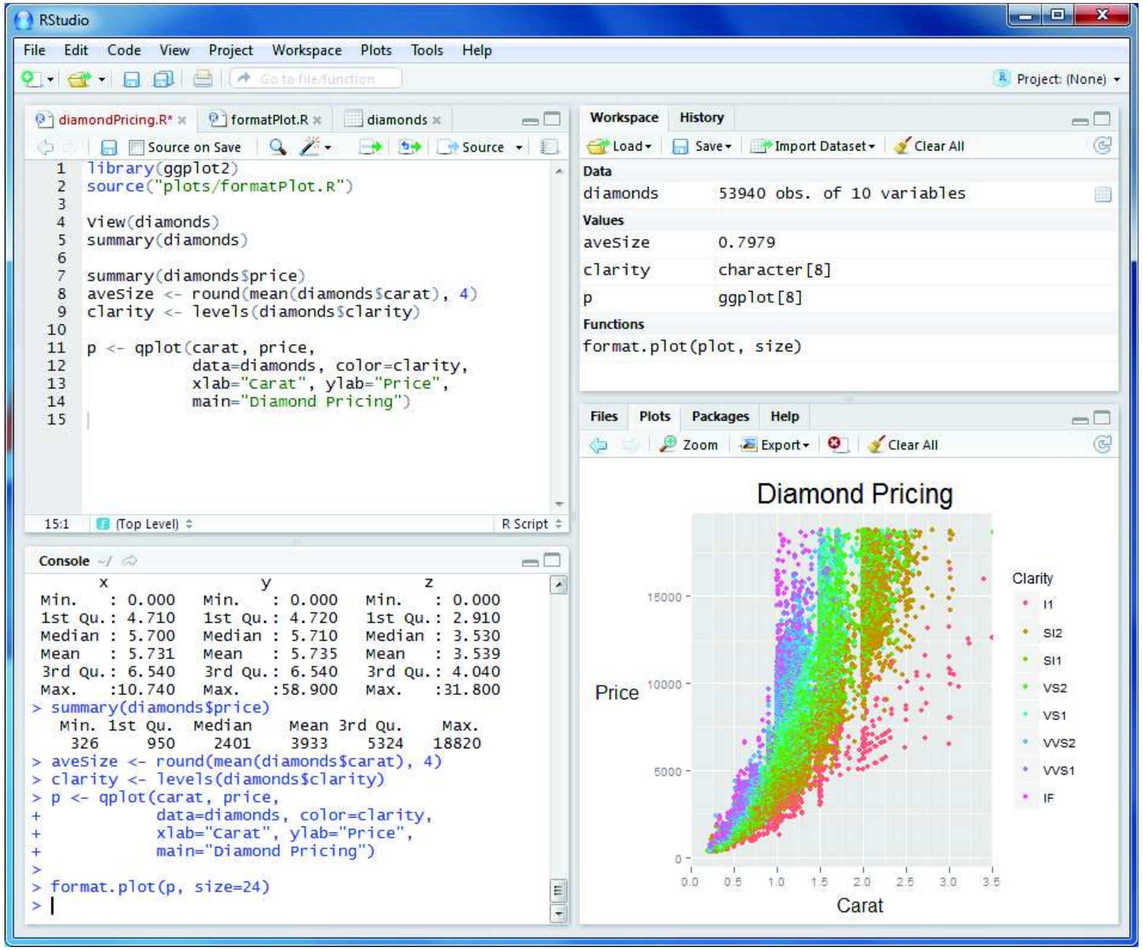
RStudio™

Free & Open-Source IDE for R



RStudio Kennenlernen

- Code
- Console
- Workspace
- History
- Files
- Plots
- Packages
- Help
- Auto-Completion
- Data Import



Statistik

1. Einführung
   - Berühmte Leute zur Statistik
   - Wie lügt man mit Statistik?
   - Gute und schlechte Grafiken
   - Begriff Statistik
   - Grundbegriffe der Datenerhebung
   - R und RStudio
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik

Quellen

Tabellen

Daten einlesen und Struktur anschauen


```r
# Arbeitsverzeichnis setzen (alternativ über Menü)
setwd("C:/ste/work/vorlesungen/2015SS_HSA_Statistik")

# Daten einlesen aus einer csv-Datei (Excel)
MyData = read.csv2(file="../_genericFiles/Daten/Umfrage_HSA_2015_03.csv", header=TRUE)
```

```r
# inspect structure of data
str(MyData)
```

```
## 'data.frame': 377 obs. of  17 variables:
##  $ Jahrgang       : int  2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 ...
##  $ Alter          : int  20 25 19 21 25 20 25 20 23 21 ...
##  $ Groesse        : int  174 157 163 185 178 170 165 175 180 161 ...
##  $ Geschlecht     : Factor w/ 2 levels "Frau","Mann": 1 1 1 2 2 1 1 2 2 1 ...
##  $ AlterV         : int  55 54 51 52 60 50 60 52 56 70 ...
##  $ AlterM         : int  53 61 49 50 63 55 60 49 50 55 ...
##  $ GroesseV       : int  187 185 178 183 170 183 185 175 175 180 ...
##  $ GroesseM       : int  169 160 168 165 160 160 170 169 170 165 ...
##  $ Geschwister    : int  3 1 1 4 2 2 4 1 1 2 ...
##  $ Farbe          : Factor w/ 6 levels "blau","gelb",..: 4 6 4 4 1 6 1 6 4 4 ...
##  $ AusgKomm       : num  240 119 270 40 550 ...
##  $ AnzSchuhe      : int  25 30 25 6 5 65 10 7 10 22 ...
##  $ AusgSchuhe     : int  450 300 100 100 80 250 150 400 150 300 ...
##  $ Essgewohnheiten: Factor w/ 5 levels "carnivor","fruktarisch",..: 1 1 1 1 1 1 5 1 1 1 ...
##  $ Raucher        : Factor w/ 2 levels "ja","nein": NA 2 2 2 1 2 2 2 2 1 ...
##  $ NoteMathe      : num  2.3 3.3 1.7 2 4 4 3.3 2.7 3.7 3.3 ...
##  $ MatheZufr      : Ord.factor w/ 4 levels "unzufrieden"<..: 2 2 2 2 2 2 2 2 2 2 ...
```


1. Einführung
- Berühmte Leute zur Statistik
- Wie lügt man mit Statistik?
- Gute und schlechte Grafiken
- Begriff Statistik
- Grundbegriffe der Datenerhebung
- R und RStudio

2. Deskriptive Statistik

3. W-Theorie

4. Induktive Statistik

Quellen

Tabellen

Erste Zeilen der Datentabelle

```r
# Erste Zeilen in Datentabelle
head(MyData, 6)

##   Jahrgang Alter Groesse Geschlecht AlterV AlterM GroesseV GroesseM Geschwister   Farbe AusgKomm
## 1     2015    20     174       Frau     55     53      187      169           3 schwarz    240.0
## 2     2015    25     157       Frau     54     61      185      160           1   weiss    119.4
## 3     2015    19     163       Frau     51     49      178      168           1 schwarz    270.0
## 4     2015    21     185       Mann     52     50      183      165           4 schwarz     40.0
## 5     2015    25     178       Mann     60     63      170      160           2    blau    550.0
## 6     2015    20     170       Frau     50     55      183      160           2   weiss    420.0
##   AnzSchuhe AusgSchuhe Essgewohnheiten Raucher NoteMathe MatheZufr
## 1        25        450        carnivor    <NA>       2.3   geht so
## 2        30        300        carnivor    nein       3.3   geht so
## 3        25        100        carnivor    nein       1.7   geht so
## 4         6        100        carnivor    nein       2.0   geht so
## 5         5         80        carnivor      ja       4.0   geht so
## 6        65        250        carnivor    nein       4.0   geht so
```

```r
# lege MyData als den "Standard"-Datensatz fest
attach(MyData)
```

```r
# Wie Viele Objekte gibt's im Datensatz?
nrow(MyData)

## [1] 377

# Wie Viele Merkmale?
ncol(MyData)

## [1] 17
```


1. Einführung
Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung
R und RStudio
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Daten kennenlernen

```
# Auswahl spezieller Objekte und Merkmale über [Zeile, Spalte]
MyData[1:3, 2:5]

##   Alter Groesse Geschlecht AlterV
## 1    20     174       Frau     55
## 2    25     157       Frau     54
## 3    19     163       Frau     51
```

```
# Auswahl von Objekten über logische Ausdrücke
Auswahl = (MyData$Geschlecht=="Mann" & MyData$Alter < 19)
# zeige die ersten Einträge
head(Auswahl, 30)

##  [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [17] FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

# Ausgabe der Auswahl: Alter, Alter des Vaters und der Mutter
MyData[Auswahl,                       # Objektauswahl
      c("Alter", "AlterM", "AlterV")]  # Welche Merkmale?

##     Alter AlterM AlterV
## 23     18     44     48
## 268    18     46     52
```


1. Einführung
Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung
R und RStudio
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Daten kennenlernen

```
# Zeige die Männer, die mehr als 1000 Euro für Schuhe
# und Mobilfunk zusammen ausgegeben haben
MyData[MyData$Geschlecht=="Mann" & MyData$AusgSchuhe + MyData$AusgKomm > 1000,
       c("Alter", "Geschwister", "Farbe", "AusgSchuhe", "AusgKomm")]

##     Alter Geschwister   Farbe AusgSchuhe AusgKomm
## 19     20           0   weiss        300      924
## 23     18           1  silber        300     1000
## 33     25           1 schwarz        300     1000
## 42     24           1 schwarz       1000      600
## 81     25           2  silber        200     1900
## 106    21           1 schwarz        200      860
## 121    22           0  silber        300     1100
## 142    20           2 schwarz        290     1570
## 161    19           1 schwarz        600      800
## 168    21           2    blau        600      505
## 179    21           0  silber        300      825
## 211    23           1 schwarz        450      630
## 223    20           1     rot        400      815
## 227    20           1 schwarz        200     1250
## 249    20           1    blau       1000      350
## 256    25           0 schwarz        280     1200
## 272    24           1 schwarz        300      900
## 281    19           2 schwarz        500      720
## 315    21           1   weiss        200     1300
## 353    20           0 schwarz        400      950
```


1. Einführung
Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung
R und RStudio
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Daten kennenlernen

```
# Auswahl spezieller Objekte und Merkmale über [Zeile, Spalte]
MyData[1:3, 2:5]

##   Alter Groesse Geschlecht AlterV
## 1    20     174       Frau     55
## 2    25     157       Frau     54
## 3    19     163       Frau     51
```

```
# Auswahl von Objekten über logische Ausdrücke
Auswahl = (MyData$Geschlecht=="Mann" & MyData$Alter < 19)
# zeige die ersten Einträge
head(Auswahl, 30)

##  [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [17] FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

# Ausgabe der Auswahl: Alter, Alter des Vaters und der Mutter
MyData[Auswahl,                        # Objektauswahl
      c("Alter", "AlterM", "AlterV")]  # Welche Merkmale?

##     Alter AlterM AlterV
## 23     18     44     48
## 268    18     46     52
## 424    17     46     50
## 456    18     52     55
## 460    18     50     57
## 464    18     40     44
## 479    18     52     44
## 501    18     51     55
## 566    18     52     57
## 620    18     49     58
```


Statistik

1. Einführung
Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung
R und RStudio

2. Deskriptive Statistik

3. W-Theorie

4. Induktive Statistik

Quellen

Tabellen

Daten kennenlernen

```r
# Zeige die Männer, die mehr als 1300 Euro für Schuhe
# und Mobilfunk zusammen ausgegeben haben
MyData.Auswahl = MyData[MyData$Geschlecht=="Mann" &
                        MyData$AusgSchuhe + MyData$AusgKomm > 1300,
                        c("Alter", "Geschwister", "Farbe",
                          "AusgSchuhe", "AusgKomm")]
```


1. Einführung

- Berühmte Leute zur Statistik
- Wie lügt man mit Statistik?
- Gute und schlechte Grafiken
- Begriff Statistik
- Grundbegriffe der Datenerhebung
- R und RStudio

2. Deskriptive Statistik

3. W-Theorie

4. Induktive Statistik

Quellen

Tabellen

Daten kennenlernen

```
# ohne NAs
MyData.Auswahl = na.exclude(MyData.Auswahl)
MyData.Auswahl
```

```
##      Alter Geschwister   Farbe AusgSchuhe AusgKomm
## 42      24         1.0 schwarz       1000      600
## 81      25         2.0  silber        200     1900
## 121     22         0.0  silber        300     1100
## 142     20         2.0 schwarz        290     1570
## 161     19         1.0 schwarz        600      800
## 227     20         1.0 schwarz        200     1250
## 249     20         1.0    blau       1000      350
## 256     25         0.0 schwarz        280     1200
## 315     21         1.0   weiss        200     1300
## 353     20         0.0 schwarz        400      950
## 415     26         1.0    blau        600     1850
## 419     21         0.0 schwarz        200     1500
## 492     23         2.0   weiss        160     1800
## 493     26         2.0 schwarz        300     2000
## 494     20         2.0 schwarz        250     1500
## 535     20         2.0   weiss       2500     1500
## 548     26         2.0 schwarz        240     1200
## 562     24         1.0 schwarz         70     4668
## 573     21         1.0 schwarz        300     1200
## 581     19         2.0  silber        500      950
## 582     20         1.0 schwarz        500     1000
## 604     24         1.0 schwarz        150     1340
## 605     21         1.0  silber        600      800
## 615     25         4.5 schwarz       1200      600
## 646     22         1.0     rot        200     2500
## 647     23         1.0 schwarz        200     2000
```

1. Einführung
Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung
R und RStudio
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Daten kennenlernen

```r
# Neue Spalte Gesamtausgaben:
MyData.Auswahl$AusgGesamt = MyData.Auswahl$AusgKomm + MyData.Auswahl$AusgSchuhe
# sortiert nach Gesamtausgaben
MyData.Auswahl[order(MyData.Auswahl$AusgGesamt), ]
```

```
##     Alter Geschwister   Farbe AusgSchuhe AusgKomm AusgGesamt
## 249    20         1.0    blau       1000      350       1350
## 353    20         0.0 schwarz        400      950       1350
## 121    22         0.0  silber        300     1100       1400
## 161    19         1.0 schwarz        600      800       1400
## 605    21         1.0  silber        600      800       1400
## 548    26         2.0 schwarz        240     1200       1440
## 227    20         1.0 schwarz        200     1250       1450
## 581    19         2.0  silber        500      950       1450
## 256    25         0.0 schwarz        280     1200       1480
## 604    24         1.0 schwarz        150     1340       1490
## 315    21         1.0   weiss        200     1300       1500
## 573    21         1.0 schwarz        300     1200       1500
## 582    20         1.0 schwarz        500     1000       1500
## 42     24         1.0 schwarz       1000      600       1600
## 653    27         2.0 schwarz        700      950       1650
## 419    21         0.0 schwarz        200     1500       1700
## 494    20         2.0 schwarz        250     1500       1750
## 615    25         4.5 schwarz       1200      600       1800
## 142    20         2.0 schwarz        290     1570       1860
## 492    23         2.0   weiss        160     1800       1960
## 663    27         2.0 schwarz        200     1800       2000
## 81     25         2.0  silber        200     1900       2100
## 647    23         1.0 schwarz        200     2000       2200
## 493    26         2.0 schwarz        300     2000       2300
## 415    26         1.0    blau        600     1850       2450
```

Statistik: Table of Contents

1 Statistik: Einführung

2 Deskriptive Statistik

3 Wahrscheinlichkeitstheorie

4 Induktive Statistik

2 Deskriptive Statistik

- Häufigkeiten
- Lage und Streuung
- Konzentration
- Zwei Merkmale
- Korrelation
- Preisindizes
- Lineare Regression

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Auswertungsmethoden für eindimensionales Datenmaterial

- Merkmal X wird an n Merkmalsträgern beobachtet ➠ **Urliste** $(x_1, \ldots, x_n)$
  Im Beispiel: $x_1 = 4,\ x_2 = 11, \ldots, x_{12} = 6$
- Urlisten sind oft unübersichtlich, z.B.:

```
##  [1] 4 5 4 1 5 4 3 4 5 6 6 5 5 4 7 4 6 5 6 4 5 4 7 5 5 6 7 3
## [29] 7 6 6 7 4 5 4 7 7 5 5 5 5 6 6 4 5 2 5 4 7 5
```

- Dann zweckmäßig: Häufigkeitsverteilungen

Ausprägung (sortiert)	a_j	1	2	3	4	5	6	7	$\sum$
absolute Häufigkeit	$h(a_j) = h_j$	1	1	2	12	17	9	8	50
kumulierte abs. H.	$H(a_j) = \sum_{i=1}^{j} h(a_i)$	1	2	4	16	33	42	50	—
relative Häufigkeit	$f(a_j) = h(a_j)/n$	$\frac{1}{50}$	$\frac{1}{50}$	$\frac{2}{50}$	$\frac{12}{50}$	$\frac{17}{50}$	$\frac{9}{50}$	$\frac{8}{50}$	1
kumulierte rel. H.	$F(a_j) = \sum_{i=1}^{j} f(a_i)$	$\frac{1}{50}$	$\frac{2}{50}$	$\frac{4}{50}$	$\frac{16}{50}$	$\frac{33}{50}$	$\frac{42}{50}$	1	—

Beispiel: Alter von Studierenden in Statistik Vorlesung

	$h(a_j)$	$H(a_j) = \sum_{i=1}^{j} h(a_i)$	$f(a_j) = \frac{h(a_j)}{n}$	$F(a_j) = \sum_{i=1}^{j} f(a_i)$
17	1	1	0.0015	0.0015
18	37	38	0.0552	0.0567
19	113	151	0.1687	0.2254
20	113	264	0.1687	0.3940
21	94	358	0.1403	0.5343
22	71	429	0.1060	0.6403
23	67	496	0.1000	0.7403
24	49	545	0.0731	0.8134
25	24	569	0.0358	0.8493
26	25	594	0.0373	0.8866
27	19	613	0.0284	0.9149
28	20	633	0.0299	0.9448
29	11	644	0.0164	0.9612
30	5	649	0.0075	0.9687
31	5	654	0.0075	0.9761
32	7	661	0.0104	0.9866
33	2	663	0.0030	0.9896
34	3	666	0.0045	0.9940
35	2	668	0.0030	0.9970
36	2	670	0.0030	1.0000

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

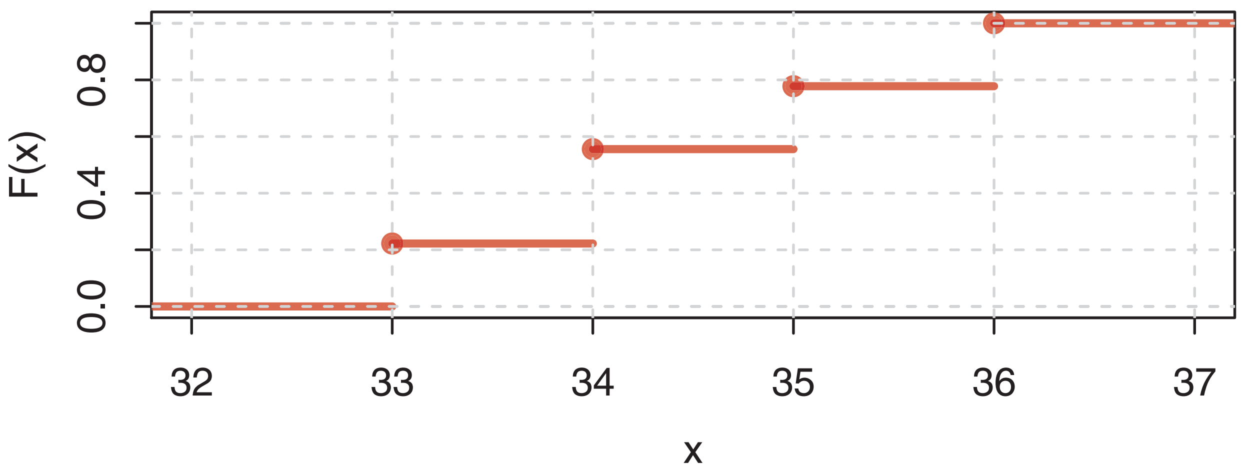
Empirische Verteilungsfunktion

- für metrische Merkmale
- Anteil der Ausprägungen, die höchstens so hoch sind wie x.
- Exakt:

$$F(x) = \sum_{a_i \leqslant x} f(a_i)$$

Beispiel

```
Studenten.ueber.32 = sort(MyData$Alter[MyData$Alter > 32])
Studenten.ueber.32

## [1] 33 33 34 34 34 35 35 36 36

# empirical cumulative distribution function (ecdf)
Studenten.F = ecdf(Studenten.ueber.32)
plot(Studenten.F, col=rgb(0.8,0,0,.7), lwd=3, main="", xlab="x", ylab="F(x)")
grid(lty=2) # Gitternetz
```



1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Empirische Quantile

- für metrische Merkmale; Voraussetzung: sortierte Urliste
- Umkehrung der Verteilungsfunktion
- Anteil p gegeben, gesucht: $F^{-1}(p)$, falls vorhanden.
- Definition p-Quantil:

$$\tilde{x}_p = \begin{cases} \frac{1}{2}(x_{n \cdot p} + x_{n \cdot p + 1}), & \text{wenn } n \cdot p \in \mathbb{N}_0 \\ x_{\lceil n \cdot p \rceil}, & \text{sonst} \end{cases}$$

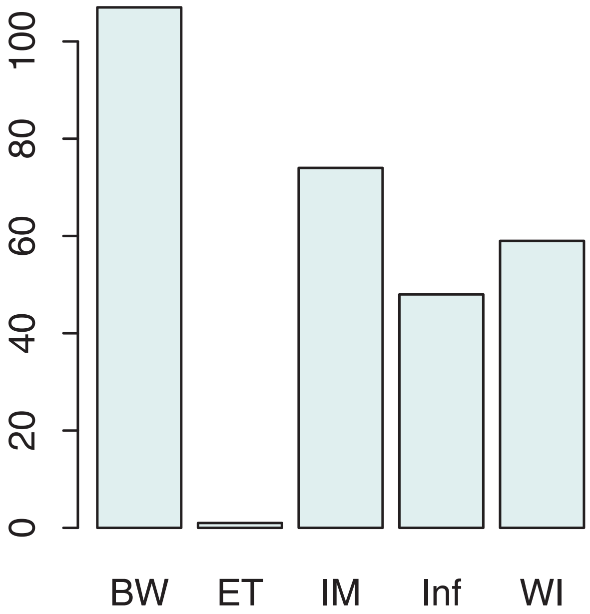
Beispiel

```
## [1] 33 33 34 34 34 35 35 36 36

n = length(Studenten.ueber.32)
p = c(0.05, 2/n, 0.3, 0.5, 0.75, 0.9)

quantile(Studenten.ueber.32, probs=p, type=2)

##        5% 22.22222%       30%       50%       75%       90%
##      33.0      33.5      34.0      34.0      35.0      36.0
```


1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

❶ Balkendiagramm

```
M.t = table(MyData$Studiengang)
M.t

##
##   BW  ET  IM Inf  WI
## 107   1  74  48  59

barplot(M.t, col="azure2")
```



(Höhe proportional zu Häufigkeit)

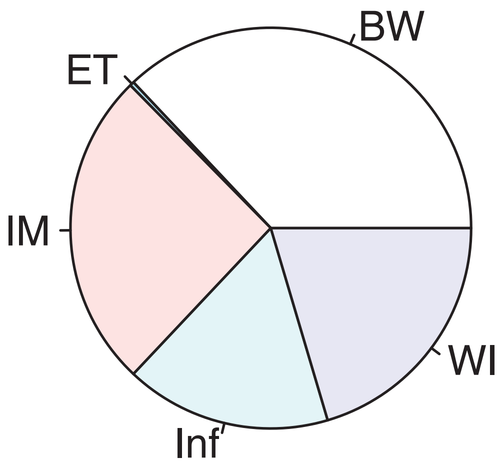
❷ Kreissektorendiagramm

Winkel: $w_j = 360° \cdot f(a_j)$

z.B. $w_{\text{BW}} = 360° \cdot \frac{107}{289} \approx 133.2°$

z.B. $w_{\text{IM}} = 360° \cdot \frac{74}{289} \approx 93.6°$

```
pie(M.t)
```

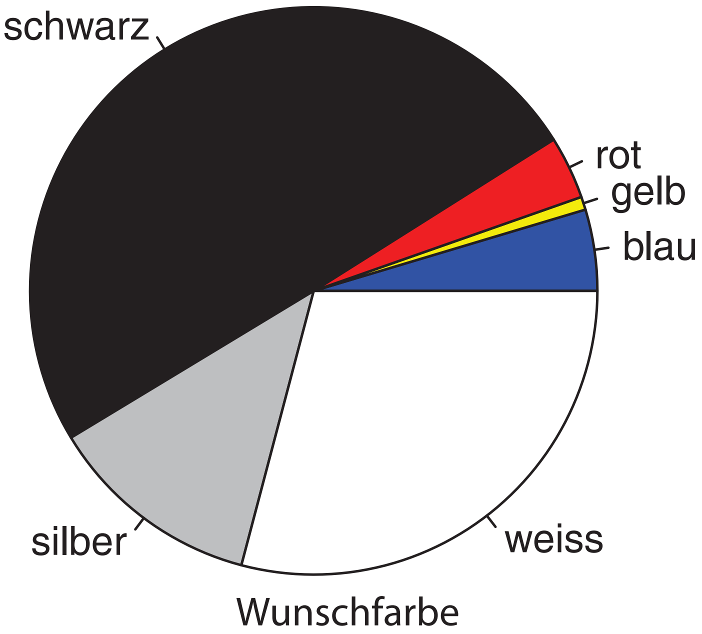


(Fläche proportional zu Häufigkeit)

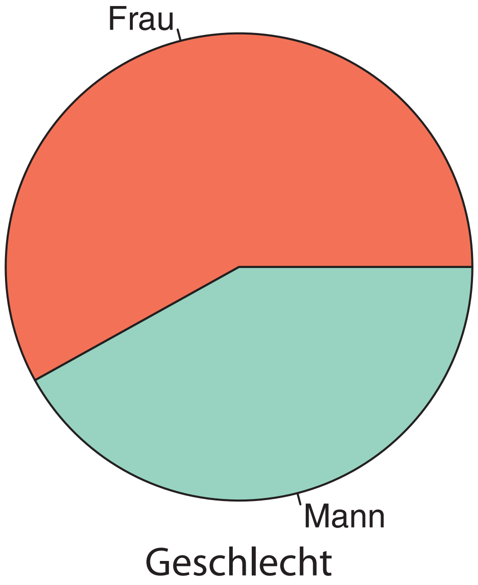
1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Kreissektorendiagramm

```
pie(table(MyData$Farbe),
    col=c("blue", "yellow", "red",
          "black", "grey", "white"))
```



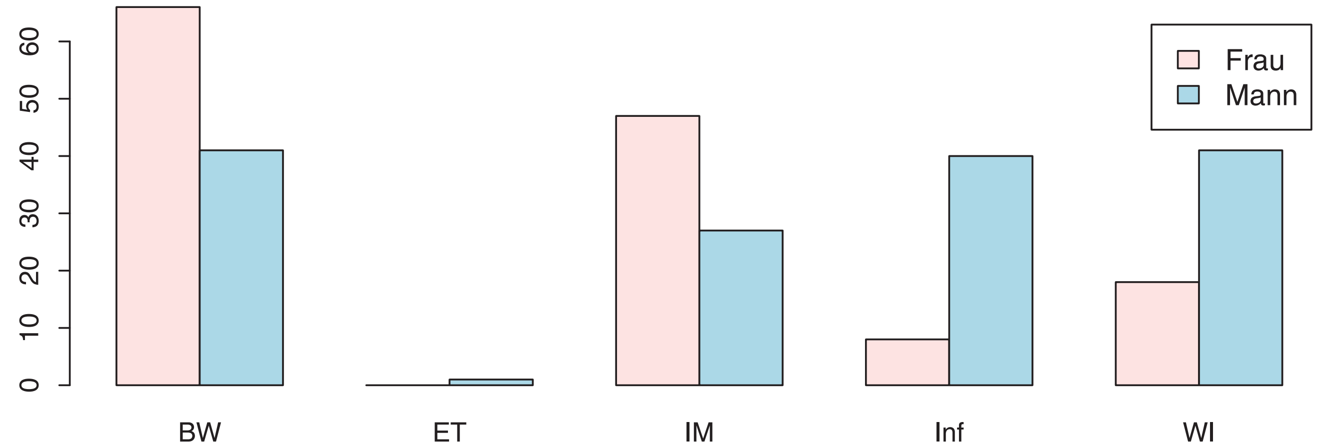

```
pie(table(MyData$Geschlecht),
    col=c("coral1", "aquamarine"))
```



1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

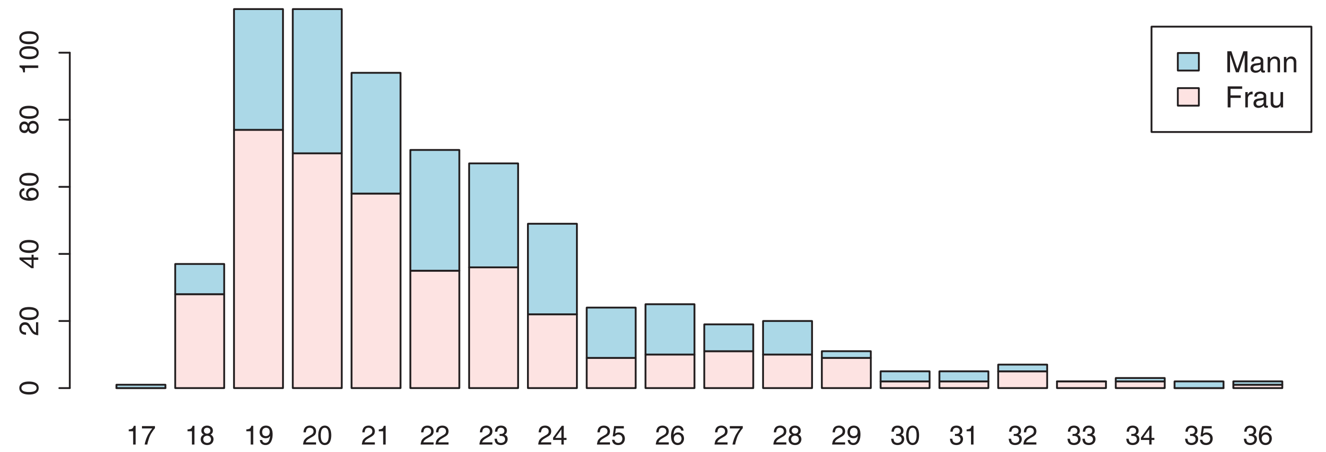
Balkendiagramm, Klassen getrennt oder gestapelt

```
barplot(xtabs(~ Geschlecht + Studiengang),
        legend=TRUE, beside=TRUE, col=c("mistyrose", "lightblue"))
```



Studiengang versus Geschlecht

```
barplot(xtabs(~ Geschlecht + Alter),
        legend=TRUE, beside=FALSE, col=c("mistyrose", "lightblue"))
```



Alter versus Geschlecht

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

❸ Histogramm

- für klassierte Daten
- Fläche proportional zu Häufigkeit:

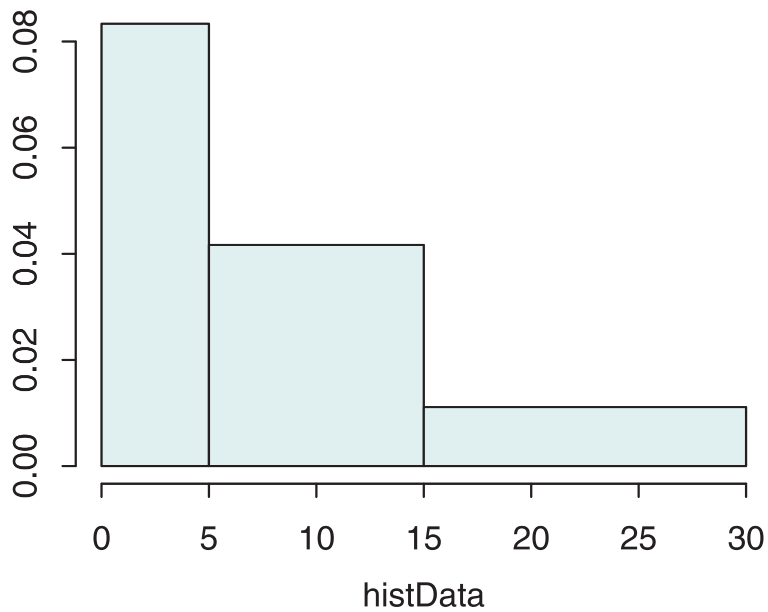
$$\text{Höhe}_j \cdot \text{Breite}_j = c \cdot h(a_j)$$

$$\Rightarrow \boxed{\text{Höhe}_j = c \cdot \frac{h(a_j)}{\text{Breite}_j}}$$

- Im Beispiel mit $c = \frac{1}{12}$:

Klasse	$[0;5)$	$[5;15)$	$[15;30]$
$h(a_j)$	5	5	2
Breite_j	5	10	15
Höhe_j	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{90}$

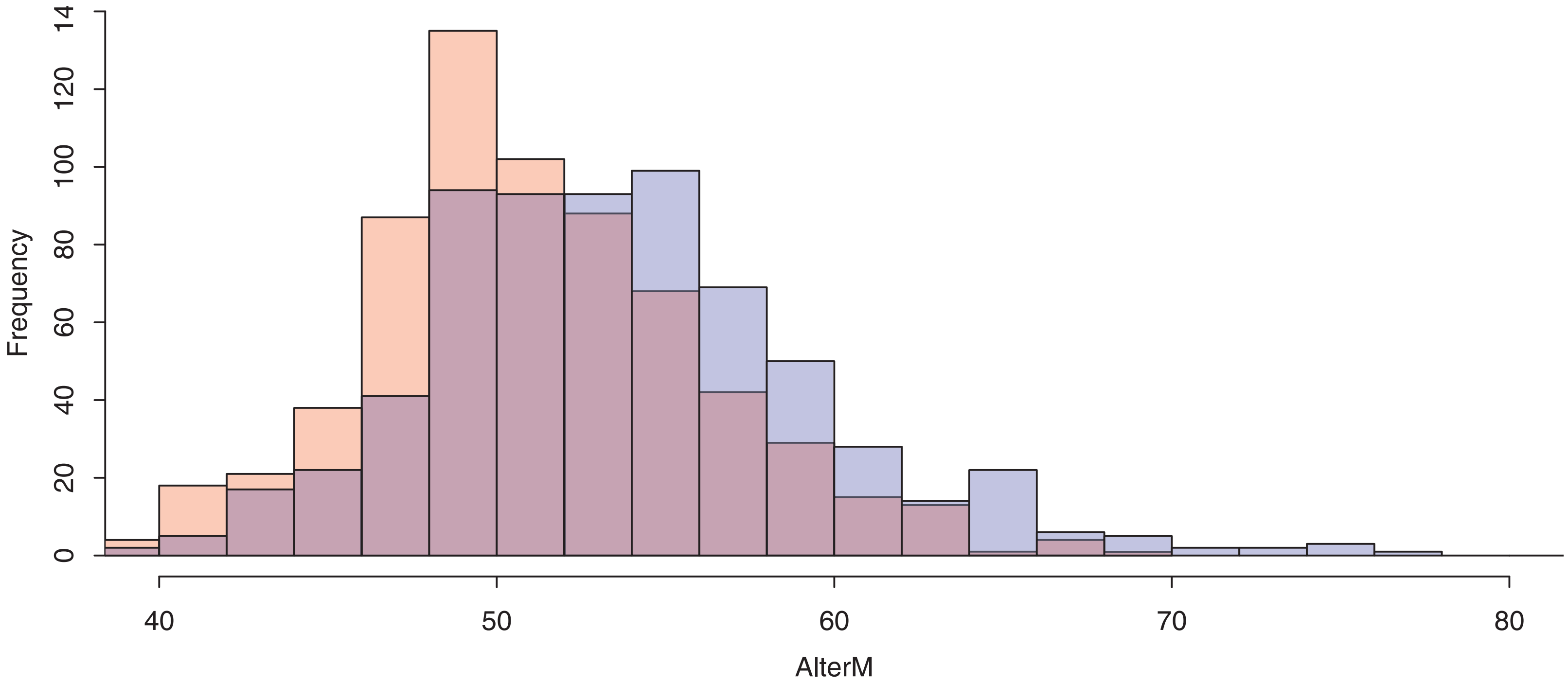
```
require(MASS)
histData <- c(0,1,2,3,4,
              5,6,7,10,14,
              15,30)
truehist(histData,
         breaks=c(0, 4.999, 14.999, 30),
         col="azure2", ylab='')
```



1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Histogramm

```
plot(hist(AlterM, plot=F, breaks=20),
          col=rgb(1,0,0,1/4), # make red transparent
          main="",
          xlim=c(40,80)) # draw from 40 to 80
plot(hist(AlterV, plot=F, breaks=20),
          col=rgb(0,0,1,1/4),
          add=TRUE)
```



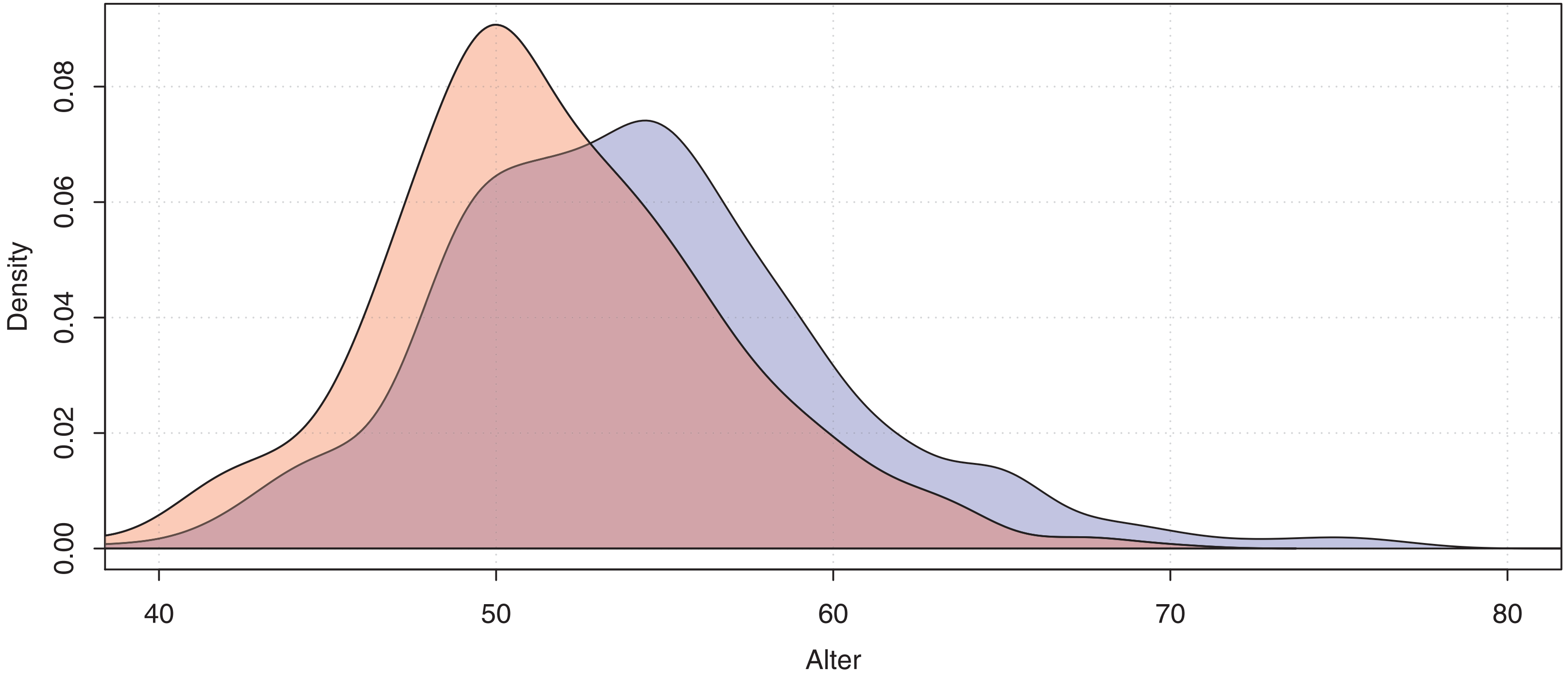
Histogramm: Alter der Väter (blau) und Mütter (rosa)

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Dichteplot

```
densMutter = density(na.exclude(AlterM))
densVater = density(na.exclude(AlterV))
plot(densMutter, main="", xlab="Alter",
xlim=c(40,80), # draw from 40 to 80
panel.first=grid()) # draw a grid
polygon(densVater, density=-1, col=rgb(0,0,1,1/4))
polygon(densMutter, density=-1, col=rgb(1,0,0,1/4))
```



Dichteplot: Alter der Väter (blau) und Mütter (rosa)

Lageparameter

"Sollen wir das arithmetische Mittel als durchschnittliche Körpergröße nehmen und den Gegner erschrecken, oder wollen wir ihn einlullen und nehmen den Median?"

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Lageparameter

Modus x_{Mod}: häufigster Wert

Beispiel:

a_j	1	2	4
$h(a_j)$	4	3	1

$\Rightarrow x_{\text{Mod}} = 1$

Sinnvoll bei allen Skalenniveaus.

Median x_{Med}: ‚mittlerer Wert', d.h.

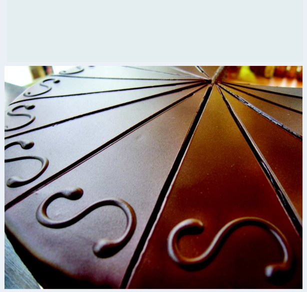
1. Urliste aufsteigend sortieren: $x_1 \leqq x_2 \leqq \cdots \leqq x_n$
2. Dann

$$x_{\text{Med}} \begin{cases} = x_{\frac{n+1}{2}}, & \text{falls } n \text{ ungerade} \\ \in [x_{\frac{n}{2}}; x_{\frac{n}{2}+1}], & \text{falls } n \text{ gerade (meist } x_{\text{Med}} = \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1})) \end{cases}$$

Im Beispiel oben:
$1,\ 1,\ 1,\ 1,\ 2,\ 2,\ 2,\ 4 \Rightarrow x_{\text{Med}} \in [1;2]$, z.B. $x_{\text{Med}} = 1{,}5$

Sinnvoll ab ordinalem Skalenniveau.

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen



Lageparameter (2)

- Arithmetisches Mittel $\bar{x}$: Durchschnitt, d.h.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n} \sum_{j=1}^{k} a_j \cdot h(a_j)$$

Im Beispiel:

$$\bar{x} = \tfrac{1}{8} \cdot (\underbrace{1+1+1+1}_{1 \cdot 4} + \underbrace{2+2+2}_{2 \cdot 3} + \underbrace{\ 4\ }_{4 \cdot 1}) = 1{,}75$$

Sinnvoll nur bei kardinalem Skalenniveau.
Bei klassierten Daten:

$$\bar{x}^* = \tfrac{1}{n} \sum \text{Klassenmitte} \cdot \text{Klassenhäufigkeit}$$

Im Beispiel:

$$\bar{x}^* = \tfrac{1}{12} \cdot (2{,}5 \cdot 5 + 10 \cdot 5 + 22{,}5 \cdot 2) = 8{,}96 \neq 7{,}5 = \bar{x}$$

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Lageparameter

Ausgaben für Schuhe

```r
median(na.exclude(AusgSchuhe))
## [1] 200
mean(na.exclude(AusgSchuhe))
## [1] 270.4529
```

Alter

```r
median(Alter)
## [1] 21
mean(Alter)
## [1] 22.12537
```

```r
summary(Geschlecht)
## Frau Mann
##  389  281
```

Alter der Mutter

```r
median(na.exclude(AlterM))
## [1] 51
mean(na.exclude(AlterM))
## [1] 51.63677
```


1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

- Voraussetzung: kardinale Werte $x_1, \ldots, x_n$
- **Beispiel:**

$$\left.\begin{array}{ll|ccc} \text{a)} & x_i & 1950 & 2000 & 2050 \\ \hline \text{b)} & x_i & 0 & 0 & 6000 \end{array}\right\} \text{je } \bar{x} = 2000$$

- Spannweite: $\mathrm{SP} = \max_i x_i - \min_i x_i$

  Im Beispiel:

$$\begin{array}{lll} \text{a)} & \mathrm{SP} = 2050 - 1950 & = 100 \\ \text{b)} & \mathrm{SP} = 6000 - 0 & = 6000 \end{array}$$

- Mittlere quadratische Abweichung:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \underbrace{\frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2}_{\text{Verschiebungssatz}}$$

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

- **Mittlere quadratische Abweichung** im Beispiel:

$$\text{a)}\quad s^2 = \tfrac{1}{3}\cdot\left(50^2+0^2+50^2\right) = \tfrac{1}{3}\cdot\left(1950^2+2000^2+2050^2\right) - 2000^2 = 1666{,}67$$

$$\text{b)}\quad s^2 = \tfrac{1}{3}\cdot\left(2000^2+2000^2+4000^2\right) = \tfrac{1}{3}\cdot\left(0^2+0^2+6000^2\right) - 2000^2 = 8000000$$

- **Standardabweichung:** $s = \sqrt{s^2}$

  Im Beispiel:

$$\text{a)}\quad s = \sqrt{1666{,}67} = 40{,}82$$

$$\text{b)}\quad s = \sqrt{8000000} = 2828{,}43$$

- **Variationskoeffizient:** $V = \dfrac{s}{\bar{x}}$ (maßstabsunabhängig)

  Im Beispiel:

$$\text{a)}\quad V = \frac{40{,}82}{2000} = 0{,}02\ (\widehat{=}\,2\,\%)$$

$$\text{b)}\quad V = \frac{2828{,}43}{2000} = 1{,}41\ (\widehat{=}\,141\,\%)$$

Lage und Streuung: Überblick

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

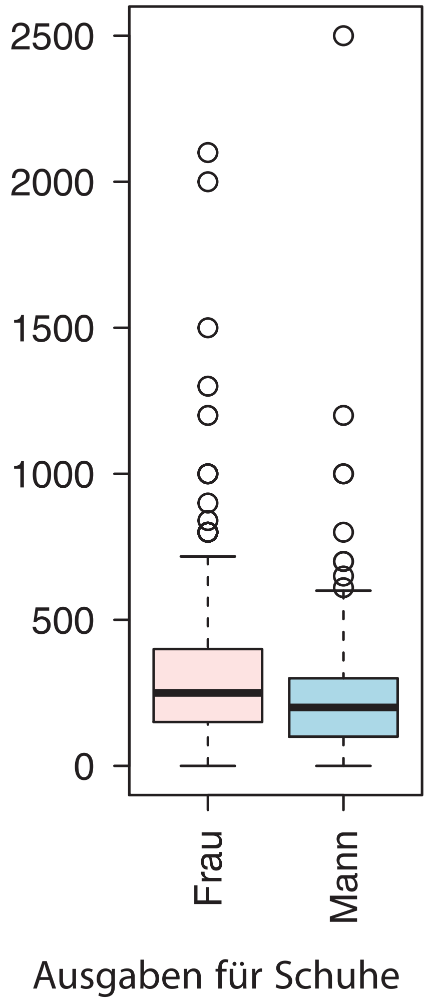

```
LageStreuung = function(x) {
  x=na.omit(x)  # ignoriere fehlende Werte
  n = length(x) # Anzahl nicht fehlender Werte
  popV = var(x)*(n-1)/n # var() ist nicht mittl. qu. Abweichung
  return(list(mean=mean(x),
              median=median(x),
              Variance=popV,
              StdDev=sqrt(popV),
              VarCoeff=sqrt(popV)/mean(x)))
}
mat1 = sapply(MyData[c("Alter","AlterV","AlterM", # sapply: pro Spalte anwenden
                       "Geschwister", "AnzSchuhe", "AusgSchuhe")],
              LageStreuung)
```

	Alter	AlterV	AlterM	Geschwister	AnzSchuhe	AusgSchuhe
mean	22.13	54.28	51.64	1.51	21.22	270.45
median	21.00	54.00	51.00	1.00	16.00	200.00
Variance	11.36	35.35	25.74	1.18	415.51	56333.39
StdDev	3.37	5.95	5.07	1.08	20.38	237.35
VarCoeff	0.15	0.11	0.10	0.72	0.96	0.88

Lage und Streuung als Grafik: Boxplot

- Graphische Darstellung von Lage und Streuung
- Box: Oberer/Unterer Rand: 3. bzw. 1. Quartil ($\tilde{x}_{0,75}$ bzw. $\tilde{x}_{0,25}$),
- Linie in Mitte: Median
- Whiskers: Länge: Max./Min Wert, aber beschränkt durch das 1,5-fache des Quartilsabstands (falls größter/kleinster Wert größeren/kleineren Abstand von Box: Länge Whiskers durch größten/kleinsten Wert innerhalb dieser Schranken)
- Ausreißer: Alle Objekte außerhalb der Whisker-Grenzen

```r
boxplot(AusgSchuhe ~ Geschlecht,
        col=c("mistyrose", "lightblue"),
        data=MyData, main="", las=2)
```



