

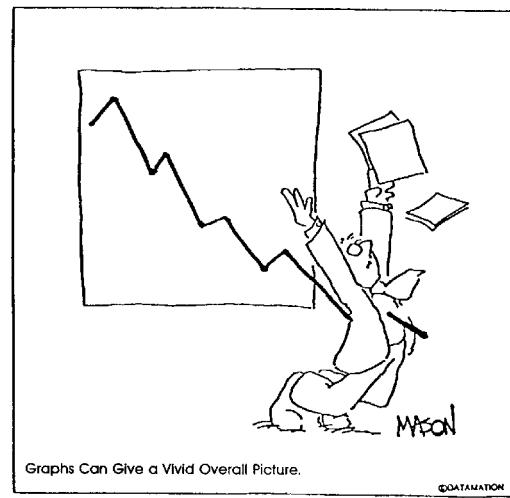


Grundbegriffe der Biostatistik

Theo Gasser & Burkhardt Seifert

Abteilung Biostatistik

Universität Zürich



[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
[Testen](#)
[Regression](#)
[Testfragen](#)
[Index](#)

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

[Seite 1 von 100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

Seite 2 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

1 Einführung	4
1.1 Was bietet die Statistik ?	5
1.2 Grundgesamtheit und Stichprobe	7
1.3 Empfohlene Literatur	9
2 Deskriptive Statistik	10
2.1 Haupttypen von Daten	12
2.2 Darstellung von diskreten Daten	13
2.3 Darstellung von Verläufen	17
2.4 Darstellung von stetigen Daten – das Histogramm	19
2.5 Lage- und Streumasse	24
3 Wahrscheinlichkeit	39
3.1 Ereignisse und ihre Wahrscheinlichkeiten	39
3.2 Bedingte Wahrscheinlichkeit und Unabhängigkeit	41
3.3 Wahrscheinlichkeitsverteilungen	43
3.4 Einige wichtige Verteilungen	50
3.5 Gesetze der grossen Zahlen	56
3.6 Transformationen von Daten und Verteilungen	58
3.7 Schätzverfahren für statistische Kennwerte	59
3.8 Versuchsplanung	67
4 Testen	76
4.1 Was ist ein statistischer Test ?	76
4.2 Tests auf Mittelwertsunterschiede	94
4.3 Tests für Proportionen oder Wahrscheinlichkeiten	106
4.4 Der χ^2-Test	107
4.5 Multiples Testen	113
4.6 Konfidenzintervalle (Vertrauensbereiche)	116
5 Regression	128
5.1 Bivariate Daten	129
5.2 Korrelation und ihre Eigenschaften	131
5.3 Tests auf linearen Zusammenhang und Konfidenzintervalle	138



5.4 Ausreisser und Gefahren der Korrelationsrechnung	140
5.5 Einfache lineare Regression	145
5.6 Multiple Regression	160
Testfragen	165
Index	182

Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)



Seite 3 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



1. Einführung

Die Medizin ist — neben den Sozialwissenschaften — derjenige universitäre Bereich, der am intensivsten moderne statistische Methodik benutzt. Auch wenn sich diese nur wenig von der anderer Wissenschaftszweige unterscheidet, wird sie doch oft als Bio- oder Medizinstatistik bezeichnet.

Das Spektrum der modernen Statistik reicht von einfachen quantitativen und grafischen Methoden bis hin zu komplexen Modellen, die nur mit Hilfe der höheren Mathematik behandelt werden können. Dabei stellt man in allen Anwendungen eine Verschiebung zugunsten des Einsatzes immer raffinierterer Methoden fest (z. B. logistische Regression oder Survivalanalyse). Dies hängt damit zusammen, dass von der medizinischen Forschung auch immer anspruchsvollere Fragestellungen und Projekte angegangen werden. Der Erfolg der Statistik wäre aber auch nicht denkbar ohne die Entwicklung günstiger und leistungsfähiger Computer und die Entwicklung von statistischen Programmpaketen. Die Programmpakete sind heute sehr leistungsfähig und benutzerfreundlich, vorausgesetzt, man versteht etwas von statistischen Methoden. In der biomedizinischen Forschung Tätige werden diese Methoden in verschiedenem Ausmass und verschiedener Vollständigkeit beherrschen (müssen). Ein Mediziner, der überwiegend in der Forschung tätig ist, sollte die Statistik besser beherrschen als ein überwiegend klinisch tätiger. Jeder Mediziner muss heute zumindest die Grundbegriffe der Statistik kennen und verstehen, um überhaupt medizinische Literatur lesen zu können.

Sie haben hier eine erste Version unseres Internet-Kurses in Biostatistik vor sich. Zur Vertiefung und Illustration des Stoffes dienen verschiedene Links:

[Animation](#)

[Datenanalyse](#)

[Selbstkontrolle](#)

(Bitte anklicken.)

März 2004

[Einführung](#)

[Deskriptive Statistik](#)

[Wahrscheinlichkeit](#)

[Testen](#)

[Regression](#)

[Testfragen](#)

[Index](#)

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 4 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

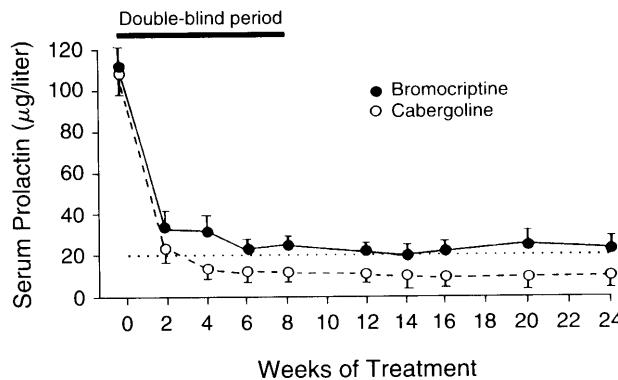


1.1. Was bietet die Statistik?

Statistische Methoden ermöglichen

- die Hervorhebung wesentlicher Zusammenhänge durch Datenreduktion und graphische Darstellungen.