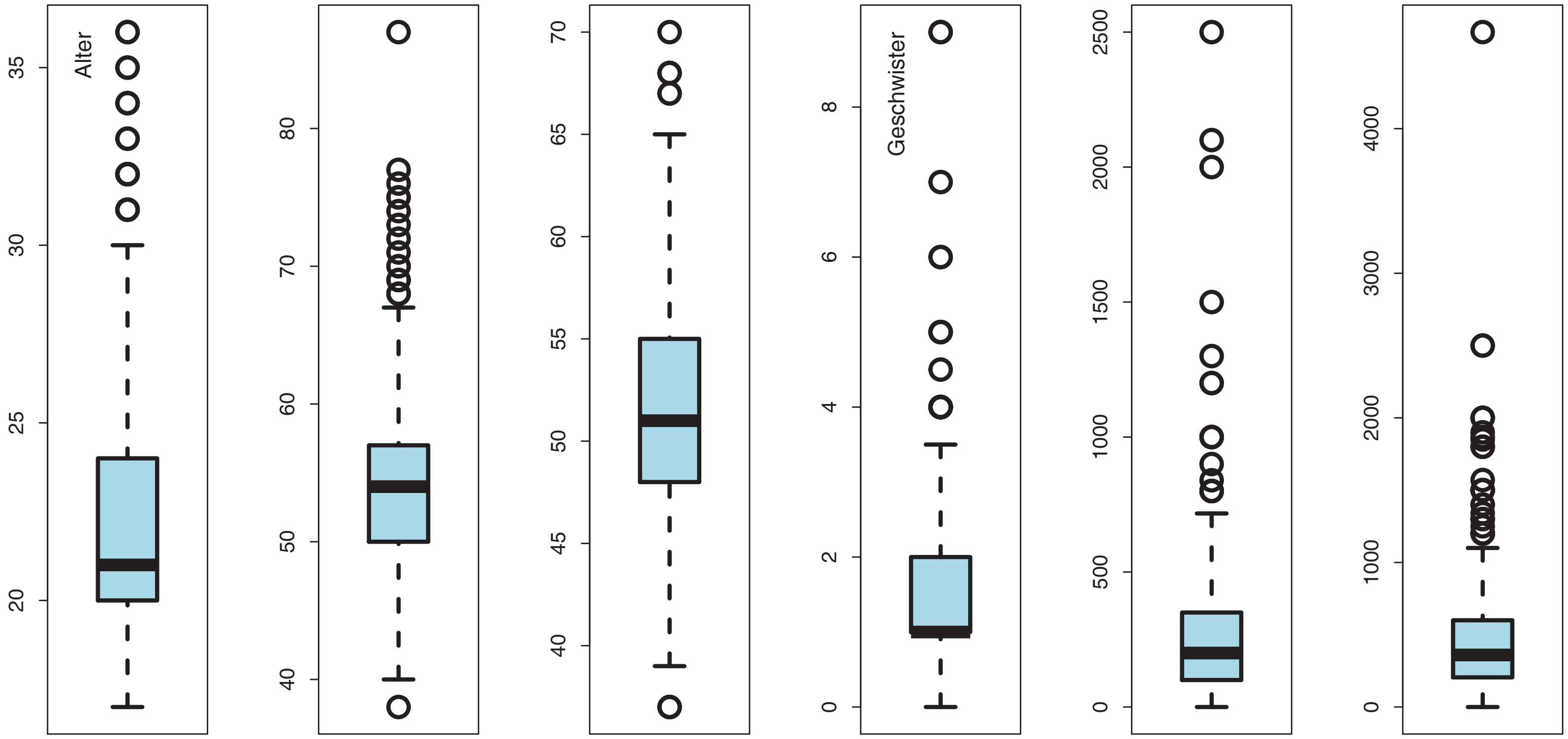
```r
summary(MyData)
```

```
##     Jahrgang        Alter          Groesse      Geschlecht     AlterV          AlterM
##  Min.   :2014   Min.   :17.00   Min.   :150.0   Frau:389   Min.   :38.00   Min.   :37.00
##  1st Qu.:2014   1st Qu.:20.00   1st Qu.:166.0   Mann:281   1st Qu.:50.00   1st Qu.:48.00
##  Median :2015   Median :21.00   Median :172.0              Median :54.00   Median :51.00
##  Mean   :2015   Mean   :22.13   Mean   :173.1              Mean   :54.28   Mean   :51.64
##  3rd Qu.:2016   3rd Qu.:24.00   3rd Qu.:180.0              3rd Qu.:57.00   3rd Qu.:55.00
##  Max.   :2016   Max.   :36.00   Max.   :198.0              Max.   :87.00   Max.   :70.00
##                                                            NA's   :1       NA's   :1
##     GroesseV        GroesseM      Geschwister        Farbe         AusgKomm        AnzSchuhe
##  Min.   :160.0   Min.   : 76.0   Min.   :0.000   blau   : 31   Min.   :   0.0   Min.   :  2.00
##  1st Qu.:175.0   1st Qu.:162.0   1st Qu.:1.000   gelb   :  5   1st Qu.: 207.5   1st Qu.:  8.00
##  Median :180.0   Median :165.0   Median :1.000   rot    : 24   Median : 360.0   Median : 16.00
##  Mean   :179.1   Mean   :166.2   Mean   :1.509   schwarz:333   Mean   : 458.1   Mean   : 21.22
##  3rd Qu.:183.0   3rd Qu.:170.0   3rd Qu.:2.000   silber : 82   3rd Qu.: 600.0   3rd Qu.: 30.00
##  Max.   :204.0   Max.   :192.0   Max.   :9.000   weiss  :195   Max.   :4668.0   Max.   :275.00
##  NA's   :11      NA's   :8                                     NA's   :2
##    AusgSchuhe         Essgewohnheiten Raucher      NoteMathe              MatheZufr   Studiengang
##  Min.   :   0.0   carnivor    :420    ja  : 81   Min.   :1.000   unzufrieden   :185   BW  :107
##  1st Qu.: 100.0   fruktarisch :  1    nein:381   1st Qu.:2.650   geht so       :151   ET  :  1
##  Median : 200.0   pescetarisch: 26    NA's:208   Median :3.300   zufrieden     :114   IM  : 74
##  Mean   : 270.5   vegan       :  3               Mean   :3.233   sehr zufrieden: 74   Inf : 48
##  3rd Qu.: 350.0   vegetarisch : 15               3rd Qu.:4.000   NA's          :146   WI  : 59
##  Max.   :2500.0   NA's        :205               Max.   :5.000                        NA's:381
##  NA's   :1                                       NA's   :162
```


1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Boxplots

```
for(attribute in c("Alter", "AlterV", "AlterM", "Geschwister",
                   "AusgSchuhe", "AusgKomm")) {
  data=MyData[, attribute]
  boxplot(data, # all rows, column of attribute
          col="lightblue",     # fill color
          lwd=3,               # line width
          cex=2,                # character size
          oma=c(1,1,2,1)
          )
  text(0.7,max(data), attribute, srt=90, adj=1)
}
```



1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Streuungsparameter (2)

- **Mittlere quadratische Abweichung** im Beispiel:

a) $s^2 = \frac{1}{3} \cdot (50^2 + 0^2 + 50^2)$

$= \frac{1}{3} \cdot (1950^2 + 2000^2 + 2050^2) - 2000^2 = 1666{,}67$

b) $s^2 = \frac{1}{3} \cdot (2000^2 + 2000^2 + 4000^2)$

$= \frac{1}{3} \cdot (0^2 + 0^2 + 6000^2) - 2000^2 = 8000000$

- **Standardabweichung:** $s = \sqrt{s^2}$

Im Beispiel:

a) $s = \sqrt{1666{,}67} = 40{,}82$

b) $s = \sqrt{8000000} = 2828{,}43$

- **Variationskoeffizient:** $V = \frac{s}{\bar{x}}$ (maßstabsunabhängig)

Im Beispiel:

a) $V = \frac{40{,}82}{2000} = 0{,}02$ ($\widehat{=}$ 2 %)

b) $V = \frac{2828{,}43}{2000} = 1{,}41$ ($\widehat{=}$ 141 %)

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Lage und Streuung: Überblick

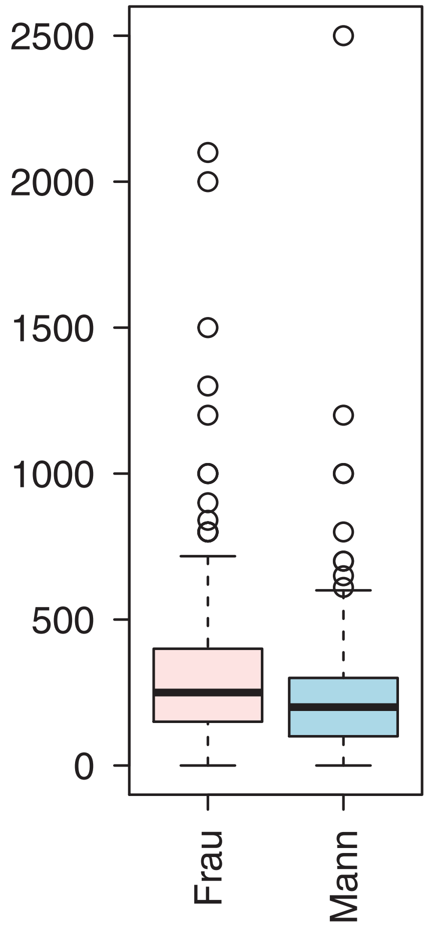
```r
LageStreuung = function(x) {
  x=na.omit(x)  # ignoriere fehlende Werte
  n = length(x) # Anzahl nicht fehlender Werte
  popV = var(x)*(n-1)/n # var() ist nicht mittl. qu. Abweichung
  return(list(mean=mean(x),
              median=median(x),
              Variance=popV,
              StdDev=sqrt(popV),
              VarCoeff=sqrt(popV)/mean(x)))
}
mat1 = sapply(MyData[c("Alter","AlterV","AlterM", # sapply: pro Spalte anwenden
                       "Geschwister", "AnzSchuhe", "AusgSchuhe")],
              LageStreuung)
```

	Alter	AlterV	AlterM	Geschwister	AnzSchuhe	AusgSchuhe
mean	22.13	54.28	51.64	1.51	21.22	270.45
median	21.00	54.00	51.00	1.00	16.00	200.00
Variance	11.36	35.35	25.74	1.18	415.51	56333.39
StdDev	3.37	5.95	5.07	1.08	20.38	237.35
VarCoeff	0.15	0.11	0.10	0.72	0.96	0.88

Lage und Streuung als Grafik: Boxplot

- Graphische Darstellung von Lage und Streuung
- Box: Oberer/Unterer Rand: 3. bzw. 1. Quartil ($\tilde{x}_{0,75}$ bzw. $\tilde{x}_{0,25}$),
- Linie in Mitte: Median
- Whiskers: Länge: Max./Min Wert, aber beschränkt durch das 1,5-fache des Quartilsabstands (falls größter/kleinster Wert größeren/kleineren Abstand von Box: Länge Whiskers durch größten/kleinsten Wert innerhalb dieser Schranken)
- Ausreißer: Alle Objekte außerhalb der Whisker-Grenzen

```
boxplot(AusgSchuhe ~ Geschlecht,
        col=c("mistyrose", "lightblue"),
        data=MyData, main="", las=2)
```



Ausgaben für Schuhe

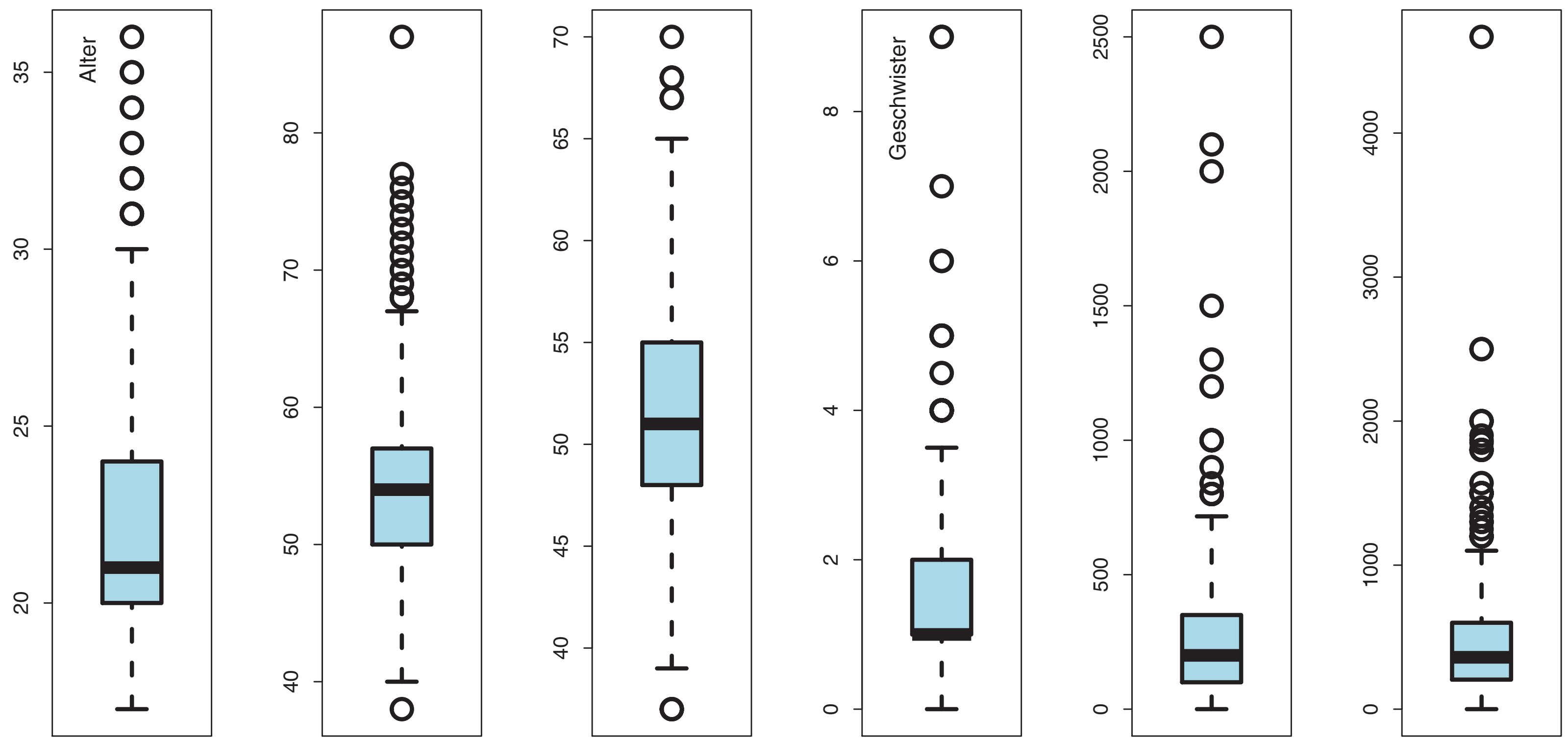
1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

```r
summary(MyData)
```

```
##     Jahrgang        Alter          Groesse      Geschlecht     AlterV          AlterM     
##  Min.   :2014   Min.   :17.00   Min.   :150.0   Frau:389   Min.   :38.00   Min.   :37.00  
##  1st Qu.:2014   1st Qu.:20.00   1st Qu.:166.0   Mann:281   1st Qu.:50.00   1st Qu.:48.00  
##  Median :2015   Median :21.00   Median :172.0              Median :54.00   Median :51.00  
##  Mean   :2015   Mean   :22.13   Mean   :173.1              Mean   :54.28   Mean   :51.64  
##  3rd Qu.:2016   3rd Qu.:24.00   3rd Qu.:180.0              3rd Qu.:57.00   3rd Qu.:55.00  
##  Max.   :2016   Max.   :36.00   Max.   :198.0              Max.   :87.00   Max.   :70.00  
##                                                            NA's   :1       NA's   :1      
##     GroesseV        GroesseM      Geschwister        Farbe        AusgKomm        AnzSchuhe     
##  Min.   :160.0   Min.   : 76.0   Min.   :0.000   blau   : 31   Min.   :   0.0   Min.   :  2.00  
##  1st Qu.:175.0   1st Qu.:162.0   1st Qu.:1.000   gelb   :  5   1st Qu.: 207.5   1st Qu.:  8.00  
##  Median :180.0   Median :165.0   Median :1.000   rot    : 24   Median : 360.0   Median : 16.00  
##  Mean   :179.1   Mean   :166.2   Mean   :1.509   schwarz:333   Mean   : 458.1   Mean   : 21.22  
##  3rd Qu.:183.0   3rd Qu.:170.0   3rd Qu.:2.000   silber : 82   3rd Qu.: 600.0   3rd Qu.: 30.00  
##  Max.   :204.0   Max.   :192.0   Max.   :9.000   weiss  :195   Max.   :4668.0   Max.   :275.00  
##  NA's   :11      NA's   :8                                     NA's   :2                        
##    AusgSchuhe         Essgewohnheiten Raucher      NoteMathe              MatheZufr   Studiengang
##  Min.   :   0.0   carnivor    :420    ja  : 81   Min.   :1.000   unzufrieden   :185   BW  :107   
##  1st Qu.: 100.0   fruktarisch :  1    nein:381   1st Qu.:2.650   geht so       :151   ET  :  1   
##  Median : 200.0   pescetarisch: 26    NA's:208   Median :3.300   zufrieden     :114   IM  : 74   
##  Mean   : 270.5   vegan       :  3               Mean   :3.233   sehr zufrieden: 74   Inf : 48   
##  3rd Qu.: 350.0   vegetarisch : 15               3rd Qu.:4.000   NA's          :146   WI  : 59   
##  Max.   :2500.0   NA's        :205               Max.   :5.000                        NA's:381   
##  NA's   :1                                       NA's   :162                                     
```

Boxplots

```
for(attribute in c("Alter", "AlterV", "AlterM", "Geschwister",
                   "AusgSchuhe", "AusgKomm")) {
  data=MyData[, attribute]
  boxplot(data, # all rows, column of attribute
          col="lightblue",     # fill color
          lwd=3,               # line width
          cex=2,                # character size
          oma=c(1,1,2,1)
          )
  text(0.7,max(data), attribute, srt=90, adj=1)
}
```



1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Konzentrationsmaße

- Gegeben: kardinale Werte $0 \leqq x_1 \leqq x_2 \leqq \cdots \leqq x_n$
- **Achtung!** Die Werte <u>müssen</u> aufsteigend sortiert werden!
- Lorenzkurve:

> Wieviel Prozent der Merkmalssumme entfällt auf die x Prozent kleinsten Merkmalsträger?

- **Beispiel:** Die 90 % ärmsten besitzen 20 % des Gesamtvermögens.
- Streckenzug: $(0{,}0),\ (u_1, v_1),\ \ldots,\ (u_n, v_n),\quad (1{,}1)$ mit

$$v_k = \text{Anteil der } k \text{ kleinsten MM-Träger an der MM-Summe} = \frac{\sum_{i=1}^{k} x_i}{\sum_{i=1}^{n} x_i}$$

$$u_k = \text{Anteil der } k \text{ kleinsten an der Gesamtzahl der MM-Träger} = \frac{k}{n}$$

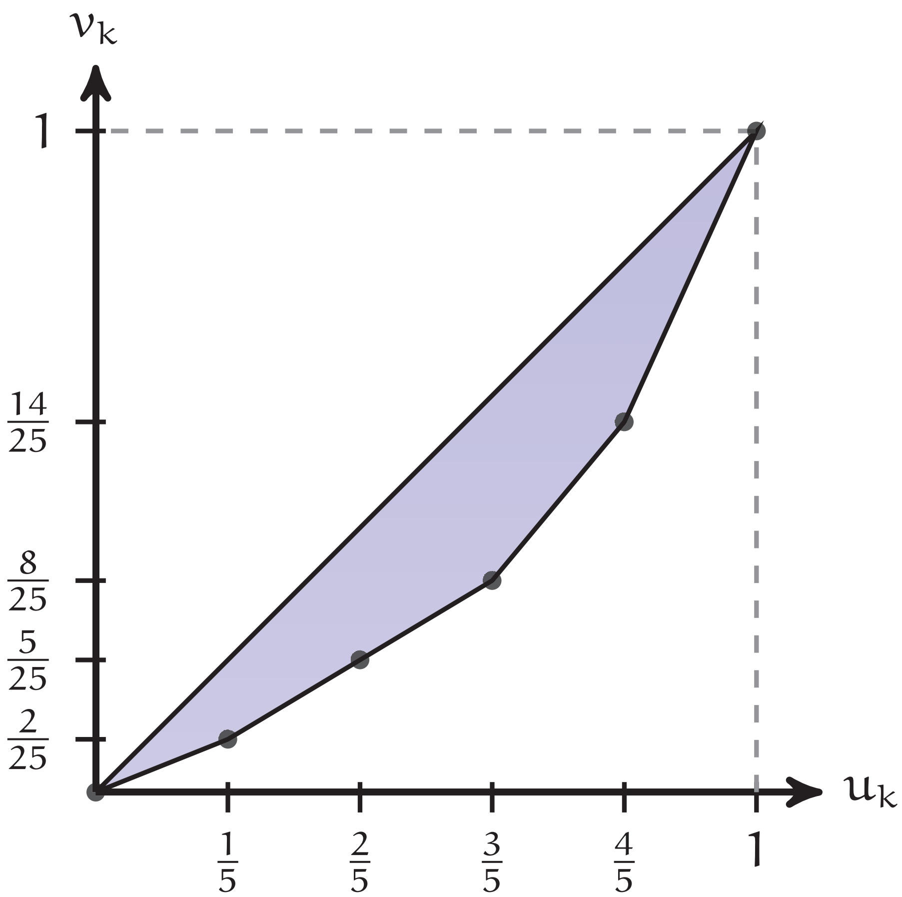
1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Lorenzkurve: Beispiel

Markt mit fünf Unternehmen; Umsätze: 6, 3, 11, 2, 3 (Mio. €)

$$\Rightarrow n = 5, \ \sum_{k=1}^{5} x_k = 25$$

k	1	2	3	4	5
x_k	2	3	3	6	11
p_k	$\frac{2}{25}$	$\frac{3}{25}$	$\frac{3}{25}$	$\frac{6}{25}$	$\frac{11}{25}$
v_k	$\frac{2}{25}$	$\frac{5}{25}$	$\frac{8}{25}$	$\frac{14}{25}$	1
u_k	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{4}{5}$	1



1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

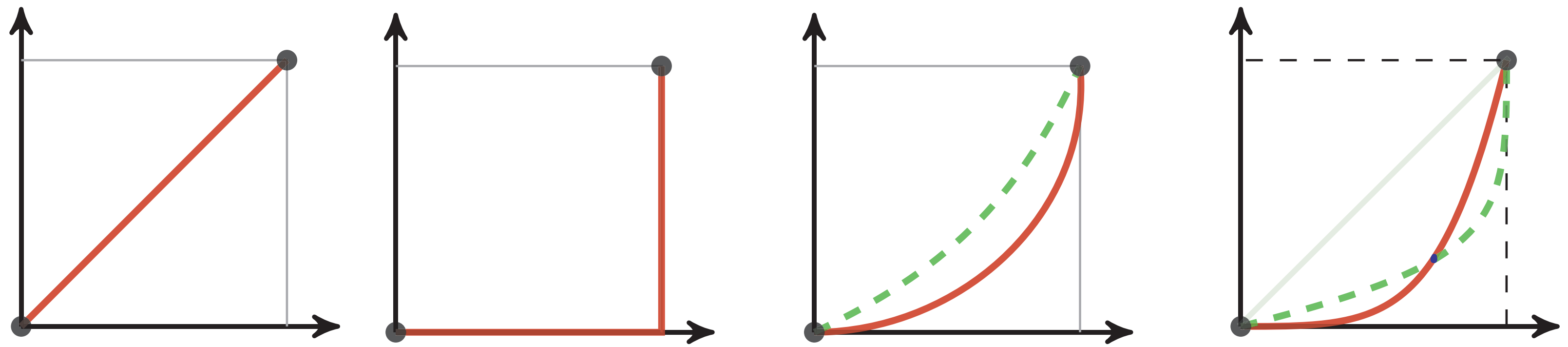
Lorenzkurve

Knickstellen:

- Bei i-tem Merkmalsträger $\iff x_{i+1} > x_i$
- Empirische Verteilungsfunktion liefert Knickstellen:

a_j	2	3	6	11
$h(a_j)$	1	2	1	1
$f(a_j)$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{1}{5}$
$F(a_j)$	$\frac{1}{5}$	$\frac{3}{5}$	$\frac{4}{5}$	1

Vergleich von Lorenzkurven:



Gleichverteilung

extreme Konzentration

stärker konzentriert als

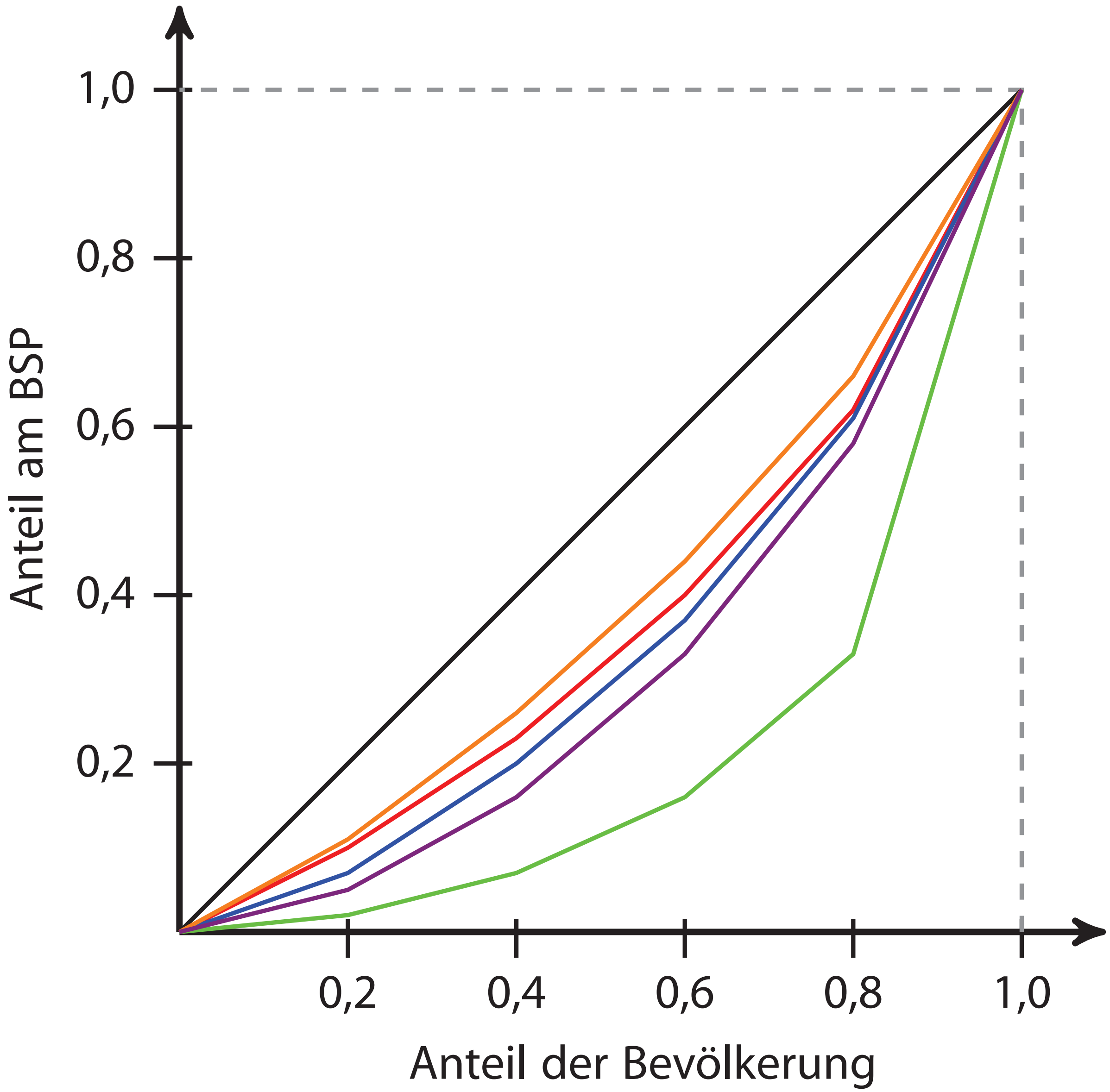
schwer vergleichbar

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Lorenzkurve: Beispiel Bevölkerungsanteil gegen BSP

Bangladesch
Brasilien
Deutschland
Ungarn
USA

(Stand 2000)

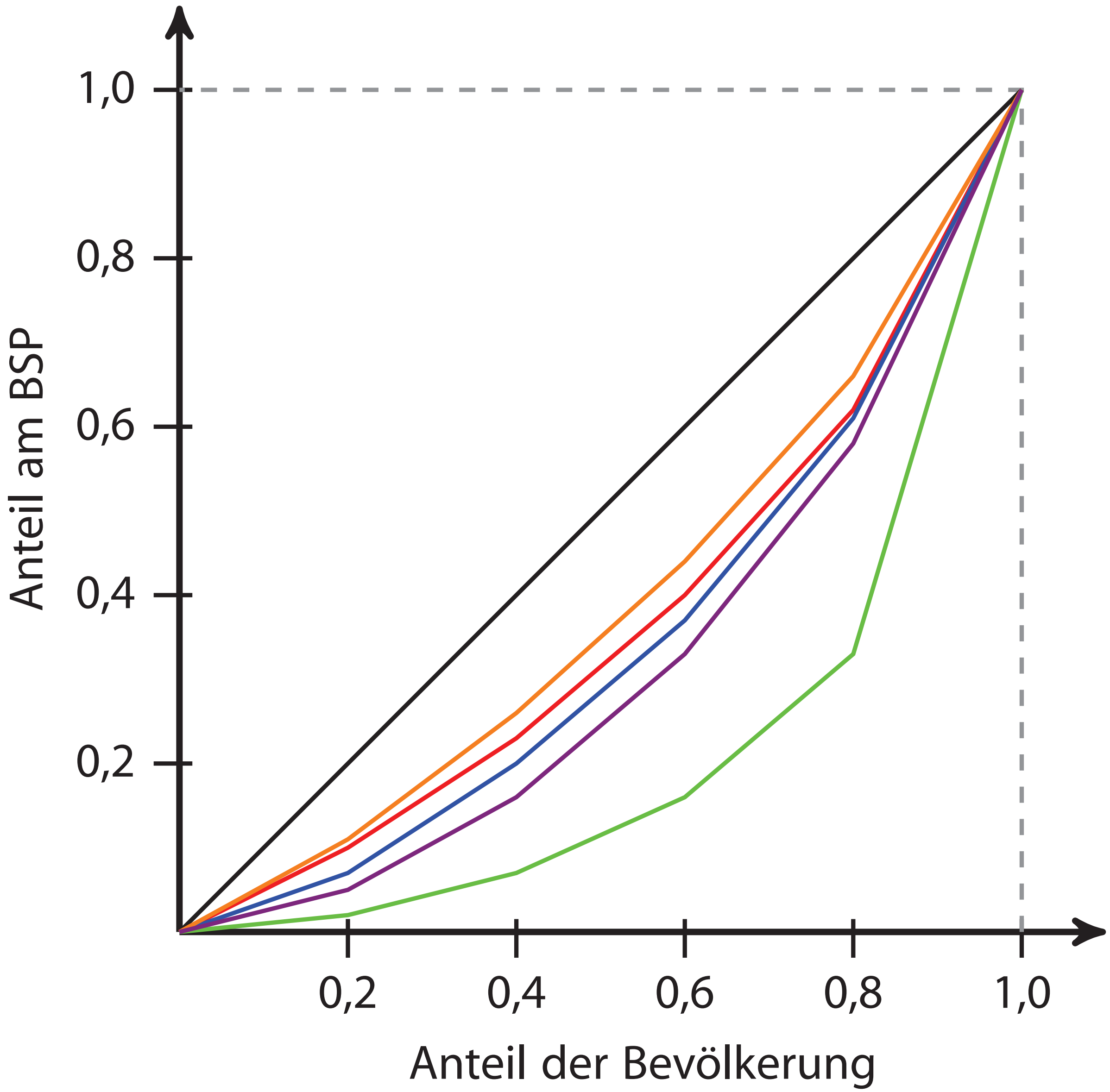


1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Lorenzkurve: Beispiel Bevölkerungsanteil gegen BSP

Bangladesch
Brasilien
Deutschland
Ungarn
USA

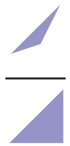
(Stand 2000)



Gini-Koeffizient

- Numerisches Maß der Konzentration: **Gini-Koeffizient** G

$$G = \frac{\text{Fläche zwischen } 45^\circ\text{-Linie und } L}{\text{Fläche unter } 45^\circ\text{-Linie}} =$$



- Aus den Daten:

$$G = \frac{2\sum_{i=1}^{n} i\,x_i - (n+1)\sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i} = \frac{2\sum_{i=1}^{n} i\,p_i - (n+1)}{n} \quad \text{wobei} \quad p_i = \frac{x_i}{\sum_{i=1}^{n} x_i}$$

- Problem: $G_{\text{max}} = \frac{n-1}{n}$

⇒ **Normierter Gini-Koeffizient:**

$$G_* = \frac{n}{n-1} \cdot G \in [0;1]$$

Statistik

1. Einführung
2. Deskriptive Statistik
   - Häufigkeiten
   - Lage und Streuung
   - Konzentration
   - Zwei Merkmale
   - Korrelation
   - Preisindizes
   - Lineare Regression
3. W-Theorie
4. Induktive Statistik

Quellen

Tabellen

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Beispiel:

i	1	2	3	4	$\sum$
x_i	1	2	2	15	20
p_i	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	$\frac{15}{20}$	1

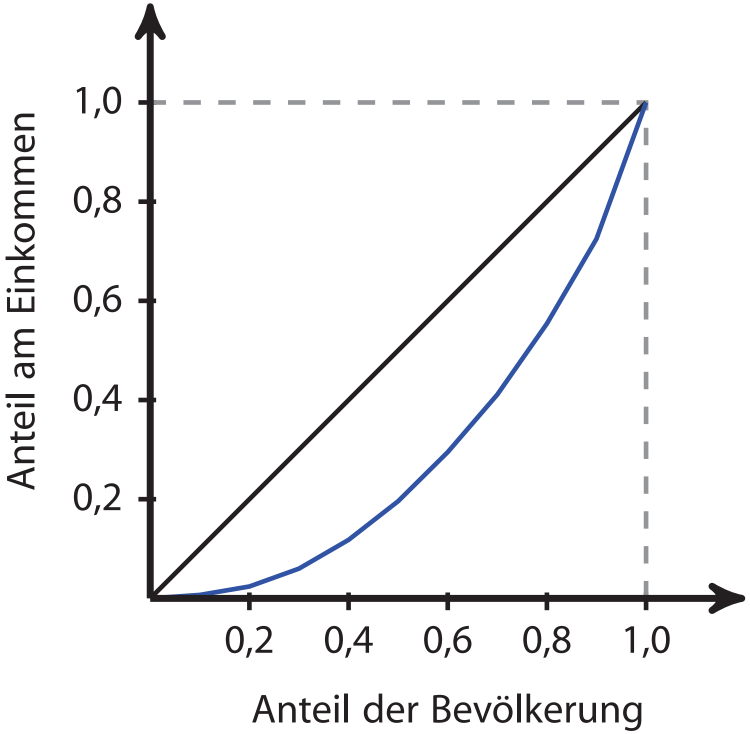
$$G = \frac{2 \cdot \left(1 \cdot \frac{1}{20} + 2 \cdot \frac{2}{20} + 3 \cdot \frac{2}{20} + 4 \cdot \frac{15}{20}\right) - (4+1)}{4} = 0{,}525$$

Mit $G_{\text{max}} = \frac{4-1}{4} = 0{,}75$ folgt

$$G_* = \frac{4}{4-1} \cdot 0{,}525 = 0{,}7$$

Armutsbericht der Bundesregierung 2008

- Verteilung der Bruttoeinkommen in Preisen von 2000
- aus unselbständiger Arbeit der Arbeitnehmer/-innen insgesamt

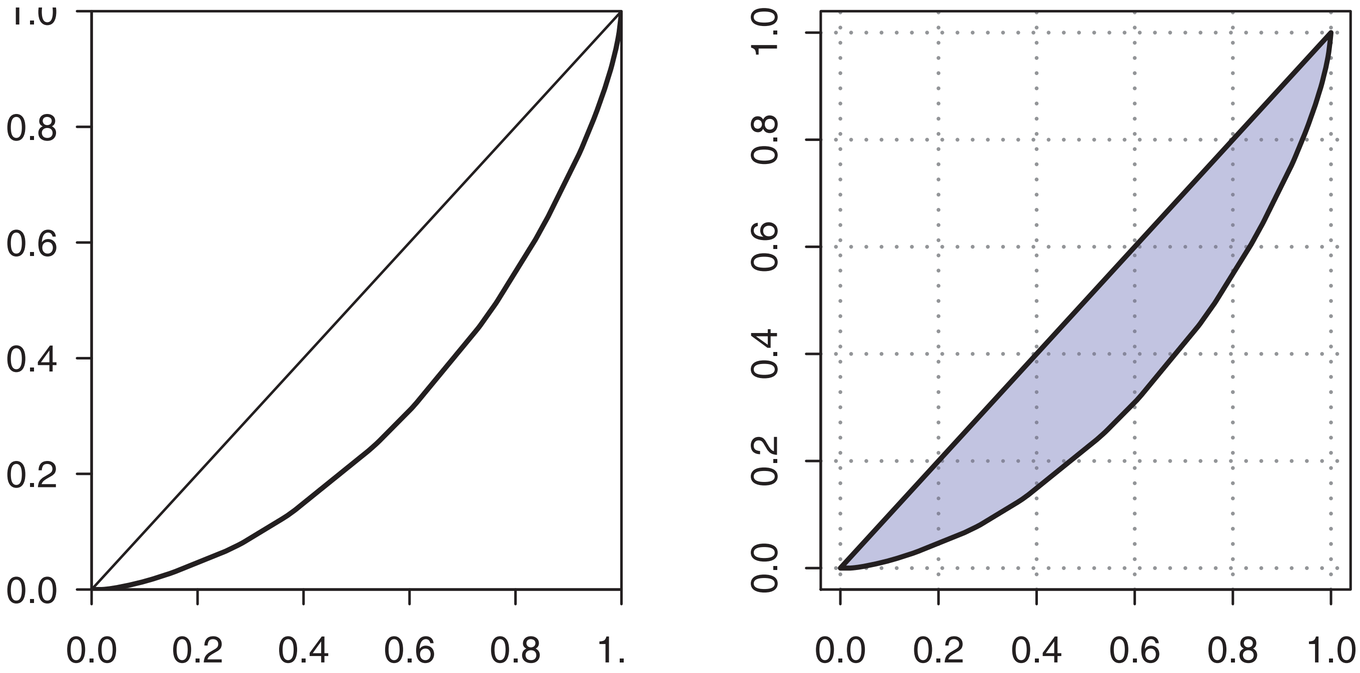


	2002	2003	2004	2005
Arithmetisches Mittel	24.873	24.563	23.987	23.648
Median	21.857	21.531	20.438	20.089
Gini-Koeffizient	0,433	0,441	0,448	0,453

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Lorenzkurve mit R

```r
require(ineq) # inequality Paket
Lorenz = Lc(na.exclude(MyData$AusgSchuhe))
   plot(Lorenz, xlab="", ylab="", main="") # Standard plot
   plot(c(0,1), c(0,1), type="n", # bisschen netter
       panel.first=grid(lwd=1.5, col=rgb(0,0,0,1/2)),
       xlab="", main="", ylab="")
   polygon(Lorenz$p, Lorenz$L, density=-1, col=rgb(0,0,1,1/4), lwd=2)
```




```r
Gini(na.exclude(AusgSchuhe)) # Gini-Koeffizient

## [1] 0.4069336
```


1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Weitere Konzentrationsmaße

- **Konzentrationskoeffizient:**

$$CR_g = \text{Anteil, der auf die } g \text{ größten entfällt} = \sum_{i=n-g+1}^{n} p_i = 1 - v_{n-g}$$

- **Herfindahl-Index:**

$$H = \sum_{i=1}^{n} p_i^2 \qquad (\in [\tfrac{1}{n}; 1])$$

Es gilt: $H = \frac{1}{n}(V^2 + 1)$ bzw. $V = \sqrt{n \cdot H - 1}$

- **Exponentialindex:**

$$E = \prod_{i=1}^{n} p_i^{p_i} \qquad (\in [\tfrac{1}{n}; 1]) \qquad \text{wobei} \qquad 0^0 = 1$$

- Im Beispiel mit $x = (1, 2, 2, 15)$:

$$CR_2 = \frac{17}{20} = 0{,}85$$

$$H = \left(\frac{1}{20}\right)^2 + \cdots + \left(\frac{15}{20}\right)^2 = 0{,}59$$

$$E = \left(\frac{1}{20}\right)^{\frac{1}{20}} \cdots \left(\frac{15}{20}\right)^{\frac{15}{20}} = 0{,}44$$

Weitere Konzentrationsmaße

- **Konzentrationskoeffizient:**

$$CR_g = \text{Anteil, der auf die } g \text{ größten entfällt} = \sum_{i=n-g+1}^{n} p_i = 1 - v_{n-g}$$

- **Herfindahl-Index:**

$$H = \sum_{i=1}^{n} p_i^2 \qquad (\in [\tfrac{1}{n}; 1])$$

Es gilt: $H = \frac{1}{n}(V^2 + 1)$ bzw. $V = \sqrt{n \cdot H - 1}$

- **Exponentialindex:**

$$E = \prod_{i=1}^{n} p_i^{p_i} \qquad (\in [\tfrac{1}{n}; 1]) \qquad \text{wobei} \qquad 0^0 = 1$$

- Im Beispiel mit $x = (1, 2, 2, 15)$:

$$CR_2 = \frac{17}{20} = 0{,}85$$

$$H = \left(\frac{1}{20}\right)^2 + \cdots + \left(\frac{15}{20}\right)^2 = 0{,}59$$

$$E = \left(\frac{1}{20}\right)^{\frac{1}{20}} \cdots \left(\frac{15}{20}\right)^{\frac{15}{20}} = 0{,}44$$

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Zweidimensionale Urliste

Urliste vom Umfang n zu **zwei** Merkmalen X und Y:

$$(x_1, y_1), \quad (x_2, y_2), \quad \ldots, \quad (x_n, y_n)$$

Kontingenztabelle:

Sinnvoll bei wenigen Ausprägungen bzw. bei klassierten Daten.

	Ausprägungen von Y			
Ausprägungen von X	b_1	b_2	$\ldots$	b_l
a_1	h_{11}	h_{12}	$\ldots$	h_{1l}
a_2	h_{21}	h_{22}	$\ldots$	h_{2l}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
a_k	h_{k1}	h_{k2}	$\ldots$	h_{kl}

Unterscheide:

- **Gemeinsame Häufigkeiten:**

$$h_{ij} = h(a_i, b_j)$$

- **Randhäufigkeiten:**

$$h_{i\cdot} = \sum_{j=1}^{l} h_{ij} \qquad \text{und} \qquad h_{\cdot j} = \sum_{i=1}^{k} h_{ij}$$

- **Bedingte (relative) Häufigkeiten:**

$$f_1(a_i \mid b_j) = \frac{h_{ij}}{h_{\cdot j}} \qquad \text{und} \qquad f_2(b_j \mid a_i) = \frac{h_{ij}}{h_{i\cdot}}$$

Statistik

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Beispiel: 400 unfallbeteiligte Autoinsassen:

	leicht verletzt ($= b_1$)	schwer verletzt ($= b_2$)	tot ($= b_3$)	
angegurtet ($= a_1$)	264 ($= h_{11}$)	90 ($= h_{12}$)	6 ($= h_{13}$)	360 ($= h_{1\cdot}$)
nicht angegurtet ($= a_2$)	2 ($= h_{21}$)	34 ($= h_{22}$)	4 ($= h_{23}$)	40 ($= h_{2\cdot}$)
	266 ($= h_{\cdot 1}$)	124 ($= h_{\cdot 2}$)	10 ($= h_{\cdot 3}$)	400 ($= n$)

$f_2(b_3 \mid a_2) = \frac{4}{40} = 0{,}1$ (10 % der nicht angegurteten starben.)

$f_1(a_2 \mid b_3) = \frac{4}{10} = 0{,}4$ (40 % der Todesopfer waren nicht angegurtet.)

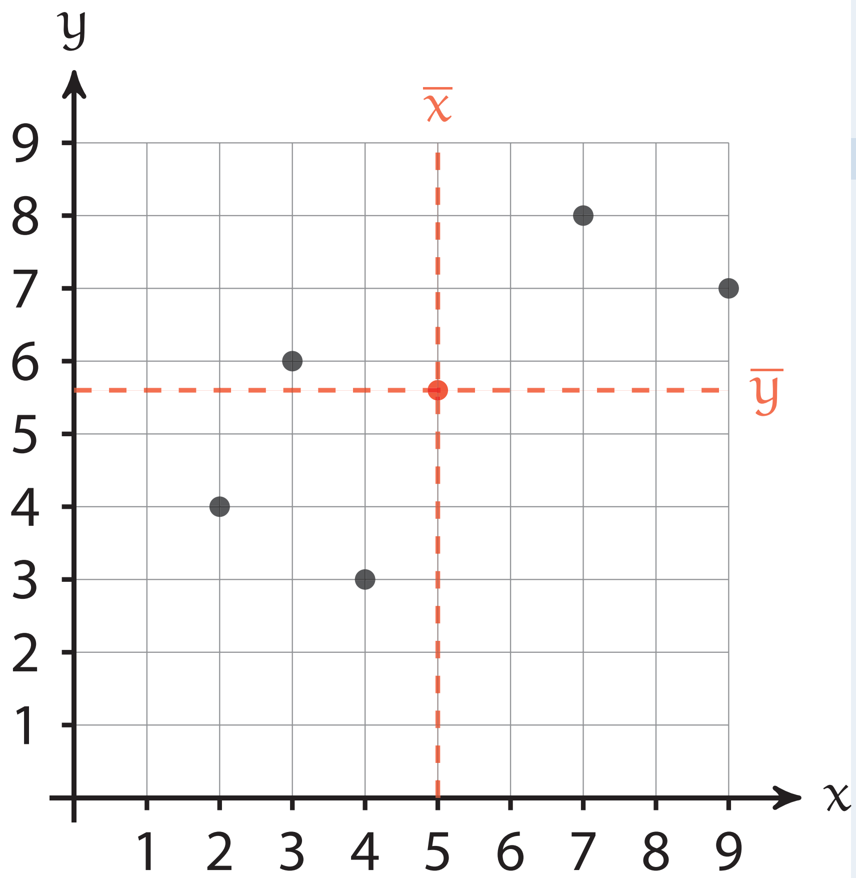
Streuungsdiagramm sinnvoll bei vielen verschiedenen Ausprägungen (z.B. stetige Merkmale)

➠ Alle (x_i, y_i) sowie $(\bar{x}, \bar{y})$ in Koordinatensystem eintragen.

Beispiel:

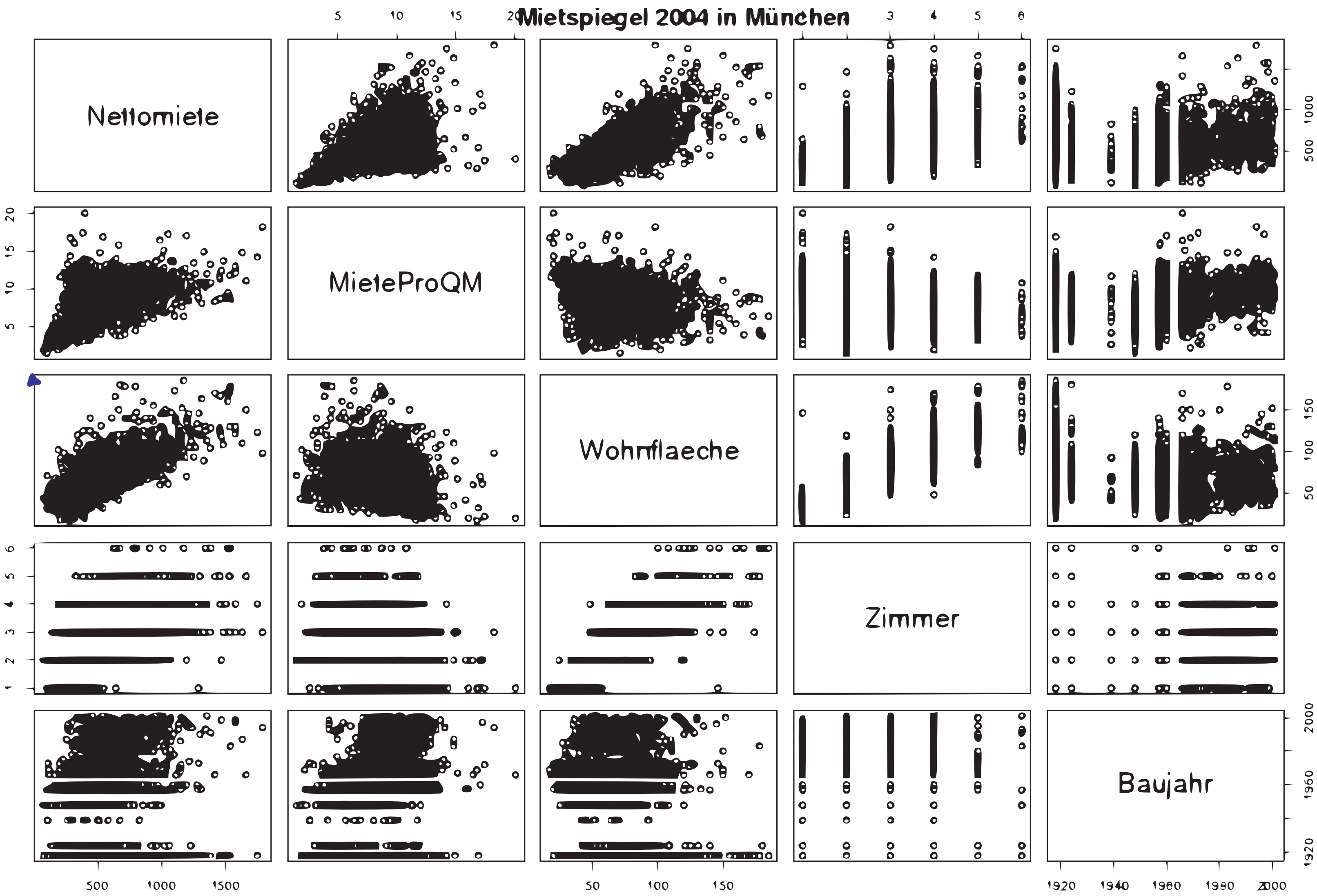
i	1	2	3	4	5	$\sum$
x_i	2	4	3	9	7	25
y_i	4	3	6	7	8	28

$$\Rightarrow \bar{x} = \frac{25}{5} = 5$$
$$\bar{y} = \frac{28}{5} = 5{,}6$$



1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Beispiel Streuungsdiagramm

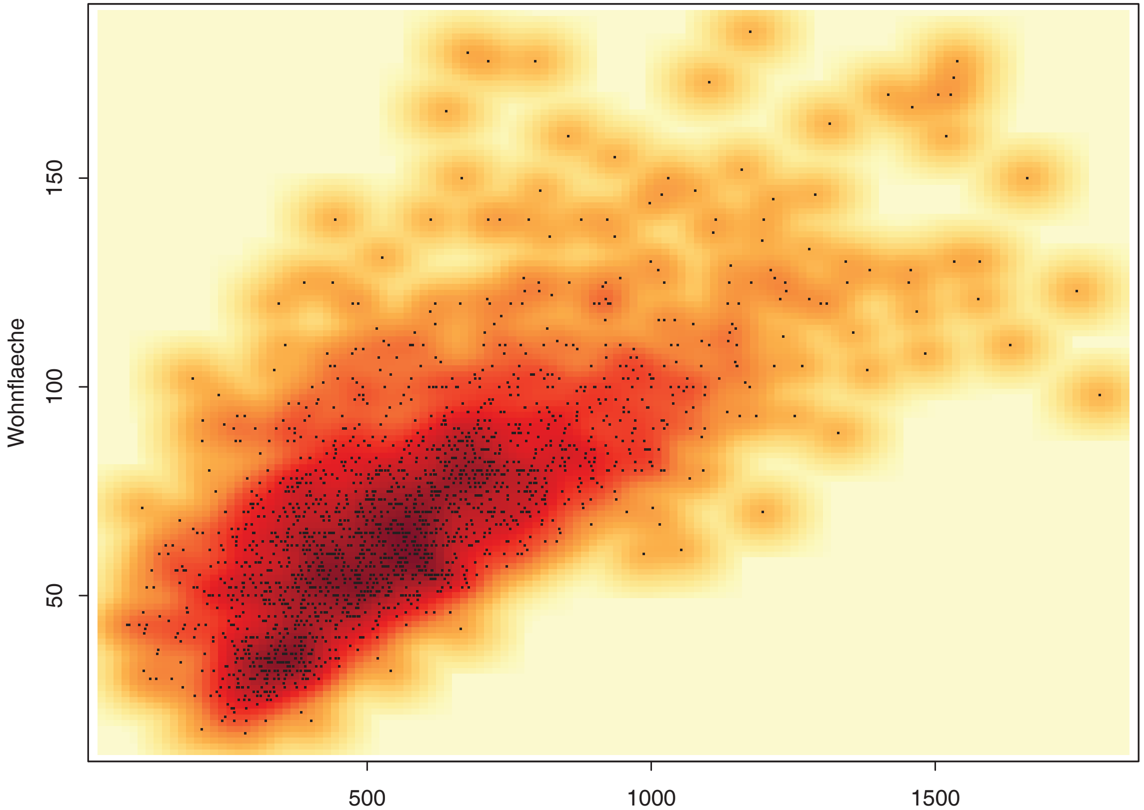


(Datenquelle: Fahrmeir u. a. (2009))

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

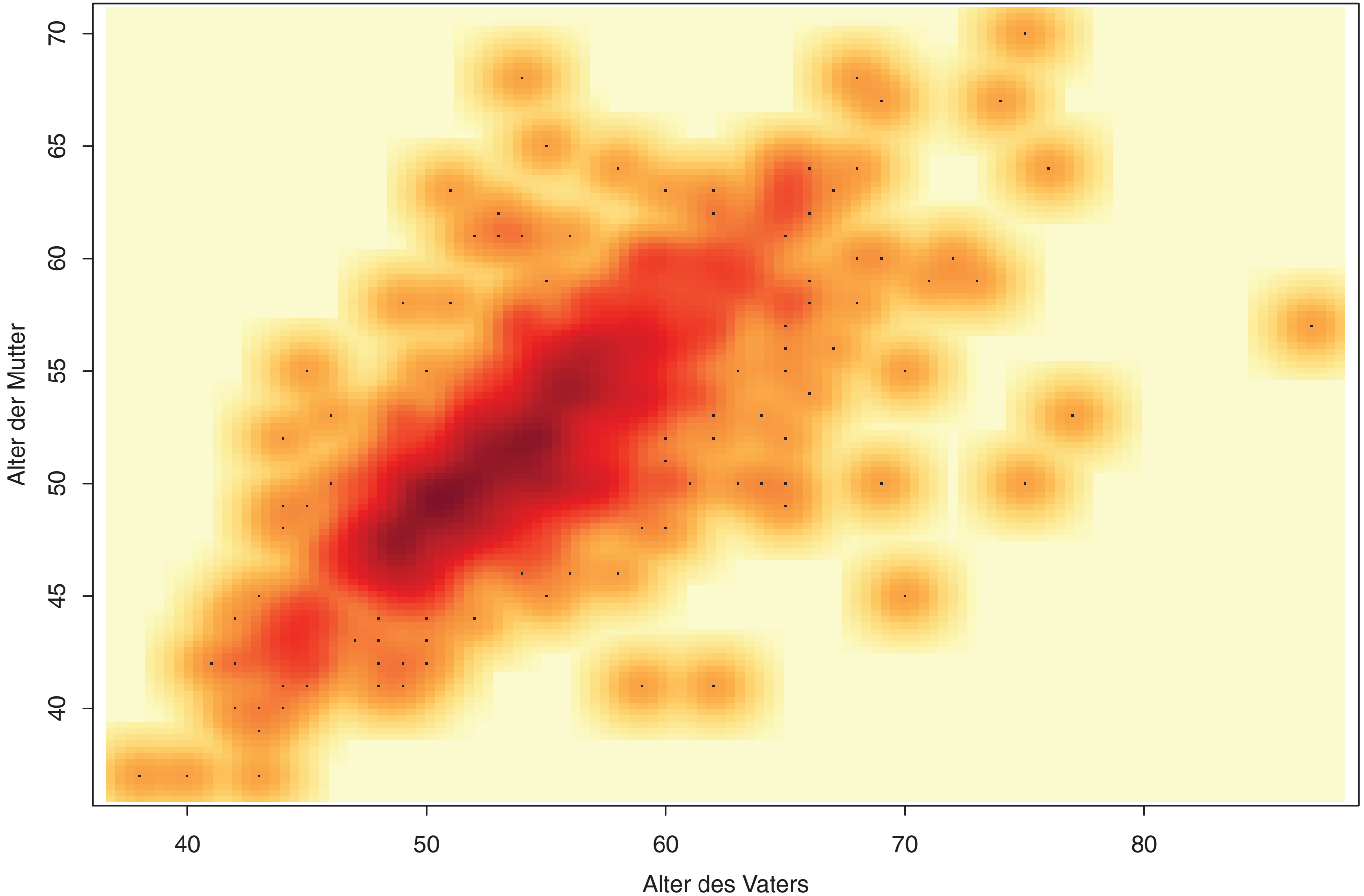
Beispiel Streuungsdiagramm

```
if (!require("RColorBrewer")) {
install.packages("RColorBrewer")
library(RColorBrewer)
}
mieten <- read.table('http://goo.gl/jhpJW4', header=TRUE, sep='\t',
                     check.names=TRUE, fill=TRUE, na.strings=c('',''))
x <- cbind(Nettomieten=mieten$nm, Wohnflaeche=mieten$wfl)

library("geneplotter")  ## from BioConductor
smoothScatter(x, nrpoints=Inf,
              colramp=colorRampPalette(brewer.pal(9,"YlOrRd")),
              bandwidth=c(30,3))
```



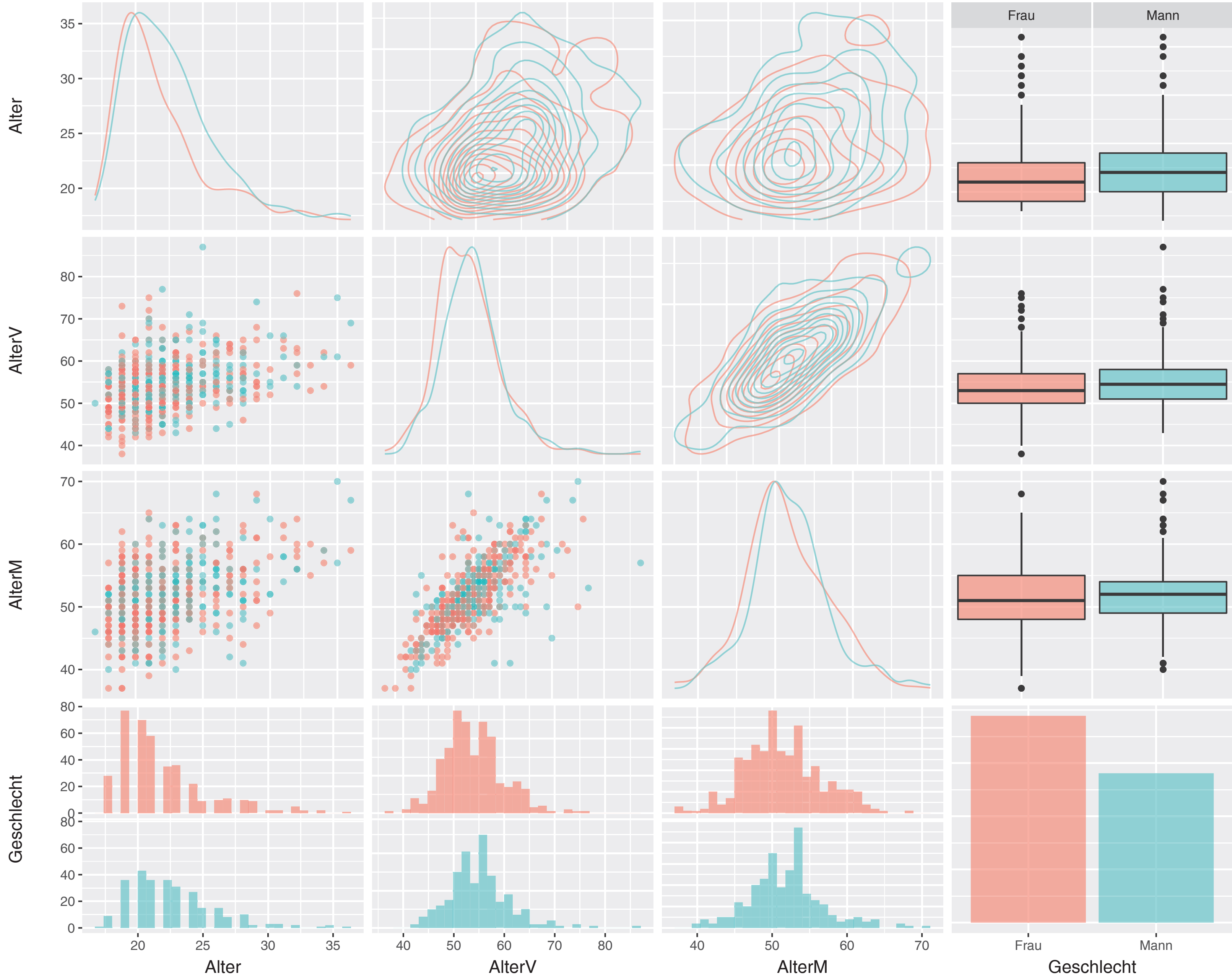
Beispiel Streuungsdiagramm

```
x = cbind("Alter des Vaters"=AlterV, "Alter der Mutter"=AlterM)
require("geneplotter") ## from BioConductor
smoothScatter(x, colramp=colorRampPalette(brewer.pal(9,"YlOrRd")) )
```



1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

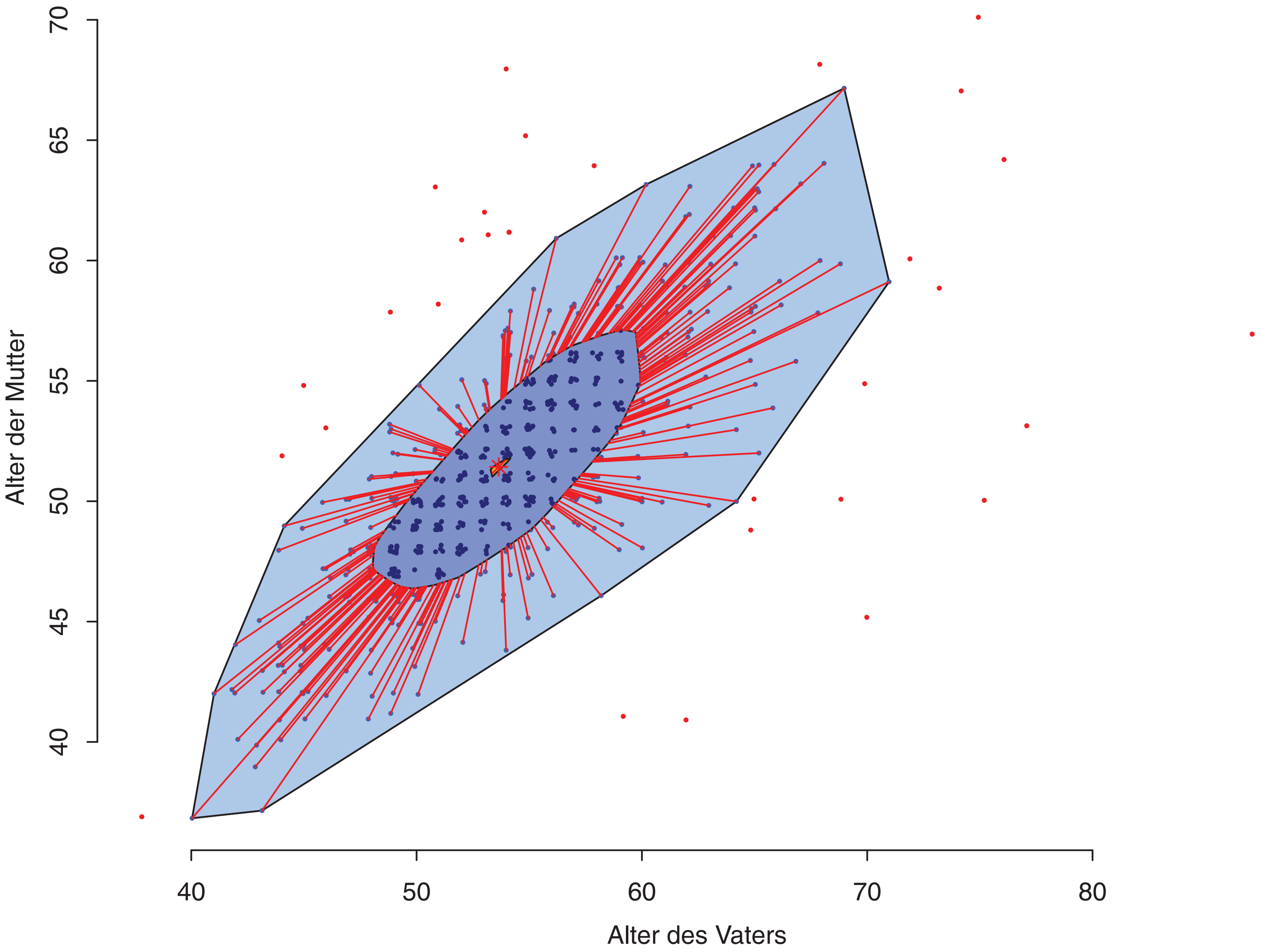
```
require(GGally)
ggpairs(MyData[, c("Alter", "AlterV", "AlterM", "Geschlecht")],
        upper = list(continuous = "density", combo = "box"),
       color='Geschlecht', alpha=0.5)
```



1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Bagplot: Boxplot in 2 Dimensionen

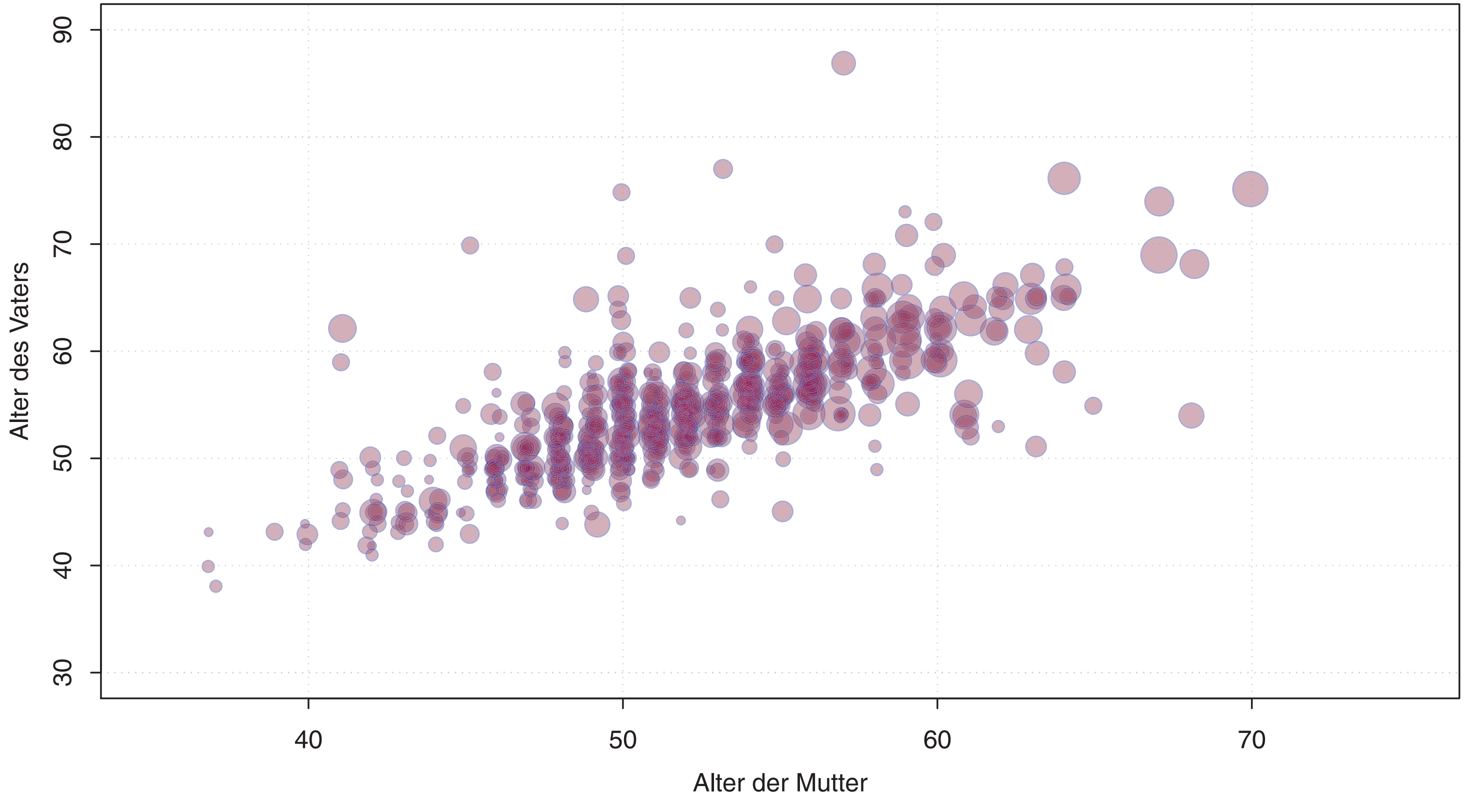
```
require(aplpack)
bagplot(jitter(AlterV), jitter(AlterM), xlab="Alter des Vaters", ylab="Alter der Mutter")

## [1] "Warning: NA elements have been exchanged by median values!!"
```



1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Bubbleplot: 3 metrische Variablen

```
require(DescTools)
My.ohne.NA = na.exclude(MyData[,c("AlterM", "AlterV", "Alter")])
with(My.ohne.NA, {
  Alter.skaliert = (Alter-min(Alter))/(max(Alter)-min(Alter))
  PlotBubble(jitter(AlterM), jitter(AlterV), Alter.skaliert,
             col=SetAlpha("deeppink4",0.3),
             border=SetAlpha("darkblue",0.3),
             xlab="Alter der Mutter", ylab="Alter des Vaters",
             panel.first=grid(),
             main="")
  })
```

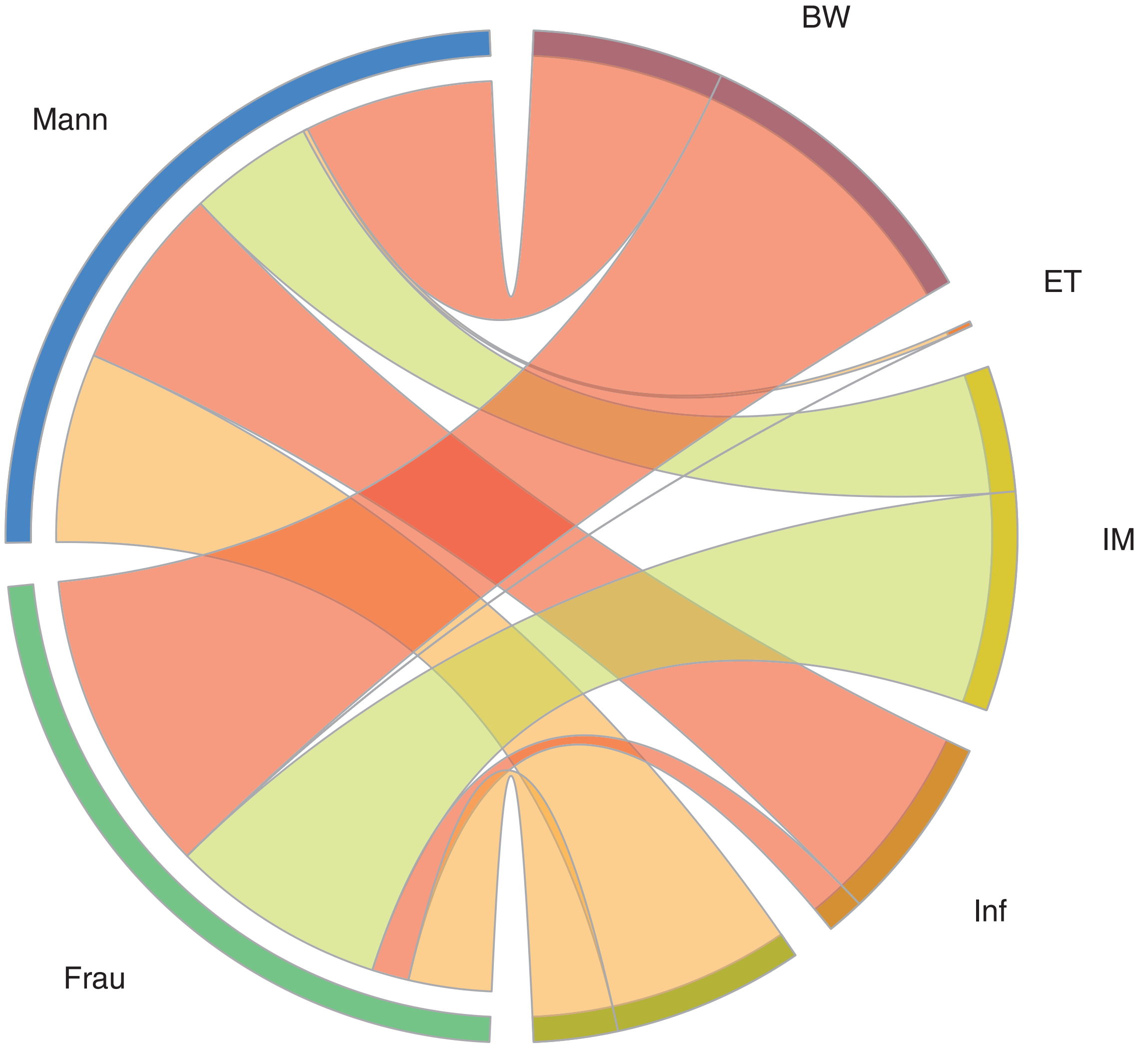


Größe der Blasen: Alter zwischen 0 (Jüngster) und 1 (Ältester)

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Circular Plots: Assoziationen

```
require(DescTools)
with(MyData, {
PlotCirc(table(Studiengang, Geschlecht),
acol=c("dodgerblue","seagreen2","limegreen","olivedrab2","goldenrod2","tomato2"),
rcol=SetAlpha(c("red","orange","olivedrab1"), 0.5)
)})
```



1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Korrelationsrechnung

- Frage: Wie stark ist der Zusammenhang zwischen X und Y?
- Dazu: Korrelationskoeffizienten
- Verschiedene Varianten: Wahl abhängig vom Skalenniveau von X und Y:

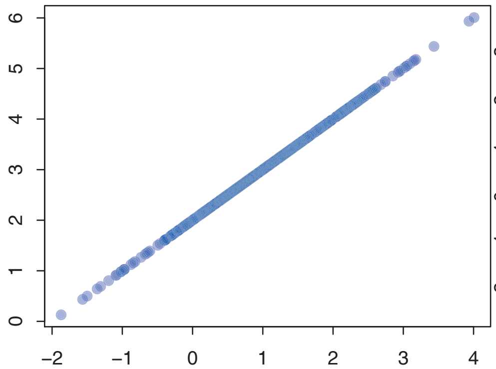
	Skalierung von Y		
Skalierung von X	kardinal	ordinal	nominal
kardinal	Bravais-Pearson-Korrelations-koeffizient	Rangkorrelations-koeffizient von Spearman	Kontingenz-koeffizient
ordinal	Rangkorrelations-koeffizient von Spearman	Rangkorrelations-koeffizient von Spearman	Kontingenz-koeffizient
nominal	Kontingenz-koeffizient	Kontingenz-koeffizient	Kontingenz-koeffizient

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

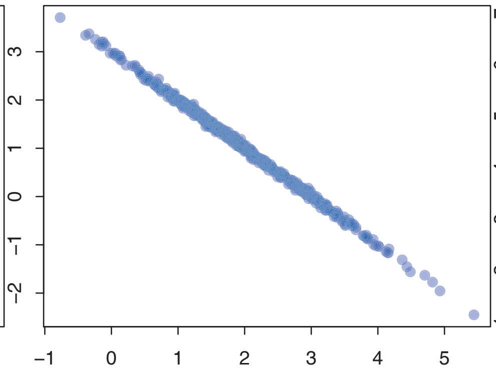
Bravais-Pearson-Korrelationskoeffizient

Voraussetzung: X, Y kardinalskaliert

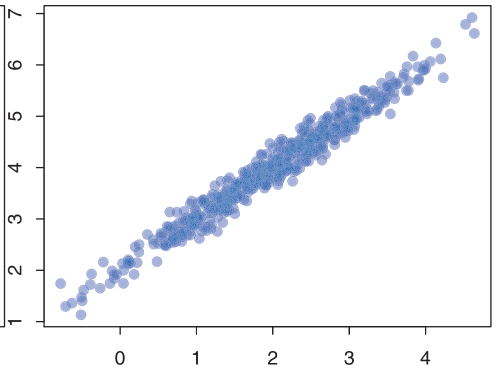
$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\cdot(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sqrt{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}\sqrt{\sum_{i=1}^{n} y_i^2 - n\bar{y}^2}} \in [-1;+1]$$



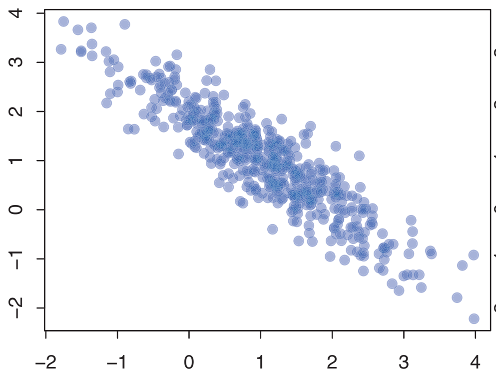
$r = 1$



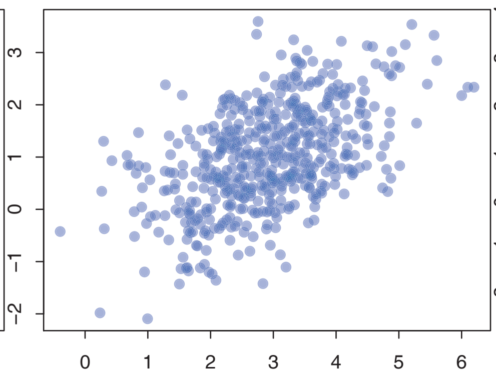
$r = -0{,}999$



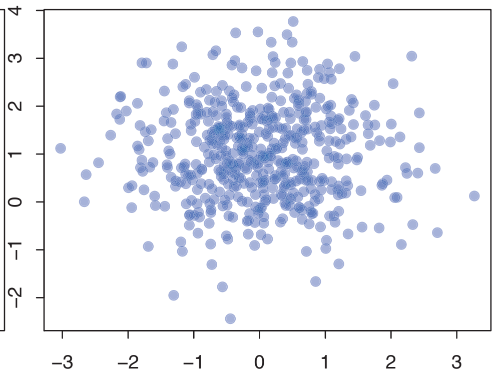
$r = 0{,}981$



$r = -0{,}903$



$r = 0{,}516$



$r = -0{,}014$

1. Einführung

2. Deskriptive Statistik
- Häufigkeiten
- Lage und Streuung
- Konzentration
- Zwei Merkmale
- Korrelation
- Preisindizes
- Lineare Regression

3. W-Theorie

4. Induktive Statistik

Quellen

Tabellen

Bravais-Pearson-Korrelationskoeffizient

Im Beispiel:

i	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	2	4	4	16	8
2	4	3	16	9	12
3	3	6	9	36	18
4	9	7	81	49	63
5	7	8	49	64	56
$\sum$	25	28	159	174	157

$$\Rightarrow \quad \bar{x} = 25/5 = 5, \qquad \bar{y} = 28/5 = 5{,}6$$

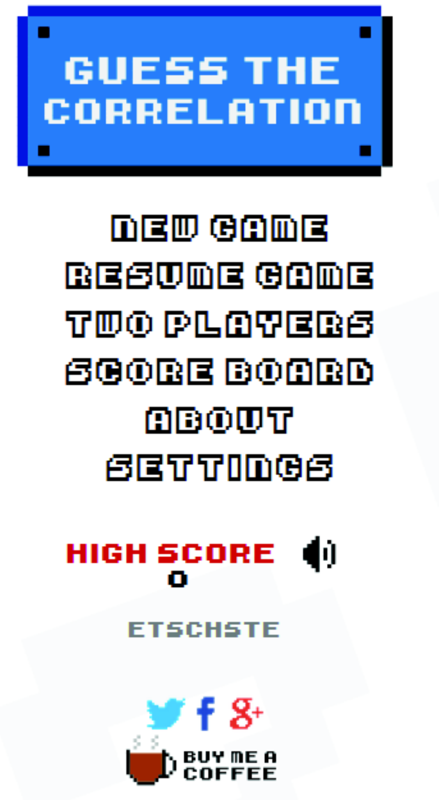
$$r = \frac{157 - 5 \cdot 5 \cdot 5{,}6}{\sqrt{159 - 5 \cdot 5^2}\sqrt{174 - 5 \cdot 5{,}6^2}} = 0{,}703$$

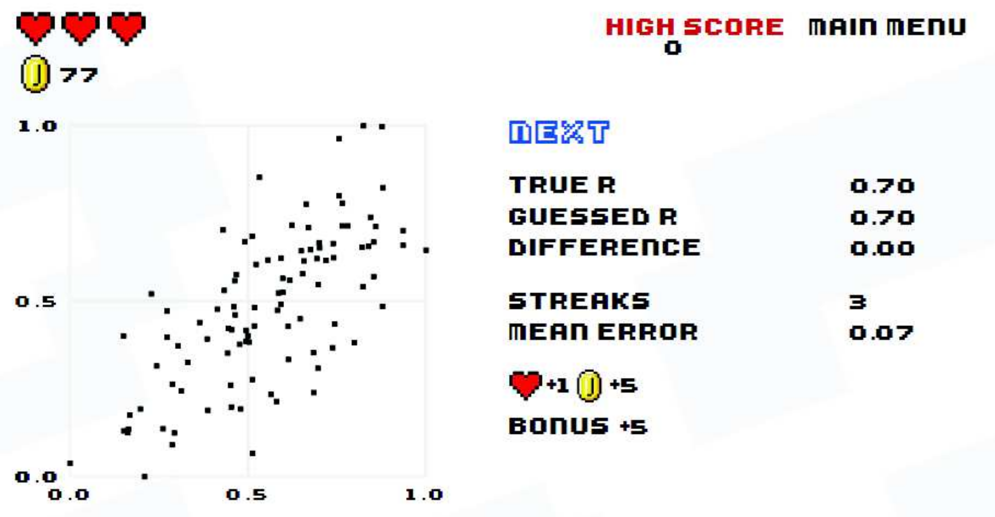
(deutliche positive Korrelation)

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Guess The Correlation

guessthecorrelation.com





Go for the Highscore!

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Rangkorrelationskoeffizient von Spearman

- Voraussetzungen: X, Y (mindestens) ordinalskaliert, Ränge eindeutig (keine Doppelbelegung von Rängen)
- Vorgehensweise:
  1. Rangnummern R_i (X) bzw. R'_i (Y) mit $R_i^{(\prime)} = 1$ bei größtem Wert usw.
  2. Berechne

$$r_{SP} = 1 - \frac{6 \sum_{i=1}^{n} (R_i - R'_i)^2}{(n-1)\,n\,(n+1)} \in [-1; +1]$$

- Hinweise:
  - $r_{SP} = +1$ wird erreicht bei $R_i = R'_i \quad \forall i = 1, \ldots, n$
  - $r_{SP} = -1$ wird erreicht bei $R_i = n + 1 - R'_i \quad \forall i = 1, \ldots, n$
- Falls Ränge nicht eindeutig: Bindungen, dann Berechnung von r_{SP} über Ränge und Formel des Korr.-Koeff. von Bravais-Pearson

Im Beispiel:

x_i	R_i	y_i	R'_i
2	5	4	4
4	3	3	5
3	4	6	3
9	1	7	2
7	2	8	1

$$r_{SP} = 1 - \frac{6 \cdot [(5-4)^2 + (3-5)^2 + (4-3)^2 + (1-2)^2 + (2-1)^2]}{(5-1) \cdot 5 \cdot (5+1)} = 0{,}6$$

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Kontingenzkoeffizient

- Gegeben: Kontingenztabelle mit k Zeilen und l Spalten (vgl. hier)
- Vorgehensweise:
  1. Ergänze Randhäufigkeiten

$$h_{i\cdot} = \sum_{j=1}^{l} h_{ij} \qquad \text{und} \qquad h_{\cdot j} = \sum_{i=1}^{k} h_{ij}$$

  2. Berechne theoretische Häufigkeiten

$$\tilde{h}_{ij} = \frac{h_{i\cdot} \cdot h_{\cdot j}}{n}$$

  3. Berechne

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(h_{ij} - \tilde{h}_{ij})^2}{\tilde{h}_{ij}}$$

χ^2 hängt von n ab! $(h_{ij} \mapsto 2 \cdot h_{ij} \Rightarrow \chi^2 \mapsto 2 \cdot \chi^2)$

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

④ **Kontingenzkoeffizient:**

$$K = \sqrt{\frac{\chi^2}{n + \chi^2}} \quad \in [0; K_{max}]$$

wobei

$$K_{max} = \sqrt{\frac{M-1}{M}} \quad \text{mit} \quad M = \min\{k, l\}$$

⑤ **Normierter Kontingenzkoeffizient:**

$$K_* = \frac{K}{K_{max}} \quad \in [0; 1]$$

$K_* = +1 \iff$
bei Kenntnis von x_i kann y_i erschlossen werden u.u.

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Beispiel

X : Staatsangehörigkeit (d,a)
Y : Geschlecht (m,w)

h_{ij}	m	w	$h_{i\cdot}$
d	30	30	60
a	10	30	40
$h_{\cdot j}$	40	60	100

$\Rightarrow$

$\tilde{h}_{ij}$	m	w
d	24	36
a	16	24

wobei $\tilde{h}_{11} = \frac{60\cdot 40}{100} = 24$ usw.

$$\chi^2 = \frac{(30-24)^2}{24} + \frac{(30-36)^2}{36} + \frac{(10-16)^2}{16} + \frac{(30-24)^2}{24} = 6{,}25$$

$$K = \sqrt{\frac{6{,}25}{100+6{,}25}} = 0{,}2425; \qquad M = \min\{2{,}2\} = 2; \qquad K_{max} = \sqrt{\frac{2-1}{2}} = 0{,}7071$$

$$K_* = \frac{0{,}2425}{0{,}7071} = 0{,}3430$$

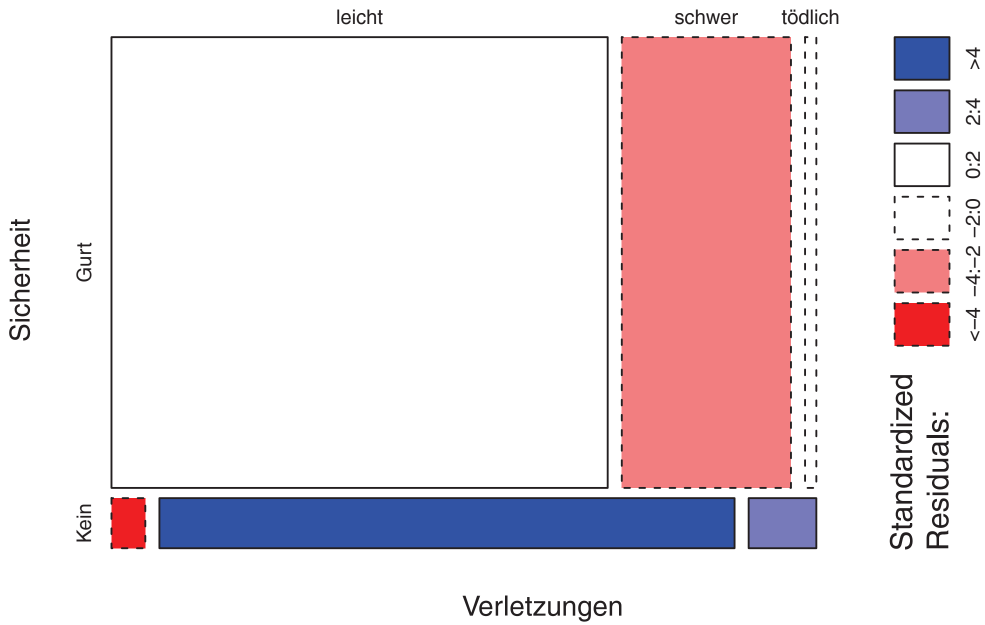
Statistik

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Beispiel Autounfälle

	Verletzung			
	leicht	schwer	tödlich	
angegurtet	264	90	6	360
nicht angegurtet	2	34	4	40
	266	124	10	400



Mosaikplot Autounfälle

1. Einführung

2. Deskriptive Statistik
- Häufigkeiten
- Lage und Streuung
- Konzentration
- Zwei Merkmale
- Korrelation
- Preisindizes
- Lineare Regression

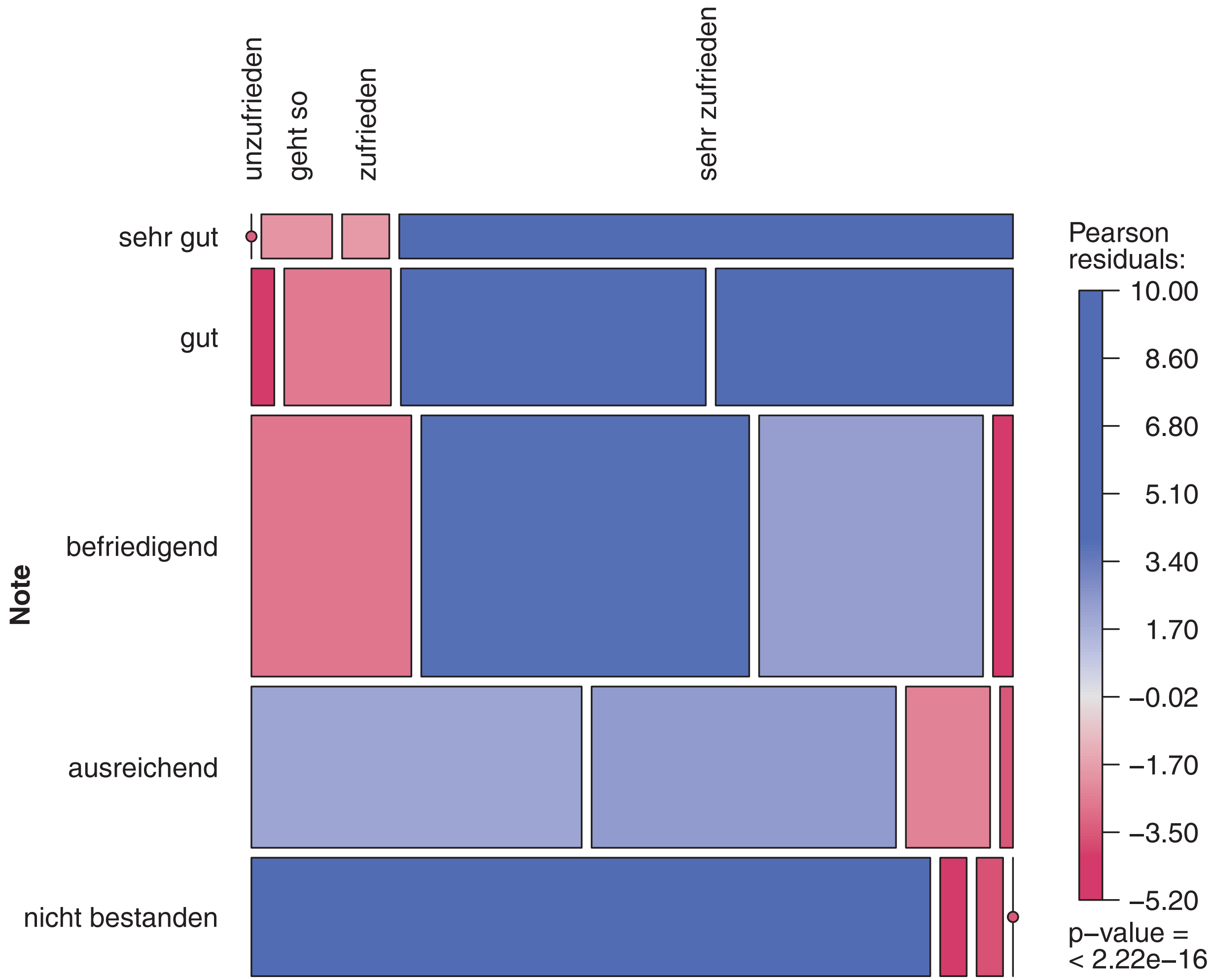
3. W-Theorie

4. Induktive Statistik

Quellen

Tabellen


```
Data.complete = na.omit(MyData[,c("MatheZufr", "NoteMathe")])
Noten.complete =
  ordered(cut(Data.complete$NoteMathe, breaks=c(0,1.5,2.5,3.5,4.1,5.0)),
          labels=c("sehr gut", "gut", "befriedigend", "ausreichend", "nicht bestanden"))
tab = table("Note"=Noten.complete, "Zufrieden mit Leistung"=Data.complete$MatheZufr)
require(vcd)
mosaic(tab, shade = TRUE, gp_args = list(interpolate = function(x) pmin(x/4, 1)), labeling_args =
   list(rot_labels = c(90,0,0,0), just_labels = c("left", "left", "right", "right"),
        offset_varnames = c(left = 5, top=5.5), offset_labels = c(right = 3)),
 margins = c(right = 1, bottom = 3, left=6, top=5))
```

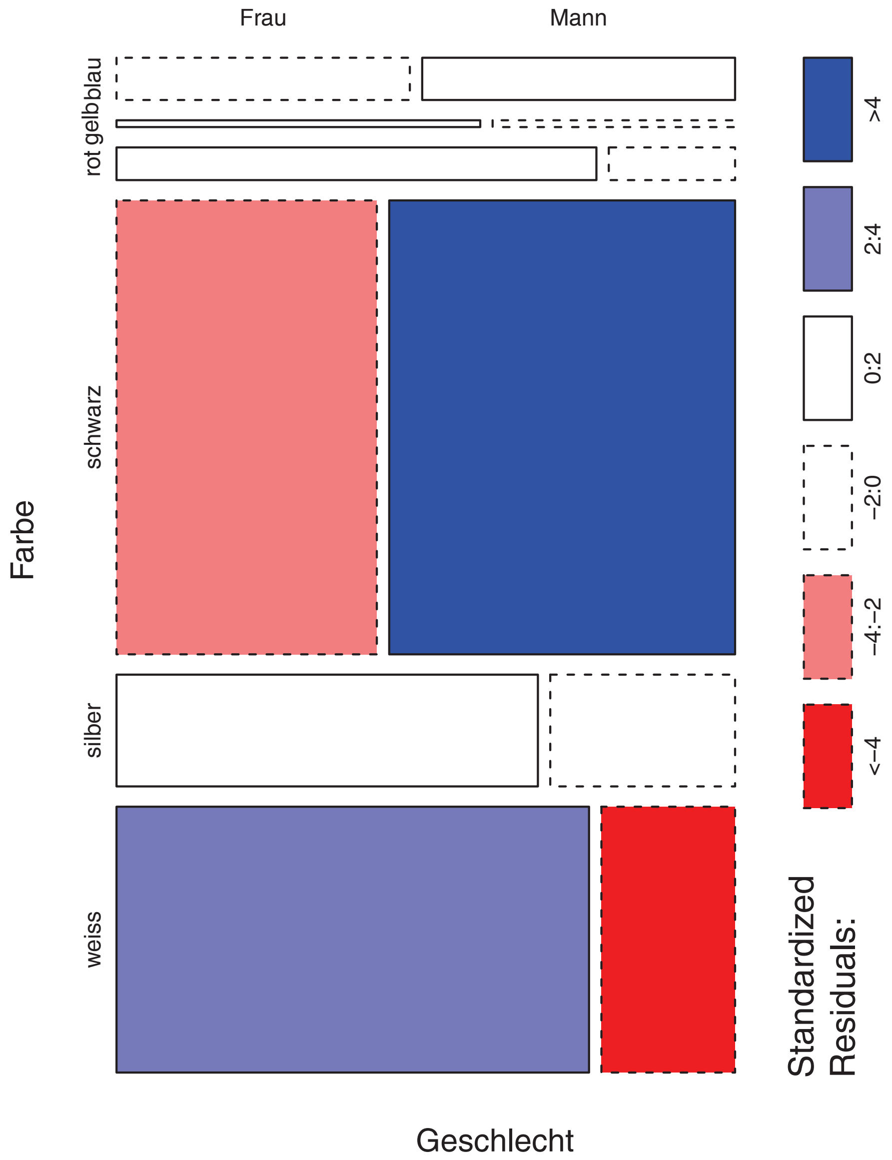


„Note in Matheklausur“ gegen „Zufrieden mit Leistung“

Mosaicplot Geschlecht, Wunschfarbe für Smartphone

```
tab = table(Farbe, Geschlecht)
tab

##          Geschlecht
## Farbe     Frau Mann
##   blau      15   16
##   gelb       3    2
##   rot       19    5
##   schwarz  143  190
##   silber    57   25
##   weiss    152   43
```

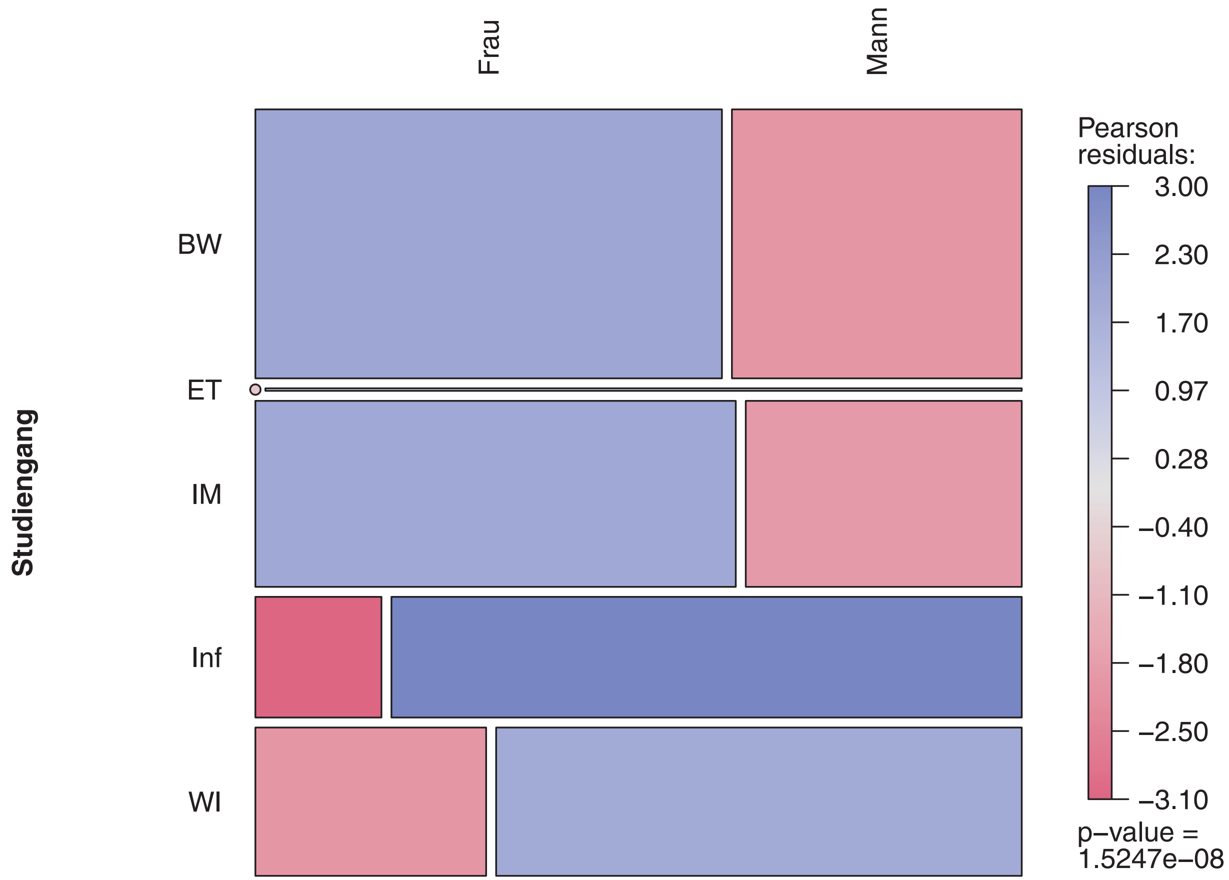
```
mosaicplot(t(tab), shade = TRUE,
           sort=2:1, main="")
```



Statistik

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

```r
require(vcd)
Data.complete = na.omit(MyData[,c("Geschlecht", "Studiengang")])
with(Data.complete, {
  tab = table("Studiengang"=Studiengang, "Geschlecht"=Geschlecht)
mosaic(tab, shade = TRUE, gp_args = list(interpolate = function(x) pmin(x/4, 1)), labeling_args =
   list(rot_labels = c(90,0,0,0), just_labels = c("left", "left", "right", "right"),
        offset_varnames = c(left = 5, top=5.5), offset_labels = c(right = 3)),
 margins = c(right = 1, bottom = 3, left=6, top=5))
})
```



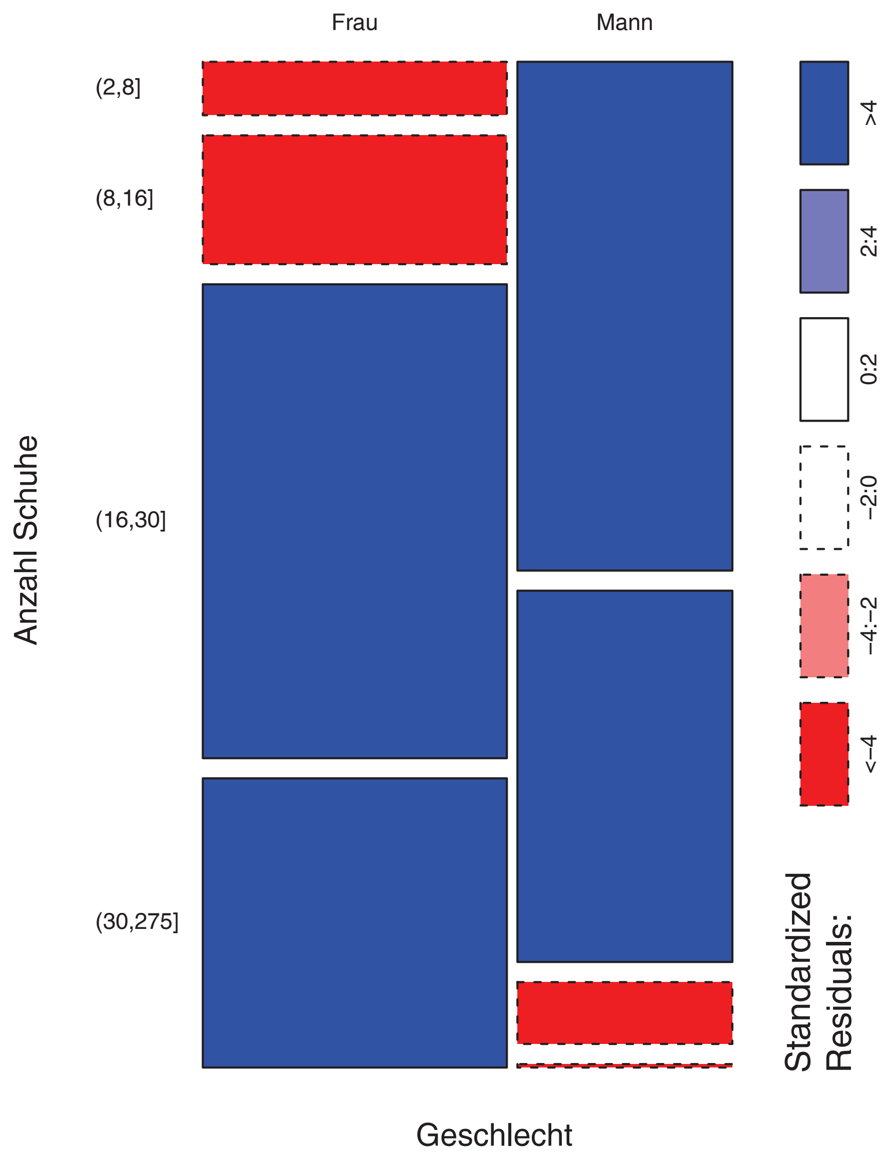
Studiengang vs. Geschlecht

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Mosaicplot Geschlecht, Anzahl Schuhe

```r
tab = table(
  "Anzahl Schuhe" =
  cut(AnzSchuhe,
      breaks =
        quantile(
          AnzSchuhe,
          probs = (0:4)/4
        )
  ),
  Geschlecht)
tab

##                 Geschlecht
## Anzahl Schuhe Frau Mann
##      (2,8]      22  148
##      (8,16]     53  108
##      (16,30]   195   18
##      (30,275]  119    1
```

```r
mosaicplot(t(tab), shade = TRUE,
           main="", las=1)
```



Preisindizes

- Preismesszahl: Misst Preisveränderung eines einzelnen Gutes:

$$\frac{\text{Preis zum Zeitpunkt } j}{\text{Preis zum Zeitpunkt } i}$$

  dabei: j: Berichtsperiode, i: Basisperiode

- Preisindex: Misst Preisveränderung mehrerer Güter (Aggregation von Preismesszahlen durch Gewichtung)
- Notation:

$$\begin{array}{rlll} p_0(i): & \text{Preis} & \text{des } i\text{-ten Gutes in Basisperiode} & 0 \\ p_t(i): & \text{Preis} & \text{des } i\text{-ten Gutes in Berichtsperiode} & t \\ q_0(i): & \text{Menge} & \text{des } i\text{-ten Gutes in Basisperiode} & 0 \\ q_t(i): & \text{Menge} & \text{des } i\text{-ten Gutes in Berichtsperiode} & t \end{array}$$

Preisindizes

- Gleichgewichteter Preisindex:

$$P^G_{0t} = \frac{1}{n}\sum_{i=1}^{n}\frac{p_t(i)}{p_0(i)} = \sum_{i=1}^{n}\frac{p_t(i)}{p_0(i)}\cdot g(i) \qquad \text{mit} \qquad g(i) = \frac{1}{n}$$

Nachteil: Auto und Streichhölzer haben gleiches Gewicht
Lösung: Preise mit Mengen gewichten!

- Preisindex von Laspeyres:

$$P^L_{0t} = \frac{\sum\limits_{i=1}^{n} p_t(i)q_0(i)}{\sum\limits_{i=1}^{n} p_0(i)q_0(i)} = \sum_{i=1}^{n}\frac{p_t(i)}{p_0(i)}\cdot g_0(i) \qquad \text{mit} \qquad g_0(i) = \frac{p_0(i)\,q_0(i)}{\sum\limits_{j=1}^{n} p_0(j)\,q_0(j)}$$

- Preisindex von Paasche:

$$P^P_{0t} = \frac{\sum\limits_{i=1}^{n} p_t(i)q_t(i)}{\sum\limits_{i=1}^{n} p_0(i)q_t(i)} = \sum_{i=1}^{n}\frac{p_t(i)}{p_0(i)}\cdot g_t(i) \qquad \text{mit} \qquad g_t(i) = \frac{p_0(i)\,q_t(i)}{\sum\limits_{j=1}^{n} p_0(j)\,q_t(j)}$$

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Warenkorb: Kartoffeln und Kaffee

	1950		2013	
	Preis (€)	Menge pro Woche	Preis (€)	Menge pro Woche
1 kg Kartoffeln	0,04	3,58	1,10	1,25
100 g Kaffeebohnen	3,00	0,25	0,70	1,31

$$P^{L}_{1950,\,2013} = \frac{1{,}10 \cdot 3{,}58 + 0{,}70 \cdot 0{,}25}{0{,}04 \cdot 3{,}58 + 3{,}00 \cdot 0{,}25} \approx 4{,}6048$$

$$P^{P}_{1950,\,2013} = \frac{1{,}10 \cdot 1{,}25 + 0{,}70 \cdot 1{,}31}{0{,}04 \cdot 1{,}25 + 3{,}00 \cdot 1{,}31} \approx 0{,}5759$$

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Weitere Preisindizes

Idealindex von Fisher:

$$P_{0t}^{F} = \sqrt{P_{0t}^{L} P_{0t}^{P}}$$

Marshall-Edgeworth-Index:

$$P_{0t}^{ME} = \frac{\sum\limits_{i=1}^{n} p_t(i)[q_0(i) + q_t(i)]}{\sum\limits_{i=1}^{n} p_0(i)[q_0(i) + q_t(i)]}$$

Preisindex von Lowe:

$$P_{0t}^{LO} = \frac{\sum\limits_{i=1}^{n} p_t(i)q(i)}{\sum\limits_{i=1}^{n} p_0(i)q(i)} \qquad \text{wobei} \qquad q(i) \triangleq \begin{cases} \text{Durchschn. Menge von} \\ \text{Gut } i \text{ über alle (bekannten)} \\ \text{Perioden} \end{cases}$$

1. Einführung

2. Deskriptive Statistik
- Häufigkeiten
- Lage und Streuung
- Konzentration
- Zwei Merkmale
- Korrelation
- Preisindizes
- Lineare Regression

3. W-Theorie

4. Induktive Statistik

Quellen

Tabellen

Warenkorb: Kartoffeln und Kaffee

	1950		2013	
	Preis (€)	Menge pro Woche	Preis (€)	Menge pro Woche
1 kg Kartoffeln	0,04	3,58	1,10	1,25
100 g Kaffeebohnen	3,00	0,25	0,70	1,31

$$P^{F}_{1950,2013} \approx \sqrt{4{,}6048 \cdot 0{,}5759} = 1{,}6284$$

$$P^{ME}_{1950,2013} = \frac{1{,}10 \cdot (3{,}58 + 1{,}25) + 0{,}70 \cdot (0{,}25 + 1{,}31)}{0{,}04 \cdot (3{,}58 + 1{,}25) + 3{,}00 \cdot (0{,}25 + 1{,}31)} = 1{,}3143$$

$$P^{Lo}_{1950,2013} = \frac{1{,}10 \cdot 2{,}5 + 0{,}70 \cdot 0{,}75}{0{,}04 \cdot 2{,}5 + 3{,}00 \cdot 0{,}75} = 1{,}3936$$

Annahme bei P^{LO}: Durchschn. Mengen bei Kartoffeln bzw. Kaffebohnen von 1950 bis 2013 sind 2,5 bzw. 0,75.

Statistik

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

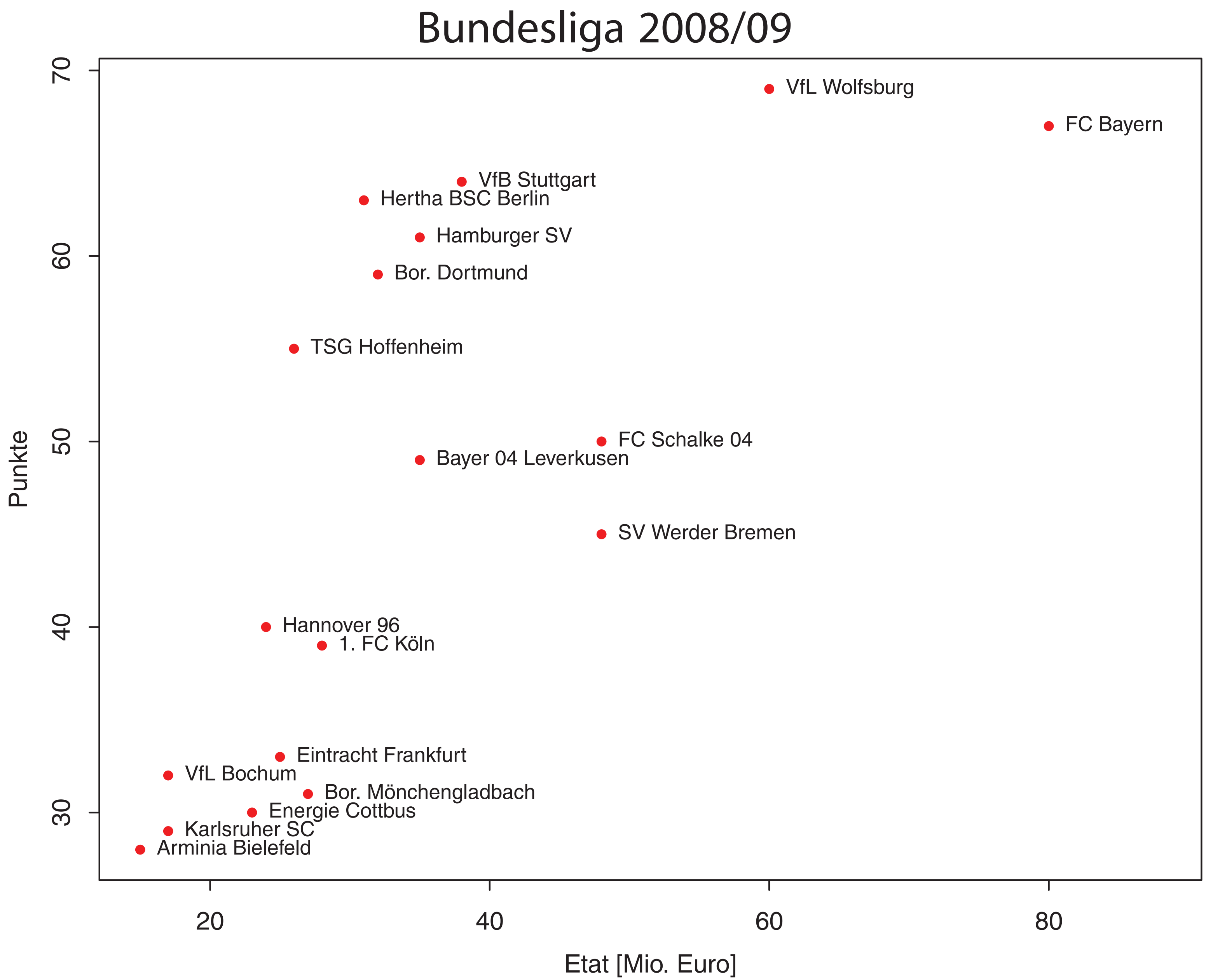
Bundesliga 2008/2009

- Gegeben: Daten zu den 18 Vereinen der ersten Bundesliga in der Saison 2008/09
- Merkmale: **Vereinssetat** für Saison (nur direkte Gehälter und Spielergehälter)
- und **Ergebnispunkte** in Tabelle am Ende der Saison

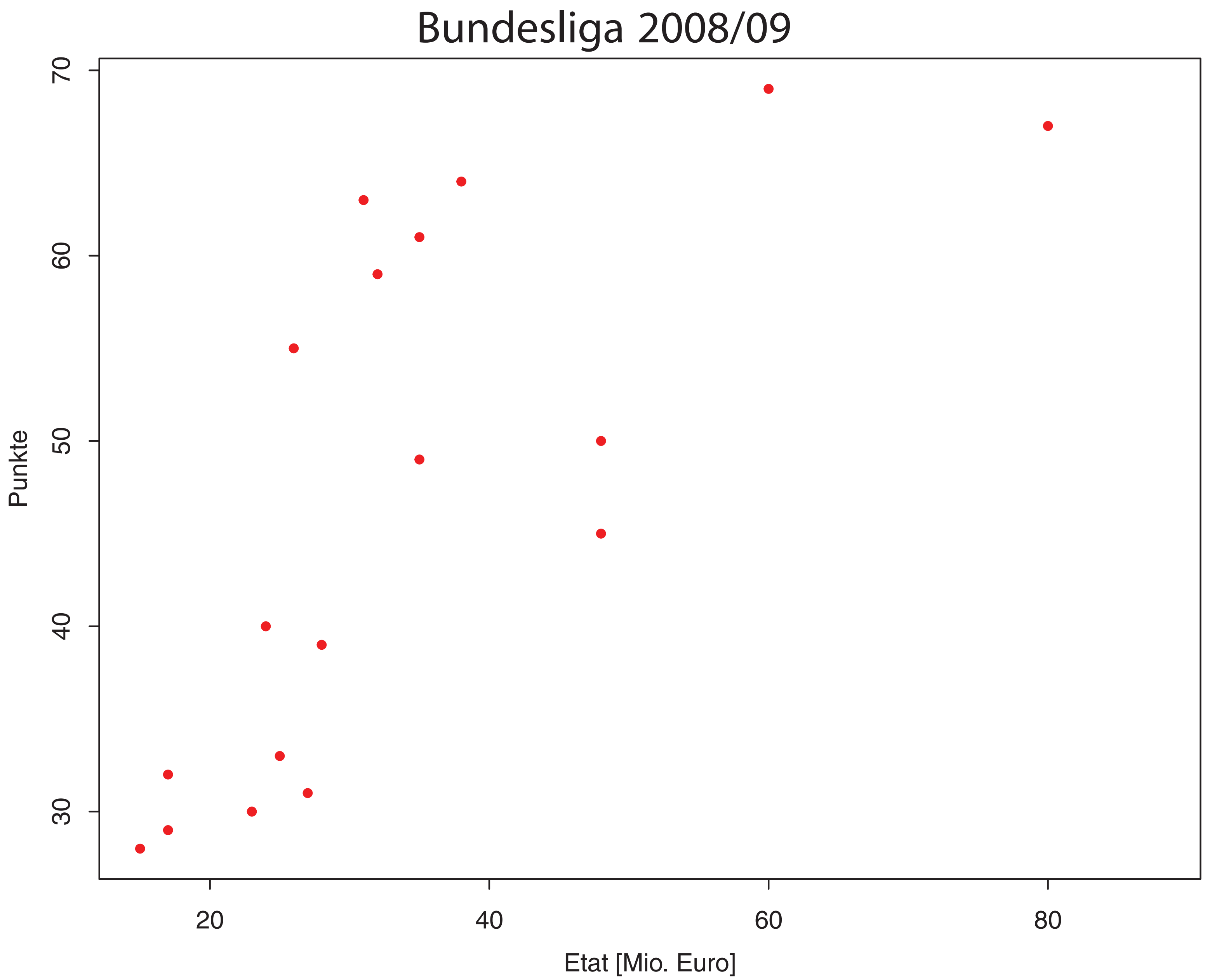
	Etat	Punkte
FC Bayern	80	67
VfL Wolfsburg	60	69
SV Werder Bremen	48	45
FC Schalke 04	48	50
VfB Stuttgart	38	64
Hamburger SV	35	61
Bayer 04 Leverkusen	35	49
Bor. Dortmund	32	59
Hertha BSC Berlin	31	63
1. FC Köln	28	39
Bor. Mönchengladbach	27	31
TSG Hoffenheim	26	55
Eintracht Frankfurt	25	33
Hannover 96	24	40
Energie Cottbus	23	30
VfL Bochum	17	32
Karlsruher SC	17	29
Arminia Bielefeld	15	28

(Quelle: Welt)

Darstellung der Daten in Streuplot



Darstellung der Daten in Streuplot



1. Einführung

2. Deskriptive Statistik
- Häufigkeiten
- Lage und Streuung
- Konzentration
- Zwei Merkmale
- Korrelation
- Preisindizes
- Lineare Regression

3. W-Theorie

4. Induktive Statistik

Quellen

Tabellen

Trend als lineares Modell

- Kann man die **Tabellenpunkte** näherungsweise über einfache Funktion **in Abhängigkeit des Vereinsetats** darstellen?
- Allgemein: Darstellung einer Variablen Y als Funktion von X:

$$y = f(x)$$

- Dabei:
  - X heißt Regressor bzw. unabhängige Variable
  - Y heißt Regressand bzw. abhängige Variable
- Wichtiger (und einfachster) Spezialfall: f beschreibt einen linearen Trend:

$$y = a + b\,x$$

- Dabei anhand der Daten zu schätzen: a (Achsenabschnitt) und b (Steigung)
- Schätzung von a und b: Lineare Regression

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Fehlerquadratsumme

- Pro Datenpunkt gilt mit Regressionsmodell:

$$y_i = a + bx_i + \epsilon_i$$

- Dabei: ϵ_i ist jeweils Fehler (der Grundgesamtheit),
- mit $e_i = y_i - (\hat{a} + \hat{b}x_i)$: Abweichung (Residuen) zwischen gegebenen Daten der Stichprobe und durch Modell geschätzten Werten
- Modell gut wenn alle Residuen e_i zusammen möglichst klein
- Einfache Summe aber nicht möglich, denn e_i positiv oder negativ
- Deswegen: Summe der Quadrate von e_i
- Prinzip der kleinsten Quadrate: Wähle a und b so, dass

$$Q(a, b) = \sum_{i=1}^{n} [y_i - (a + bx_i)]^2 \to \text{min}$$

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Beste Lösung

- Beste und eindeutige Lösung:

$$\hat{b} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

$$\hat{a} = \bar{y} - \hat{b}\,\bar{x}$$

- Regressionsgerade:

$$\hat{y} = \hat{a} + \hat{b}\,x$$

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Bundesligabeispiel

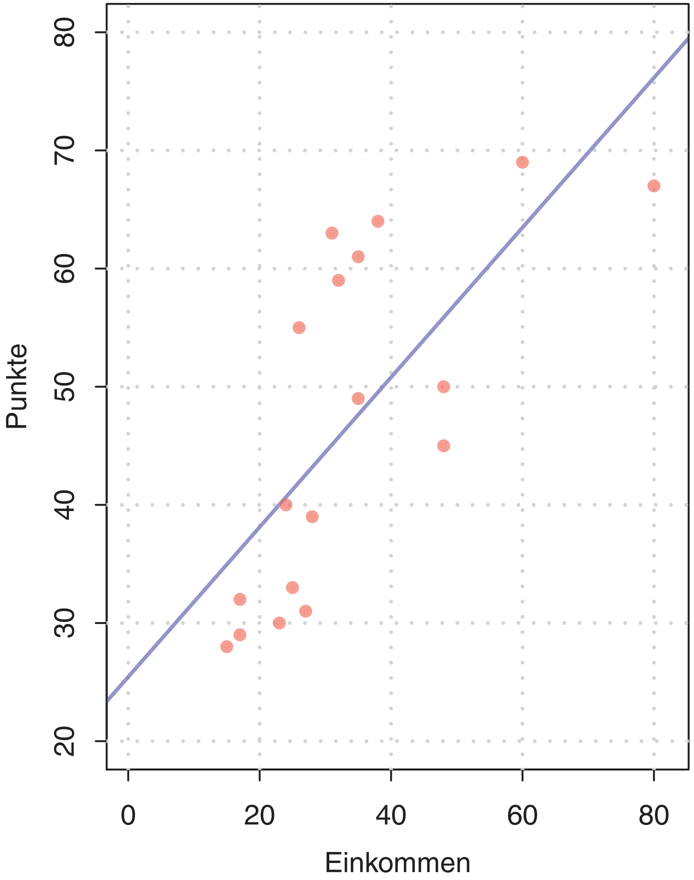
- Berechnung eines linearen Modells der Bundesligadaten
- dabei: Punkte $\triangleq y$ und Etat $\triangleq x$:

$\overline{x}$	33,83
$\overline{y}$	46,89
$\sum x_i^2$	25209
$\sum x_i y_i$	31474
n	18

$$\Rightarrow \hat{b} = \frac{31474 - 18 \cdot 33{,}83 \cdot 46{,}89}{25209 - 18 \cdot 33{,}83^2} \approx 0{,}634$$

$$\Rightarrow \hat{a} = 46{,}89 - \hat{b} \cdot 33{,}83 \approx 25{,}443$$

- Modell: $\hat{y} = 25{,}443 + 0{,}634 \cdot x$



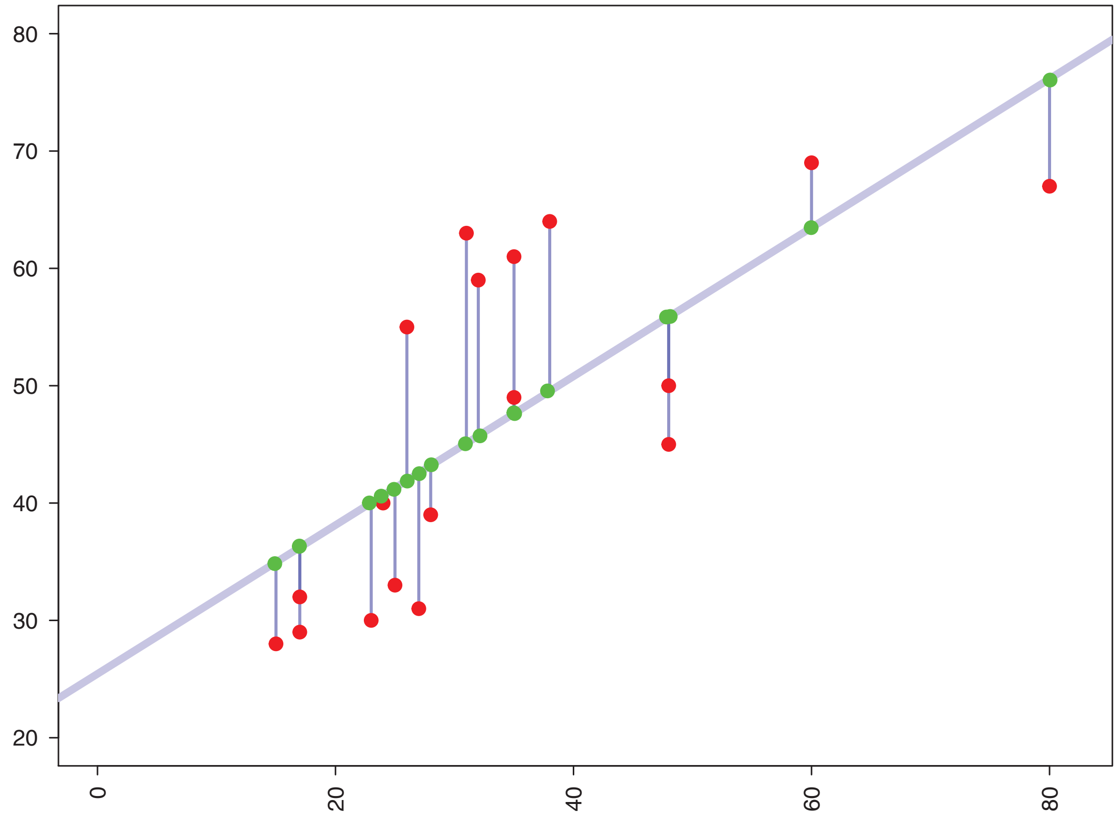
- Prognosewert für Etat $= 30$:

$$\hat{y}(30) = 25{,}443 + 0{,}634 \cdot 30 \approx 44{,}463$$

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Varianz und Information

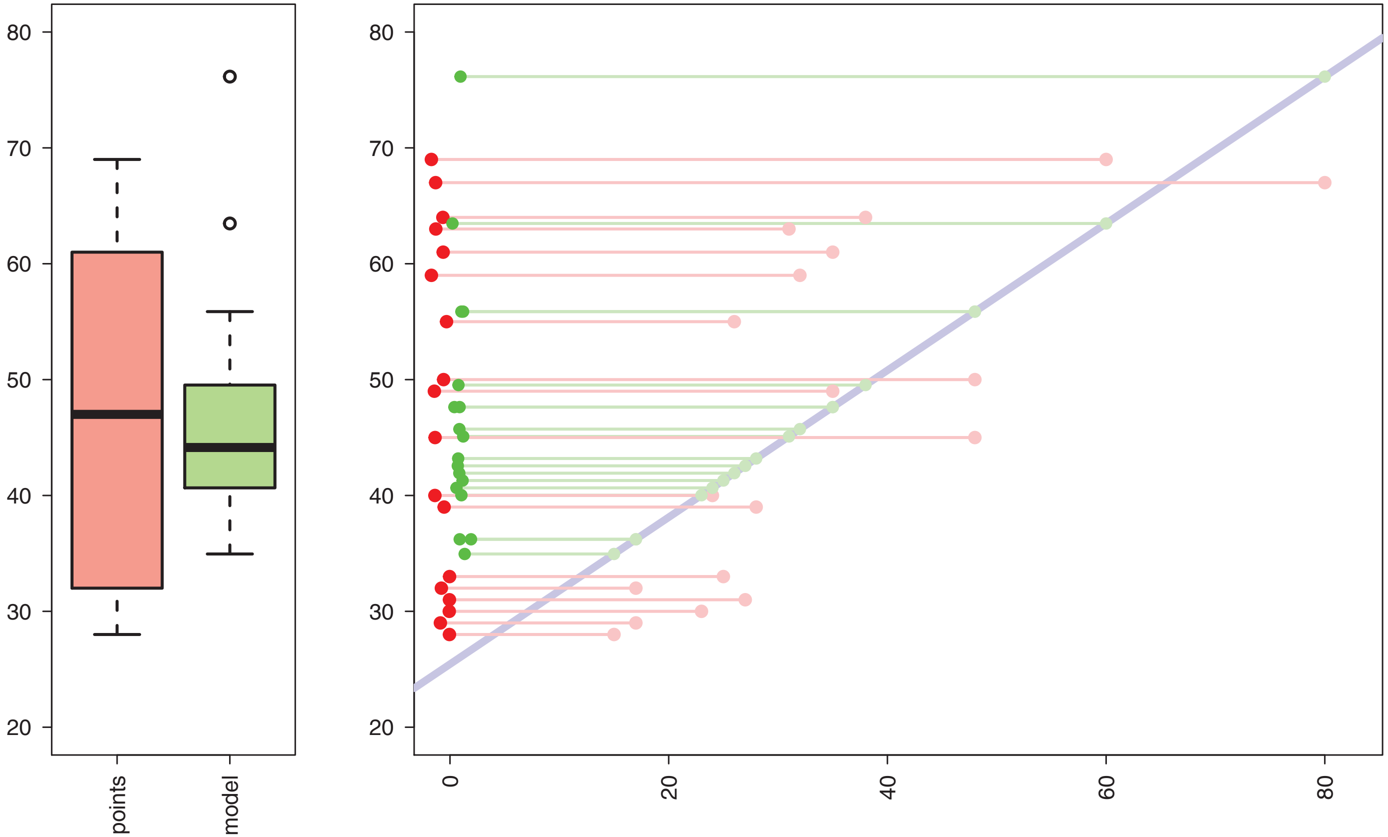
- **Varianz** der Daten in abhängiger Variablen y_i als Repräsentant des **Information**sgehalts
- Ein Bruchteil davon kann in Modellwerten $\hat{y}_i$ abgebildet werden



1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Varianz und Information

- **Varianz** der Daten in abhängiger Variablen y_i als Repräsentant des **Information**sgehalts
- Ein Bruchteil davon kann in Modellwerten $\hat{y}_i$ abgebildet werden



- Empirische Varianz (mittlere quadratische Abweichung) für „rot" bzw. „grün" ergibt jeweils

$$\frac{1}{18}\sum_{i=1}^{18}(y_i-\overline{y})^2 \approx 200{,}77 \qquad \text{bzw.} \qquad \frac{1}{18}\sum_{i=1}^{18}(\hat{y}_i-\overline{y})^2 \approx 102{,}78$$

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Determinationskoeffizient

- Gütemaß für die Regression: Determinationskoeffizient (Bestimmtheitskoeffizient):

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = \frac{\sum_{i=1}^{n} \hat{y}_i^2 - n\bar{y}^2}{\sum_{i=1}^{n} y_i^2 - n\bar{y}^2} = r^2 \in [0;1]$$

- Mögliche Interpretation von R^2:
  Durch die Regression erklärter Anteil der Varianz
- $R^2 = 0$ wird erreicht wenn X, Y unkorreliert
  $R^2 = 1$ wird erreicht wenn $\hat{y}_i = y_i \; \forall i$ (alle Punkte auf Regressionsgerade)
- Im (Bundesliga-)Beispiel:

$$R^2 = \frac{\sum_{i=1}^{18} (\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{18} (y_i - \overline{y})^2} \approx \frac{102{,}78}{200{,}77} \approx 51{,}19\,\%$$

Regression: 4 eindimensionale Beispiele

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

- Berühmte Daten aus den 1970er Jahren:

i	x_{1i}	x_{2i}	x_{3i}	x_{4i}	y_{1i}	y_{2i}	y_{3i}	y_{4i}
1	10	10	10	8	8,04	9,14	7,46	6,58
2	8	8	8	8	6,95	8,14	6,77	5,76
3	13	13	13	8	7,58	8,74	12,74	7,71
4	9	9	9	8	8,81	8,77	7,11	8,84
5	11	11	11	8	8,33	9,26	7,81	8,47
6	14	14	14	8	9,96	8,10	8,84	7,04
7	6	6	6	8	7,24	6,13	6,08	5,25
8	4	4	4	19	4,26	3,10	5,39	12,50
9	12	12	12	8	10,84	9,13	8,15	5,56
10	7	7	7	8	4,82	7,26	6,42	7,91
11	5	5	5	8	5,68	4,74	5,73	6,89

(Quelle: Anscombe (1973))

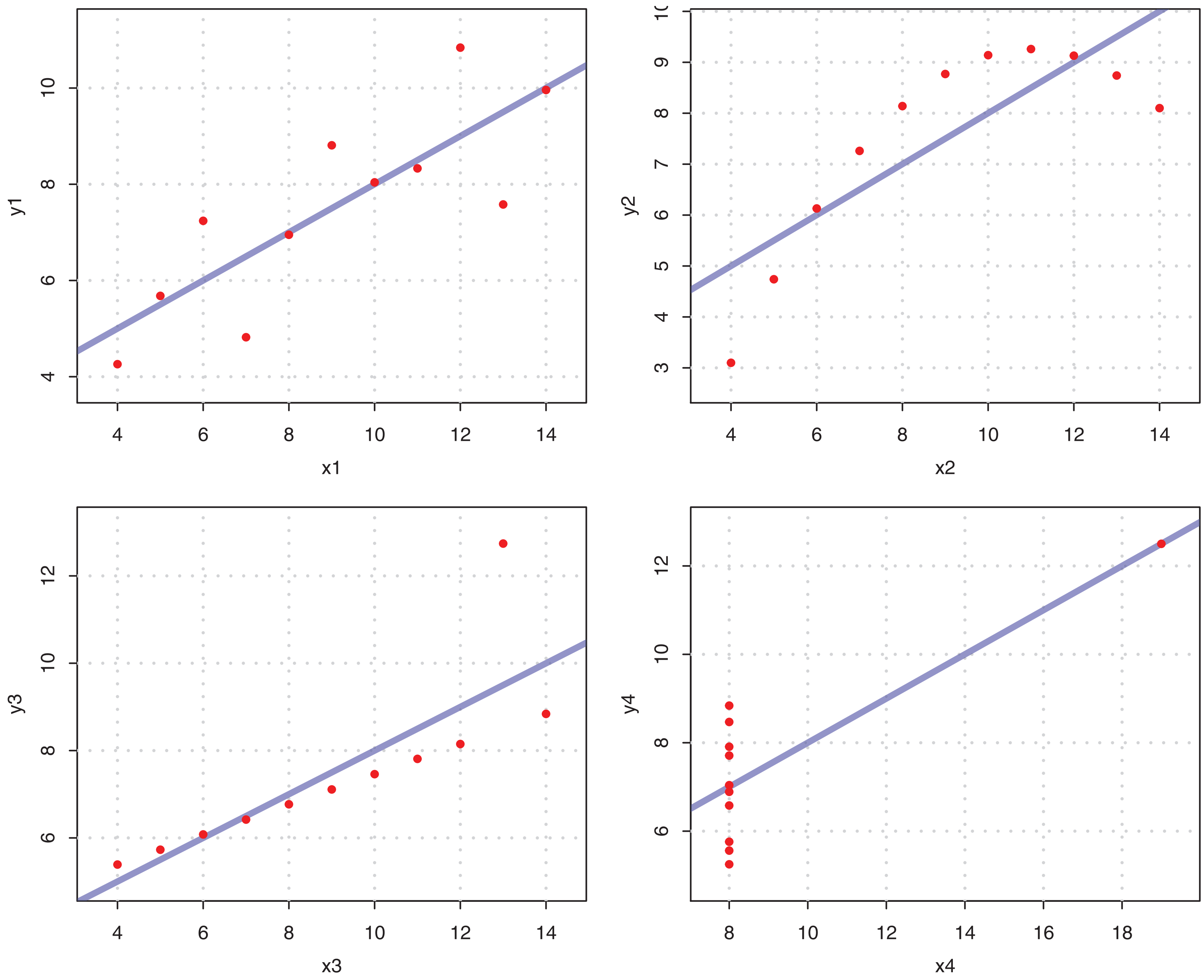
Regression: 4 eindimensionale Beispiele

- In folgender Tabelle: Jeweils Ergebnisse der linearen Regressionsanalyse
- dabei: x_k unabhängige Variable und y_k abhängige Variable
- Modell jeweils: $y_k = a_k + b_k x_k$

k	$\hat{a}_k$	$\hat{b}_k$	R_k^2
1	3,0001	0,5001	0,6665
2	3,0010	0,5000	0,6662
3	3,0025	0,4997	0,6663
4	3,0017	0,4999	0,6667

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Plot der Anscombe-Daten



1. Einführung

2. Deskriptive Statistik

Häufigkeiten

Lage und Streuung

Konzentration

Zwei Merkmale

Korrelation

Preisindizes

Lineare Regression

3. W-Theorie

4. Induktive Statistik

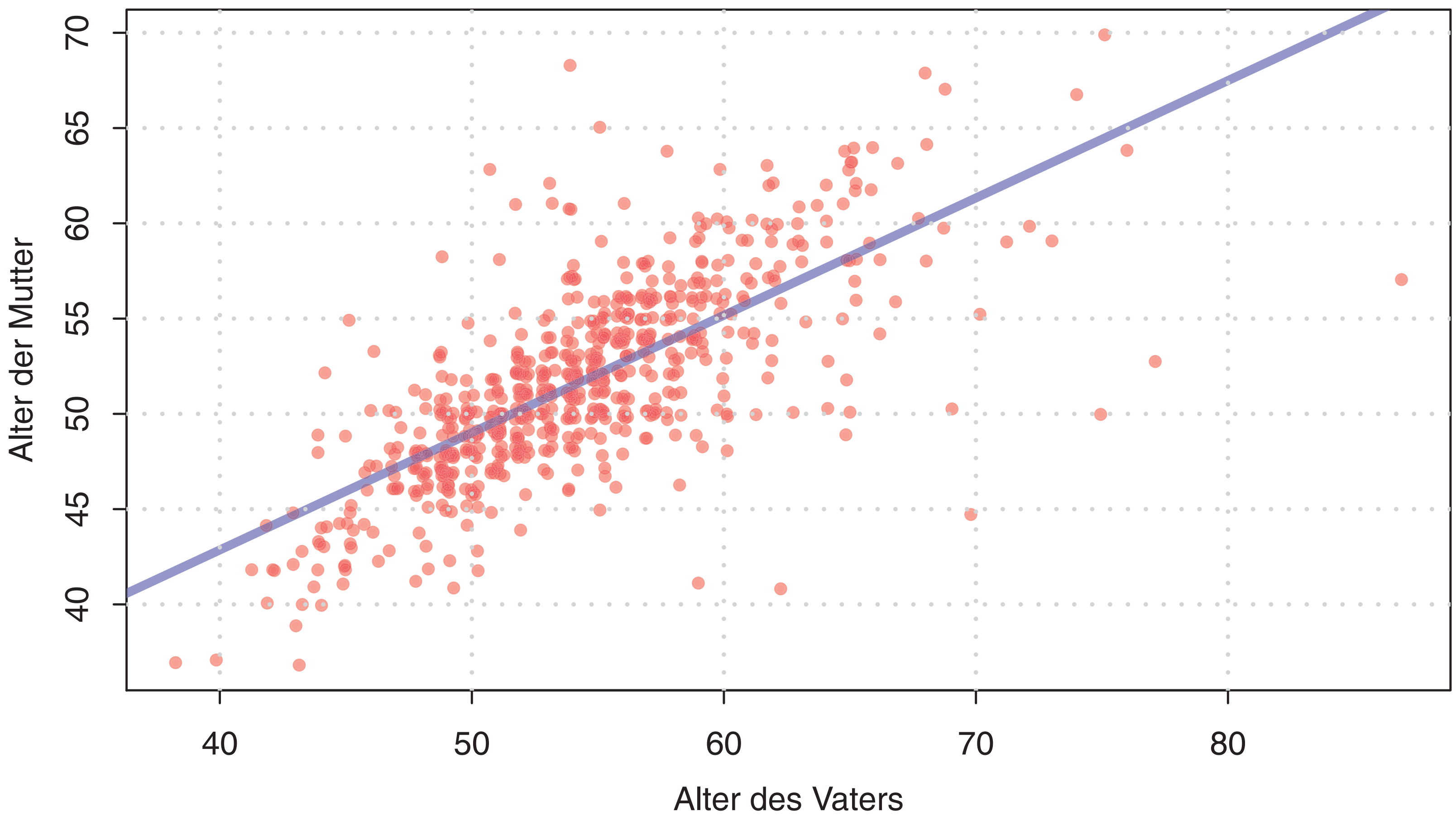
Quellen

Tabellen

Beispieldaten

```
meineRegression = lm(AlterM ~ AlterV)
meineRegression

plot(AlterV, AlterM,
     xlab="Alter des Vaters",
     ylab="Alter der Mutter")
abline(meineRegression)
```

```
##
## Call:
## lm(formula = AlterM ~ AlterV)
##
## Coefficients:
## (Intercept)       AlterV
##      18.2234       0.6159
```



1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Cook's Distanz

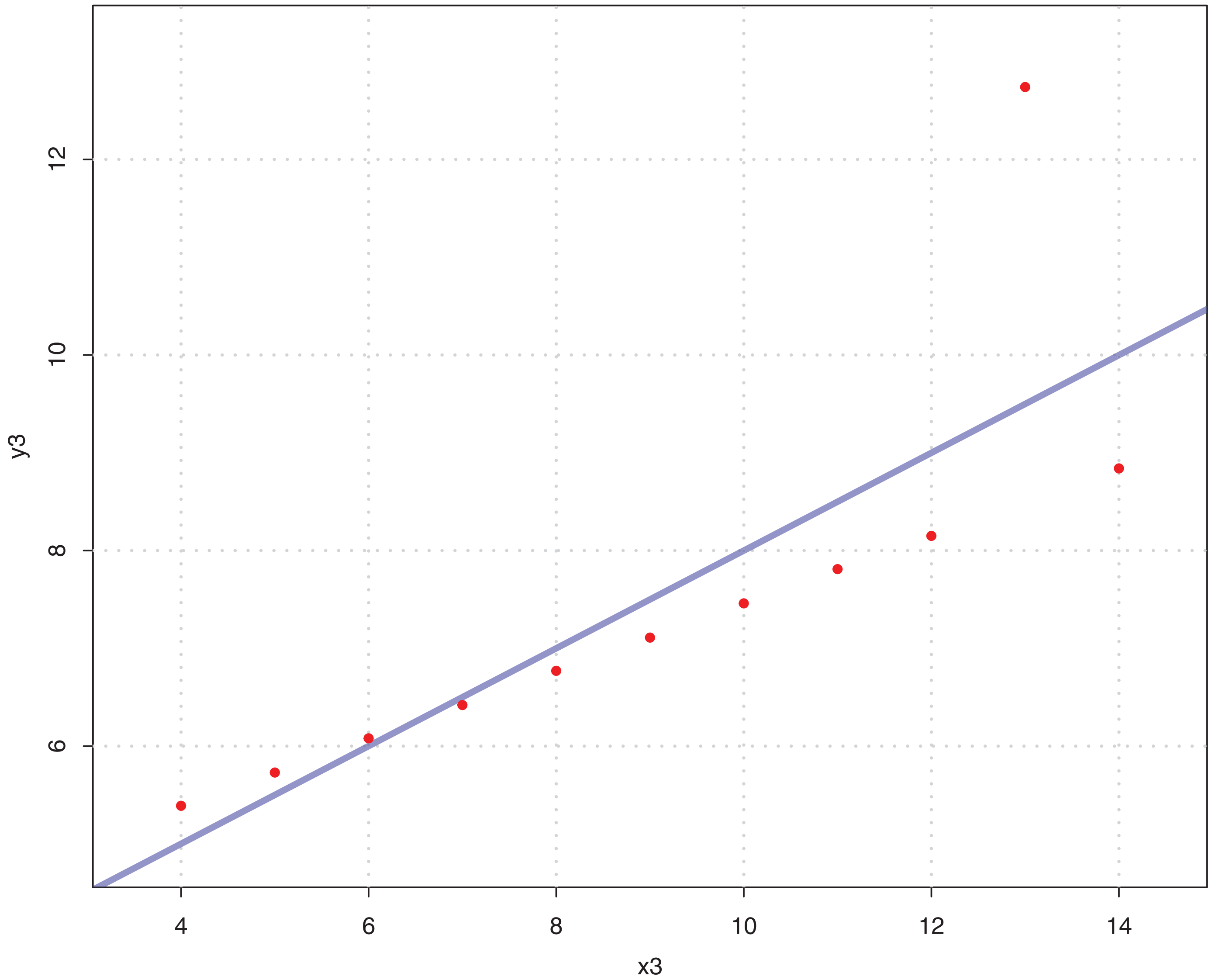
- Oft Kritisch: Einzelne Punkte, die Modell stark beeinflussen
- Idee: Was würde sich ändern, wenn solche Punkte weggelassen würden?
- Cook-Distanz: Misst den Effekt eines gelöschten Objekts
- Formel für ein lineares Modell mit einem unabh. Merkmal:

$$D_i = \frac{\sum_{j=1}^{n} (\hat{y}_j - \hat{y}_{j(\text{ohne } i)})^2}{\text{MSE}}$$

- Dabei bedeutet:
    - $\hat{y}_j$: Prognosewert des kompletten Modells für das j-te Objekt
    - $\hat{y}_{j(\text{ohne } i)}$: Prognosewert des Modells ohne Objekt i für das j-te Objekt
    - $\text{MSE} = \frac{1}{n} \cdot \sum (\hat{y}_i - y_i)^2$: Normierender Term (Schätzwert für Fehlerstreuung)

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

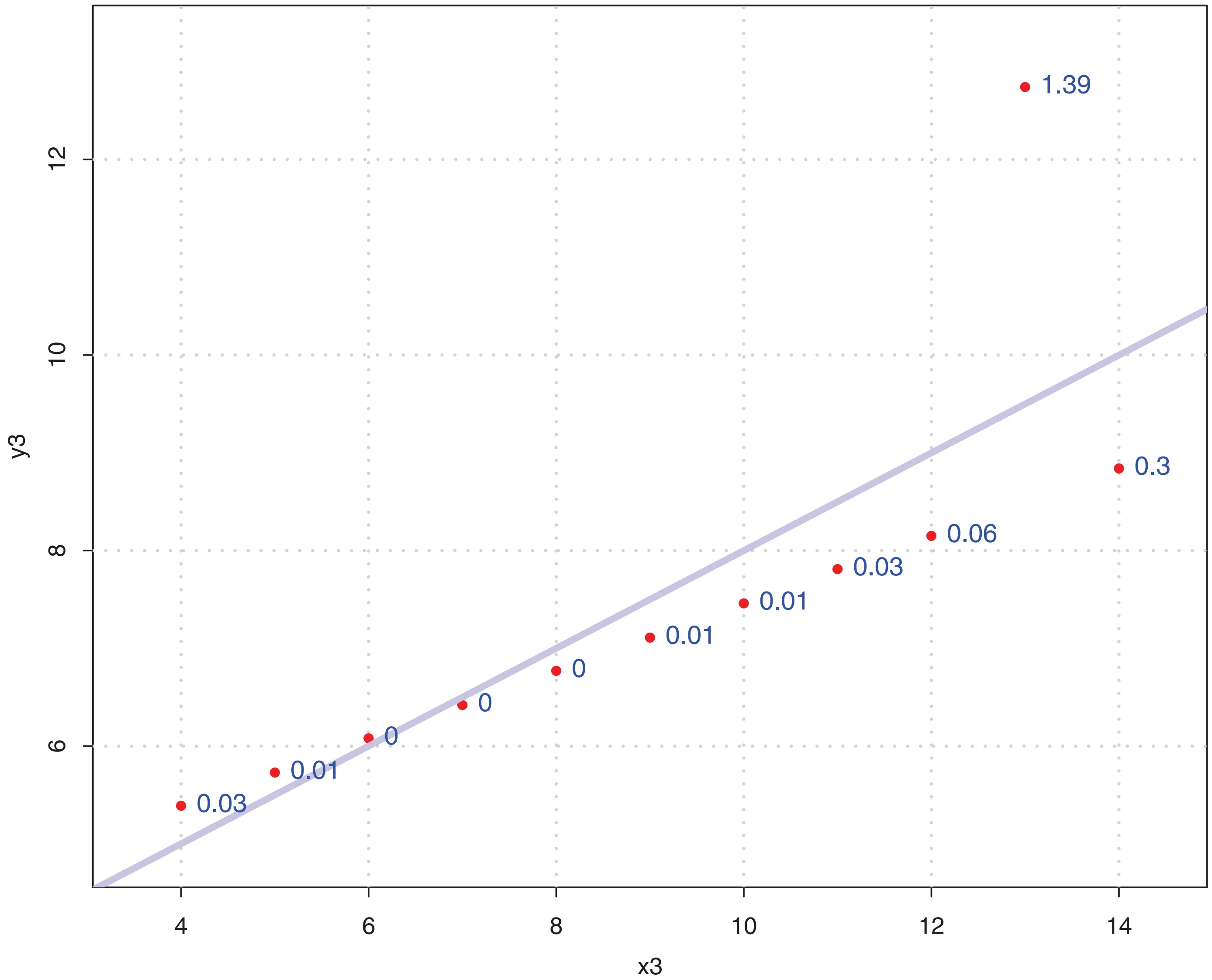
- Anscombe-Daten: Regressionsmodell Nr. 3



1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Ausreißer?

- Anscombe-Daten: Regressionsmodell Nr. 3
- Darstellung der Cook-Distanz neben Punkten
- Faustformel: Werte über 1 sollten genau untersucht werden



1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

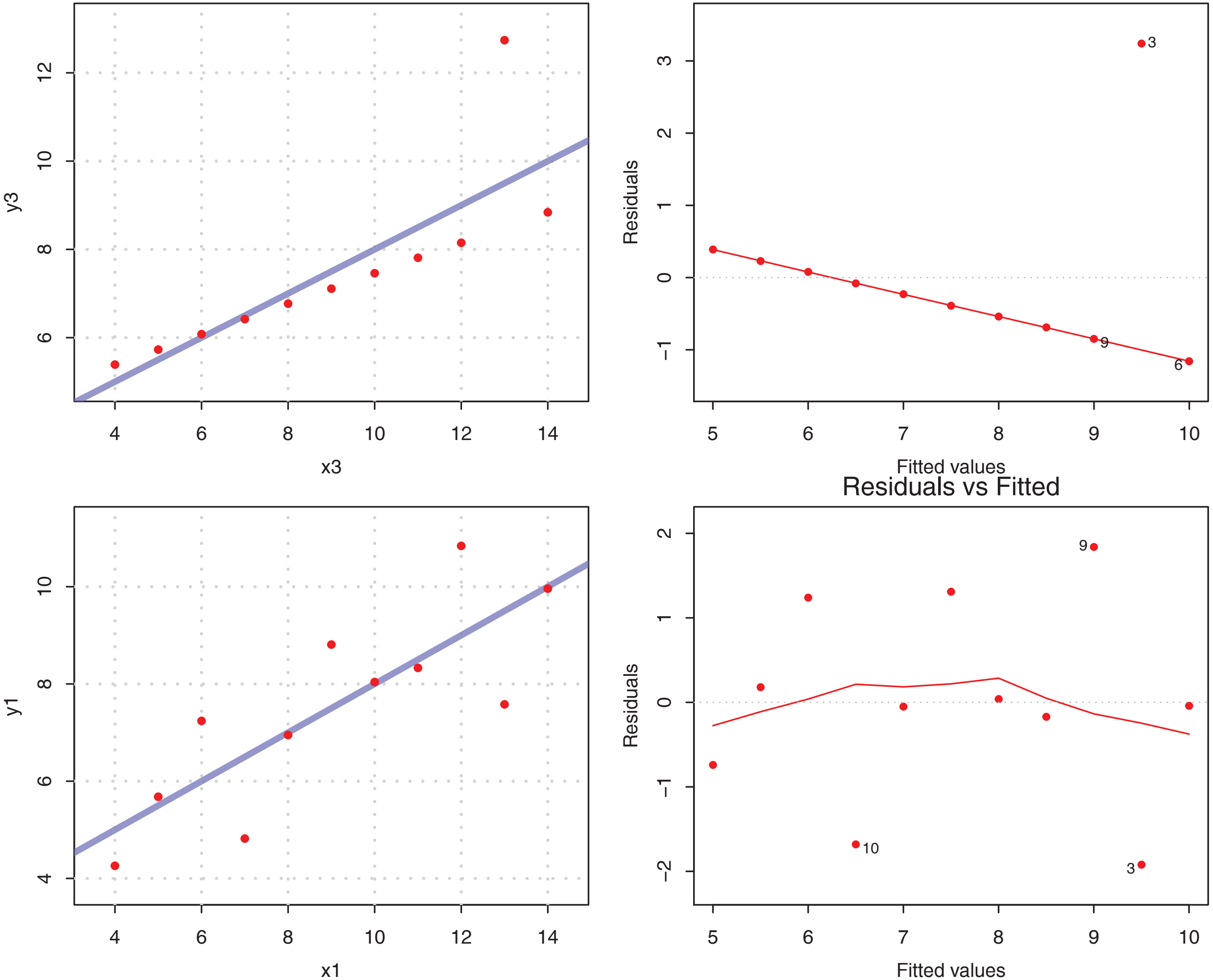
Residualanalyse

- Oft aufschlussreich: Verteilung der Residuen e_i
- Verbreitet: Graphische Darstellungen der Residuen
- Z.B.: e_i über $\hat{y}_i$

1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Residualanalyse

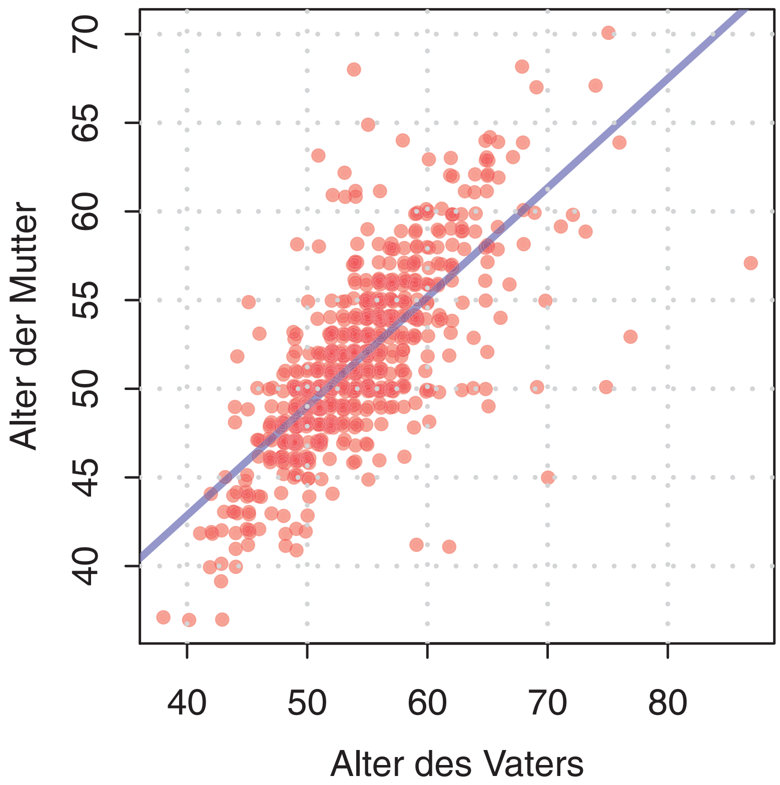
- Oft aufschlussreich: Verteilung der Residuen e_i
- Verbreitet: Graphische Darstellungen der Residuen
- Z.B.: e_i über $\hat{y}_i$

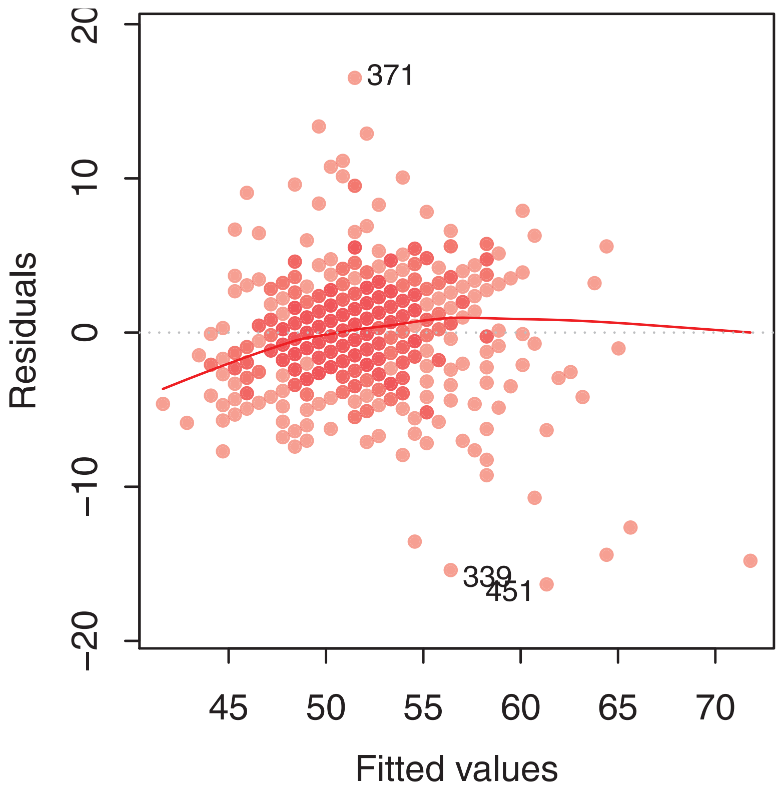


Residualanalyse

Wichtige Eigenschaften der Residuenverteilung

- Möglichst keine systematischen Muster
- Keine Änderung der Varianz in Abhängigkeit von $\hat{y}_i$ (Homoskedastizität)
- Nötig für inferentielle Analysen: Näherungsweise Normalverteilung der Residuen (q-q-plots)





1. Einführung

2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression

3. W-Theorie

4. Induktive Statistik

Quellen

Tabellen

Kausalität versus Korrelation

Exkurs: Kausalität vs. Korrelation

- Meist wichtig für sinnvolle Regressionsanalysen:
- Kausale Verbindung zwischen unabhängigem und abhängigem Merkmal
- Sonst bei Änderung der unabhängigen Variablen keine sinnvollen Prognosen möglich
- Oft: Latente Variablen im Hintergrund

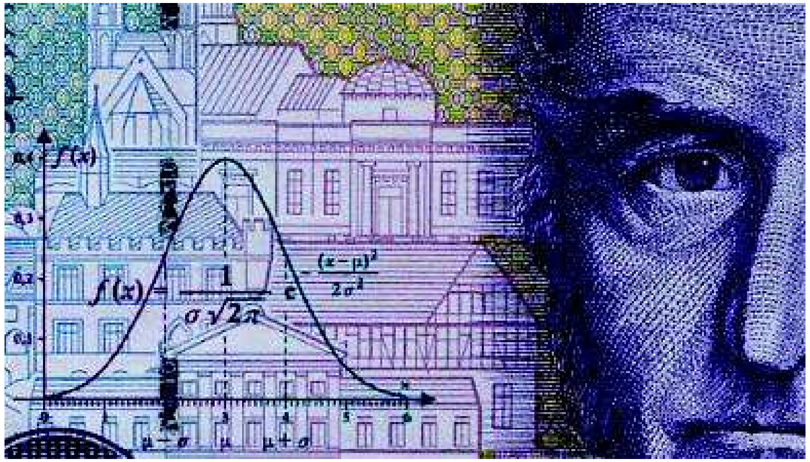
1. Einführung
2. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

1 Statistik: Einführung

2 Deskriptive Statistik

3 Wahrscheinlichkeitstheorie

4 Induktive Statistik



3 Wahrscheinlichkeitstheorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter

Kombinatorik: Anzahl von Kombinationen bei Auswahl

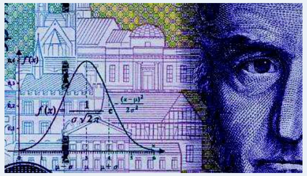
2-mal Würfeln, das heißt Auswahl von $k = 2$ aus $n = 6$ Zahlen.

$\rightarrow$

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

- mit WH, mit RF: alle Möglichkeiten, $6^2 = 36$
- ohne WH, mit RF: Diagonale entfällt, $36 - 6 = 30 = 6 \cdot 5 = \frac{6!}{(6-2)!}$
- ohne WH, ohne RF: Hälfte des letzten Ergebnisses: $\frac{30}{2} = 15 = \frac{6!}{4!2!} = \binom{6}{2}$
- mit WH, ohne RF: Letztes Ergebnis plus Diagonale, $15 + 6 = 21 = \binom{7}{2}$

Auswahl von k aus n Dingen		
	mit Wiederholung	ohne Wiederholung
mit Reihenfolge	n^k	$\frac{n!}{(n-k)!}$
ohne Reihenfolge	$\binom{n+k-1}{k}$	$\binom{n}{k}$



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
4. Induktive Statistik
Quellen
Tabellen

- **Zufallsvorgang:** Geschehen mit ungewissem Ausgang, z.B. Münzwurf
- **Elementarereignis** ω: Ein möglicher Ausgang, z.B. „ Kopf “ Elementarereignisse schließen sich gegenseitig aus („ Kopf “ oder „ Zahl “)!
- **Ergebnismenge** Ω: Menge aller ω
- **Beispiel:** Werfen zweier Würfel:

$$\Omega : \left\{ \begin{matrix} (1,1) & (1,2) & \cdots & (1,6) \\ (2,1) & (2,2) & \cdots & (2,6) \\ \vdots & \vdots & \ddots & \vdots \\ (6,1) & (6,2) & \cdots & (6,6) \end{matrix} \right\}$$

$$\Rightarrow \Omega = \{(x_1, x_2) : x_1, x_2 \in \{1, \ldots, 6\}\}$$

- Ereignis A: Folgeerscheinung eines Elementarereignisses
- Formal:

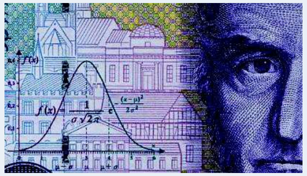
$$A \subset \Omega$$

- Ereignisse schließen sich nicht gegenseitig aus!
- **Beispiel:** Werfen zweier Würfel:

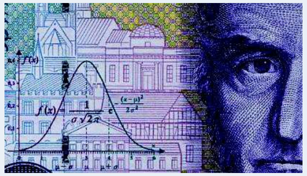
Ereignis	verbal	formal
A	Augensumme $= 4$	$\{(1{,}3),\ (2{,}2),\ (3{,}1)\}$
B	Erste Zahl $= 2$	$\{(2{,}1),\ (2{,}2),\ \ldots,\ (2{,}6)\}$

- Wahrscheinlichkeit $P(A)$: Chance für das Eintreten von A
- Laplace-Wahrscheinlichkeit:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der für } A \text{ günstigen Fälle}}{\text{Anzahl aller möglichen Fälle}}$$



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
4. Induktive Statistik
Quellen
Tabellen



Laplace Wahrscheinlichkeit und Urnenmodell

- **Beispiel:** Werfen zweier Würfel:

$$\text{Augensumme} = 4: \quad A = \{(1,3),\ (2,2),\ (3,1)\}$$

$$|\Omega| = 36, \quad |A| = 3 \Rightarrow P(A) = \frac{3}{36} = \frac{1}{12} = 0{,}083$$

- **Urnenmodell:** Ziehe n Objekte aus einer Menge mit N Objekten
  Anzahl Möglichkeiten:

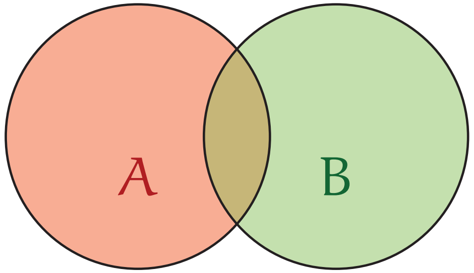
  mit Zurücklegen: N^n

  ohne Zurücklegen: $N \cdot (N-1) \cdots (N-(n-1)) = \frac{N!}{(N-n)!}$

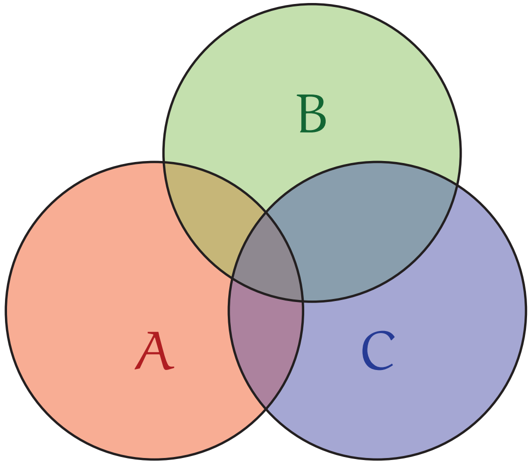
- **Beispiel:**
  Wie groß ist die Wahrscheinlichkeit, aus einem gut gemischten 32-er Kartenblatt bei viermaligem Ziehen vier Asse zu bekommen?
  a) Ziehen mit Zurücklegen,
  b) Ziehen ohne Zurücklegen

Rechenregeln für Wahrscheinlichkeiten

- Wichtige Rechenregeln:

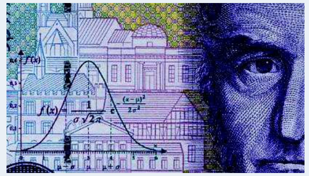


1. $P(A) \leqq 1$
2. $P(\emptyset) = 0$
3. $A \subset B \Rightarrow P(A) \leqq P(B)$
4. $P(\bar{A}) = 1 - P(A)$
5. $P(A \cup B) =$
   $P(A) + P(B) - P(A \cap B)$



- **Beispiel:**

$$P(\text{„Augenzahl} \leqq 5\text{“}) = 1 - P(\text{„Augenzahl} = 6\text{“}) = 1 - \frac{1}{6} = \frac{5}{6}$$

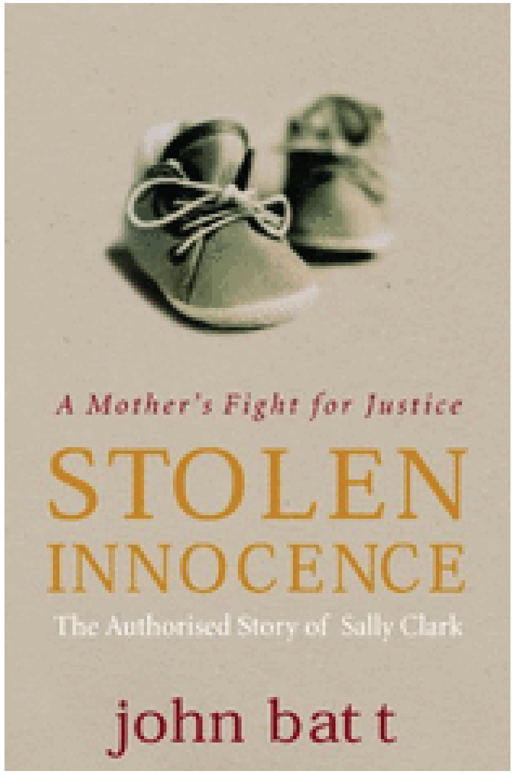


Der Fall Sally Clark

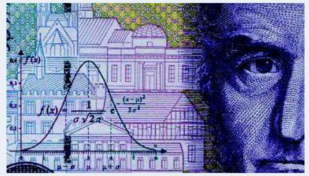
- Sally Clarks Söhne Christopher und Harry sterben 1996 und 1997 beide kurz nach der Geburt an plötzlichem Kindstod.
- Kinderarzt: „Wahrscheinlich Mord, da 2 maliger plötzlicher Kindstod sehr unwahrscheinlich!“ (ohne konkrete Hinweise)
- Gerichtliche Untersuchung
- Hauptargument der Anklage gestützt durch Gerichtsgutachter Sir Roy Meadow (renommierter Facharzt für Kinderheilkunde): Wahrscheinlichkeit für plötzlichen Kindstod ist 1:8500, d.h. Wahrscheinlichkeit für 2 maliges Auftreten in einer Familie

$$p = \left(\frac{1}{8500}\right)^2 \approx 1 : 72\,000\,000$$

- Urteil: Doppelmord; Strafe: 2 mal lebenslang; Inhaftierung von Sally Clark 1999



Statistik



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
4. Induktive Statistik
Quellen
Tabellen

Der Fall Sally Clark

- Problem: Es gibt sehr viele Familien mit 2 Kindern
- Europa: ca. 80 Mio Familien mit Kindern, davon ca. 50% mit mindestens zwei Kindern, also ca. 40 Mio.

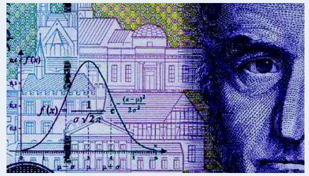
Gene find casts doubt on double 'cot death' murders

An expert said there was a one in 73 million chance Sally Clark's babies died naturally – and a jury agreed. Now new genetic research could help to clear her. **John Sweeney** and **Bill Law** investigate

- Wahrscheinlichkeit, dass in einer solchen Familie kein zweifacher plötzlicher Kindstod auftritt: $1 - \left(\frac{1}{8500}\right)^2$
  Annahmen:
  - Jede dieser Familien hat genau 2 Kinder; in Wirklichkeit: ca. 20% dieser Familien haben mindestens 3 Kinder
  - Zweiter plötzlicher Kindstod unabhängig von erstem (nicht untersucht)
- Wahrscheinlichkeit, dass in 40 Mio. Familien mindestens ein zweifacher plötzlicher Kindstod auftritt:

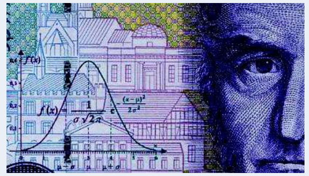
$$1 - \left(1 - \left(\frac{1}{8500}\right)^2\right)^{40\ 000\ 000} \approx 42{,}5\%$$

- 2001: Royal Statistical Society interveniert
- 2003: Sally Clark wird nach Revision freigesprochen
- 2007 findet man sie tot in ihrer Wohnung auf - gestorben an einer akuten Alkoholvergiftung. Sie habe sich, so ihre Familie, von dem Justizirrtum nie erholt.



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
4. Induktive Statistik
Quellen
Tabellen

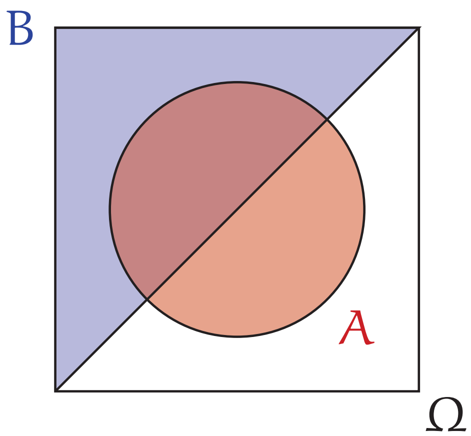
Bedingte Wahrscheinlichkeiten

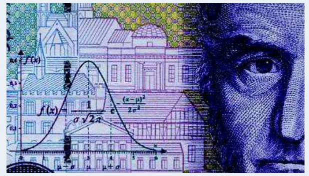


- Wahrscheinlichkeit von A hängt von anderem Ereignis B ab. (B kann zeitlich vor A liegen, muss aber nicht!)
- **Beispiel:** Wahrscheinlichkeit für Statistiknote hängt von Mathenote ab.
- Formal:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

- Im Venndiagramm:





1. Einführung
2. Deskriptive Statistik
3. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
4. Induktive Statistik
Quellen
Tabellen

Zufallsvariablen und Verteilungen

- Beschreibung von Ereignissen durch reelle Zahlen
- Formal: Zufallsvariable ist Abbildung von Ereignisraum in reelle Zahlen:

$$X: \ \Omega \to \mathbb{R}$$

- **Nach** Durchführung des Zufallsvorgangs:

Realisation: $\qquad x = X(\omega)$

- **Vor** Durchführung des Zufallsvorgangs:

Wertebereich: $\qquad X(\Omega) = \{x : \ x = X(\omega), \ \omega \in \Omega\}$

- **Beispiel:** Würfeln, X: Augenzahl, $X(\Omega) = \{1,2,\ldots,6\}$, $x = 4$ (z.B.)

$$P(X = 4) = \tfrac{1}{6}, \quad P(X \leqq 3) = \tfrac{3}{6} = \tfrac{1}{2}$$

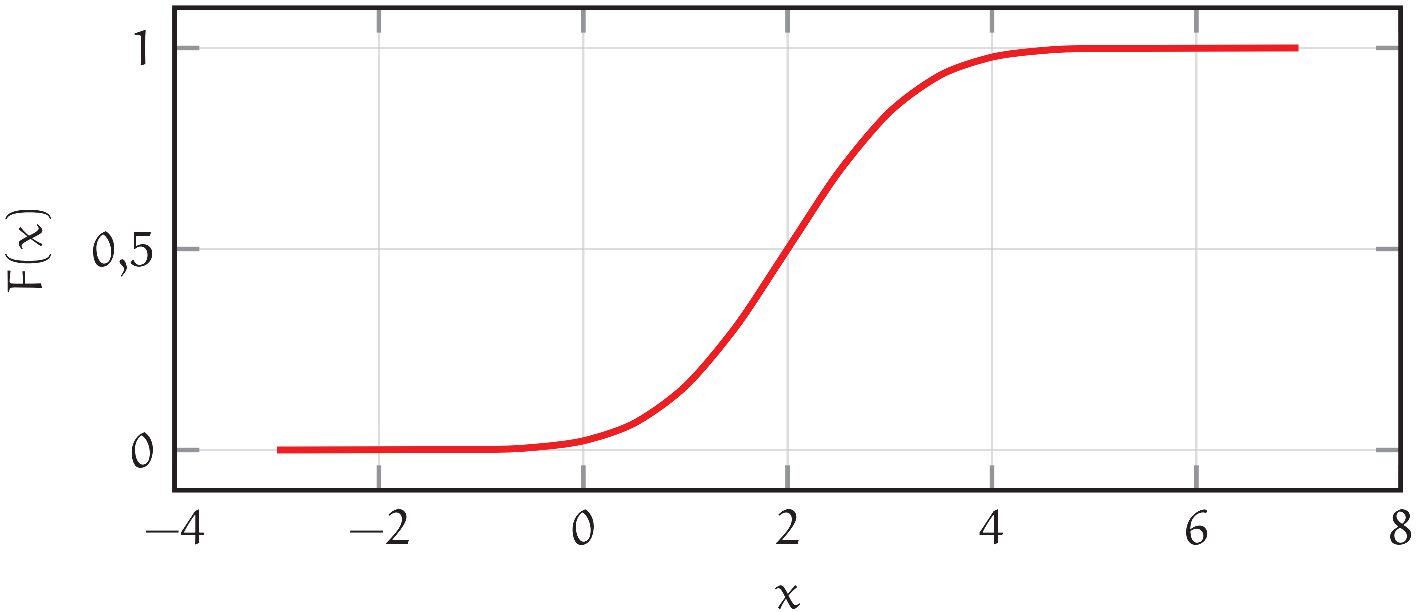
Verteilungsfunktion

- Zuweisung von Wahrscheinlichkeiten zu Realisationen
- Formal:

$$F(x) = P(X \leqq x)$$

- Eigenschaften der Verteilungsfunktion:
  - $F(x) \in [0; 1]$
  - Definitionsbereich: $\mathbb{R}$ mit $F(-\infty) = 0$, $F(\infty) = 1$
  - monoton wachsend, d.h. $x_1 < x_2 \Rightarrow F(x_1) \leqq F(x_2)$
  - Es gilt:

$$P(a < X \leqq b) = F(b) - F(a)$$



Beispiel einer Verteilungsfunktion

Diskrete Zufallsvariablen

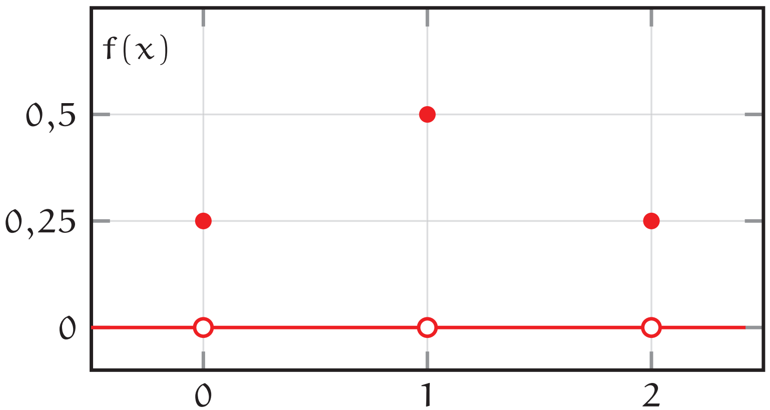
- X heißt diskret, wenn $X(\Omega) = \{x_1, x_2, \dots\}$ endlich ist.
- Wahrscheinlichkeitsfunktion dann:

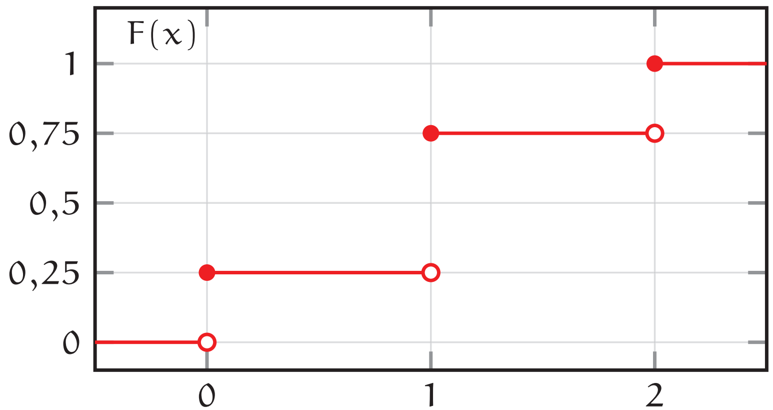
$$f(x) = P(X = x)$$

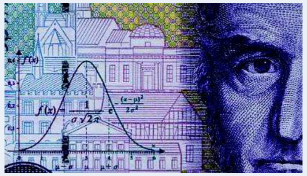
Beispiel: Münze 2 mal werfen; X: Anzahl "'Kopf"'

	(Z,Z)	$(Z,K),\ (K,Z)$	(K,K)
x_i	0	1	2
$f(x_i)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

$$F(x) = \begin{cases} 0, & \text{falls } \quad x < 0 \\ \frac{1}{4}, & \text{falls } 0 \leqq x < 1 \\ \frac{3}{4}, & \text{falls } 1 \leqq x < 2 \\ 1, & \text{falls } \quad x \geqq 2 \end{cases}$$







1. Einführung
2. Deskriptive Statistik
3. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
4. Induktive Statistik
Quellen
Tabellen

Binomialverteilung

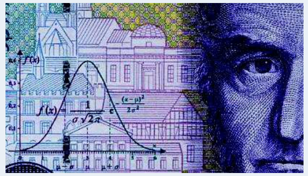
- **Wiederholter** Zufallsvorgang
- n Durchführungen (jeweils unabhängig)
- Pro Durchführung: A oder $\bar{A}$ mit $P(A) = p$ ($\widehat{=}$ Ziehen mit Zurücklegen)
- Schreibe:

$$X_i = \begin{cases} 1, & \text{falls } A \text{ bei } i\text{-ter Durchführung eintritt} \\ 0, & \text{falls } \bar{A} \text{ bei } i\text{-ter Durchführung eintritt} \end{cases}$$

- Dann gibt

$$X = \sum_{i=1}^{n} X_i$$

an, wie oft A eintritt.
- Gesucht: Wahrscheinlichkeitsfunktion von X



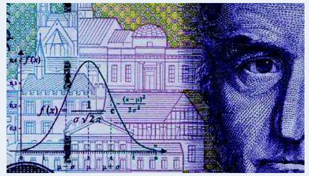
Binomialverteilung

- Herleitung:
  1) $P(X_i = 1) = P(A) = p, \;\; P(X_i = 0) = P(\bar{A}) = 1 - p$
  2) $\sum_{i=1}^{n} x_i = x$ entspricht "'x mal Ereignis A und $n - x$ mal $\bar{A}$"'
     Wahrscheinlichkeit (bei Unabhängigkeit): $p^x \cdot (1-p)^{n-x}$
  3) Aber: Reihenfolge irrelevant! Anzahl Anordnungen: $\binom{n}{x}$

➠ Wahrscheinlichkeitsfunktion der Binomialverteilung:

$$f(x) = \begin{cases} \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}, & \text{falls } x \in \{0, 1, \ldots, n\} \\ 0, & \text{sonst} \end{cases}$$

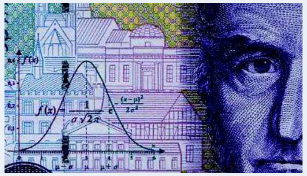
- Kurzschreibweise: $X \sim B(n; p)$
  X ist binomialverteilt mit Parametern n und p
- Tabellen zeigen meist $F(x)$
- für $f(x)$ gilt: $f(x) = F(x) - F(x-1)$



$X \sim B(n, 0.25)$, Tabelle der Binomialverteilung $F(x) = P(X \leqslant x)$

$x \backslash n$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0.7500	0.5625	0.4219	0.3164	0.2373	0.1780	0.1335	0.1001	0.0751	0.0563	0.0422	0.0317	0.0238	0.0178	0.0134
1	1.0000	0.9375	0.8438	0.7383	0.6328	0.5339	0.4450	0.3671	0.3003	0.2440	0.1971	0.1584	0.1267	0.1010	0.0802
2		1.0000	0.9844	0.9492	0.8965	0.8306	0.7564	0.6786	0.6007	0.5256	0.4552	0.3907	0.3326	0.2811	0.2361
3			1.0000	0.9961	0.9844	0.9624	0.9295	0.8862	0.8343	0.7759	0.7133	0.6488	0.5843	0.5213	0.4613
4				1.0000	0.9990	0.9954	0.9871	0.9727	0.9511	0.9219	0.8854	0.8424	0.7940	0.7415	0.6865
5					1.0000	0.9998	0.9987	0.9958	0.9900	0.9803	0.9657	0.9456	0.9198	0.8883	0.8516
6						1.0000	0.9999	0.9996	0.9987	0.9965	0.9924	0.9858	0.9757	0.9617	0.9434
7							1.0000	1.0000	0.9999	0.9996	0.9988	0.9972	0.9944	0.9897	0.9827
8								1.0000	1.0000	1.0000	0.9999	0.9996	0.9990	0.9979	0.9958
9									1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9992
10										1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
11											1.0000	1.0000	1.0000	1.0000	1.0000

$x \backslash n$	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
0	0.0100	0.0075	0.0056	0.0042	0.0032	0.0024	0.0018	0.0013	0.0010	0.0008	0.0006	0.0004	0.0003	0.0002	0.0002
1	0.0635	0.0501	0.0395	0.0310	0.0243	0.0190	0.0149	0.0116	0.0090	0.0070	0.0055	0.0042	0.0033	0.0025	0.0020
2	0.1971	0.1637	0.1353	0.1114	0.0913	0.0745	0.0607	0.0492	0.0398	0.0321	0.0258	0.0208	0.0166	0.0133	0.0106
3	0.4050	0.3530	0.3057	0.2631	0.2252	0.1917	0.1624	0.1370	0.1150	0.0962	0.0802	0.0666	0.0551	0.0455	0.0375
4	0.6302	0.5739	0.5187	0.4654	0.4149	0.3674	0.3235	0.2832	0.2467	0.2138	0.1844	0.1583	0.1354	0.1153	0.0979
5	0.8104	0.7653	0.7175	0.6678	0.6172	0.5666	0.5168	0.4685	0.4222	0.3783	0.3372	0.2990	0.2638	0.2317	0.2026
6	0.9205	0.8929	0.8610	0.8251	0.7858	0.7436	0.6994	0.6537	0.6074	0.5611	0.5154	0.4708	0.4279	0.3869	0.3481
7	0.9729	0.9598	0.9431	0.9226	0.8982	0.8701	0.8385	0.8037	0.7662	0.7265	0.6852	0.6427	0.5998	0.5568	0.5143
8	0.9925	0.9876	0.9807	0.9713	0.9591	0.9439	0.9254	0.9037	0.8787	0.8506	0.8196	0.7860	0.7502	0.7126	0.6736
9	0.9984	0.9969	0.9946	0.9911	0.9861	0.9794	0.9705	0.9592	0.9453	0.9287	0.9092	0.8868	0.8616	0.8337	0.8034
10	0.9997	0.9994	0.9988	0.9977	0.9961	0.9936	0.9900	0.9852	0.9787	0.9703	0.9599	0.9472	0.9321	0.9145	0.8943
11	1.0000	0.9999	0.9998	0.9995	0.9991	0.9983	0.9971	0.9954	0.9928	0.9893	0.9845	0.9784	0.9706	0.9610	0.9494
12	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9993	0.9988	0.9979	0.9966	0.9948	0.9922	0.9888	0.9842	0.9784
13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9997	0.9995	0.9991	0.9985	0.9976	0.9962	0.9944	0.9918
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9993	0.9989	0.9982	0.9973
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9995	0.9992
16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998
17		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
18			1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000



Beispiel

Aus einem 32-er Kartenblatt wird 3-mal eine Karte mit Zurücklegen gezogen.
Wie wahrscheinlich ist es, 2-mal Herz zu ziehen?

$$X_i = \begin{cases} 1, & \text{falls } i\text{-te Karte Herz} \\ 0, & \text{sonst} \end{cases} \quad \Rightarrow \quad X_i \sim B(1; \tfrac{8}{32})$$

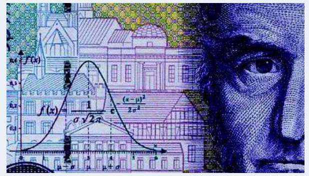
$$X = \sum_{i=1}^{n} X_i = X_1 + X_2 + X_3 \quad \Rightarrow \quad X \sim B(3; \tfrac{1}{4})$$

Mithilfe der Wahrscheinlichkeitsfunktion:

$$P(X = 2) = f(2) = \binom{3}{2} \cdot 0{,}25^2 \cdot 0{,}75^1 = 0{,}1406$$

Mithilfe der Tabelle ($n = 3$):

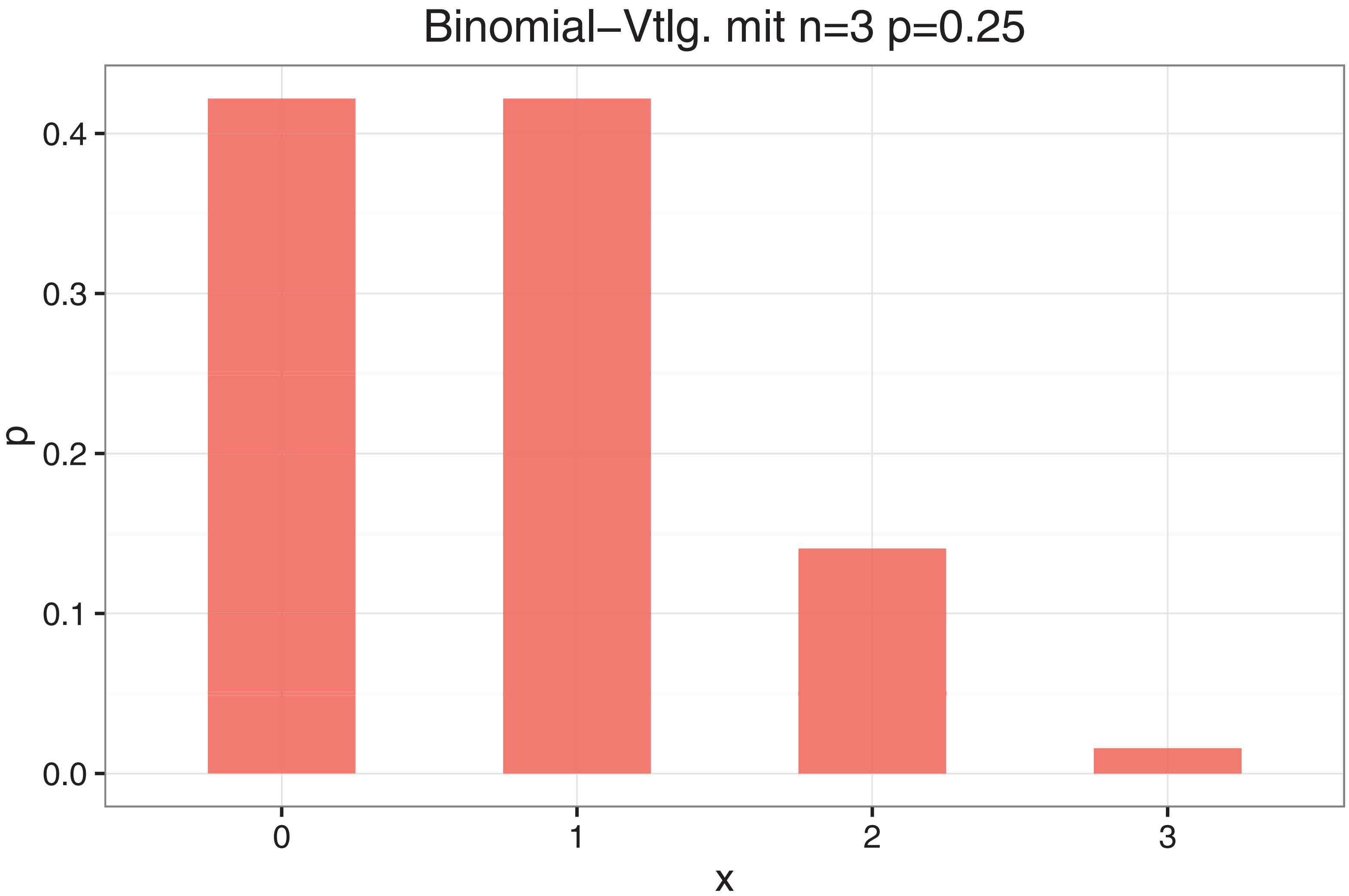
$$P(X = 2) = F(2) - F(1) = 0{,}9844 - 0{,}8438 = 0{,}1406$$

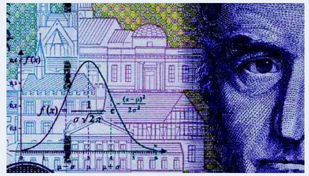


1. Einführung
2. Deskriptive Statistik
3. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
4. Induktive Statistik
Quellen
Tabellen

Binomialverteilung: Wahrscheinlichkeitsfunktion

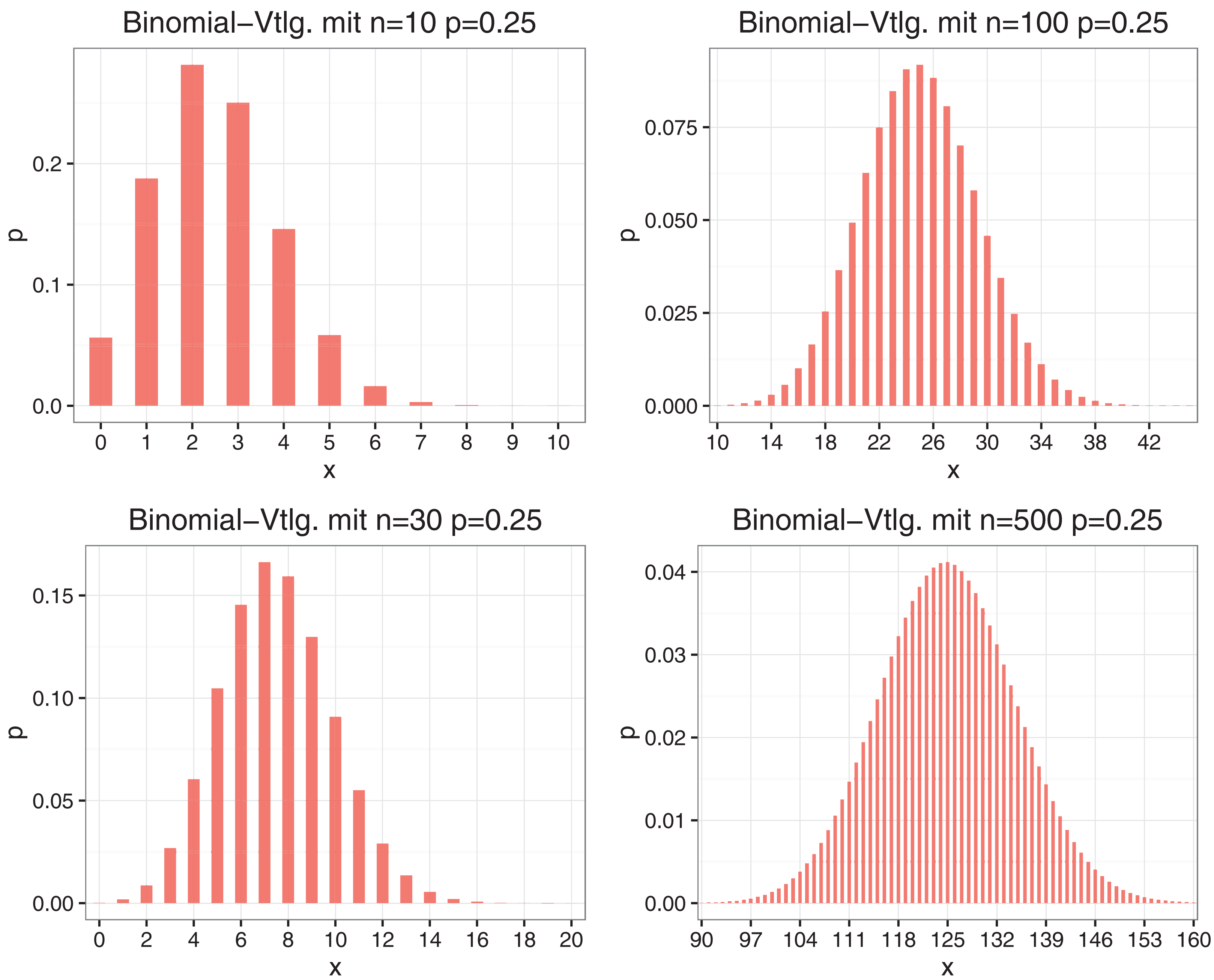
- $X \sim B(3, \frac{1}{4})$

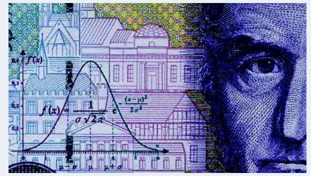




1. Einführung
2. Deskriptive Statistik
3. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
4. Induktive Statistik
Quellen
Tabellen

Binomialverteilung: Wahrscheinlichkeitsfunktion





Hypergeometrische Verteilung

- n-faches Ziehen **ohne** Zurücklegen aus N Objekten, davon M markiert.

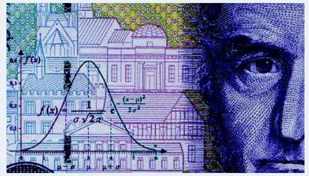
$$X = \text{Anzahl gezogener Objekte mit Markierung}$$

heißt hypergeometrisch verteilt mit den Parametern N, M, n.

- Kurzschreibweise: $X \sim \text{Hyp}(N; M; n)$
- Wahrscheinlichkeitsfunktion:

$$f(x) = \begin{cases} \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, & \text{falls } x \text{ möglich} \\ 0, & \text{sonst} \end{cases}$$

- Ist $n \leq \frac{N}{20}$, so gilt: $\text{Hyp}(N; M; n) \approx B(n; \frac{M}{N})$



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
4. Induktive Statistik
Quellen
Tabellen

Beispiel: Hypergeometrische Verteilung

- Aus einem 32-Kartenblatt wird 3-mal eine Karte ohne Zurücklegen gezogen.
- Wie wahrscheinlich ist es, 2-mal "‘Herz"’ zu ziehen?
- D.h.: $N = 32,\ M = 8,\ n = 3,\ x = 2.$

$$P(X = 2) = f(2) = \frac{\binom{8}{2}\binom{32-8}{3-2}}{\binom{32}{3}} = \frac{\binom{8}{2}\binom{24}{1}}{\binom{32}{3}} = \frac{\frac{8!}{2!\cdot 6!}\cdot 24}{\frac{32!}{3!\cdot 29!}}$$

$$= \frac{29!\cdot 8!\cdot 3!\cdot 24}{32!\cdot 6!\cdot 2!} = \frac{8\cdot 7\cdot 3\cdot 24}{32\cdot 31\cdot 30} = \frac{4032}{29760} = \frac{21}{155}$$

$$= 0{,}1355$$

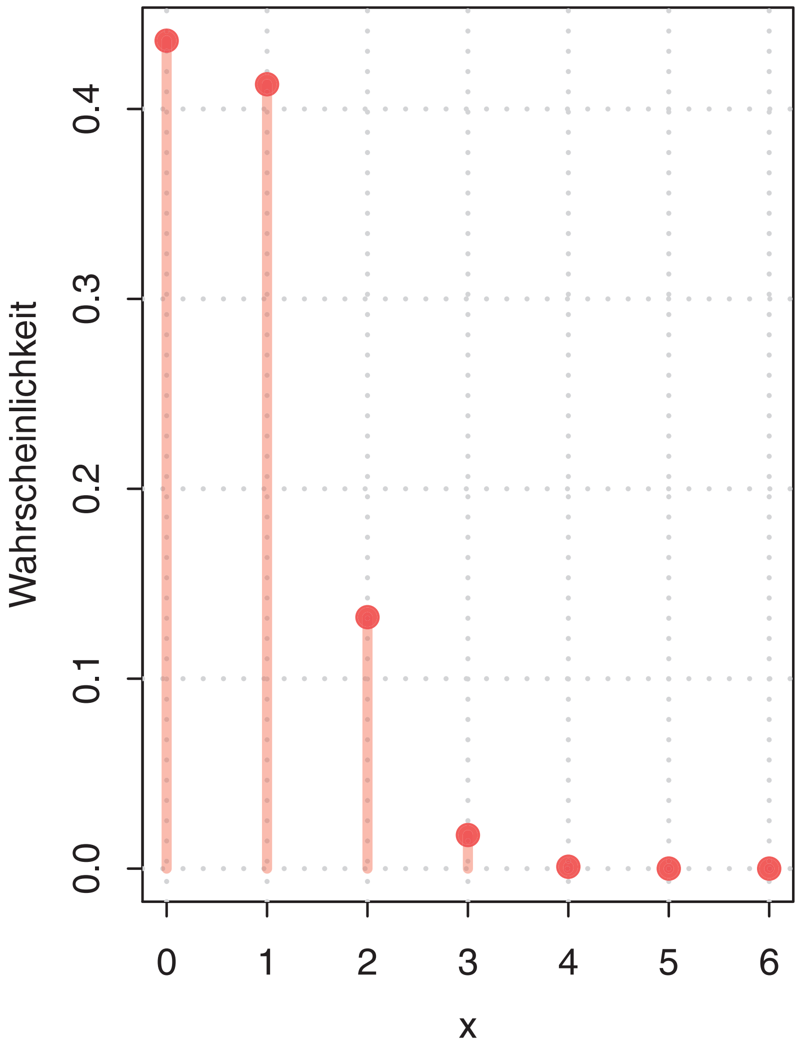
Dabei wurde verwendet:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad \text{und} \quad \binom{n}{1} = n.$$

Beispiel: x Treffer im Lotto 6 aus 49

- $X \sim \text{Hyp}(49, 6, 6)$

x	$P(X = x)$ (in %)
0	43.596498
1	41.301945
2	13.237803
3	1.765040
4	0.096862
5	0.001845
6	0.000007



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
4. Induktive Statistik
Quellen
Tabellen

Poisson-Verteilung

- Approximation für $B(n;p)$ und $\mathrm{Hyp}(N;M;n)$
- Geeignet, wenn
  p klein ($\leqq 0{,}1$), n groß ($\geqq 50$) und $np \leqq 10$.

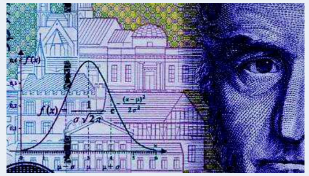
⇒ „Verteilung der seltenen Ereignisse“
(z.B. Anzahl 6-er pro Lottoausspielung)

- X ist poissonverteilt mit Parameter λ: $X \sim P(\lambda)$
- Wahrscheinlichkeitsfunktion:

$$f(x) = \begin{cases} \dfrac{\lambda^x}{x!} \cdot e^{-\lambda}, & \text{falls } x = 0,1,2,\ldots \\ 0, & \text{sonst} \end{cases}$$

- $F(x)$ in Tabelle
- Überblick: Approximation

$$\mathrm{Hyp}(N;M;n) \xrightarrow{p = \frac{M}{N}} B(n;p) \xrightarrow{\lambda = np = n\frac{M}{N}} P(\lambda)$$

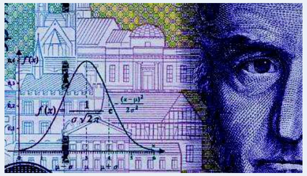


1. Einführung
2. Deskriptive Statistik
3. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
4. Induktive Statistik
Quellen
Tabellen

Poissonverteilung: $X \sim P(\lambda)$, Tabelle der Verteilungsfunktionen

$x \backslash \lambda$	1.6	1.7	1.8	1.9	2	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3
0	0.2019	0.1827	0.1653	0.1496	0.1353	0.1225	0.1108	0.1003	0.0907	0.0821	0.0743	0.0672	0.0608	0.0550	0.0498
1	0.5249	0.4933	0.4628	0.4338	0.4060	0.3796	0.3546	0.3309	0.3085	0.2873	0.2674	0.2487	0.2311	0.2146	0.1992
2	0.7834	0.7572	0.7306	0.7037	0.6767	0.6496	0.6227	0.5960	0.5697	0.5438	0.5184	0.4936	0.4695	0.4460	0.4232
3	0.9212	0.9068	0.8913	0.8747	0.8571	0.8387	0.8194	0.7994	0.7787	0.7576	0.7360	0.7141	0.6919	0.6696	0.6472
4	0.9763	0.9704	0.9636	0.9559	0.9474	0.9379	0.9275	0.9163	0.9041	0.8912	0.8774	0.8629	0.8477	0.8318	0.8153
5	0.9940	0.9920	0.9896	0.9868	0.9834	0.9796	0.9751	0.9700	0.9643	0.9580	0.9510	0.9433	0.9349	0.9258	0.9161
6	0.9987	0.9981	0.9974	0.9966	0.9955	0.9941	0.9925	0.9906	0.9884	0.9858	0.9828	0.9794	0.9756	0.9713	0.9665
7	0.9997	0.9996	0.9994	0.9992	0.9989	0.9985	0.9980	0.9974	0.9967	0.9958	0.9947	0.9934	0.9919	0.9901	0.9881
8	1.0000	0.9999	0.9999	0.9998	0.9998	0.9997	0.9995	0.9994	0.9991	0.9989	0.9985	0.9981	0.9976	0.9970	0.9962
9	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9998	0.9997	0.9996	0.9995	0.9993	0.9992	0.9989
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9998	0.9998	0.9997
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

$x \backslash \lambda$	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4	4.1	4.2	4.3	4.4	4.5
0	0.0451	0.0408	0.0369	0.0334	0.0302	0.0273	0.0247	0.0224	0.0203	0.0183	0.0166	0.0150	0.0136	0.0123	0.0111
1	0.1847	0.1712	0.1586	0.1469	0.1359	0.1257	0.1162	0.1074	0.0992	0.0916	0.0845	0.0780	0.0719	0.0663	0.0611
2	0.4012	0.3799	0.3594	0.3397	0.3209	0.3028	0.2854	0.2689	0.2531	0.2381	0.2238	0.2102	0.1974	0.1852	0.1736
3	0.6248	0.6025	0.5803	0.5584	0.5366	0.5152	0.4942	0.4735	0.4533	0.4335	0.4142	0.3954	0.3772	0.3595	0.3423
4	0.7982	0.7806	0.7626	0.7442	0.7255	0.7064	0.6872	0.6679	0.6484	0.6288	0.6093	0.5898	0.5704	0.5512	0.5321
5	0.9057	0.8946	0.8829	0.8706	0.8576	0.8441	0.8301	0.8156	0.8006	0.7851	0.7693	0.7532	0.7367	0.7199	0.7029
6	0.9612	0.9554	0.9490	0.9422	0.9347	0.9267	0.9182	0.9091	0.8995	0.8893	0.8787	0.8675	0.8558	0.8437	0.8311
7	0.9858	0.9832	0.9802	0.9769	0.9733	0.9692	0.9648	0.9599	0.9546	0.9489	0.9427	0.9361	0.9290	0.9214	0.9134
8	0.9953	0.9943	0.9931	0.9917	0.9901	0.9883	0.9863	0.9840	0.9815	0.9786	0.9755	0.9721	0.9683	0.9642	0.9598
9	0.9986	0.9982	0.9978	0.9973	0.9967	0.9960	0.9952	0.9942	0.9931	0.9919	0.9905	0.9889	0.9871	0.9851	0.9829
10	0.9996	0.9995	0.9994	0.9992	0.9990	0.9987	0.9984	0.9981	0.9977	0.9972	0.9966	0.9959	0.9952	0.9943	0.9933
11	0.9999	0.9999	0.9998	0.9998	0.9997	0.9996	0.9995	0.9994	0.9993	0.9991	0.9989	0.9986	0.9983	0.9980	0.9976
12	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9998	0.9998	0.9997	0.9997	0.9996	0.9995	0.9994	0.9992
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000



Beispiel

- $X \sim B(10\,000; 0{,}0003)$; In Tabelle der Binomialverteilung nicht vertafelt! Approximation:

$$\left.\begin{array}{rlrl} p &= 0{,}0003 &< 0{,}1 \\ n &= 10\,000 &> 50 \\ np &= 3 &< 10 \end{array}\right\} \Rightarrow B(10\,000; 0{,}0003) \approx P(3)$$

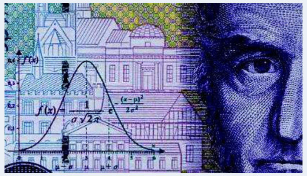
- Mithilfe der Wahrscheinlichkeitsfunktion:

$$P(X = 5) = \frac{3^5}{5!} \cdot e^{-3} = 0{,}1008188$$

- Mithilfe der Tabelle der Poissonverteilung:

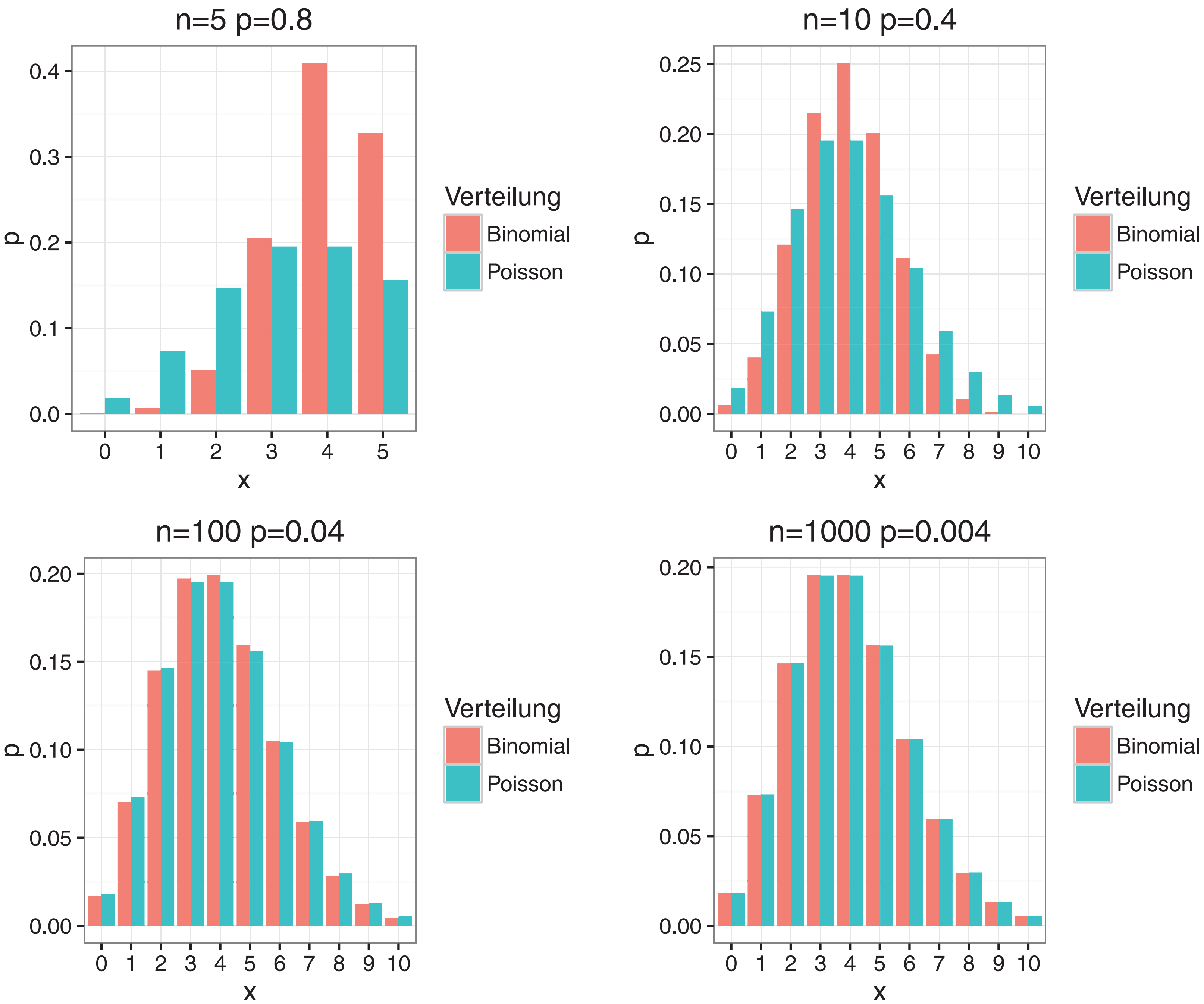
$$P(X = 5) = F(5) - F(4) = 0{,}9161 - 0{,}8153 = 0{,}1008$$

- Exakter Wert: $P(X = 5) = 0{,}1008239$

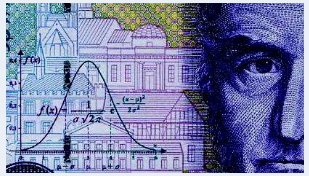


1. Einführung
2. Deskriptive Statistik
3. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
4. Induktive Statistik
Quellen
Tabellen

Poisson- versus Binomialverteilung: Vergleich



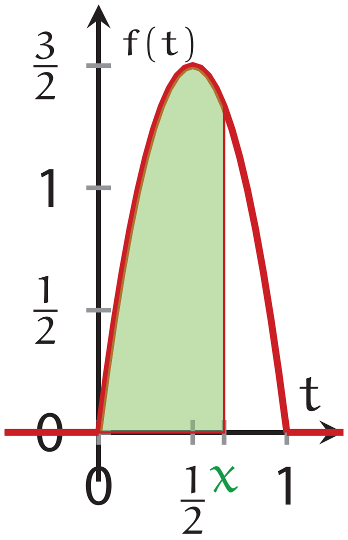
Stetige Zufallsvariablen

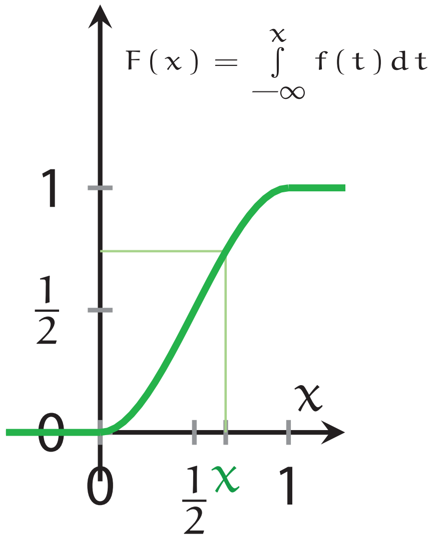


- X heißt stetig, wenn $F(x)$ stetig ist.
- Dann existiert ein $f(t)$ mit:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

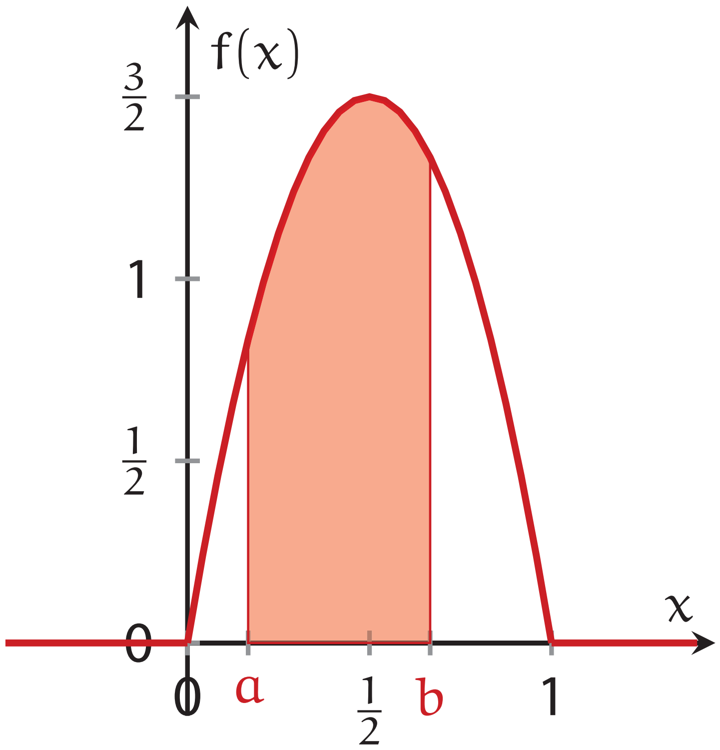
$f(x)$ heißt Dichtefunktion von X.





- Dann:

$$\begin{aligned} P(a < X < b) &= P(a \leqq X < b) \\ &= P(a < X \leqq b) \\ &= P(a \leqq X \leqq b) \\ &= \int_a^b f(x)\,dx \\ &= F(b) - F(a) \end{aligned}$$



Dichtefunktion

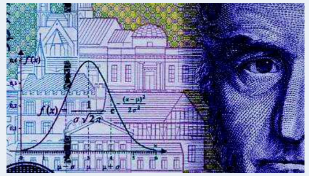
Eigenschaften der Dichtefunktion

- $f(x) \geqq 0$ für alle $x \in \mathbb{R}$
- Wegen $F(\infty) = 1$ muss stets gelten:

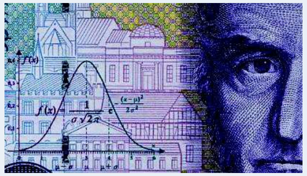
$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

- $P(X = x) = 0$ für alle $x \in \mathbb{R}$
- $f(x) > 1$ ist möglich
- für $x \in \mathbb{R}$ ist $F(x)$ differenzierbar $\Rightarrow F'(x) = f(x)$.
- Intervallgrenzen spielen keine Rolle:

$$\begin{aligned} P(X \in [a;b]) &= P(X \in (a;b]) \\ &= P(X \in [a;b)) \\ &= P(X \in (a;b)) \\ &= F(b) - F(a) \end{aligned}$$



Dichtefunktion: Beispiel



Beispiel

$$f(x) = \begin{cases} 0, & \text{falls} \quad x < 0 \\ \frac{1}{10}, & \text{falls} \quad 0 \leqq x \leqq 10 \\ 0, & \text{falls} \quad x > 10 \end{cases}$$

Verteilungsfunktion:

$$\int_0^x f(t)\,dt = \int_0^x \frac{1}{10}\,dt = \left[\frac{t}{10}\right]_0^x = \frac{x}{10} \Rightarrow$$

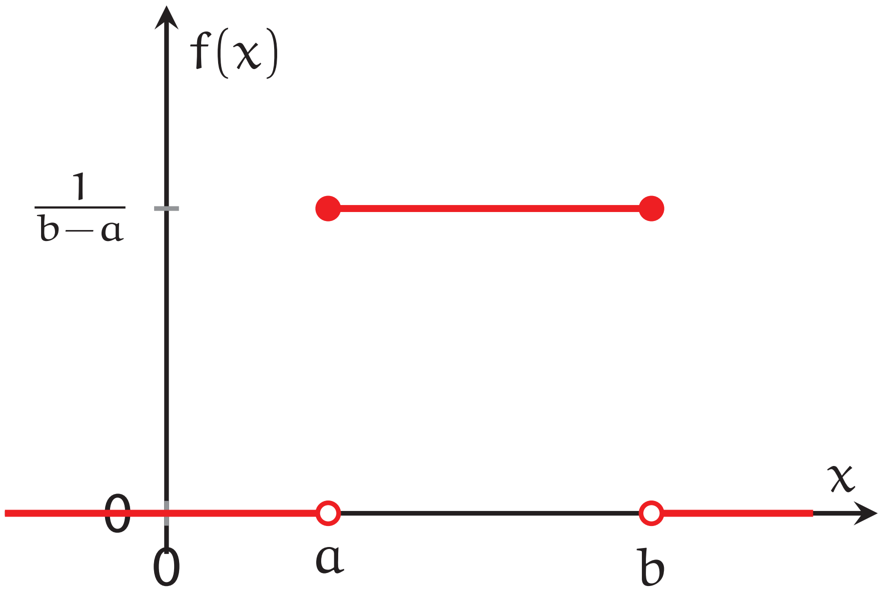
$$F(x) = \begin{cases} 0, & \text{falls} \quad x < 0 \\ \frac{x}{10}, & \text{falls} \quad 0 \leqq x \leqq 10 \\ 1, & \text{falls} \quad x > 10 \end{cases}$$

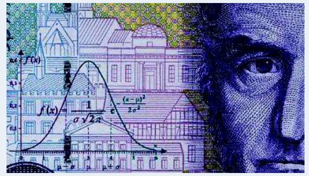
Gleichverteilung

Eine Zufallsvariable X mit

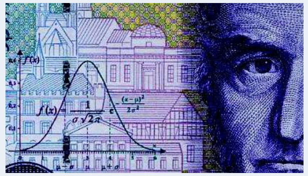
$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{falls } a \leqq x \leqq b \\ 0, & \text{sonst} \end{cases}$$

heißt gleichverteilt im Intervall $[a; b]$.





1. Einführung
2. Deskriptive Statistik
3. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
4. Induktive Statistik
Quellen
Tabellen



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
4. Induktive Statistik
Quellen
Tabellen

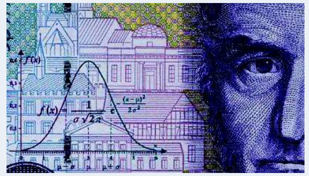
- **Verteilungsfunktion der Gleichverteilung**:

$$F(x) = \begin{cases} 0 \quad , & \text{falls} \quad x < a \\ \dfrac{x-a}{b-a}, & \text{falls} \quad a \leqq x \leqq b \\ 1 \quad , & \text{falls} \quad x > b \end{cases}$$

- **Beispiel:** X gleichverteilt in $[1;20]$

$$P(2 \leqq X \leqq 12) = F(12) - F(2) = \frac{12-1}{20-1} - \frac{2-1}{20-1}$$
$$= \frac{12-2}{20-1} = \frac{10}{19}$$
$$= 0{,}5263$$

Normalverteilung

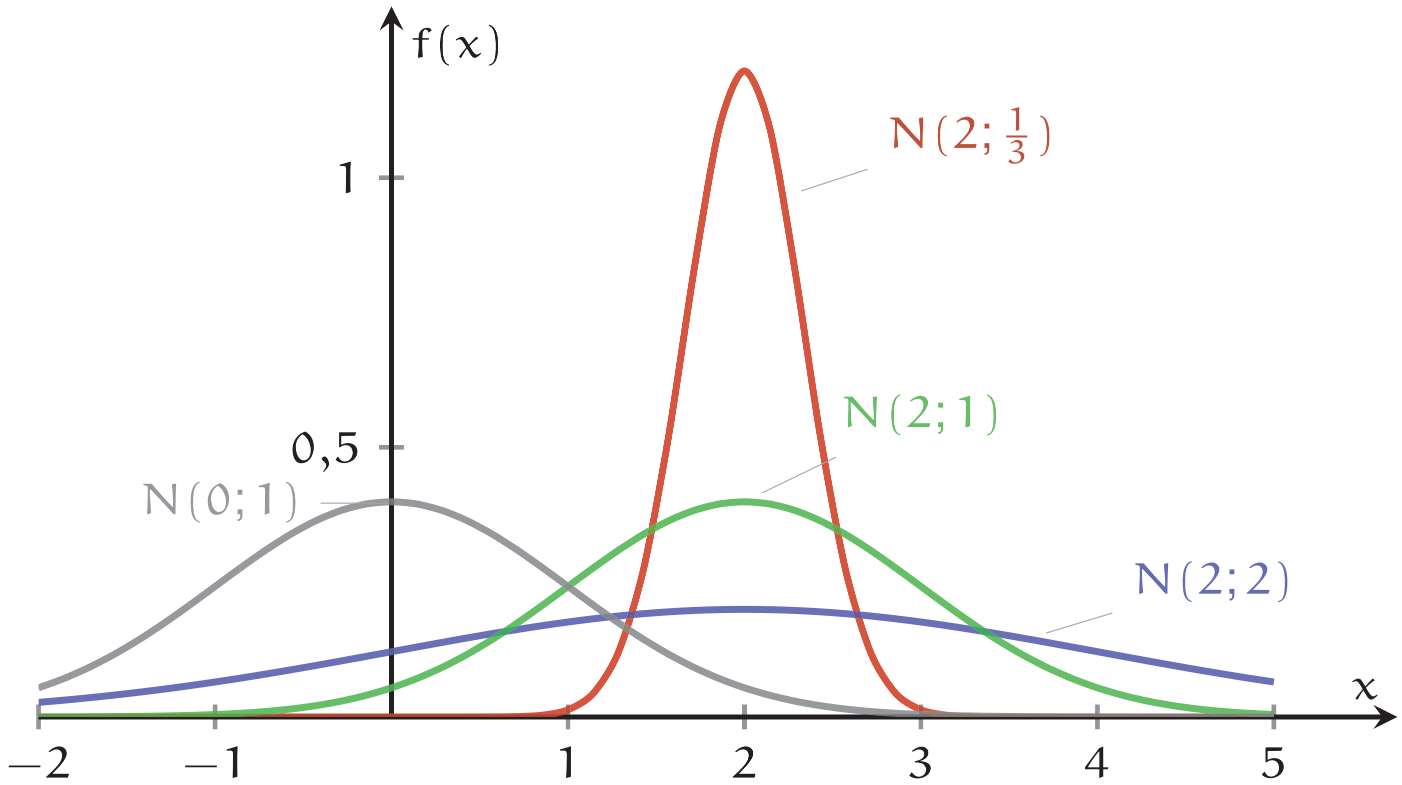


1. Einführung
2. Deskriptive Statistik
3. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
4. Induktive Statistik
Quellen
Tabellen

Eine Zufallsvariable X mit einer Dichtefunktion

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}$$

und $\sigma > 0$ heißt normalverteilt.

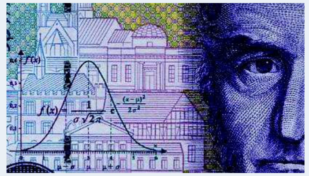


Kurzschreibweise: $X \sim N(\mu; \sigma)$

Normalverteilung: Gaußkurve

Normalverteilung

C.F. Gauß



1. Einführung

2. Deskriptive Statistik

3. W-Theorie

Kombinatorik

Zufall und Wahrscheinlichkeit

Zufallsvariablen und Verteilungen

Verteilungsparameter

4. Induktive Statistik

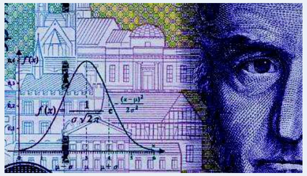
Quellen

Tabellen

Verteilungsfunktion Φ der Standardnormalverteilung

Dabei bedeutet $\Phi(x)$ zum Beispiel: $\Phi(2{,}13) = \Phi(2{,}1 + 0{,}03) = 0{,}9834$. Diesen Wert findet man in der Zeile mit $x_1 = 2{,}1$ und der Spalte mit $x_2 = 0{,}03$.

$x_1 \setminus x_2$	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5754
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6737	0.6773	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7020	0.7054	0.7089	0.7123	0.7157	0.7191	0.7224
0.6	0.7258	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7518	0.7549
0.7	0.7580	0.7612	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7882	0.7910	0.7939	0.7967	0.7996	0.8023	0.8051	0.8079	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8290	0.8315	0.8340	0.8365	0.8389
1	0.8414	0.8438	0.8461	0.8485	0.8508	0.8532	0.8554	0.8577	0.8599	0.8622
1.1	0.8643	0.8665	0.8687	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9083	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9193	0.9207	0.9222	0.9237	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9358	0.9370	0.9382	0.9394	0.9406	0.9418	0.9430	0.9441
1.6	0.9452	0.9463	0.9474	0.9485	0.9495	0.9505	0.9516	0.9526	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9600	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9679	0.9686	0.9693	0.9700	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9762	0.9767
2	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9865	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9914	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9933	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9975	0.9975	0.9976	0.9977	0.9978	0.9978	0.9979	0.9980	0.9980	0.9981
2.9	0.9981	0.9982	0.9983	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995	0.9995



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
4. Induktive Statistik
Quellen
Tabellen

Eigenschaften der Normalverteilung

- Dichte ist symmetrisch zu μ:

$$f(\mu - x) = f(\mu + x)$$

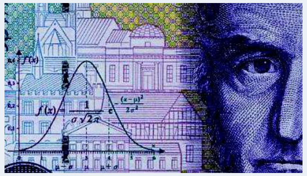
➠ μ ist Lage-, σ ist Streuungsparameter

- Standardnormalverteilung:
  $N(0;1)$ mit Verteilungsfunktion $\Phi(x)$ ($\rightarrow$ Tabelle 3)
- Kenntnis von $\Phi(x)$, μ und σ genügt, denn:
  $X \sim N(\mu;\sigma) \iff \frac{X-\mu}{\sigma} \sim N(0;1) \quad \Rightarrow$

$$\boxed{F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)}$$

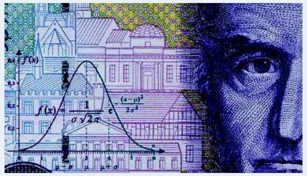
- Tabelle enthält nur positive x: Deswegen

$$\boxed{\Phi(-x) = 1 - \Phi(x)}$$



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
4. Induktive Statistik
Quellen
Tabellen

Normalverteilung: Beispiel



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
4. Induktive Statistik
Quellen
Tabellen

Beispiel:

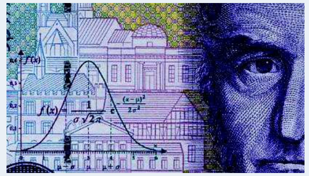
Projektdauer $X \sim N(39; 2)$.

Wahrscheinlichkeit für Projektdauer zwischen 37 und 41 Wochen?

Lösung:

$$
\begin{aligned}
P(37 \leqq X \leqq 41) &= F(41) - F(37) \\
&= \Phi\left(\tfrac{41-39}{2}\right) - \Phi\left(\tfrac{37-39}{2}\right) \\
&= \Phi(1) - \Phi(-1) \\
&= \Phi(1) - [1 - \Phi(1)] \\
&= 2 \cdot \Phi(1) - 1 \\
&= 2 \cdot 0{,}8413 - 1 \\
&= 0{,}6826
\end{aligned}
$$

Lageparameter



a) **Modus** x_{Mod}: $f(x_{\text{Mod}}) \geqq f(x)$ für alle x
(i.A. nicht eindeutig, z.B. Gleichverteilung)

Beispiele:

- Normalverteilung: $x_{\text{Mod}} = \mu$
- Diskrete Verteilung mit:

x	0	1	2
$f(x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

$\Rightarrow x_{\text{Mod}} = 1$

b) **Median** x_{Med}: $F(x_{\text{Med}}) = \frac{1}{2}$ bzw. kleinstes x mit $F(x) > \frac{1}{2}$

Beispiele:

- Normalverteilung: $x_{\text{Med}} = \mu$
- Diskrete Verteilung
  oben: $F(0) = \frac{1}{4} < \frac{1}{2}$, $F(1) = \frac{3}{4} > \frac{1}{2} \Rightarrow x_{\text{Med}} = 1$

Lageparameter: Fraktile

c) **α-Fraktil** x_α: $F(x_\alpha) = \alpha$ (für stetige Verteilungen)

Beispiel: $X \sim N(0;1)$, $Y \sim N(3;2)$

$$\begin{aligned} x_{0,975} &= & &\ 1{,}96 \qquad \text{(Tab. 3)} \\ x_{0,025} &= -x_{0,975} & &= -1{,}96 \\ y_{0,025} &= 2 \cdot x_{0,025} + 3 & &= -0{,}92 \end{aligned}$$

Hinweise:

- $x_{\text{Med}} = x_{0,5}$
- Wenn x_α nicht vertafelt $\rightarrow$ Interpolation:

$$\boxed{x_\alpha \approx x_a + (x_b - x_a) \cdot \frac{\alpha - a}{b - a}}$$

mit a : größte vertafelte Zahl $< \alpha$
b : kleinste vertafelte Zahl $> \alpha$

Beispiel: $X \sim N(0;1)$; $x_{0,6} \approx$
$0{,}25 + (0{,}26 - 0{,}25) \cdot \frac{0{,}6 - 0{,}5987}{0{,}6026 - 0{,}5987} = 0{,}2533$

Lageparameter: Erwartungswert

d) Erwartungswert $E(X)$ bzw. μ:

$$E(X) = \begin{cases} \sum\limits_i x_i f(x_i), & \text{falls } X \text{ diskret} \\ \int\limits_{-\infty}^{\infty} x f(x)\, dx, & \text{falls } X \text{ stetig} \end{cases}$$

Beispiel: Diskrete Verteilung mit

x	0	1	2
$f(x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

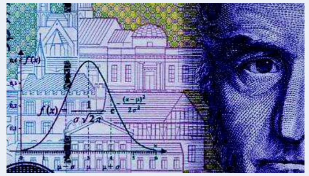
$$\Rightarrow \quad E(X) = 0 \cdot \tfrac{1}{4} + 1 \cdot \tfrac{1}{2} + 2 \cdot \tfrac{1}{4} = 1$$

Beispiel: Für eine exponentialverteilte Zufallsvariable X mit der Dichte

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & \text{für } x \geqslant 0 \\ 0 & \text{sonst} \end{cases} \qquad \text{folgt}$$

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx = \lambda \int_0^{\infty} x \cdot e^{-\lambda x}\, dx = \lambda \left[-\frac{1}{\lambda} x e^{-\lambda x} - \int_0^{\infty} 1 \cdot \left(-\frac{1}{\lambda} e^{-\lambda x} \right) dx \right]$$

$$= -x e^{-\lambda x} - \frac{1}{\lambda} e^{-\lambda x} \Big|_0^{\infty} = -0 - \left(-0 - \frac{1}{\lambda} \right) = \frac{1}{\lambda}$$



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
4. Induktive Statistik
Quellen
Tabellen

Rechenregeln für den Erwartungswert

- Ist f symmetrisch bzgl. a, so gilt $E(X) = a$
  Beispiel: f der Gleichverteilung symmetrisch bzgl. $\frac{a+b}{2} \Rightarrow E(X) = \frac{a+b}{2}$
- Lineare Transformation:

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

- Summenbildung:

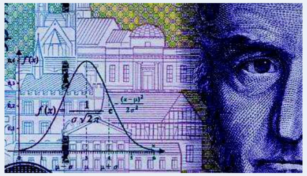
$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$

Beispiel: X gleichverteilt in $[0; 10]$, $Y \sim N(1; 1)$; $Z = X + 5Y$

$$E(Z) = E(X + 5Y) = E(X) + E(5Y) = E(X) + 5 \cdot E(Y) = \frac{10+0}{2} + 5 \cdot 1 = 10$$

- Unabhängigkeit:

$$X, Y \text{ unabhängig} \Rightarrow E(X \cdot Y) = E(X) \cdot E(Y)$$



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
4. Induktive Statistik
Quellen
Tabellen

Streuungsparameter

- Varianz $\mathrm{Var}(X)$ bzw. σ^2:

$$\mathrm{Var}(X) = E([X - E(X)]^2) = \begin{cases} \sum_i [x_i - E(X)]^2 f(x_i), & \text{wenn } X \text{ diskret} \\ \int_{-\infty}^{\infty} [x - E(X)]^2 f(x)\,dx, & \text{wenn } X \text{ stetig} \end{cases}$$

- Standardabweichung $\mathrm{Sta}(X)$ bzw. σ:

$$\mathrm{Sta}(X) = \sqrt{\mathrm{Var}(X)}$$

Beispiel: Diskrete Verteilung

x	0	1	2
$f(x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

:

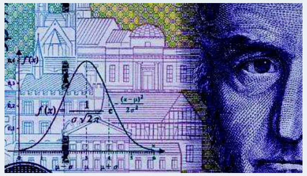
$$\mathrm{Var}(X) = (0-1)^2 \cdot \frac{1}{4} + (1-1)^2 \cdot \frac{1}{2} + (2-1)^2 \cdot \frac{1}{4} = \frac{1}{2}$$

Beispiel: Für eine exponentialverteilte Zufallsvariable X (Dichte siehe Erwartungswert) folgt

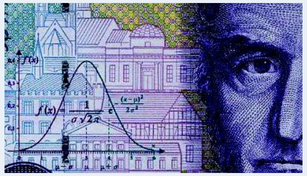
$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - E(X)) f(x)\,dx = \lambda \int_0^{\infty} \left(x - \tfrac{1}{\lambda}\right)^2 \cdot e^{-\lambda x}\,dx$$

$$= e^{-\lambda x} \left(-x^2 + \tfrac{2x}{\lambda} - \left(\tfrac{1}{\lambda}\right)^2 - \tfrac{2}{\lambda^2} - \tfrac{2x}{\lambda} + \tfrac{2}{\lambda^2} \right) \Big|_0^{\infty}$$

$$= 0 - \left(-0^2 - \left(\tfrac{1}{\lambda}\right)^2 \right) = \frac{1}{\lambda^2}$$



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
4. Induktive Statistik
Quellen
Tabellen



1. Einführung

2. Deskriptive Statistik

3. W-Theorie

Kombinatorik

Zufall und Wahrscheinlichkeit

Zufallsvariablen und Verteilungen

Verteilungsparameter

4. Induktive Statistik

Quellen

Tabellen

Rechenregeln für die Varianz

- **Verschiebungssatz**:

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2$$

Beispiel: Diskrete Verteilung

x	0	1	2
$f(x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

:

$$\begin{aligned} E(X^2) &= 0^2 \cdot \frac{1}{4} + 1^2 \cdot \frac{1}{2} + 2^2 \cdot \frac{1}{4} \\ &= \frac{3}{2} \\ \Rightarrow \quad E(X^2) - [E(X)]^2 &= \frac{3}{2} - 1^2 = \frac{1}{2} = \mathrm{Var}(X) \end{aligned}$$

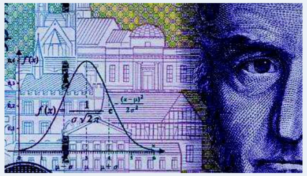
- **Lineare Transformation**:

$$\mathrm{Var}(a + bX) = b^2\,\mathrm{Var}(X)$$

- **Summenbildung** gilt nur, wenn die X_i unabhängig! Dann:

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

Erwartungswerte und Varianzen wichtiger Verteilungen



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
4. Induktive Statistik
Quellen
Tabellen

Verteilung von X	$E(X)$	$Var(X)$
Binomialverteilung $B(n;p)$	np	$np(1-p)$
Hypergeometrische Verteilung mit den Parametern N, M, n	$n\frac{M}{N}$	$n\frac{M}{N}\frac{N-M}{N}\frac{N-n}{N-1}$
Poisson-Verteilung $P(\lambda)$	λ	λ
Gleichverteilung in $[a;b]$ mit $a < b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normalverteilung $N(\mu;\sigma)$	μ	σ^2

Anwendung: Ungleichung von Tschebyschow

- Für beliebige Zufallsvariablen X und $\varepsilon > 0$ gilt die Ungleichung von Tschebyschow:

$$P\big(|X - E[X]| \geqslant \varepsilon\big) \leqslant \frac{\mathrm{Var}[X]}{\varepsilon^2}$$

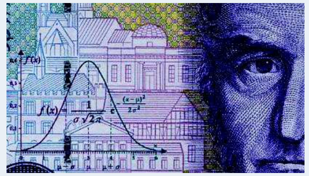
Beispiele:

- X ist gleichverteilt mit Parametern a, b und $\varepsilon = \frac{1}{3}(a-b)$, also $E[X] = \frac{1}{2}(a+b)$ und $\mathrm{Var}[X] = \frac{1}{12}(a-b)^2$

$$\Rightarrow P\Big(\big|X - \tfrac{1}{2}(a+b)\big| \geqslant \tfrac{1}{3}(a-b)\Big) \leqslant \frac{(a-b)^2}{12} \cdot \frac{3^2}{(a-b)^2} = 3/4$$

- $X \sim B(100; 0{,}2)$ und $\varepsilon = 10$
damit: $E[X] = 100 \cdot 0{,}2 = 20$ und $\mathrm{Var}[X] = 100 \cdot 0{,}2 \cdot (1 - 0{,}2) = 16$

$$\Rightarrow P\big(|X - 20| \geqslant 10\big) \leqslant \frac{16}{10^2} = 0{,}16$$



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter
4. Induktive Statistik
Quellen
Tabellen

Kovarianz und Korrelation

- **Kovarianz:**

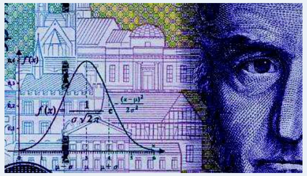
$$\begin{aligned} \mathrm{Cov}(X,Y) &= E[(X-E(X))(Y-E(Y))] \\ &= E(X\cdot Y) - E(X)\cdot E(Y) \\ &\text{(Verschiebungssatz)} \end{aligned}$$

- **Korrelationskoeffizient:**

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\cdot \mathrm{Var}(Y)}}$$

- **Bemerkungen:**
  - ρ ist r nachgebildet $\Rightarrow \rho \in [-1;1]$
  - $|\rho| = 1 \iff Y = a + bX$ (mit $b \neq 0$)
  - $\rho = 0 \iff X,\ Y$ **unkorreliert**
- Varianz einer Summe zweier ZV:

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$$



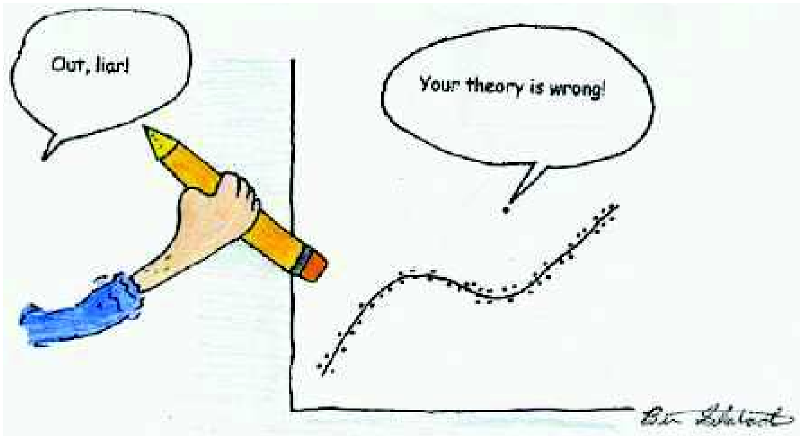
Statistik: Table of Contents

1 Statistik: Einführung

2 Deskriptive Statistik

3 Wahrscheinlichkeitstheorie

4 Induktive Statistik



4 Induktive Statistik

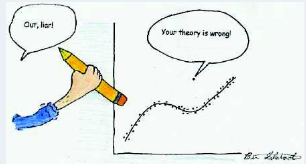
- Grundlagen
- Punkt-Schätzung
- Intervall-Schätzung
- Signifikanztests

- Vollerhebung of unmöglich,
- Deshalb: Beobachte Teilgesamtheit und schließe auf Grundgesamtheit

Beispiel

Warensendung von 1000 Stück; darunter M Stück Ausschuss.
M ist unbekannt.
$\rightarrow$ Zufällige Entnahme von $n = 30$ Stück („Stichprobe").
Darunter 2 Stück Ausschuss.
Denkbare Zielsetzungen:

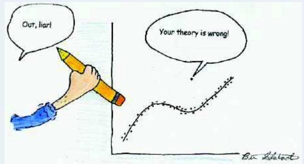
- Schätze M durch eine Zahl (z.B. $\frac{2}{30} \cdot 1000 = 66{,}67$)
- Schätze ein Intervall für M (z.B. $M \in [58; 84]$)
- Teste die Hypothese, dass $M > 50$ ist.



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Grundbegriffe

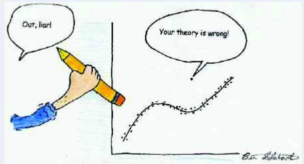
- Grundgesamtheit (G): Menge aller relevanten Merkmalsträger.
- Verteilung von G: $F(x) = P(X \leq x) =$ Wahrscheinlichkeit, dass ein Merkmalsträger ausgewählt wird, der beim untersuchten Merkmal maximal die Ausprägung x aufweist.
- Uneingeschränkte (reine) Zufallsauswahl:
  Jedes Element von G hat die selbe Chance, ausgewählt zu werden.
- Stichprobenumfang (n): Anzahl der Merkmalsträger in der Stichprobe.
- Einfache Stichprobe:
  Uneingeschränkte Zufallsauswahl und unabhängige Ziehung.
  $\rightarrow$ Alle Stichprobenvariablen $X_1, \ldots, X_n$ sind iid.
- Stichprobenergebnis:
  n-Tupel der Realisationen der Stichprobenvariablen, $(x_1, \ldots, x_n)$.



Wichtige Stichprobenfunktionen

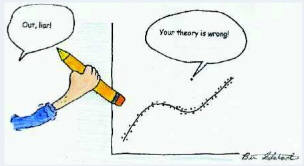
- Gegeben: Einfache Stichprobe $X_1, \ldots, X_n$, Beliebige Verteilung, mit $E(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2$

Stichprobenfunktion V	Bezeichnung	E(V)	Var(V)
$\sum_{i=1}^{n} X_i$	Merkmalssumme	$n\mu$	$n\sigma^2$
$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$	Stichprobenmittel	μ	$\frac{\sigma^2}{n}$
$\frac{\bar{X} - \mu}{\sigma} \sqrt{n}$	Gauß-Statistik	0	1
$\frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2$	mittlere quadratische Abweichung bezüglich μ	σ^2	
$\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$	mittlere quadratische Abweichung	$\frac{n-1}{n} \sigma^2$	
$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$	Stichprobenvarianz	σ^2	
$S = \sqrt{S^2}$	Stichproben-Standardabweichung		
$\frac{\bar{X} - \mu}{S} \sqrt{n}$	t-Statistik		



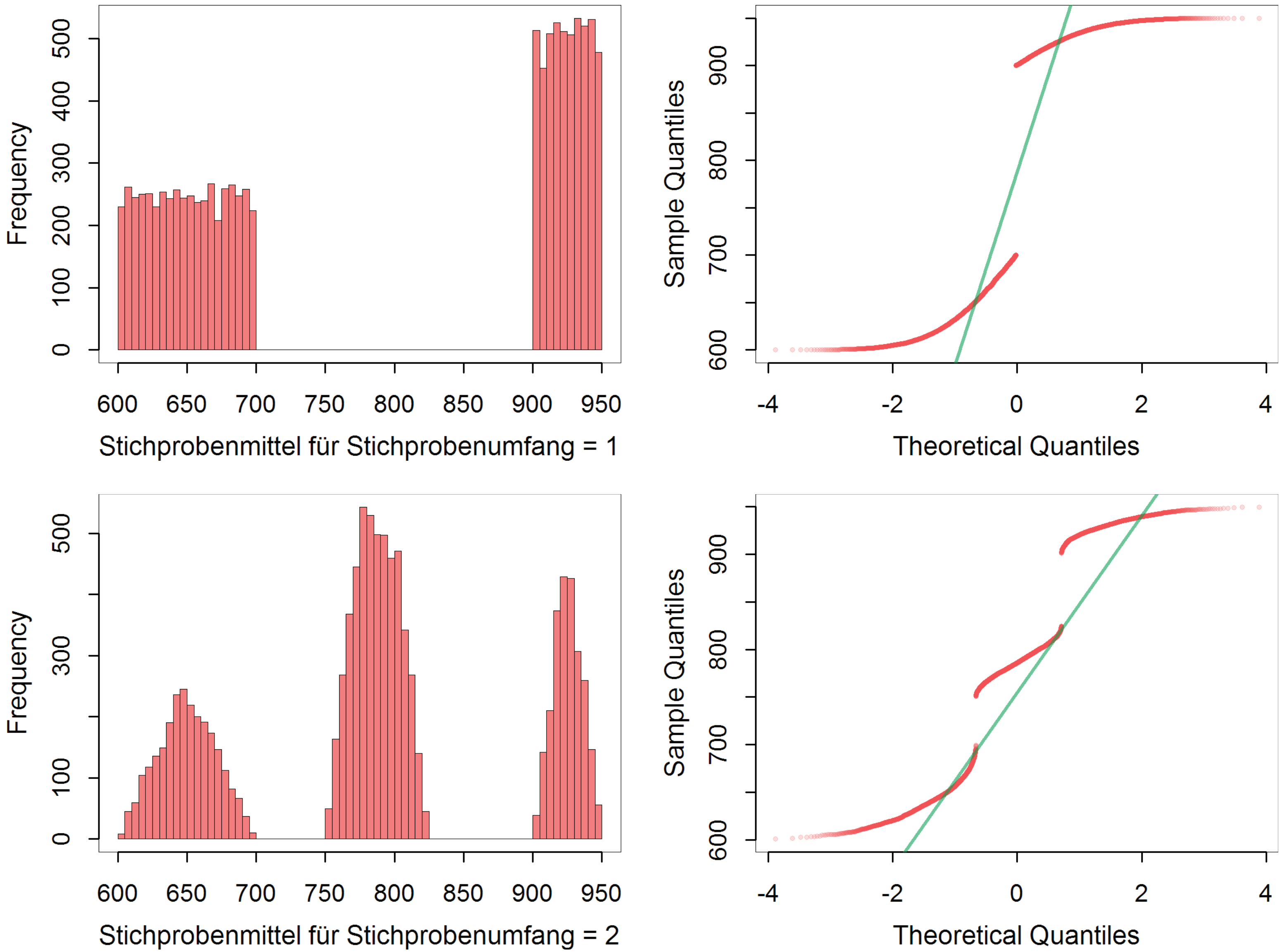
1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Auswirkungen der Stichprobengröße

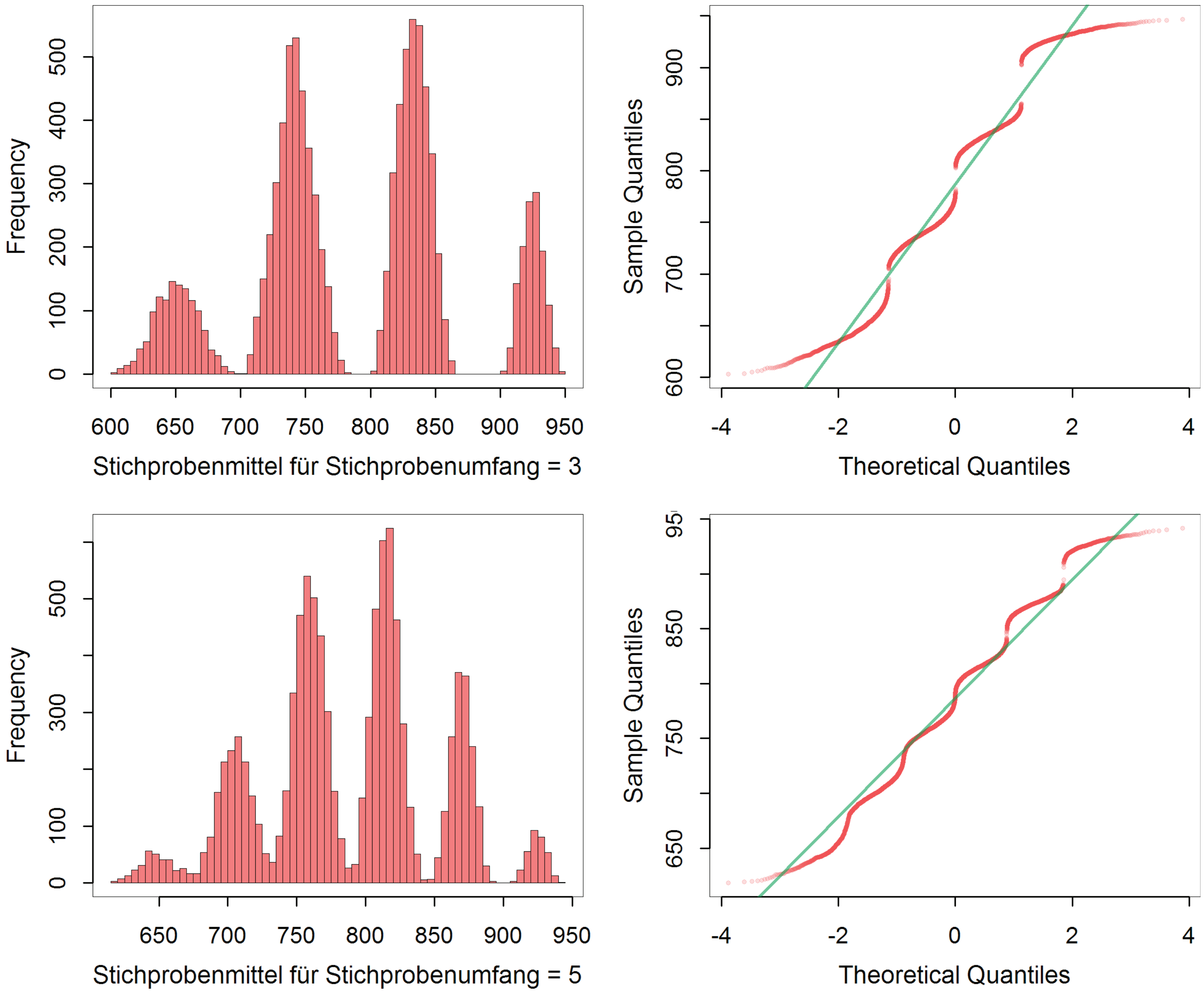


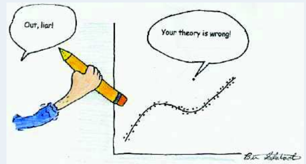
1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

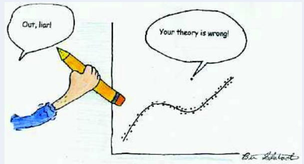
Ziehen von 10.000 Stichproben (jeweils vom Umfang n) und Berechnung der Stichprobenmittel (Verteilung: zwei überlagerte Gleichverteilungen):



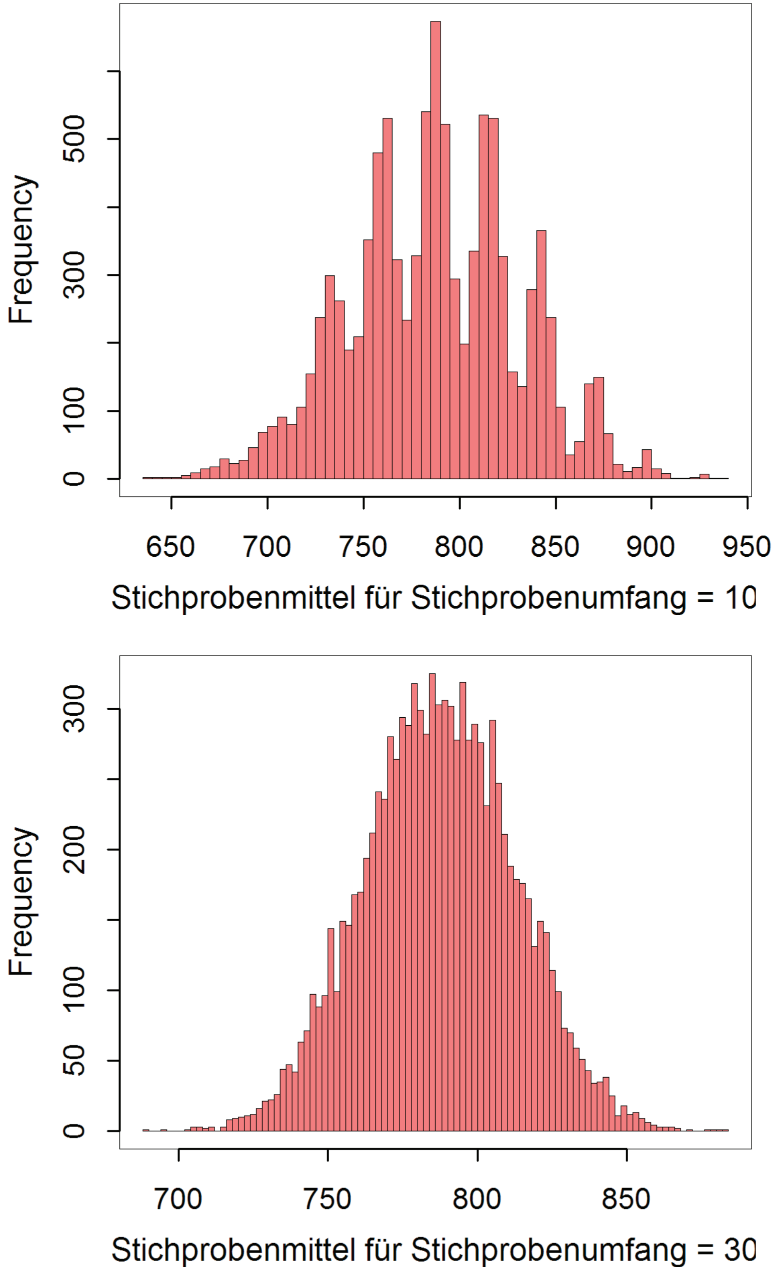
Auswirkungen der Stichprobengröße







Auswirkungen der Stichprobengröße



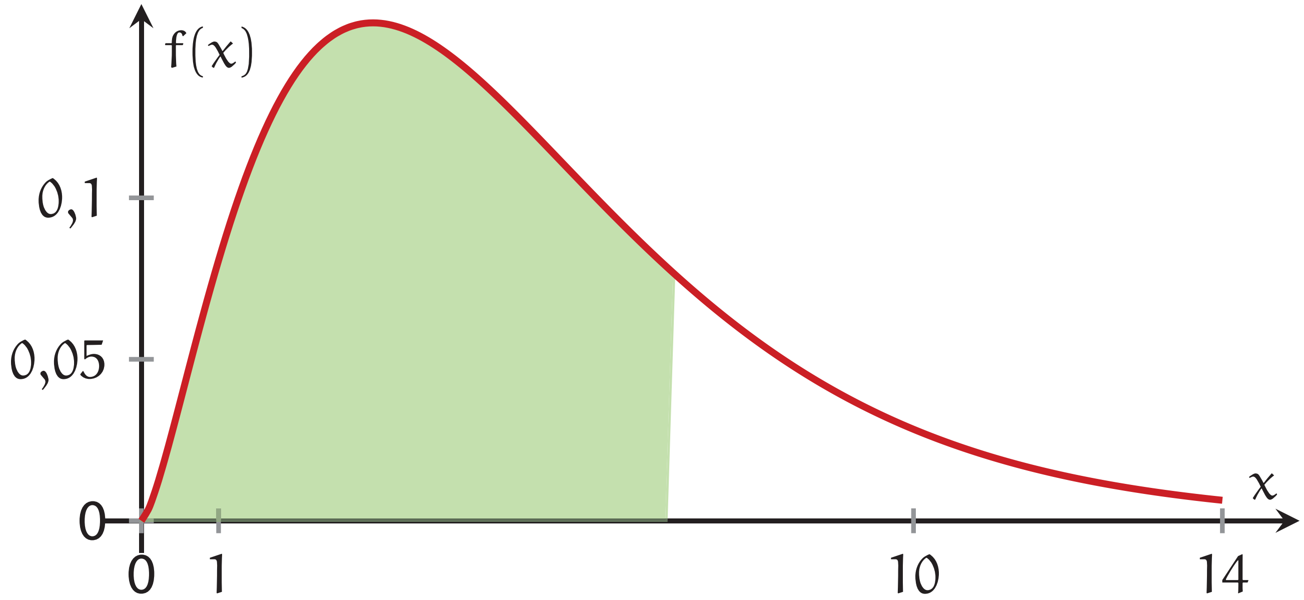
1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Chi-Quadrat-Verteilung

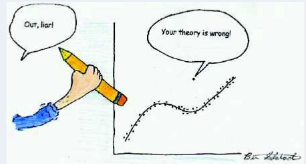
- Sind $X_1, \dots, X_n$ iid $N(0;1)$-verteilte Zufallsvariablen, so wird die Verteilung von

$$Z = \sum_{i=1}^{n} X_i^2$$

als Chi-Quadrat-Verteilung mit n Freiheitsgraden bezeichnet.



- Kurzschreibweise: $Z \sim \chi^2(n)$
- **Beispiel:** $\chi^2(30)$: $x_{0,975} = 46{,}98$

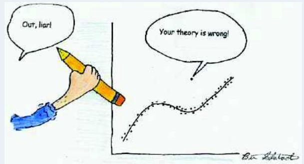


1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Quantilstabelle der χ^2-Verteilung mit n Freiheitsgraden

$\alpha \setminus n$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0.005	0.00	0.01	0.07	0.21	0.41	0.68	0.99	1.34	1.73	2.16	2.60	3.07	3.56	4.07	4.60
0.01	0.00	0.02	0.11	0.30	0.55	0.87	1.24	1.65	2.09	2.56	3.05	3.57	4.11	4.66	5.23
0.025	0.00	0.05	0.22	0.48	0.83	1.24	1.69	2.18	2.70	3.25	3.82	4.40	5.01	5.63	6.26
0.05	0.00	0.10	0.35	0.71	1.15	1.64	2.17	2.73	3.33	3.94	4.57	5.23	5.89	6.57	7.26
0.1	0.02	0.21	0.58	1.06	1.61	2.20	2.83	3.49	4.17	4.87	5.58	6.30	7.04	7.79	8.55
0.2	0.06	0.45	1.01	1.65	2.34	3.07	3.82	4.59	5.38	6.18	6.99	7.81	8.63	9.47	10.31
0.25	0.10	0.58	1.21	1.92	2.67	3.45	4.25	5.07	5.90	6.74	7.58	8.44	9.30	10.17	11.04
0.4	0.28	1.02	1.87	2.75	3.66	4.57	5.49	6.42	7.36	8.30	9.24	10.18	11.13	12.08	13.03
0.5	0.45	1.39	2.37	3.36	4.35	5.35	6.35	7.34	8.34	9.34	10.34	11.34	12.34	13.34	14.34
0.6	0.71	1.83	2.95	4.04	5.13	6.21	7.28	8.35	9.41	10.47	11.53	12.58	13.64	14.69	15.73
0.75	1.32	2.77	4.11	5.39	6.63	7.84	9.04	10.22	11.39	12.55	13.70	14.85	15.98	17.12	18.25
0.8	1.64	3.22	4.64	5.99	7.29	8.56	9.80	11.03	12.24	13.44	14.63	15.81	16.98	18.15	19.31
0.9	2.71	4.61	6.25	7.78	9.24	10.64	12.02	13.36	14.68	15.99	17.27	18.55	19.81	21.06	22.31
0.95	3.84	5.99	7.81	9.49	11.07	12.59	14.07	15.51	16.92	18.31	19.68	21.03	22.36	23.68	25.00
0.975	5.02	7.38	9.35	11.14	12.83	14.45	16.01	17.53	19.02	20.48	21.92	23.34	24.74	26.12	27.49
0.99	6.63	9.21	11.34	13.28	15.09	16.81	18.48	20.09	21.67	23.21	24.73	26.22	27.69	29.14	30.58
0.995	7.88	10.60	12.84	14.86	16.75	18.55	20.28	21.95	23.59	25.19	26.76	28.30	29.82	31.32	32.80

$\alpha \setminus n$	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
0.005	5.14	5.70	6.26	6.84	7.43	8.03	8.64	9.26	9.89	10.52	11.16	11.81	12.46	13.12	13.79
0.01	5.81	6.41	7.01	7.63	8.26	8.90	9.54	10.20	10.86	11.52	12.20	12.88	13.56	14.26	14.95
0.025	6.91	7.56	8.23	8.91	9.59	10.28	10.98	11.69	12.40	13.12	13.84	14.57	15.31	16.05	16.79
0.05	7.96	8.67	9.39	10.12	10.85	11.59	12.34	13.09	13.85	14.61	15.38	16.15	16.93	17.71	18.49
0.1	9.31	10.09	10.86	11.65	12.44	13.24	14.04	14.85	15.66	16.47	17.29	18.11	18.94	19.77	20.60
0.2	11.15	12.00	12.86	13.72	14.58	15.44	16.31	17.19	18.06	18.94	19.82	20.70	21.59	22.48	23.36
0.25	11.91	12.79	13.68	14.56	15.45	16.34	17.24	18.14	19.04	19.94	20.84	21.75	22.66	23.57	24.48
0.4	13.98	14.94	15.89	16.85	17.81	18.77	19.73	20.69	21.65	22.62	23.58	24.54	25.51	26.48	27.44
0.5	15.34	16.34	17.34	18.34	19.34	20.34	21.34	22.34	23.34	24.34	25.34	26.34	27.34	28.34	29.34
0.6	16.78	17.82	18.87	19.91	20.95	21.99	23.03	24.07	25.11	26.14	27.18	28.21	29.25	30.28	31.32
0.75	19.37	20.49	21.60	22.72	23.83	24.93	26.04	27.14	28.24	29.34	30.43	31.53	32.62	33.71	34.80
0.8	20.47	21.61	22.76	23.90	25.04	26.17	27.30	28.43	29.55	30.68	31.79	32.91	34.03	35.14	36.25
0.9	23.54	24.77	25.99	27.20	28.41	29.62	30.81	32.01	33.20	34.38	35.56	36.74	37.92	39.09	40.26
0.95	26.30	27.59	28.87	30.14	31.41	32.67	33.92	35.17	36.41	37.65	38.89	40.11	41.34	42.56	43.77
0.975	28.85	30.19	31.53	32.85	34.17	35.48	36.78	38.08	39.36	40.65	41.92	43.19	44.46	45.72	46.98
0.99	32.00	33.41	34.81	36.19	37.57	38.93	40.29	41.64	42.98	44.31	45.64	46.96	48.28	49.59	50.89
0.995	34.27	35.72	37.16	38.58	40.00	41.40	42.80	44.18	45.56	46.93	48.29	49.64	50.99	52.34	53.67



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

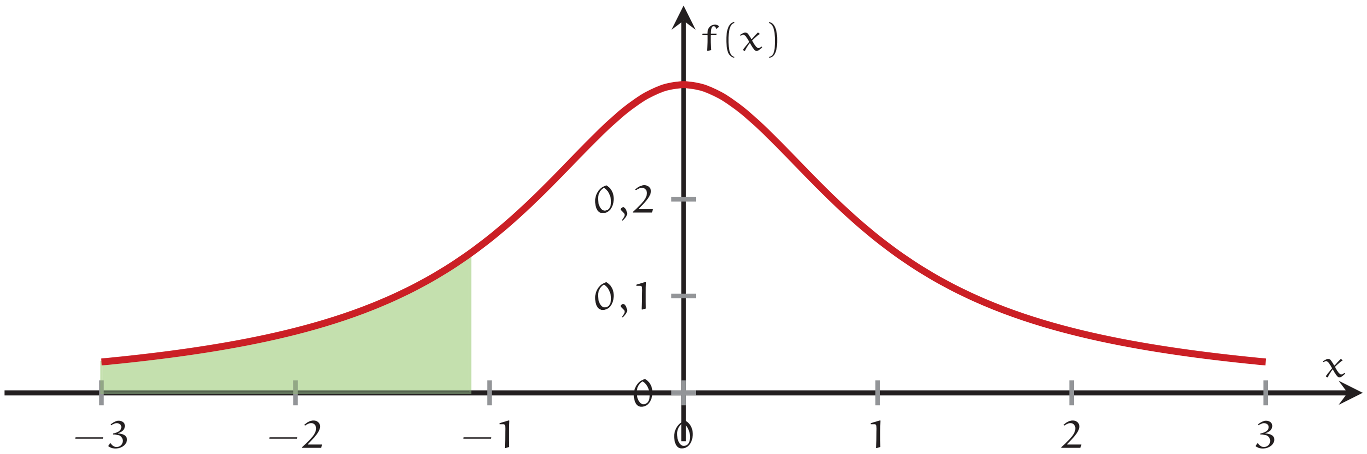
Testverteilungen: t-Verteilung

- Ist $X \sim N(0; 1)$, $Z \sim \chi^2(n)$, X, Z unabhängig, so wird die Verteilung von

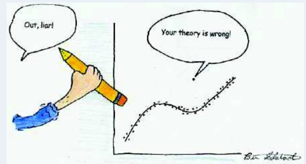
$$T = \frac{X}{\sqrt{\frac{1}{n} Z}}$$

als t-Verteilung mit n Freiheitsgraden bezeichnet.

William Sealy Gosset
1876 – 1937



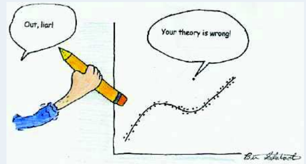
- Kurzschreibweise: $T \sim t(n)$
- **Beispiel:** $t(10)$ $x_{0,6} = 0{,}260$, $x_{0,5} = 0$, $x_{0,1} = -x_{0,9} = -1{,}372$

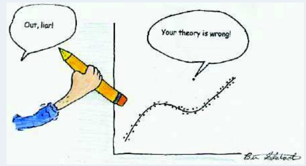


1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Quantilstabelle der t-Verteilung mit n Freiheitsgraden

$\alpha \setminus n$	0.6	0.75	0.8	0.9	0.95	0.975	0.99	0.995
1	0.325	1.000	1.376	3.078	6.314	12.706	31.820	63.657
2	0.289	0.816	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.765	0.979	1.638	2.353	3.183	4.541	5.841
4	0.271	0.741	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.727	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.718	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.711	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.706	0.889	1.397	1.860	2.306	2.897	3.355
9	0.261	0.703	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.700	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.698	0.875	1.363	1.796	2.201	2.718	3.106
12	0.259	0.696	0.873	1.356	1.782	2.179	2.681	3.054
13	0.259	0.694	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.692	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.691	0.866	1.341	1.753	2.131	2.603	2.947
16	0.258	0.690	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.689	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.688	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.688	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.687	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.686	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.686	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.685	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.685	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.684	0.856	1.316	1.708	2.059	2.485	2.787
26	0.256	0.684	0.856	1.315	1.706	2.055	2.479	2.779
27	0.256	0.684	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.683	0.855	1.312	1.701	2.048	2.467	2.763
29	0.256	0.683	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.683	0.854	1.310	1.697	2.042	2.457	2.750

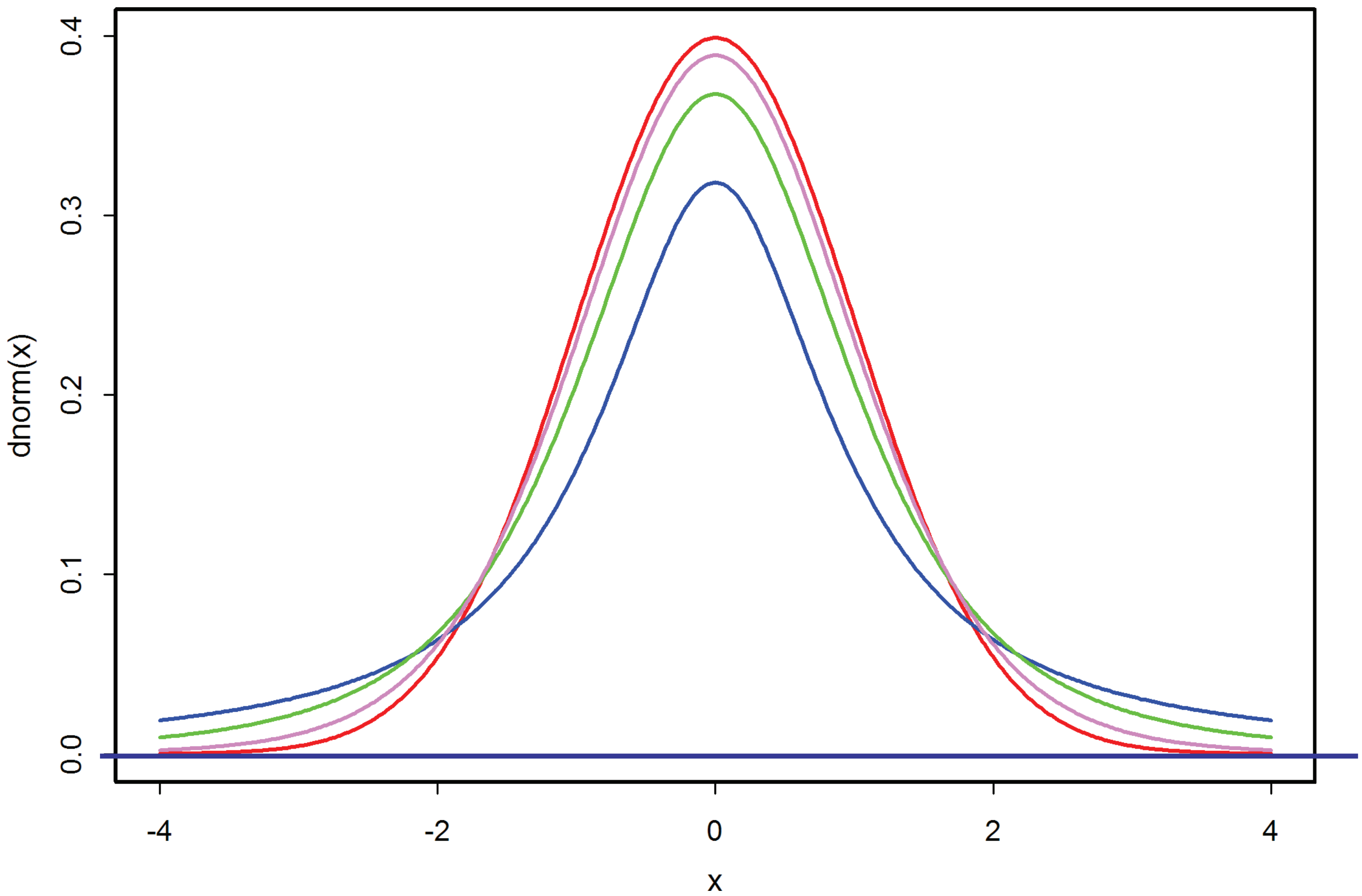




t-Verteilung vs. Normalverteilung

Dichtefunktion

- t-Verteilung mit 1 (blau), 3 (grün) und 10 (lila) Freiheitsgraden
- Standardnormalverteilung (rot)



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Punkt-Schätzung

- Ein unbekannter Parameter ϑ der Verteilung von G soll auf Basis einer Stichprobe geschätzt werden.
- Zum Beispiel: σ von $N(10;\sigma)$
- Schätzwert: $\hat{\vartheta}$
- Vorgehen: Verwendung einer Schätzfunktion

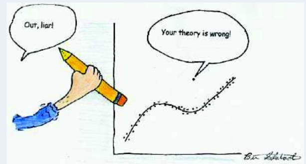
$$\hat{\Theta} = g(X_1, \ldots, X_n)$$

Beachte: Der Schätzwert $\hat{\vartheta}$ ist die Realisierung der ZV (!) $\hat{\Theta}$.

- Frage: Welche Stichprobenfunktion ist zur Schätzung geeignet?

➠ Kriterien für die Beurteilung/Konstruktion von Schätzfunktionen!

- Im Folgenden: Vorliegen einer einfachen Stichprobe, d.h. $X_1, \ldots, X_n$ iid.

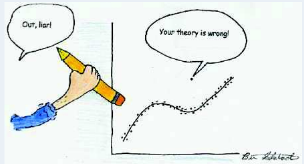


1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Beispiel

- Schätzen des Mittelwertes einer Grundgesamtheit
- dazu: Einfache Stichprobe vom Umfang 5
- und den beiden Stichprobenfunktionen

$$\hat{\Theta}_1 = \frac{1}{n} \sum_{i=1}^{n} X_i, \qquad \hat{\Theta}_2 = \frac{1}{n-1} \sum_{i=1}^{n} X_i$$

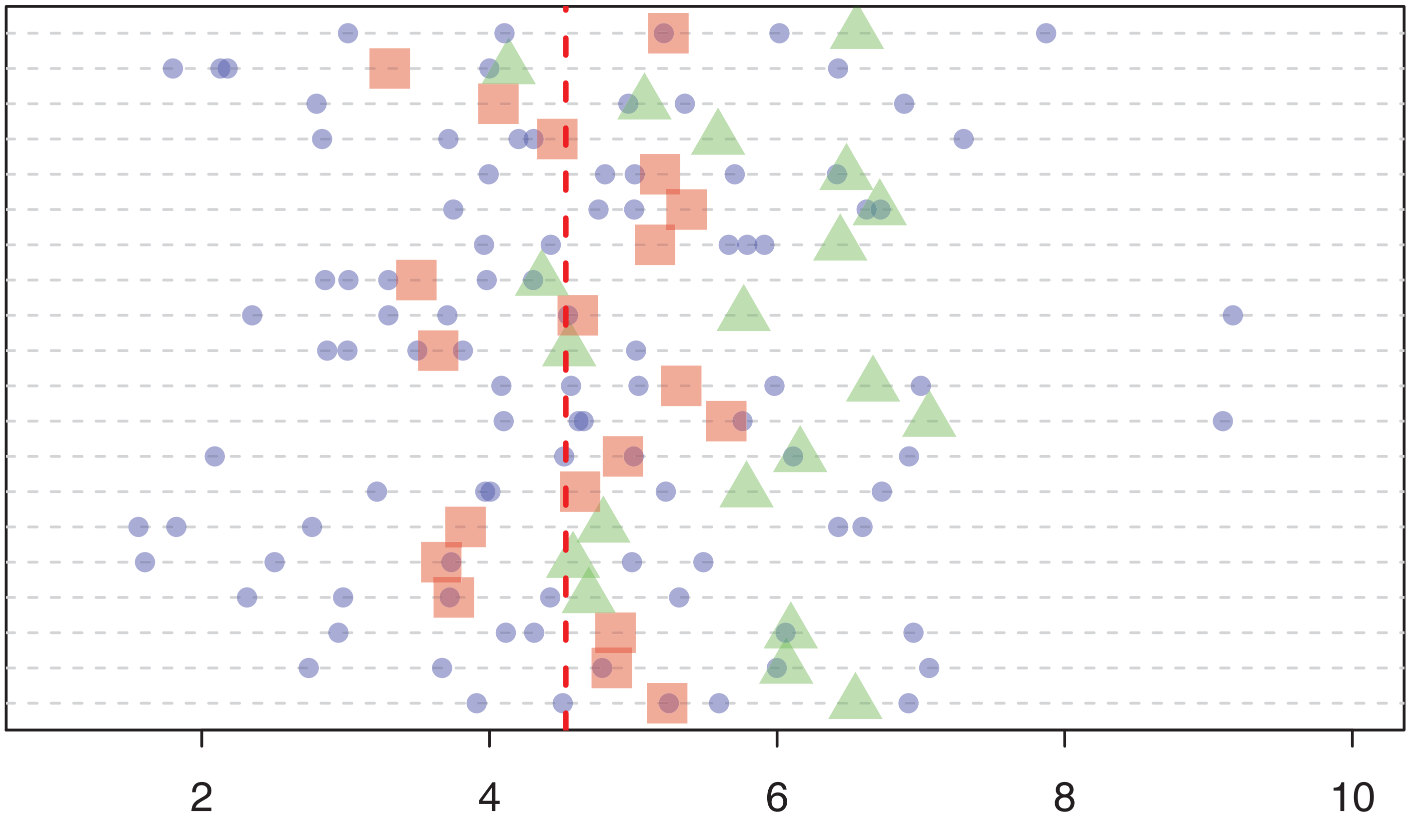


1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Beispiel

- Schätzen des Mittelwertes einer Grundgesamtheit
- dazu: Einfache Stichprobe vom Umfang 5
- und den beiden Stichprobenfunktionen

$$\hat{\Theta}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \hat{\Theta}_2 = \frac{1}{n-1}\sum_{i=1}^{n} X_i$$



Mittelwert Grundgesamtheit = 4.53

- Eine Schätzfunktion $\widehat{\Theta} = g(X_1, \ldots, X_n)$ heißt erwartungstreu oder unverzerrt für ϑ, wenn unabhängig vom numerischen Wert von ϑ gilt:

$$E(\widehat{\Theta}) = \vartheta$$

Beispiel

Sind $\widehat{\Theta}_1 = \bar{X}$, $\widehat{\Theta}_2 = \frac{X_1 + X_n}{2}$, $\widehat{\Theta}_3 = \frac{1}{n-1} \sum_{i=1}^{n} X_i$ erwartungstreu für μ?

a) $\widehat{\Theta}_1$: $\quad E(\bar{X}) = \mu$

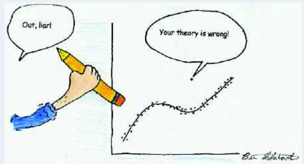
$\Rightarrow \widehat{\Theta}_1$ ist erwartungstreu.

b) $\widehat{\Theta}_2$: $\quad E\left(\frac{X_1 + X_n}{2}\right) = \frac{1}{2}[E(X_1) + E(X_n)] = \frac{1}{2}(\mu + \mu) = \mu$

$\Rightarrow \widehat{\Theta}_2$ ist erwartungstreu.

c) $\widehat{\Theta}_3$: $E\left(\frac{1}{n-1} \sum_{i=1}^{n} X_i\right) = \frac{1}{n-1} \sum_{i=1}^{n} E(X_i) = \frac{1}{n-1} \sum_{i=1}^{n} \mu = \frac{n}{n-1} \mu \neq \mu$

$\Rightarrow \widehat{\Theta}_3$ ist nicht erwartungstreu



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

- Welche der erwartungstreuen Schätzfunktionen $\hat{\Theta}_1, \hat{\Theta}_2$ ist „besser"?
- Von zwei erwartungstreuen Schätzfunktionen $\hat{\Theta}_1$, $\hat{\Theta}_2$ für ϑ heißt $\hat{\Theta}_1$ wirksamer als $\hat{\Theta}_2$, wenn unabhängig vom numerischen Wert von ϑ gilt:

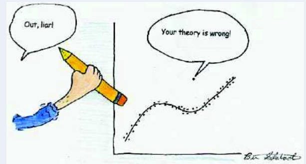
$$\mathrm{Var}(\hat{\Theta}_1) < \mathrm{Var}(\hat{\Theta}_2)$$

Beispiel: ($\hat{\Theta}_1 = \bar{X}$, $\hat{\Theta}_2 = \frac{X_1 + X_n}{2}$)
Wegen

$$\left.\begin{aligned} \mathrm{Var}(\hat{\Theta}_1) &= \mathrm{Var}(\bar{X}) && = \frac{\sigma^2}{n} \\ \mathrm{Var}(\hat{\Theta}_2) &= \mathrm{Var}\left(\frac{X_1 + X_n}{2}\right) = \frac{1}{4}(\sigma^2 + \sigma^2) && = \frac{\sigma^2}{2} \end{aligned}\right\} \Rightarrow \mathrm{Var}(\hat{\Theta}_1) < \mathrm{Var}(\hat{\Theta}_2)$$

(falls $n > 2$) ist $\hat{\Theta}_1$ wirksamer als $\hat{\Theta}_2$.

Statistik



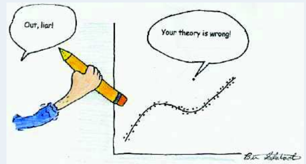
1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Intervall-Schätzung

- Für einen unbekannten Verteilungsparameter ϑ soll auf Basis einer Stichprobe ein Intervall geschätzt werden.
- Verwendung der Stichprobenfunktionen $V_u, V_o,$ so dass $V_u \leqq V_o$ und

$$P(V_u \leqq \vartheta \leqq V_o) = 1 - \alpha$$

  stets gelten.
  $[V_u; V_o]$ heißt Konfidenzintervall (KI) für ϑ zum Konfidenzniveau $1 - \alpha$.
- Beachte: Das Schätzintervall $[v_u; v_o]$ ist Realisierung der Zufallsvariablen (!) V_u, V_o.
  ➠ Irrtumswahrscheinlichkeit α (klein, i.d.R. $\alpha \leqq 0{,}1$)
- Frage: Welche Konfidenzintervalle sind zur Schätzung geeignet?
  ➠ Hängt von Verteilung von G sowie vom unbekannten Parameter (μ, σ^2) ab!
- Im Folgenden: Einfache Stichprobe $X_1, \ldots, X_n$ mit $E(X_i) = \mu$, $Var(X_i) = \sigma^2$

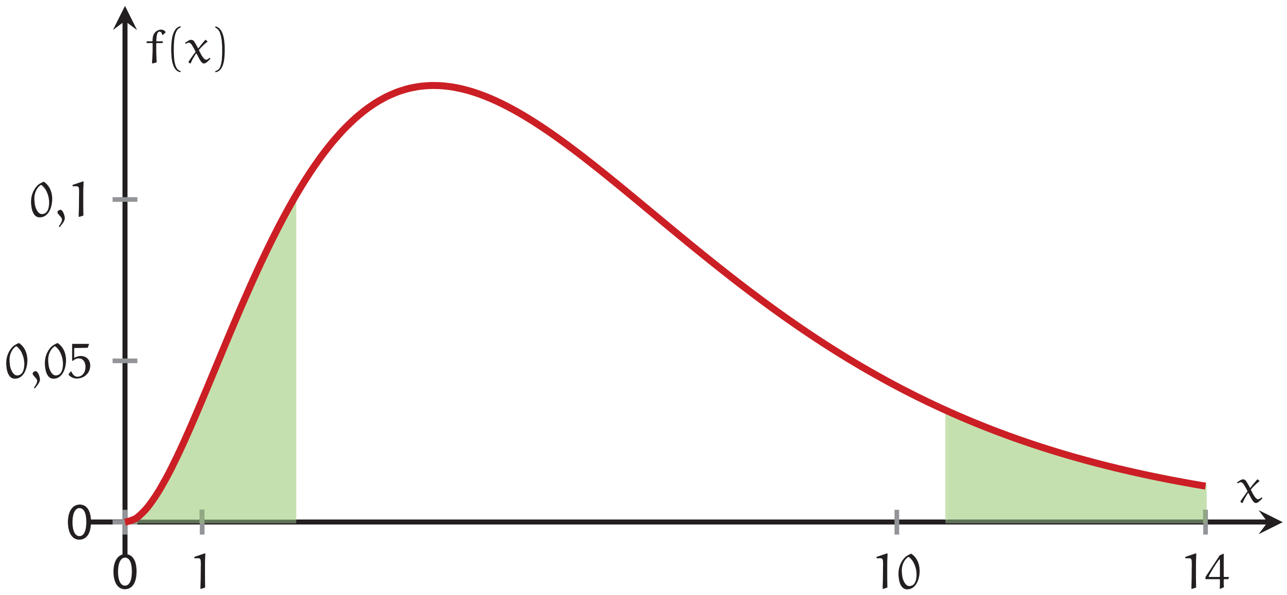


1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

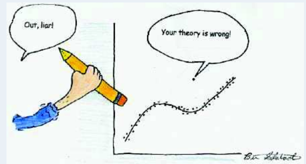
Wichtiger Spezialfall: Symmetrische Konfidenzintervalle

- Symmetrisch heißt **nicht**, dass die Dichte symmetrisch ist, sondern
- übereinstimmende Wahrscheinlichkeiten für Über-/Unterschreiten des Konfidenzintervalls, d.h.

$$P(V_u > \vartheta) = P(V_o < \vartheta) = \frac{\alpha}{2}$$

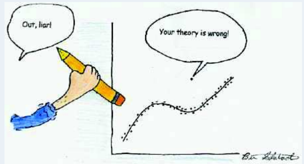


- **Wichtig:** Eine Verkleinerung von α bewirkt eine Vergrößerung des Konfidenzintervalls.



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Konfidenzintervall für μ bei Normalverteilung mit bekanntem σ^2



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Vorgehensweise:

1. Festlegen des Konfidenzniveaus $1 - \alpha$
2. Bestimmung des $\left(1 - \frac{\alpha}{2}\right)$-Fraktils c der $\mathrm{N}(0, 1)$-Verteilung
3. Berechnen des Stichprobenmittels $\bar{x}$
4. Berechnen des Wertes $\frac{\sigma c}{\sqrt{n}}$
5. Ergebnis der Intervall-Schätzung:

$$\left[\bar{x} - \frac{\sigma c}{\sqrt{n}}\,;\quad \bar{x} + \frac{\sigma c}{\sqrt{n}}\right]$$

Intervallschätzung: Beispiel

Beispiel

Normalverteilung mit $\sigma = 2{,}4$
$(x_1, \ldots, x_9)$ = (184.2, 182.6, 185.3, 184.5, 186.2, 183.9, 185.0, 187.1, 184.4)

Gesucht: Konfidenzintervall für μ zum Konfidenzniveau
$1 - \alpha = 0{,}99$

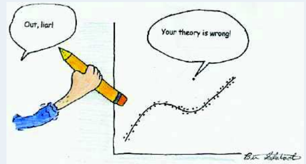
1. $1 - \alpha = 0{,}99$
2. $N(0;1)$: $c = x_{1-\frac{\alpha}{2}} = x_{1-\frac{0{,}01}{2}} = x_{0{,}995} = 2{,}576$ (Tab. 3; Interpolation)
3. $\bar{x} = \frac{1}{9}\,(184{,}2 + \cdots + 184{,}4) = 184{,}8$
4. $\frac{\sigma c}{\sqrt{n}} = \frac{2{,}4 \cdot 2{,}576}{\sqrt{9}} = 2{,}06$
5. $\mathrm{KI} = [184{,}8 - 2{,}06; 184{,}8 + 2{,}06] = [182{,}74; 186{,}86]$

Interpretation: Mit 99 % Wahrscheinlichkeit ist $\mu \in [182{,}74; 186{,}86]$.

Wichtige $N(0;1)$-Fraktilswerte:

α	x_α
0,9	1,281552
0,95	1,644854
0,975	1,959964
0,99	2,326348
0,995	2,575829

(I.d.R. genügen drei Nachkommastellen.)



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Intervalllänge

- Bei bekannter Standardabweichung gilt offenkundig

$$L = V_o - V_u = \frac{2\sigma c}{\sqrt{n}}$$

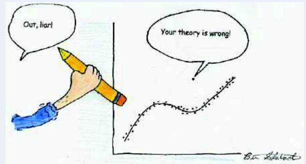
- Welcher Stichprobenumfang n sichert eine vorgegebene (Maximal-)Länge L? $\Rightarrow$ Nach n auflösen! $\Rightarrow$

$$\boxed{n \geqq \left(\frac{2\sigma c}{L}\right)^2}$$

- Eine Halbierung von L erfordert eine Vervierfachung von n!
- Angewendet auf letztes **Beispiel**:

$$L = 4 \Rightarrow n \geqq \left(\frac{2 \cdot 2{,}4 \cdot 2{,}576}{4}\right)^2 = 9{,}556 \Rightarrow n \geqq 10$$

$$L = 2 \Rightarrow n \geqq \left(\frac{2 \cdot 2{,}4 \cdot 2{,}576}{2}\right)^2 = 38{,}222 \Rightarrow n \geqq 39$$



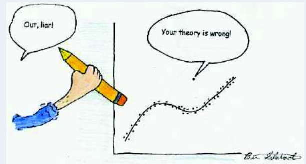
1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Konfidenzintervall für μ bei Normalverteilung mit unbekanntem σ^2

- Vorgehensweise:
  1. Festlegen des Konfidenzniveaus $1 - \alpha$
  2. Bestimmung des $\left(1 - \frac{\alpha}{2}\right)$-Fraktils c der $t(n-1)$-Verteilung
  3. Berechnen des Stichprobenmittels $\bar{x}$ und der Stichproben-Standardabweichung s
  4. Berechnen des Wertes $\frac{sc}{\sqrt{n}}$
  5. Ergebnis der Intervall-Schätzung:

$$\left[\bar{x} - \frac{sc}{\sqrt{n}} \,;\quad \bar{x} + \frac{sc}{\sqrt{n}}\right]$$

- Zu Schritt 2: Falls $n - 1 > 30$ wird die $N(0;1)$-Verteilung verwendet.

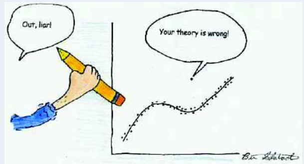


1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Beispiel:
Wie das letzte Beispiel, jedoch σ unbekannt.

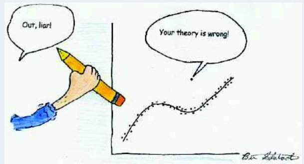
1. $1 - \alpha = 0{,}99$
2. $t(8)$: $c = x_{1-\frac{\alpha}{2}} = x_{1-\frac{0{,}01}{2}} = x_{0{,}995} = 3{,}355$ (Tab. 4)
3. $\bar{x} = \frac{1}{9}(184{,}2 + \cdots + 184{,}4) = 184{,}8$
   $s = \sqrt{\frac{1}{8}\left[(184{,}2^2 + \cdots + 184{,}4^2) - 9 \cdot 184{,}8^2\right]} = 1{,}31$
4. $\frac{sc}{\sqrt{n}} = \frac{1{,}31 \cdot 3{,}355}{\sqrt{9}} = 1{,}47$
5. $\mathrm{KI} = [184{,}8 - 1{,}47; 184{,}8 + 1{,}47] = [183{,}33; 186{,}27]$

Interpretation: Mit 99 % Wahrscheinlichkeit ist $\mu \in [183{,}33; 186{,}27]$.



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

```
x <- c(184.2, 182.6, 185.3, 184.5, 186.2,
       183.9, 185.0, 187.1, 184.4)
t.test(x,conf.level=.99)

##
##  One Sample t-test
##
## data:  x
## t = 422.1129, df = 8, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 99 percent confidence interval:
##  183.331 186.269
## sample estimates:
## mean of x
##     184.8
```

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Konfidenzintervall für μ bei beliebiger Verteilung

- Voraussetzung: $n > 30$, bzw. falls G dichotom: $5 \leqq \sum_{i=1}^{n} x_i \leqq n - 5$
- Vorgehensweise:

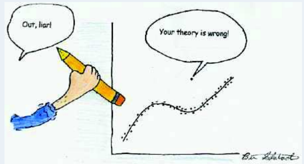
  1. Festlegen des Konfidenzniveaus $1 - \alpha$
  2. Bestimmung des $(1 - \frac{\alpha}{2})$-Fraktils c der Standardnormalverteilung $N(0; 1)$
  3. Berechnung des Stichprobenmittels $\bar{x}$ sowe eines Schätzwertes $\hat{\sigma}$ für die Standardabweichung σ der GG mittels

$$\hat{\sigma} = \begin{cases} \sigma, & \text{falls } \sigma \text{ bekannt} \\ \sqrt{\bar{x}(1 - \bar{x})}, & \text{falls GG dichotom} \\ s, & \text{sonst} \end{cases}$$

  4. Berechnung von $\frac{\hat{\sigma}c}{\sqrt{n}}$
  5. Ergebnis der Intervallschätzung:

$$\left[\bar{x} - \frac{\hat{\sigma}c}{\sqrt{n}}; \bar{x} + \frac{\hat{\sigma}c}{\sqrt{n}}\right]$$

- Zu Schritt 3: Manchmal kann anderer Schätzwert $\hat{\sigma}$ sinnvoller sein.



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Konfidenzintervall für μ bei beliebiger Verteilung

Beispiel:

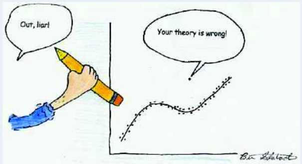
Poisson-Verteilung mit $\lambda\ (= \mu = \sigma^2)$ unbekannt.

$(x_1, \ldots, x_{40}) = (3; 8; \ldots; 6)$

Gesucht: KI für λ zum Konfidenzniveau $1 - \alpha = 0{,}9$

1. $1 - \alpha = 0{,}9$
2. $N(0;1):\ c = x_{1-\frac{\alpha}{2}} = x_{1-\frac{0,1}{2}} = x_{0,95} = 1{,}645$
3. $\bar{x} = \dfrac{1}{40}(3 + 8 + \cdots + 6) = 6{,}5$

   $\hat{\sigma} = \sqrt{\bar{x}} = \sqrt{6{,}5} = 2{,}55$ (da $\sigma^2 = \lambda$)
4. $\dfrac{\hat{\sigma} c}{\sqrt{n}} = \dfrac{2{,}55 \cdot 1{,}645}{\sqrt{40}} = 0{,}66$
5. $\mathrm{KI} = [6{,}5 - 0{,}66; 6{,}5 + 0{,}66] = [5{,}84; 7{,}16]$



1. Einführung

2. Deskriptive Statistik

3. W-Theorie

4. Induktive Statistik

Grundlagen

Punkt-Schätzung

Intervall-Schätzung

Signifikanztests

Quellen

Tabellen

Konfidenzintervall für σ^2 bei Normalverteilung

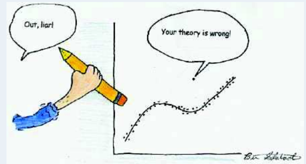
Vorgehensweise

1. Festlegen eines Konfidenzniveaus $1 - \alpha$
2. Bestimmung der $\frac{\alpha}{2}$- bzw. $(1 - \frac{\alpha}{2})$-Fraktile (c_1 bzw. c_2) der $\chi^2(n-1)$-Verteilung
3. Aus der Stichprobe: Berechnung der Größe

$$(n-1)s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

4. Berechnung des Konfidenzintervalls

$$\left[\frac{(n-1)s^2}{c_2} ; \frac{(n-1)s^2}{c_1} \right]$$



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

KI für σ^2 bei Normalverteilung

Beispiel:

$G \sim N(\mu; \sigma)$;

$(x_1, \ldots, x_5) = (1, 1.5, 2.5, 3, 2)$

Gesucht: KI für σ^2 zum Konfidenzniveau $1 - \alpha = 0{,}99$

1. $1 - \alpha = 0{,}99$
2. $\chi^2(5-1): \; c_1 = x_{\frac{\alpha}{2}} = x_{0{,}005} = 0{,}21$

   $c_2 = x_{1-\frac{\alpha}{2}} = x_{0{,}995} = 14{,}86$
3. $\bar{x} = \frac{1}{5}(1 + 1{,}5 + 2{,}5 + 3 + 2) = 2$

   $$\sum_{i=1}^{5} x_i^2 - 5 \cdot \bar{x}^2 = 1^2 + 1{,}5^2 + 2{,}5^2 + 3^2 + 2^2 - 5 \cdot 2^2 = 2{,}5$$
4. $$\text{KI} = \left[\frac{2{,}5}{14{,}86}; \frac{2{,}5}{0{,}21}\right] = [0{,}17; 11{,}9]$$

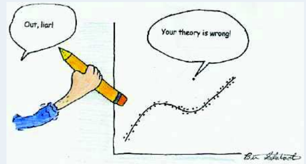
(Extrem groß, da n klein.)

Signifikanztests

- Vorliegen einer Hypothese über die Verteilung(en) der Grundgesamtheit(en).
- Beispiele:
  - „Der Würfel ist fair."
  - „Die Brenndauern zweier unterschiedlicher Glühbirnensorten sind gleich."
- Hypothese soll anhand einer Stichprobe überprüft werden.
- Prinzip:
  - Hypothese verwerfen, wenn „signifikanter" Widerspruch zur Stichprobe.
  - Ansonsten: Hypothese nicht verwerfen.
- Eine verworfene Hypothese gilt als statistisch widerlegt.
- Nicht-Verwerfung ist dagegen ein „Freispruch aus Mangel an Beweisen".

Zu Beachten:

Nicht-Verwerfung ist **kein** „statistischer Beweis", dass Hypothese wahr ist!
(„Trick": Hypothese falsch $\iff$ Gegenhypothese wahr!)



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

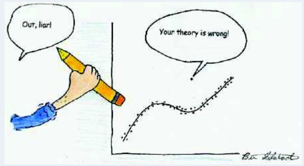
Test des Erwartungswertes bei bekannter Varianz in der Grundgesamtheit

- Zunächst:
  - $G \sim N(\mu; \sigma)$ mit σ bekannt
  - Einfache Stichprobe $X_1, \ldots, X_n$
  - (Null-)Hypothese $H_0 : \mu = \mu_0$
- **Beispiel:**
  $X_1, \ldots, X_{25}$ mit $X_i =$ Füllmenge der i-ten Flasche $\sim N(\mu; 1{,}5)$
  Nullhypothese $H_0 : \mu = 500$, d.h. $\mu_0 = 500$
- Je nach Interessenlage sind unterschiedliche **Gegenhypothesen** möglich:

$$
\begin{array}{ll}
\text{a)} & H_1 : \mu \neq \mu_0 \\
\text{b)} & H_1 : \mu < \mu_0 \\
\text{c)} & H_1 : \mu > \mu_0
\end{array}
$$

- Entscheidung:

	$H_0 : \mu = \mu_0$	wird abgelehnt gegenüber
a)	$H_1 : \mu \neq \mu_0$,	wenn $\lvert\bar{x} - \mu_0\rvert$ „sehr groß" ist
b)	$H_1 : \mu < \mu_0$,	wenn $\bar{x}$ „weit kleiner" als μ_0 ist
c)	$H_1 : \mu > \mu_0$,	wenn $\bar{x}$ „weit größer" als μ_0 ist



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Test des Erwartungswertes bei bekannter Varianz in der Grundgesamtheit

Entscheidungskriterium aus Stichprobe:

$$v = \frac{\bar{x} - \mu_0}{\sigma} \sqrt{n}$$

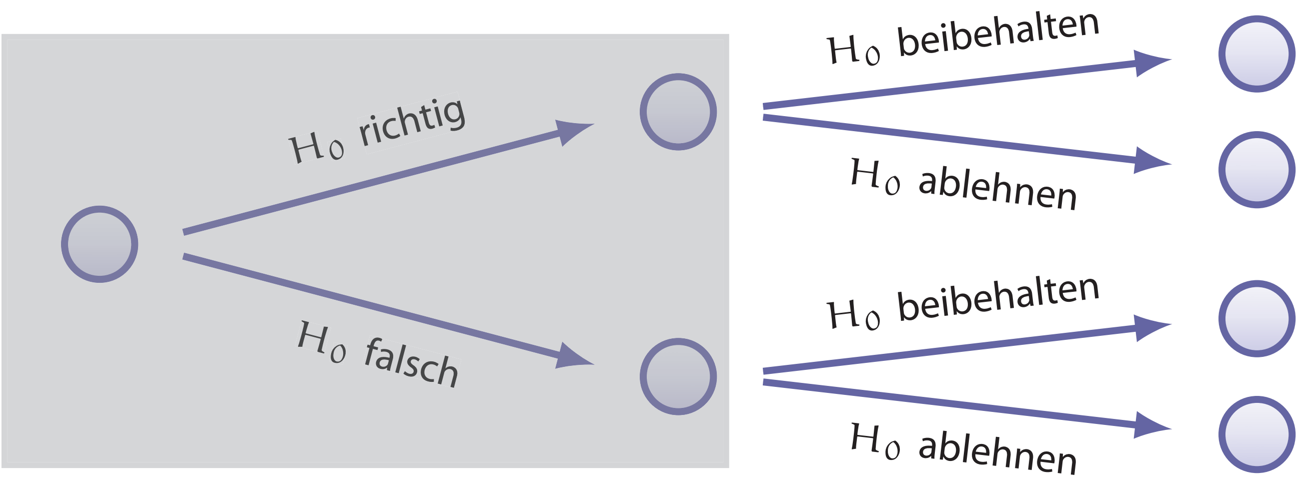
- Vorteil: Verteilung bekannt: $N(0;1)$
- Dann:

$H_0 : \mu = \mu_0$ wird abgelehnt gegenüber

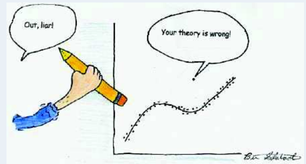
a)	$H_1 : \mu \neq \mu_0$,	wenn	$\lvert v \rvert$	„sehr groß“ ist
b)	$H_1 : \mu < \mu_0$,	wenn	v	„sehr negativ“ ist
c)	$H_1 : \mu > \mu_0$,	wenn	v	„sehr positiv“ ist

Mögliche Fehlentscheidungen

- **Ablehnung von H_0**, obwohl H_0 richtig ist: Fehler 1. Art
- **Nicht-Ablehnung von H_0**, obwohl H_0 falsch ist: Fehler 2. Art



- Signifikanzniveau α: Maximal erlaubte Wahrscheinlichkeit für einen Fehler 1. Art.



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

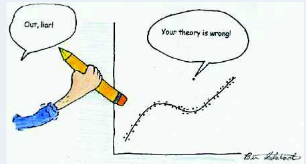
Test des Erwartungswertes bei bekannter Varianz in der Grundgesamtheit

- Mithilfe von α und V kann geklärt werden, was „sehr groß“ usw. heißt:
  Wahrscheinlichkeit für Fehler 1. Art im Fall a): $|v| > x$, obwohl H_0 richtig:

$$
\begin{aligned}
P(|V| > x) &= P(V > x) + P(V < -x) \\
&= 2 \cdot P(V > x) \qquad \text{(Symmetrie der Normalverteilung)} \\
&= 2 \cdot [1 - P(V \leqq x)] = 2 \cdot [1 - \Phi(x)] \stackrel{!}{=} \alpha
\end{aligned}
$$

$$
\begin{aligned}
&\iff \Phi(x) = 1 - \tfrac{\alpha}{2} \\
&\iff x = x_{1-\frac{\alpha}{2}}
\end{aligned}
$$

  H_0 wird demnach verworfen,
  wenn $|v| > x_{1-\frac{\alpha}{2}}$ bzw. $v \in B$ ist.
  $B = (-\infty; -x_{1-\frac{\alpha}{2}}) \cup (x_{1-\frac{\alpha}{2}}; \infty)$ heißt Verwerfungsbereich.
- Analoge Vorgehensweise für die Fälle b) und c)



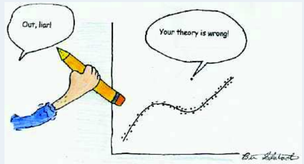
1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Rezept

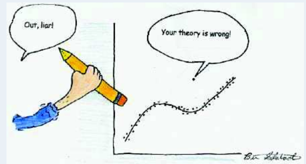
1. Ein Signifikanzniveau α wird festgelegt.
2. Der Verwerfungsbereich

$$B = (-\infty;\ -x_{1-\alpha/2}) \cup (x_{1-\alpha/2};\ \infty) \qquad \text{im Fall a)}$$
$$B = (-\infty;\ -x_{1-\alpha}) \qquad \text{im Fall b)}$$
$$B = (x_{1-\alpha};\ \infty) \qquad \text{im Fall c)}$$

   wird festgelegt, wobei $x_{1-\alpha/2}$ bzw. $x_{1-\alpha}$ das $(1-\alpha/2)$- bzw. das $(1-\alpha)$-Fraktil der $N(0,1)$-Verteilung ist. (Wichtig: Der Ablehnungsbereich ist also unabhängig von der Stichprobe)
3. Wichtig: Erst jetzt werden die Daten der Stichprobe erhoben/beachtet:
   Der Testfunktionswert $v = \frac{\overline{x} - \mu_0}{\sigma}\sqrt{n}$ wird berechnet.
4. H_0 wird genau dann verworfen, wenn $v \in B$ gilt.



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Beispiel:

$X_1, \dots, X_{25}$ mit $X_i \sim N(\mu; 1{,}5)$ und $\bar{x} = 499{,}28$

Prüfe $H_0: \; \mu = 500, \; H_1: \; \mu \neq 500$ zum Signifikanzniveau $\alpha = 0{,}01$

Lösung: Einstichproben-Gaußtest, Fall a)

1. $\alpha = 0{,}01$
2. $N(0;1): \; x_{1-\frac{\alpha}{2}} = x_{1-0{,}005} = x_{0{,}995} = 2{,}576$
   $\Rightarrow B = (-\infty; -2{,}576) \cup (2{,}576; \infty)$
3. $v = \frac{499{,}28 - 500}{1{,}5} \cdot \sqrt{25} = -2{,}4$
4. $v \notin B \Rightarrow H_0$ nicht verwerfen

Interpretation: Zum Signifikanzniveau 1 % kann der Brauerei keine Abweichung vom Sollwert $\mu_0 = 500$ nachgewiesen werden.

Gegeben:

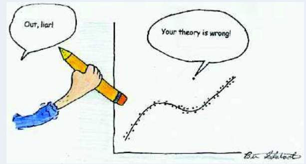
- Einfache Stichprobe $X_1, \ldots, X_n$ mit
- $E(X_i) = \mu, \quad Var(X_i) = \sigma^2$

Hypothesenpaare:

$$
\begin{array}{llll}
\text{a)} & H_0: \mu = \mu_0 & & H_1: \mu \neq \mu_0 \\
\text{b)} & H_0: \mu = \mu_0 & (\text{oder } \mu \geqq \mu_0), & H_1: \mu < \mu_0 \\
\text{c)} & H_0: \mu = \mu_0 & (\text{oder } \mu \leqq \mu_0), & H_1: \mu > \mu_0
\end{array}
$$

Voraussetzungen:

1. Normalverteilung mit σ unbekannt (Einstichproben-t-Test)
   oder
2. Beliebige Verteilung
   mit $n > 30$ bzw. $5 \leqq \sum x_i \leqq n - 5$ (bei $B(1; p)$)
   (approximativer Gaußtest)



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Ablauf:

1. Festlegen des Signifikanzniveaus α

2. Festlegen des Verwerfungsbereichs B:

   - Falls $H_1 : \mu \neq \mu_0$: $\quad B = (-\infty; -x_{1-\alpha/2}) \cup (x_{1-\alpha/2}; \infty)$
   - Falls $H_1 : \mu < \mu_0$: $\quad B = (-\infty; -x_{1-\alpha})$
   - Falls $H_1 : \mu > \mu_0$: $\quad B = (x_{1-\alpha}; \infty)$

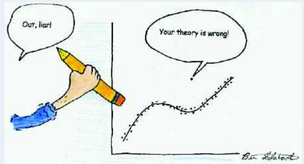
   Dabei steht $x_{1-\alpha/2}$ bzw. $x_{1-\alpha}$ für das jeweilige Fraktil

   - der $t(n-1)$-Verteilung bei $n \leqslant 29$ bzw.
   - der $N(0;1)$-Verteilung bei $n \geqslant 30$.

3. Berechnen des Testfunktionswertes:

$$v = \begin{cases} \dfrac{\bar{x} - \mu_0}{s}\sqrt{n} & \text{falls Grundgesamtheit } N(\mu;\sigma)\text{-verteilt, } \sigma \text{ unbekannt oder falls Verteilung der GG beliebig, } n > 30,\ \sigma \text{ unbekannt} \\[2ex] \dfrac{\bar{x} - \mu_0}{\sigma}\sqrt{n} & \text{falls Verteilung der GG beliebig, } n > 30,\ \sigma \text{ bekannt} \\[2ex] \dfrac{\bar{x} - \mu_0}{\sqrt{\mu_0(1-\mu_0)}}\sqrt{n} & \text{falls GG gemäß } B(1;\mu)\text{-verteilt, } n > 30 \end{cases}$$

Statistik



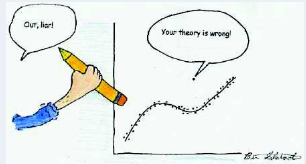
1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Einstichproben-t-Test: Beispiel

Beispiel t-Test: Energieaufnahme von Frauen

- Empfohlene täglich Energieaufnahme für Frauen: 7724 kJ (1845 kcal)
- Nehme einfache Stichprobe von 11 Frauen und teste zum Signifkanzniveau $\alpha = 0{,}05$ für
- H_0: „Der Erwartungswert der täglichen Energieaufnahme für Frauen ist 7724 kJ“ (μ_0)
- gegen H_1: $\mu \neq \mu_0$

```
daily.intake <- c(5260, 5470, 5640, 6180, 6390, 6515, 6805, 7515, 7515, 8230, 8770)
t.test(daily.intake, alternative="two.sided", mu=7724, conf.level=0.95)

##
##  One Sample t-test
##
## data:  daily.intake
## t = -2.8179, df = 10, p-value = 0.01823
## alternative hypothesis: true mean is not equal to 7724
## 95 percent confidence interval:
##  5986.348 7520.925
## sample estimates:
## mean of x
##  6753.636
```

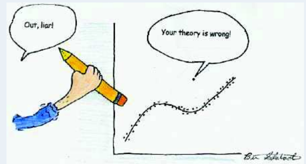
1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Einstichproben-t-Test: Beispiel

Beispiel t-Test: Energieaufnahme von Frauen

- Empfohlene täglich Energieaufnahme für Frauen: 7724 kJ (1845 kcal)
- Nehme einfache Stichprobe von 11 Frauen und teste zum Signifkanzniveau $\alpha = 0{,}05$ für
- H_0: „Der Erwartungswert der täglichen Energieaufnahme für Frauen ist 7724 kJ“ (μ_0)
- gegen H_1: $\mu \neq \mu_0$

```r
daily.intake <- c(5260, 5470, 5640, 6180, 6390, 6515, 6805, 7515, 7515, 8230, 8770)
t.test(daily.intake, alternative="two.sided", mu=7724, conf.level=0.95)

##
##  One Sample t-test
##
## data:  daily.intake
## t = -2.8179, df = 10, p-value = 0.01823
## alternative hypothesis: true mean is not equal to 7724
## 95 percent confidence interval:
##  5986.348 7520.925
## sample estimates:
## mean of x
##  6753.636
```

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Einstichproben-t-Test, approx. Gaußtest

Beispiel:

$X_1, \ldots, X_{2000} \sim B(1; p)$ mit

$$X_i = \begin{cases} 1, & \text{falls i-te Person Wähler einer bestimmten Partei} \\ 0, & \text{sonst} \end{cases}$$

Ergebnis der Stichprobe: $\sum_{i=1}^{2000} x_i = 108$

Prüfe $H_0 : p \leqq 0{,}05$ gegen $H_1 : p > 0{,}05$ zum Signifikanzniveau 2 %

Lösung:

approximativer Gaußtest bei dichotomer (zweiwertiger) Verteilung; Voraussetzung 2 erfüllt: $5 \leqq 108 \leqq 2000 - 5$

1. $\alpha = 0{,}02$
2. $N(0; 1): x_{1-\alpha} = x_{0{,}98} = 2{,}05$ (Tabelle) $\Rightarrow B = (2{,}05; \infty)$
3. $v = \dfrac{\frac{108}{2000} - 0{,}05}{\sqrt{0{,}05 \cdot (1 - 0{,}05)}} \sqrt{2000} = 0{,}82$
4. $v \notin B \Rightarrow H_0$ nicht verwerfen

Zusatzfrage: Entscheidung, falls $\alpha = 0{,}01$? $\to$ Keine Änderung!

Chi-Quadrat-Test für die Varianz

- Gegeben: Einfache Stichprobe $X_1, \ldots, X_n \sim N(\mu; \sigma)$
- Hypothesenpaare:

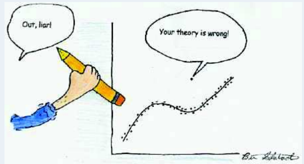
$$
\begin{array}{llll}
\text{a)} & H_0 : \sigma^2 = \sigma_0^2 & & H_1 : \sigma^2 \neq \sigma_0^2 \\
\text{b)} & H_0 : \sigma^2 = \sigma_0^2 & (\text{oder } \sigma^2 \geqq \sigma_0^2), & H_1 : \sigma^2 < \sigma_0^2 \\
\text{c)} & H_0 : \sigma^2 = \sigma_0^2 & (\text{oder } \sigma^2 \leqq \sigma_0^2), & H_1 : \sigma^2 > \sigma_0^2
\end{array}
$$

- **Vorgehensweise:**
  1. Festlegen des Signifikanzniveaus α.
  2. Festlegen des Verwerfungsbereichs:

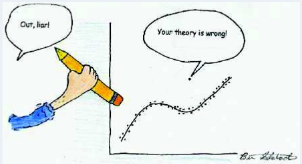
$$
\begin{array}{ll}
B = [0; x_{\alpha/2}) \cup (x_{1-\alpha/2}; \infty) & \text{im Fall a)} \\
B = [0; x_{\alpha}) & \text{im Fall b)} \\
B = (x_{1-\alpha}; \infty) & \text{im Fall c)}
\end{array}
$$

  3. Berechnung des Testfunktionswertes:

$$
v = \frac{(n-1)s^2}{\sigma_0^2} = \frac{1}{\sigma_0^2} \sum_{i=1}^{n} (x_i - \bar{x})^2
$$



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Chi-Quadrat-Test für die Varianz

Beispiel: $G \sim N(\mu; \sigma)$

$$(x_1, \ldots, x_{10}) = (2100; 2130; 2150; 2170; 2210; 2070; 2230; 2150; 2230; 2200)$$

Prüfe $H_0: \ \sigma = 40, \ H_1: \ \sigma \neq 40$ zum Signifikanzniveau $\alpha = 0{,}1$

Lösung: χ^2-Test für die Varianz, Hypothese Fall a);
Voraussetzungen sind erfüllt

1. $\alpha = 0{,}1$
2. $\chi^2(9): \ x_{\frac{\alpha}{2}} = x_{0,05} = 3{,}33; \ x_{1-\frac{\alpha}{2}} = x_{0,95} = 16{,}92$
   (Tabelle der χ^2-Verteilung)

$$\Rightarrow B = [0; 3{,}33) \cup (16{,}92; \infty)$$

3. $\bar{x} = \frac{1}{10}(2100 + 2130 + \cdots + 2200) = 2164$

   $v = \frac{1}{40^2}\left[(2100 - 2164)^2 + \cdots + (2200 - 2164)^2\right] = 16{,}65$

   $\Rightarrow v \notin B \Rightarrow H_0$ nicht verwerfen

Zwei verbundene einfache Stichproben: Kontingenztest

- Situation: In Grundgesamtheit G: Zwei verbundene einfache Stichproben, also Beobachtung zweier Merkmale X, Y
- Hypothese:

$$
\begin{aligned}
H_0 &: \text{ Die beiden Merkmale } X \text{ und } Y \text{ sind in } G \text{ unabhängig.}\\
H_1 &: \text{ } X \text{ und } Y \text{ sind in } G \text{ abhängig.}
\end{aligned}
$$

Vorgehensweise Kontingenztest:

1. Festlegen des Signifikanzniveaus α.
2. Unterteilung der x-Achse in $k \geqslant 2$ und die y-Achse in $l \geqslant 2$ disjunkte, aneinander angrenzende Intervalle $A_1, \dots, A_k$ bzw. $B_1, \dots, B_l$
3. Erstellen einer Kontingenztabelle mit Randhäufigkeiten:

$x\downarrow\ y\rightarrow$	B_1	B_2	$\cdots$	B_l	
A_1	h_{11}	h_{12}	$\cdots$	h_{1l}	$h_{1\bullet}$
A_2	h_{21}	h_{22}	$\cdots$	h_{2l}	$h_{2\bullet}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	
A_k	h_{k1}	h_{k2}	$\cdots$	h_{kl}	$h_{k\bullet}$
	$h_{\bullet 1}$	$h_{\bullet 2}$	$\cdots$	$h_{\bullet l}$	n

Vorgehensweise Kontingenztest (Fortsetzung):

4. Mit dem Fraktilswert $x_{1-\alpha}$ der χ^2-Verteilung mit $(k-1)\cdot(l-1)$ Freiheitsgraden: Berechnung des Verwerfungsbereichs

$$B = (x_{1-\alpha}; \infty)$$

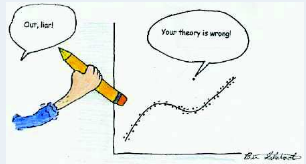
5. Zu jeder Kombination aus $i = 1, \ldots, k$ und $j = 1, \ldots, l$: Berechnung der Größe

$$\tilde{h}_{ij} = \frac{h_{i\bullet} \cdot h_{\bullet j}}{n}$$

6. Berechnung des Testfunktionswerts v:

$$v = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{\left(\tilde{h}_{ij} - h_{ij}\right)^2}{\tilde{h}_{ij}} = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{h_{ij}^2}{\tilde{h}_{ij}} - n$$

7. Ablehnung von H_0 genau dann, wenn $v \in B$.



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

Zwei verbundene einfache Stichproben: Kontingenztest

Kontingenztest: Beispiel

- 400 Erstkandidaten einer praktischen Führerscheinprüfung schneiden abhängig von der besuchten Fahrschule folgendermaßen ab:

	Fahrschule		
	A	B	C
bestanden	130	88	62
durchgefallen	70	38	12

- Zum Signifikanzniveau von 5 % soll getestet werden, ob das Bestehen der Prüfung unabhängig von der besuchten Fahrschule ist.

Testdurchführung

1. Signifikanzniveau $\alpha = 5\,\%$
2. entfällt, da Skalenniveau nominal
3. Kontingenztabelle:

	A	B	C	$\sum$
best.	130	88	62	280
durchg.	70	38	12	120
$\sum$	200	126	74	400

4. χ^2-Verteilung mit $(3-1)\cdot(2-1) = 2$ Freiheitsgraden: $x_{1-0{,}05} = x_{0{,}95} = 5{,}99$:

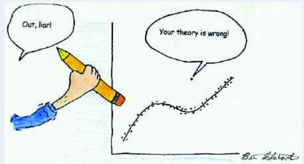
$$B = (5{,}99; \infty)$$

5. Berechnung der $\tilde{h}_{ij}$:

	A	B	C
best.	140	88,2	51,8
durchg.	60	37,8	22,2

6. $$v = \frac{(140-130)^2}{140} + \dots + \frac{(22{,}2-12)^2}{22{,}2} \approx 9{,}077$$

7. $v \in B$: Also wird H_0 abgelehnt, die Prüfungsergebnisse sind abhängig von der Fahrschule.



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests
Quellen
Tabellen

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Bücher

- Bamberg, Günter, Franz Baur und Michael Krapp (2011). **Statistik**. 16. Aufl. München: Oldenbourg Verlag. ISBN: 3486702580.
- Dalgaard, Peter (2002). **Introductory Statistics with R**. New York: Springer.
- Fahrmeir, Ludwig, Rita Künstler, Iris Pigeot und Gerhard Tutz (2009). **Statistik: Der Weg zur Datenanalyse**. 7. Aufl. Berlin, Heidelberg: Springer. ISBN: 3642019382.

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen

Quellen zu Bildern und Daten

- Anscombe, Francis (1973). „Graphs in Statistical Analysis“. In: **The American Statistician**, S. 195–199.

- Bach, Axel, Reinhard Brüning, Katrin Krieft, Hilmar Liebsch und Martin Rosenberg (2006). **Mit Zahlen lügen**. URL: `http://www.wdr.de/tv/quarks/sendungsbeitraege/2006/1017/000_zahlen.jsp`.

- Fahrmeir, Ludwig, Rita Künstler, Iris Pigeot und Gerhard Tutz (2009). **Statistik: Der Weg zur Datenanalyse**. 7. Aufl. Berlin, Heidelberg: Springer. ISBN: 3642019382.

- Kramer, Walter (2011). **So lügt man mit Statistik**. Piper Verlag. ISBN: 3492264131.

Binomialverteilung $X \sim B(n;p)$, Verteilungsfunktion $F(x) = P(X \leqslant x)$

$n = 2$

$\downarrow x \quad p \rightarrow$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.2	0.25	0.3	0.4	0.5
0	0.9801	0.9604	0.9409	0.9216	0.9025	0.8836	0.8649	0.8464	0.8281	0.8100	0.6400	0.5625	0.4900	0.3600	0.2500
1	0.9999	0.9996	0.9991	0.9984	0.9975	0.9964	0.9951	0.9936	0.9919	0.9900	0.9600	0.9375	0.9100	0.8400	0.7500

$n = 3$

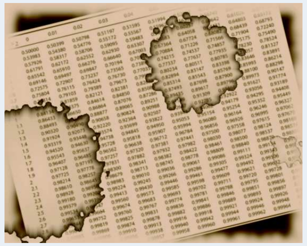
$\downarrow x \quad p \rightarrow$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.2	0.25	0.3	0.4	0.5
0	0.9703	0.9412	0.9127	0.8847	0.8574	0.8306	0.8044	0.7787	0.7536	0.7290	0.5120	0.4219	0.3430	0.2160	0.1250
1	0.9997	0.9988	0.9974	0.9953	0.9928	0.9896	0.9860	0.9818	0.9772	0.9720	0.8960	0.8438	0.7840	0.6480	0.5000
2	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9997	0.9995	0.9993	0.9990	0.9920	0.9844	0.9730	0.9360	0.8750

$n = 4$

$\downarrow x \quad p \rightarrow$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.2	0.25	0.3	0.4	0.5
0	0.9606	0.9224	0.8853	0.8493	0.8145	0.7807	0.7481	0.7164	0.6857	0.6561	0.4096	0.3164	0.2401	0.1296	0.0625
1	0.9994	0.9977	0.9948	0.9909	0.9860	0.9801	0.9733	0.9656	0.9570	0.9477	0.8192	0.7383	0.6517	0.4752	0.3125
2	1.0000	1.0000	0.9999	0.9998	0.9995	0.9992	0.9987	0.9981	0.9973	0.9963	0.9728	0.9492	0.9163	0.8208	0.6875
3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9984	0.9961	0.9919	0.9744	0.9375

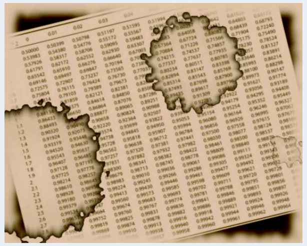
$n = 5$

$\downarrow x \quad p \rightarrow$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.2	0.25	0.3	0.4	0.5
0	0.9510	0.9039	0.8587	0.8154	0.7738	0.7339	0.6957	0.6591	0.6240	0.5905	0.3277	0.2373	0.1681	0.0778	0.0313
1	0.9990	0.9962	0.9915	0.9852	0.9774	0.9681	0.9575	0.9456	0.9326	0.9185	0.7373	0.6328	0.5282	0.3370	0.1875
2	1.0000	0.9999	0.9997	0.9994	0.9988	0.9980	0.9969	0.9955	0.9937	0.9914	0.9421	0.8965	0.8369	0.6826	0.5000
3	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9997	0.9995	0.9933	0.9844	0.9692	0.9130	0.8125
4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9990	0.9976	0.9898	0.9688



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen
Binomialverteilung
Poissonverteilung
Standardnormalverteilung
χ^2-Verteilung
t-Verteilung
F-Verteilung

Binomialverteilung $X \sim B(n; p)$, Verteilungsfunktion $F(x) = P(X \leqslant x)$



$n = 6$

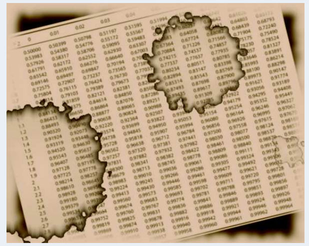
$\downarrow x$ $p \rightarrow$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.2	0.25	0.3	0.4	0.5
0	0.9415	0.8858	0.8330	0.7828	0.7351	0.6899	0.6470	0.6064	0.5679	0.5314	0.2621	0.1780	0.1176	0.0467	0.0156
1	0.9985	0.9943	0.9875	0.9784	0.9672	0.9541	0.9392	0.9227	0.9048	0.8857	0.6554	0.5339	0.4202	0.2333	0.1094
2	1.0000	0.9998	0.9995	0.9988	0.9978	0.9962	0.9942	0.9915	0.9882	0.9842	0.9011	0.8306	0.7443	0.5443	0.3438
3	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9995	0.9992	0.9987	0.9830	0.9624	0.9295	0.8208	0.6563
4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9984	0.9954	0.9891	0.9590	0.8906
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9959	0.9844

$n = 7$

$\downarrow x$ $p \rightarrow$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.2	0.25	0.3	0.4	0.5
0	0.9321	0.8681	0.8080	0.7514	0.6983	0.6485	0.6017	0.5578	0.5168	0.4783	0.2097	0.1335	0.0824	0.0280	0.0078
1	0.9980	0.9921	0.9829	0.9706	0.9556	0.9382	0.9187	0.8974	0.8745	0.8503	0.5767	0.4449	0.3294	0.1586	0.0625
2	1.0000	0.9997	0.9991	0.9980	0.9962	0.9937	0.9903	0.9860	0.9807	0.9743	0.8520	0.7564	0.6471	0.4199	0.2266
3	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9993	0.9988	0.9982	0.9973	0.9667	0.9294	0.8740	0.7102	0.5000
4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9953	0.9871	0.9712	0.9037	0.7734
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996	0.9987	0.9962	0.9812	0.9375
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9984	0.9922

$n = 8$

$\downarrow x$ $p \rightarrow$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.2	0.25	0.3	0.4	0.5
0	0.9227	0.8508	0.7837	0.7214	0.6634	0.6096	0.5596	0.5132	0.4703	0.4305	0.1678	0.1001	0.0576	0.0168	0.0039
1	0.9973	0.9897	0.9777	0.9619	0.9428	0.9208	0.8965	0.8702	0.8423	0.8131	0.5033	0.3671	0.2553	0.1064	0.0352
2	0.9999	0.9996	0.9987	0.9969	0.9942	0.9904	0.9853	0.9789	0.9711	0.9619	0.7969	0.6785	0.5518	0.3154	0.1445
3	1.0000	1.0000	0.9999	0.9998	0.9996	0.9993	0.9987	0.9978	0.9966	0.9950	0.9437	0.8862	0.8059	0.5941	0.3633
4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9997	0.9996	0.9896	0.9727	0.9420	0.8263	0.6367
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9988	0.9958	0.9887	0.9502	0.8555
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9987	0.9915	0.9648
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9961



Binomialverteilung $X \sim B(n; p)$, Verteilungsfunktion $F(x) = P(X \leqslant x)$

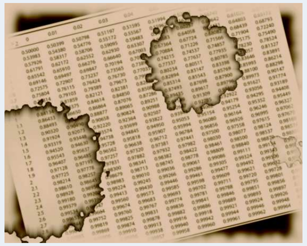
$n = 9$

$\downarrow x \quad p \rightarrow$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.2	0.25	0.3	0.4	0.5
0	0.9135	0.8337	0.7602	0.6925	0.6302	0.5730	0.5204	0.4722	0.4279	0.3874	0.1342	0.0751	0.0404	0.0101	0.0020
1	0.9966	0.9869	0.9718	0.9522	0.9288	0.9022	0.8729	0.8417	0.8088	0.7748	0.4362	0.3003	0.1960	0.0705	0.0195
2	0.9999	0.9994	0.9980	0.9955	0.9916	0.9862	0.9791	0.9702	0.9595	0.9470	0.7382	0.6007	0.4628	0.2318	0.0898
3	1.0000	1.0000	0.9999	0.9997	0.9994	0.9987	0.9977	0.9963	0.9943	0.9917	0.9144	0.8343	0.7297	0.4826	0.2539
4	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9995	0.9991	0.9804	0.9511	0.9012	0.7334	0.5000
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9969	0.9900	0.9747	0.9006	0.7461
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987	0.9957	0.9750	0.9102
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9962	0.9805
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9980

$n = 10$

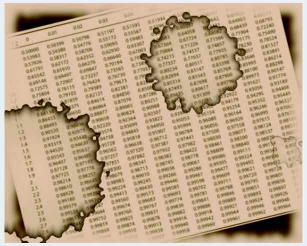
$\downarrow x \quad p \rightarrow$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.2	0.25	0.3	0.4	0.5
0	0.9044	0.8171	0.7374	0.6648	0.5987	0.5386	0.4840	0.4344	0.3894	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010
1	0.9957	0.9838	0.9655	0.9418	0.9139	0.8824	0.8483	0.8121	0.7746	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107
2	0.9999	0.9991	0.9972	0.9938	0.9885	0.9812	0.9717	0.9599	0.9460	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547
3	1.0000	1.0000	0.9999	0.9996	0.9990	0.9980	0.9964	0.9942	0.9912	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719
4	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9994	0.9990	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9877	0.9453
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990

Binomialverteilung $X \sim B(n;p)$, Verteilungsfunktion $F(x) = P(X \leqslant x)$



$n = 15$

$\downarrow x \quad p \rightarrow$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.2	0.25	0.3	0.4	0.5
0	0.8601	0.7386	0.6333	0.5421	0.4633	0.3953	0.3367	0.2863	0.2430	0.2059	0.0352	0.0134	0.0047	0.0005	0.0000
1	0.9904	0.9647	0.9270	0.8809	0.8290	0.7738	0.7168	0.6597	0.6035	0.5490	0.1671	0.0802	0.0353	0.0052	0.0005
2	0.9996	0.9970	0.9906	0.9797	0.9638	0.9429	0.9171	0.8870	0.8531	0.8159	0.3980	0.2361	0.1268	0.0271	0.0037
3	1.0000	0.9998	0.9992	0.9976	0.9945	0.9896	0.9825	0.9727	0.9601	0.9444	0.6482	0.4613	0.2969	0.0905	0.0176
4	1.0000	1.0000	0.9999	0.9998	0.9994	0.9986	0.9972	0.9950	0.9918	0.9873	0.8358	0.6865	0.5155	0.2173	0.0592
5	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9997	0.9993	0.9987	0.9978	0.9389	0.8516	0.7216	0.4032	0.1509
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9819	0.9434	0.8689	0.6098	0.3036
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9958	0.9827	0.9500	0.7869	0.5000
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9992	0.9958	0.9848	0.9050	0.6964
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9992	0.9963	0.9662	0.8491
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9907	0.9408
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9981	0.9824
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9963
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

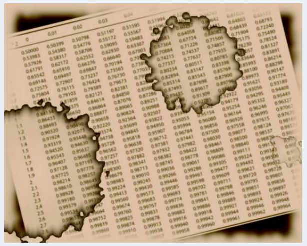


Binomialverteilung $X \sim B(n;p)$, Verteilungsfunktion $F(x) = P(X \leqslant x)$

$n = 20$

$\downarrow x \quad p \rightarrow$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.2	0.25	0.3	0.4	0.5
0	0.8179	0.6676	0.5438	0.4420	0.3585	0.2901	0.2342	0.1887	0.1516	0.1216	0.0115	0.0032	0.0008	0.0000	0.0000
1	0.9831	0.9401	0.8802	0.8103	0.7358	0.6605	0.5869	0.5169	0.4516	0.3917	0.0692	0.0243	0.0076	0.0005	0.0000
2	0.9990	0.9929	0.9790	0.9561	0.9245	0.8850	0.8390	0.7879	0.7334	0.6769	0.2061	0.0913	0.0355	0.0036	0.0002
3	1.0000	0.9994	0.9973	0.9926	0.9841	0.9710	0.9529	0.9294	0.9007	0.8670	0.4114	0.2252	0.1071	0.0160	0.0013
4	1.0000	1.0000	0.9997	0.9990	0.9974	0.9944	0.9893	0.9817	0.9710	0.9568	0.6296	0.4148	0.2375	0.0510	0.0059
5	1.0000	1.0000	1.0000	0.9999	0.9997	0.9991	0.9981	0.9962	0.9932	0.9887	0.8042	0.6172	0.4164	0.1256	0.0207
6	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9994	0.9987	0.9976	0.9133	0.7858	0.6080	0.2500	0.0577
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9679	0.8982	0.7723	0.4159	0.1316
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9900	0.9591	0.8867	0.5956	0.2517
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9974	0.9861	0.9520	0.7553	0.4119
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9994	0.9961	0.9829	0.8725	0.5881
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9991	0.9949	0.9435	0.7483
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9987	0.9790	0.8684
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9935	0.9423
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9984	0.9793
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9941
16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9987
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

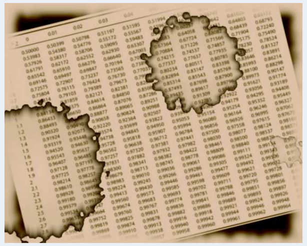
Binomialverteilung $X \sim B(n;p)$, Verteilungsfunktion $F(x) = P(X \leqslant x)$



$n = 25$

$\downarrow x \quad p \rightarrow$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.2	0.25	0.3	0.4	0.5
0	0.7778	0.6035	0.4670	0.3604	0.2774	0.2129	0.1630	0.1244	0.0946	0.0718	0.0038	0.0008	0.0001	0.0000	0.0000
1	0.9742	0.9114	0.8280	0.7358	0.6424	0.5527	0.4696	0.3947	0.3286	0.2712	0.0274	0.0070	0.0016	0.0001	0.0000
2	0.9980	0.9868	0.9620	0.9235	0.8729	0.8129	0.7466	0.6768	0.6063	0.5371	0.0982	0.0321	0.0090	0.0004	0.0000
3	0.9999	0.9986	0.9938	0.9835	0.9659	0.9402	0.9064	0.8649	0.8169	0.7636	0.2340	0.0962	0.0332	0.0024	0.0001
4	1.0000	0.9999	0.9992	0.9972	0.9928	0.9850	0.9726	0.9549	0.9314	0.9020	0.4207	0.2137	0.0905	0.0095	0.0005
5	1.0000	1.0000	0.9999	0.9996	0.9988	0.9969	0.9935	0.9877	0.9790	0.9666	0.6167	0.3783	0.1935	0.0294	0.0020
6	1.0000	1.0000	1.0000	1.0000	0.9998	0.9995	0.9987	0.9972	0.9946	0.9905	0.7800	0.5611	0.3407	0.0736	0.0073
7	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9995	0.9989	0.9977	0.8909	0.7265	0.5118	0.1536	0.0216
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9995	0.9532	0.8506	0.6769	0.2735	0.0539
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9827	0.9287	0.8106	0.4246	0.1148
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9944	0.9703	0.9022	0.5858	0.2122
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9985	0.9893	0.9558	0.7323	0.3450
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996	0.9966	0.9825	0.8462	0.5000
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9991	0.9940	0.9222	0.6550
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9982	0.9656	0.7878
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9868	0.8852
16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9957	0.9461
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9988	0.9784
18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9927
19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9980
20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995
21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
23	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
24	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

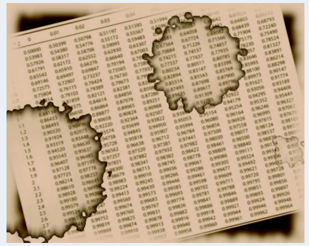
Binomialverteilung $X \sim B(n;p)$, Verteilungsfunktion $F(x) = P(X \leqslant x)$



$n = 50$

$\downarrow x$ $p \rightarrow$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.2	0.25	0.3	0.4	0.5
0	0.6050	0.3642	0.2181	0.1299	0.0769	0.0453	0.0266	0.0155	0.0090	0.0052	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.9106	0.7358	0.5553	0.4005	0.2794	0.1900	0.1265	0.0827	0.0532	0.0338	0.0002	0.0000	0.0000	0.0000	0.0000
2	0.9862	0.9216	0.8108	0.6767	0.5405	0.4162	0.3108	0.2260	0.1605	0.1117	0.0013	0.0001	0.0000	0.0000	0.0000
3	0.9984	0.9822	0.9372	0.8609	0.7604	0.6473	0.5327	0.4253	0.3303	0.2503	0.0057	0.0005	0.0000	0.0000	0.0000
4	0.9999	0.9968	0.9832	0.9510	0.8964	0.8206	0.7290	0.6290	0.5277	0.4312	0.0185	0.0021	0.0002	0.0000	0.0000
5	1.0000	0.9995	0.9963	0.9856	0.9622	0.9224	0.8650	0.7919	0.7072	0.6161	0.0480	0.0070	0.0007	0.0000	0.0000
6	1.0000	0.9999	0.9993	0.9964	0.9882	0.9711	0.9417	0.8981	0.8404	0.7702	0.1034	0.0194	0.0025	0.0000	0.0000
7	1.0000	1.0000	0.9999	0.9992	0.9968	0.9906	0.9780	0.9562	0.9232	0.8779	0.1904	0.0453	0.0073	0.0001	0.0000
8	1.0000	1.0000	1.0000	0.9999	0.9992	0.9973	0.9927	0.9833	0.9672	0.9421	0.3073	0.0916	0.0183	0.0002	0.0000
9	1.0000	1.0000	1.0000	1.0000	0.9998	0.9993	0.9978	0.9944	0.9875	0.9755	0.4437	0.1637	0.0402	0.0008	0.0000
10	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9994	0.9983	0.9957	0.9906	0.5836	0.2622	0.0789	0.0022	0.0000
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9987	0.9968	0.7107	0.3816	0.1390	0.0057	0.0000
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9990	0.8139	0.5110	0.2229	0.0133	0.0002
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.8894	0.6370	0.3279	0.0280	0.0005
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9393	0.7481	0.4468	0.0540	0.0013
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9692	0.8369	0.5692	0.0955	0.0033
16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9856	0.9017	0.6839	0.1561	0.0077
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9937	0.9449	0.7822	0.2369	0.0164
18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9975	0.9713	0.8594	0.3356	0.0325
19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9991	0.9861	0.9152	0.4465	0.0595
20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9937	0.9522	0.5610	0.1013
21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9974	0.9749	0.6701	0.1611
22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9990	0.9877	0.7660	0.2399
23	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996	0.9944	0.8438	0.3359
24	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9976	0.9022	0.4439
25	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9991	0.9427	0.5561
26	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9686	0.6641
27	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9840	0.7601
28	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9924	0.8389
29	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9966	0.8987
30	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9986	0.9405
31	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9675
32	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9836
33	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9923
34	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9967
35	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9987

Binomialverteilung $X \sim B(n;p)$, Verteilungsfunktion $F(x) = P(X \leqslant x)$



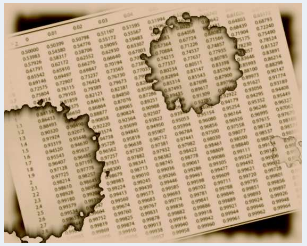
$n = 100$

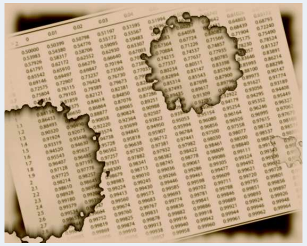
$\downarrow x \quad p \rightarrow$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.2	0.25	0.3	0.4	0.5
0	0.3660	0.1326	0.0476	0.0169	0.0059	0.0021	0.0007	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.7358	0.4033	0.1946	0.0872	0.0371	0.0152	0.0060	0.0023	0.0009	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.9206	0.6767	0.4198	0.2321	0.1183	0.0566	0.0258	0.0113	0.0048	0.0019	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.9816	0.8590	0.6472	0.4295	0.2578	0.1430	0.0744	0.0367	0.0173	0.0078	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.9966	0.9492	0.8179	0.6289	0.4360	0.2768	0.1632	0.0903	0.0474	0.0237	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.9995	0.9845	0.9192	0.7884	0.6160	0.4407	0.2914	0.1799	0.1045	0.0576	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.9999	0.9959	0.9688	0.8936	0.7660	0.6064	0.4443	0.3032	0.1940	0.1172	0.0001	0.0000	0.0000	0.0000	0.0000
7	1.0000	0.9991	0.9894	0.9525	0.8720	0.7483	0.5988	0.4471	0.3128	0.2061	0.0003	0.0000	0.0000	0.0000	0.0000
8	1.0000	0.9998	0.9968	0.9810	0.9369	0.8537	0.7340	0.5926	0.4494	0.3209	0.0009	0.0000	0.0000	0.0000	0.0000
9	1.0000	1.0000	0.9991	0.9932	0.9718	0.9225	0.8380	0.7220	0.5875	0.4513	0.0023	0.0000	0.0000	0.0000	0.0000
10	1.0000	1.0000	0.9998	0.9978	0.9885	0.9624	0.9092	0.8243	0.7118	0.5832	0.0057	0.0001	0.0000	0.0000	0.0000
11	1.0000	1.0000	1.0000	0.9993	0.9957	0.9832	0.9531	0.8972	0.8124	0.7030	0.0126	0.0004	0.0000	0.0000	0.0000
12	1.0000	1.0000	1.0000	0.9998	0.9985	0.9931	0.9776	0.9441	0.8862	0.8018	0.0253	0.0010	0.0000	0.0000	0.0000
13	1.0000	1.0000	1.0000	1.0000	0.9995	0.9974	0.9901	0.9718	0.9355	0.8761	0.0469	0.0025	0.0001	0.0000	0.0000
14	1.0000	1.0000	1.0000	1.0000	0.9999	0.9991	0.9959	0.9867	0.9659	0.9274	0.0804	0.0054	0.0002	0.0000	0.0000
15	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9942	0.9831	0.9601	0.1285	0.0111	0.0004	0.0000	0.0000
16	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9976	0.9922	0.9794	0.1923	0.0211	0.0010	0.0000	0.0000
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9991	0.9966	0.9900	0.2712	0.0376	0.0022	0.0000	0.0000
18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9986	0.9954	0.3621	0.0630	0.0045	0.0000	0.0000
19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9980	0.4602	0.0995	0.0089	0.0000	0.0000
20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9992	0.5595	0.1488	0.0165	0.0000	0.0000
21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.6540	0.2114	0.0288	0.0000	0.0000
22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.7389	0.2864	0.0479	0.0001	0.0000
23	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8109	0.3711	0.0755	0.0003	0.0000
24	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8686	0.4617	0.1136	0.0006	0.0000
25	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9125	0.5535	0.1631	0.0012	0.0000
26	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9442	0.6417	0.2244	0.0024	0.0000
27	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9658	0.7224	0.2964	0.0046	0.0000
28	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9800	0.7925	0.3768	0.0084	0.0000
29	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9888	0.8505	0.4623	0.0148	0.0000
30	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9939	0.8962	0.5491	0.0248	0.0000
31	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9969	0.9307	0.6331	0.0398	0.0001
32	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9984	0.9554	0.7107	0.0615	0.0002
33	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9993	0.9724	0.7793	0.0913	0.0004
34	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9836	0.8371	0.1303	0.0009
35	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9906	0.8839	0.1795	0.0018

Poissonverteilung $X_\lambda \sim P(\lambda)$, Verteilungsfunktionen $F_\lambda(x) = P(X_\lambda \leqslant x)$

$\downarrow x \quad \lambda \rightarrow$	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2	2.1	2.2	2.3	2.4
0	0.3679	0.3329	0.3012	0.2725	0.2466	0.2231	0.2019	0.1827	0.1653	0.1496	0.1353	0.1225	0.1108	0.1003	0.0907
1	0.7358	0.6990	0.6626	0.6268	0.5918	0.5578	0.5249	0.4932	0.4628	0.4337	0.4060	0.3796	0.3546	0.3309	0.3084
2	0.9197	0.9004	0.8795	0.8571	0.8335	0.8088	0.7834	0.7572	0.7306	0.7037	0.6767	0.6496	0.6227	0.5960	0.5697
3	0.9810	0.9743	0.9662	0.9569	0.9463	0.9344	0.9212	0.9068	0.8913	0.8747	0.8571	0.8386	0.8194	0.7993	0.7787
4	0.9963	0.9946	0.9923	0.9893	0.9857	0.9814	0.9763	0.9704	0.9636	0.9559	0.9473	0.9379	0.9275	0.9162	0.9041
5	0.9994	0.9990	0.9985	0.9978	0.9968	0.9955	0.9940	0.9920	0.9896	0.9868	0.9834	0.9796	0.9751	0.9700	0.9643
6	0.9999	0.9999	0.9997	0.9996	0.9994	0.9991	0.9987	0.9981	0.9974	0.9966	0.9955	0.9941	0.9925	0.9906	0.9884
7	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9997	0.9996	0.9994	0.9992	0.9989	0.9985	0.9980	0.9974	0.9967
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9998	0.9997	0.9995	0.9994	0.9991
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9998

$\downarrow x \quad \lambda \rightarrow$	2.5	2.75	3	3.25	3.5	3.75	4	4.25	4.5	4.75	5	5.25	5.5	5.75	6
0	0.0821	0.0639	0.0498	0.0388	0.0302	0.0235	0.0183	0.0143	0.0111	0.0087	0.0067	0.0052	0.0041	0.0032	0.0025
1	0.2873	0.2397	0.1991	0.1648	0.1359	0.1117	0.0916	0.0749	0.0611	0.0497	0.0404	0.0328	0.0266	0.0215	0.0174
2	0.5438	0.4815	0.4232	0.3696	0.3208	0.2771	0.2381	0.2037	0.1736	0.1473	0.1247	0.1051	0.0884	0.0741	0.0620
3	0.7576	0.7030	0.6472	0.5914	0.5366	0.4838	0.4335	0.3862	0.3423	0.3019	0.2650	0.2317	0.2017	0.1749	0.1512
4	0.8912	0.8554	0.8153	0.7717	0.7254	0.6775	0.6288	0.5801	0.5321	0.4854	0.4405	0.3978	0.3575	0.3199	0.2851
5	0.9580	0.9392	0.9161	0.8888	0.8576	0.8229	0.7851	0.7449	0.7029	0.6597	0.6160	0.5722	0.5289	0.4866	0.4457
6	0.9858	0.9776	0.9665	0.9523	0.9347	0.9137	0.8893	0.8617	0.8311	0.7978	0.7622	0.7248	0.6860	0.6464	0.6063
7	0.9958	0.9927	0.9881	0.9817	0.9733	0.9624	0.9489	0.9326	0.9134	0.8914	0.8666	0.8392	0.8095	0.7776	0.7440
8	0.9989	0.9978	0.9962	0.9937	0.9901	0.9852	0.9786	0.9702	0.9597	0.9470	0.9319	0.9144	0.8944	0.8719	0.8472
9	0.9997	0.9994	0.9989	0.9980	0.9967	0.9947	0.9919	0.9880	0.9829	0.9764	0.9682	0.9582	0.9462	0.9322	0.9161
10	0.9999	0.9999	0.9997	0.9994	0.9990	0.9983	0.9972	0.9956	0.9933	0.9903	0.9863	0.9812	0.9747	0.9669	0.9574
11	1.0000	1.0000	0.9999	0.9999	0.9997	0.9995	0.9991	0.9985	0.9976	0.9963	0.9945	0.9922	0.9890	0.9850	0.9799
12	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9997	0.9995	0.9992	0.9987	0.9980	0.9970	0.9955	0.9937	0.9912
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9997	0.9996	0.9993	0.9989	0.9983	0.9975	0.9964
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9996	0.9994	0.9991	0.9986
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9997	0.9995
16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999





Poissonverteilung $X_\lambda \sim P(\lambda)$, Verteilungsfunktionen $F_\lambda(x) = P(X_\lambda \leqslant x)$

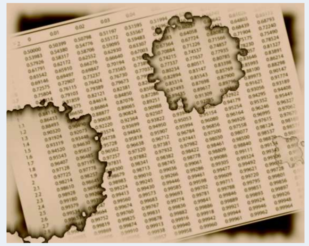
$\downarrow x \quad \lambda \rightarrow$	6.25	6.5	6.75	7	7.25	7.5	7.75	8	8.25	8.5	8.75	9	9.25	9.5	10
0	0.0019	0.0015	0.0012	0.0009	0.0007	0.0006	0.0004	0.0003	0.0003	0.0002	0.0002	0.0001	0.0001	0.0001	0.0000
1	0.0140	0.0113	0.0091	0.0073	0.0059	0.0047	0.0038	0.0030	0.0024	0.0019	0.0015	0.0012	0.0010	0.0008	0.0005
2	0.0517	0.0430	0.0357	0.0296	0.0245	0.0203	0.0167	0.0138	0.0113	0.0093	0.0076	0.0062	0.0051	0.0042	0.0028
3	0.1303	0.1118	0.0958	0.0818	0.0696	0.0591	0.0501	0.0424	0.0358	0.0301	0.0253	0.0212	0.0178	0.0149	0.0103
4	0.2530	0.2237	0.1970	0.1730	0.1514	0.1321	0.1149	0.0996	0.0862	0.0744	0.0640	0.0550	0.0471	0.0403	0.0293
5	0.4064	0.3690	0.3338	0.3007	0.2699	0.2414	0.2152	0.1912	0.1694	0.1496	0.1317	0.1157	0.1013	0.0885	0.0671
6	0.5662	0.5265	0.4876	0.4497	0.4132	0.3782	0.3449	0.3134	0.2838	0.2562	0.2305	0.2068	0.1849	0.1649	0.1301
7	0.7089	0.6728	0.6359	0.5987	0.5615	0.5246	0.4884	0.4530	0.4186	0.3856	0.3540	0.3239	0.2954	0.2687	0.2202
8	0.8204	0.7916	0.7611	0.7291	0.6960	0.6620	0.6274	0.5925	0.5577	0.5231	0.4890	0.4557	0.4232	0.3918	0.3328
9	0.8978	0.8774	0.8549	0.8305	0.8043	0.7764	0.7471	0.7166	0.6852	0.6530	0.6203	0.5874	0.5545	0.5218	0.4579
10	0.9462	0.9332	0.9183	0.9015	0.8828	0.8622	0.8399	0.8159	0.7903	0.7634	0.7352	0.7060	0.6760	0.6453	0.5830
11	0.9737	0.9661	0.9571	0.9467	0.9345	0.9208	0.9053	0.8881	0.8692	0.8487	0.8266	0.8030	0.7781	0.7520	0.6968
12	0.9880	0.9840	0.9790	0.9730	0.9658	0.9573	0.9475	0.9362	0.9234	0.9091	0.8932	0.8758	0.8568	0.8364	0.7916
13	0.9949	0.9929	0.9904	0.9872	0.9832	0.9784	0.9727	0.9658	0.9578	0.9486	0.9380	0.9261	0.9129	0.8981	0.8645
14	0.9979	0.9970	0.9958	0.9943	0.9923	0.9897	0.9866	0.9827	0.9781	0.9726	0.9661	0.9585	0.9499	0.9400	0.9165
15	0.9992	0.9988	0.9983	0.9976	0.9966	0.9954	0.9938	0.9918	0.9893	0.9862	0.9824	0.9780	0.9727	0.9665	0.9513
16	0.9997	0.9996	0.9994	0.9990	0.9986	0.9980	0.9973	0.9963	0.9950	0.9934	0.9914	0.9889	0.9859	0.9823	0.9730
17	0.9999	0.9998	0.9998	0.9996	0.9995	0.9992	0.9989	0.9984	0.9978	0.9970	0.9960	0.9947	0.9931	0.9911	0.9857
18	1.0000	0.9999	0.9999	0.9999	0.9998	0.9997	0.9996	0.9993	0.9991	0.9987	0.9982	0.9976	0.9968	0.9957	0.9928
19	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9997	0.9996	0.9995	0.9992	0.9989	0.9986	0.9980	0.9965
20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9998	0.9997	0.9996	0.9994	0.9991	0.9984
21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9998	0.9997	0.9996	0.9993
22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9997
23	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999
24	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
25	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen
Binomialverteilung
Poissonverteilung
Standardnormalverteilung
χ^2-Verteilung
t-Verteilung
F-Verteilung

Verteilungsfunktion Φ der Standardnormalverteilung

Dabei bedeutet $\Phi(x)$ zum Beispiel: $\Phi(2,13) = \Phi(2,1+0,03) = 0,9834$. Diesen Wert findet man in der Zeile mit $x_1 = 2,1$ und der Spalte mit $x_2 = 0,03$.

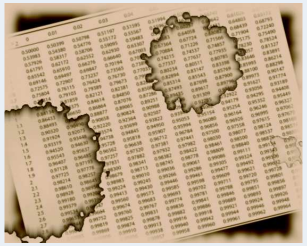
$x_1 \backslash x_2$	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1	0.84134	0.84375	0.84614	0.84850	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99897	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976



α-Fraktile der χ^2-Verteilung mit n Freiheitsgraden

$\downarrow \alpha \setminus n \rightarrow$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0.005	0.00	0.01	0.07	0.21	0.41	0.68	0.99	1.34	1.73	2.16	2.60	3.07	3.56	4.07	4.60
0.01	0.00	0.02	0.11	0.30	0.55	0.87	1.24	1.65	2.09	2.56	3.05	3.57	4.11	4.66	5.23
0.025	0.00	0.05	0.22	0.48	0.83	1.24	1.69	2.18	2.70	3.25	3.82	4.40	5.01	5.63	6.26
0.05	0.00	0.10	0.35	0.71	1.15	1.64	2.17	2.73	3.33	3.94	4.57	5.23	5.89	6.57	7.26
0.1	0.02	0.21	0.58	1.06	1.61	2.20	2.83	3.49	4.17	4.87	5.58	6.30	7.04	7.79	8.55
0.2	0.06	0.45	1.01	1.65	2.34	3.07	3.82	4.59	5.38	6.18	6.99	7.81	8.63	9.47	10.31
0.25	0.10	0.58	1.21	1.92	2.67	3.45	4.25	5.07	5.90	6.74	7.58	8.44	9.30	10.17	11.04
0.4	0.28	1.02	1.87	2.75	3.66	4.57	5.49	6.42	7.36	8.30	9.24	10.18	11.13	12.08	13.03
0.5	0.45	1.39	2.37	3.36	4.35	5.35	6.35	7.34	8.34	9.34	10.34	11.34	12.34	13.34	14.34
0.6	0.71	1.83	2.95	4.04	5.13	6.21	7.28	8.35	9.41	10.47	11.53	12.58	13.64	14.69	15.73
0.75	1.32	2.77	4.11	5.39	6.63	7.84	9.04	10.22	11.39	12.55	13.70	14.85	15.98	17.12	18.25
0.8	1.64	3.22	4.64	5.99	7.29	8.56	9.80	11.03	12.24	13.44	14.63	15.81	16.98	18.15	19.31
0.9	2.71	4.61	6.25	7.78	9.24	10.64	12.02	13.36	14.68	15.99	17.27	18.55	19.81	21.06	22.31
0.95	3.84	5.99	7.81	9.49	11.07	12.59	14.07	15.51	16.92	18.31	19.68	21.03	22.36	23.68	25.00
0.975	5.02	7.38	9.35	11.14	12.83	14.45	16.01	17.53	19.02	20.48	21.92	23.34	24.74	26.12	27.49
0.99	6.63	9.21	11.34	13.28	15.09	16.81	18.48	20.09	21.67	23.21	24.73	26.22	27.69	29.14	30.58
0.995	7.88	10.60	12.84	14.86	16.75	18.55	20.28	21.95	23.59	25.19	26.76	28.30	29.82	31.32	32.80

$\downarrow \alpha \setminus n \rightarrow$	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
0.005	5.14	5.70	6.26	6.84	7.43	8.03	8.64	9.26	9.89	10.52	11.16	11.81	12.46	13.12	13.79
0.01	5.81	6.41	7.01	7.63	8.26	8.90	9.54	10.20	10.86	11.52	12.20	12.88	13.56	14.26	14.95
0.025	6.91	7.56	8.23	8.91	9.59	10.28	10.98	11.69	12.40	13.12	13.84	14.57	15.31	16.05	16.79
0.05	7.96	8.67	9.39	10.12	10.85	11.59	12.34	13.09	13.85	14.61	15.38	16.15	16.93	17.71	18.49
0.1	9.31	10.09	10.86	11.65	12.44	13.24	14.04	14.85	15.66	16.47	17.29	18.11	18.94	19.77	20.60
0.2	11.15	12.00	12.86	13.72	14.58	15.44	16.31	17.19	18.06	18.94	19.82	20.70	21.59	22.48	23.36
0.25	11.91	12.79	13.68	14.56	15.45	16.34	17.24	18.14	19.04	19.94	20.84	21.75	22.66	23.57	24.48
0.4	13.98	14.94	15.89	16.85	17.81	18.77	19.73	20.69	21.65	22.62	23.58	24.54	25.51	26.48	27.44
0.5	15.34	16.34	17.34	18.34	19.34	20.34	21.34	22.34	23.34	24.34	25.34	26.34	27.34	28.34	29.34
0.6	16.78	17.82	18.87	19.91	20.95	21.99	23.03	24.07	25.11	26.14	27.18	28.21	29.25	30.28	31.32
0.75	19.37	20.49	21.60	22.72	23.83	24.93	26.04	27.14	28.24	29.34	30.43	31.53	32.62	33.71	34.80
0.8	20.47	21.61	22.76	23.90	25.04	26.17	27.30	28.43	29.55	30.68	31.79	32.91	34.03	35.14	36.25
0.9	23.54	24.77	25.99	27.20	28.41	29.62	30.81	32.01	33.20	34.38	35.56	36.74	37.92	39.09	40.26
0.95	26.30	27.59	28.87	30.14	31.41	32.67	33.92	35.17	36.41	37.65	38.89	40.11	41.34	42.56	43.77
0.975	28.85	30.19	31.53	32.85	34.17	35.48	36.78	38.08	39.36	40.65	41.92	43.19	44.46	45.72	46.98
0.99	32.00	33.41	34.81	36.19	37.57	38.93	40.29	41.64	42.98	44.31	45.64	46.96	48.28	49.59	50.89
0.995	34.27	35.72	37.16	38.58	40.00	41.40	42.80	44.18	45.56	46.93	48.29	49.64	50.99	52.34	53.67



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen
Binomialverteilung
Poissonverteilung
Standardnormalverteilung
χ^2-Verteilung
t-Verteilung
F-Verteilung

α-Fraktile der t-Verteilung mit n Freiheitsgraden

$\downarrow n \setminus \alpha \rightarrow$	0.6	0.75	0.8	0.9	0.95	0.975	0.99	0.995
1	0.325	1.000	1.376	3.078	6.314	12.706	31.820	63.657
2	0.289	0.816	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.765	0.979	1.638	2.353	3.183	4.541	5.841
4	0.271	0.741	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.727	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.718	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.711	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.706	0.889	1.397	1.860	2.306	2.897	3.355
9	0.261	0.703	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.700	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.698	0.875	1.363	1.796	2.201	2.718	3.106
12	0.259	0.696	0.873	1.356	1.782	2.179	2.681	3.054
13	0.259	0.694	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.692	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.691	0.866	1.341	1.753	2.131	2.603	2.947
16	0.258	0.690	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.689	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.688	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.688	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.687	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.686	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.686	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.685	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.685	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.684	0.856	1.316	1.708	2.059	2.485	2.787
26	0.256	0.684	0.856	1.315	1.706	2.055	2.479	2.779
27	0.256	0.684	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.683	0.855	1.312	1.701	2.048	2.467	2.763
29	0.256	0.683	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.683	0.854	1.310	1.697	2.042	2.457	2.750

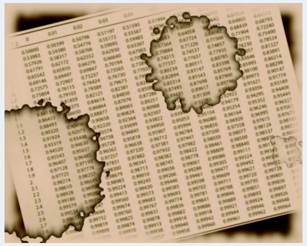
α-Fraktile der F-Verteilung mit den Freiheitsgraden ν_1 und ν_2

$\alpha = 0{,}95$

$\nu_1 \backslash \nu_2$	1	2	3	4	5	6	7	8	9	10	15	20	30	40	50	100
1	161.4	18.51	10.13	7.71	6.61	5.99	5.59	5.32	5.12	4.96	4.54	4.35	4.17	4.08	4.03	3.94
2	199.5	19.00	9.55	6.94	5.79	5.14	4.74	4.46	4.26	4.10	3.68	3.49	3.32	3.23	3.18	3.09
3	215.7	19.16	9.28	6.59	5.41	4.76	4.35	4.07	3.86	3.71	3.29	3.10	2.92	2.84	2.79	2.70
4	224.6	19.25	9.12	6.39	5.19	4.53	4.12	3.84	3.63	3.48	3.06	2.87	2.69	2.61	2.56	2.46
5	230.2	19.30	9.01	6.26	5.05	4.39	3.97	3.69	3.48	3.33	2.90	2.71	2.53	2.45	2.40	2.31
6	234.0	19.33	8.94	6.16	4.95	4.28	3.87	3.58	3.37	3.22	2.79	2.60	2.42	2.34	2.29	2.19
7	236.8	19.35	8.89	6.09	4.88	4.21	3.79	3.50	3.29	3.14	2.71	2.51	2.33	2.25	2.20	2.10
8	238.9	19.37	8.85	6.04	4.82	4.15	3.73	3.44	3.23	3.07	2.64	2.45	2.27	2.18	2.13	2.03
9	240.5	19.38	8.81	6.00	4.77	4.10	3.68	3.39	3.18	3.02	2.59	2.39	2.21	2.12	2.07	1.97
10	241.9	19.40	8.79	5.96	4.74	4.06	3.64	3.35	3.14	2.98	2.54	2.35	2.16	2.08	2.03	1.93
15	245.9	19.43	8.70	5.86	4.62	3.94	3.51	3.22	3.01	2.85	2.40	2.20	2.01	1.92	1.87	1.77
20	248.0	19.45	8.66	5.80	4.56	3.87	3.44	3.15	2.94	2.77	2.33	2.12	1.93	1.84	1.78	1.68
30	250.1	19.46	8.62	5.75	4.50	3.81	3.38	3.08	2.86	2.70	2.25	2.04	1.84	1.74	1.69	1.57
40	251.1	19.47	8.59	5.72	4.46	3.77	3.34	3.04	2.83	2.66	2.20	1.99	1.79	1.69	1.63	1.52
50	251.8	19.48	8.58	5.70	4.44	3.75	3.32	3.02	2.80	2.64	2.18	1.97	1.76	1.66	1.60	1.48
100	253.0	19.49	8.55	5.66	4.41	3.71	3.27	2.97	2.76	2.59	2.12	1.91	1.70	1.59	1.52	1.39

$\alpha = 0{,}99$

$\nu_1 \backslash \nu_2$	1	2	3	4	5	6	7	8	9	10	15	20	30	40	50	100
1	4052	98.50	34.12	21.20	16.26	13.75	12.25	11.26	10.56	10.04	8.68	8.10	7.56	7.31	7.17	6.90
2	5000	99.00	30.82	18.00	13.27	10.92	9.55	8.65	8.02	7.56	6.36	5.85	5.39	5.18	5.06	4.82
3	5403	99.17	29.46	16.69	12.06	9.78	8.45	7.59	6.99	6.55	5.42	4.94	4.51	4.31	4.20	3.98
4	5625	99.25	28.71	15.98	11.39	9.15	7.85	7.01	6.42	5.99	4.89	4.43	4.02	3.83	3.72	3.51
5	5764	99.30	28.24	15.52	10.97	8.75	7.46	6.63	6.06	5.64	4.56	4.10	3.70	3.51	3.41	3.21
6	5859	99.33	27.91	15.21	10.67	8.47	7.19	6.37	5.80	5.39	4.32	3.87	3.47	3.29	3.19	2.99
7	5928	99.36	27.67	14.98	10.46	8.26	6.99	6.18	5.61	5.20	4.14	3.70	3.30	3.12	3.02	2.82
8	5981	99.37	27.49	14.80	10.29	8.10	6.84	6.03	5.47	5.06	4.00	3.56	3.17	2.99	2.89	2.69
9	6022	99.39	27.35	14.66	10.16	7.98	6.72	5.91	5.35	4.94	3.89	3.46	3.07	2.89	2.78	2.59
10	6056	99.40	27.23	14.55	10.05	7.87	6.62	5.81	5.26	4.85	3.80	3.37	2.98	2.80	2.70	2.50
15	6157	99.43	26.87	14.20	9.72	7.56	6.31	5.52	4.96	4.56	3.52	3.09	2.70	2.52	2.42	2.22
20	6209	99.45	26.69	14.02	9.55	7.40	6.16	5.36	4.81	4.41	3.37	2.94	2.55	2.37	2.27	2.07
30	6261	99.47	26.50	13.84	9.38	7.23	5.99	5.20	4.65	4.25	3.21	2.78	2.39	2.20	2.10	1.89
40	6287	99.47	26.41	13.75	9.29	7.14	5.91	5.12	4.57	4.17	3.13	2.69	2.30	2.11	2.01	1.80
50	6303	99.48	26.35	13.69	9.24	7.09	5.86	5.07	4.52	4.12	3.08	2.64	2.25	2.06	1.95	1.74
100	6334	99.49	26.24	13.58	9.13	6.99	5.75	4.96	4.41	4.01	2.98	2.54	2.13	1.94	1.82	1.60



1. Einführung
2. Deskriptive Statistik
3. W-Theorie
4. Induktive Statistik
Quellen
Tabellen
Binomialverteilung
Poissonverteilung
Standardnormalverteilung
χ^2-Verteilung
t-Verteilung
F-Verteilung