Es folgt ein Beispiel aus einem Heft des New England Journal of Medicine (einer führenden klinischen Zeitschrift). Durch eine einfache Graphik wird die Wirkung von Behandlungen im Zeitverlauf auf einen Blick deutlich.



[Einführung](#)

[Deskriptive Statistik](#)

[Wahrscheinlichkeit](#)

[Testen](#)

[Regression](#)

[Testfragen](#)

[Index](#)

[Home Page](#)

[Titelseite](#)

◀◀

▶▶

◀

▶

Seite 5 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



[Einführung](#)

[Deskriptive Statistik](#)

[Wahrscheinlichkeit](#)

[Testen](#)

[Regression](#)

[Testfragen](#)

[Index](#)

[Home Page](#)

[Titelseite](#)



Seite 6 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Betrachten wir einen Artikel aus dem gleichen Heft (siehe Kopie auf der nächsten Seite). Ohne statistische Kenntnisse sind viele Artikel in medizinischen Zeitschriften nur schwer verständlich. Emerson et al. (1986) stellten fest, dass jemand, der nur die deskriptive Statistik beherrscht, knapp 58% der Artikel des New England Journal of Medicine verstehen kann. Das Verständnis des *t*-Tests erhöht diesen Anteil auf 67%; das zusätzliche Verständnis von Kontingenztafeln erhöht ihn auf 73%. Seit 1986 ist der Anteil höherer statistischer Methoden deutlich gestiegen.

Ein Ziel dieses Biostatistik-Kurses ist es denn auch, das Studium wissenschaftlich-medizinischer Literatur zu erleichtern.

Statistische Methoden ermöglichen es,

- aus einer Stichprobe gültige Schlussfolgerungen zu ziehen,
- die Unsicherheit der Entscheidung zu quantifizieren.

Dazu stehen gute statistische Programmpakete zur Verfügung, die zum Teil auch benutzerfreundlich sind. Mit diesen und dem Stoff der Vorlesung sollte es möglich sein, einfachere Auswertungen selber durchzuführen. Die meisten Beispiele in diesem Skript wurden mit dem Paket StatView analysiert, das auf dem Mac und auch unter Windows läuft. Ein grosser Vorteil ist die Benutzerfreundlichkeit, es ist aber weniger umfassend als andere bekannte Programmpakete wie SPSS und SAS.



INTENSIVE POSTREMISSION CHEMOTHERAPY IN ADULTS WITH ACUTE MYELOID LEUKEMIA

ROBERT J. MAYER, M.D., ROGER B. DAVIS, Sc.D., CHARLES A. SCHIFFER, M.D., DEBORAH T. BERG, R.N., BAYARD L. POWELL, M.D., PHILIP SCHULMAN, M.D., GEORGE A. OMURA, M.D., JOSEPH O. MOORE, M.D., O. ROSS MCINTYRE, M.D., AND EMIL FREI III, M.D., FOR THE CANCER AND LEUKEMIA GROUP B*

Abstract *Background.* About 65 percent of previously untreated adults with primary acute myeloid leukemia (AML) enter complete remission when treated with cytarabine and an anthracycline. However, such responses are rarely durable when conventional postremission therapy is administered. Uncontrolled trials have suggested that intensive postremission therapy may prolong these complete remissions.

Methods. We treated 1088 adults with newly diagnosed AML with three days of daunorubicin and seven days of cytarabine and randomly assigned patients who had a complete remission to receive four courses of cytarabine at one of three doses: 100 mg per square meter of body-surface area per day for five days by continuous infusion, 400 mg per square meter per day for five days by continuous infusion, or 3 g per square meter in a 3-hour infusion every 12 hours (twice daily) on days 1, 3, and 5. All patients then received four courses of monthly maintenance treatment.

Results. Of the 693 patients who had a complete remission, 596 were randomly assigned to receive post-

remission cytarabine. After a median follow-up of 52 months, the disease-free survival rates in the three treatment groups were significantly different ($P = 0.003$). Relative to the 100-mg group, the hazard ratios were 0.67 for the 3-g group (95 percent confidence interval, 0.53 to 0.86) and 0.75 for the 400-mg group (95 percent confidence interval, 0.60 to 0.94). The probability of remaining in continuous complete remission after four years for patients 60 years of age or younger was 24 percent in the 100-mg group, 29 percent in the 400-mg group, and 44 percent in the 3-g group ($P = 0.002$). In contrast, for patients older than 60, the probability of remaining disease-free after four years was 16 percent or less in each of the three postremission cytarabine groups.

Conclusions. These data support the concept of a dose-response effect for cytarabine in patients with AML who are 60 years of age or younger. The results with the high-dose schedule in this age group are comparable to those reported in similar patients who have undergone allogeneic bone marrow transplantation during a first remission. (N Engl J Med 1994;331:896-903.)

1.2. Grundgesamtheit und Stichprobe

Zwei zentrale Begriffe der Statistik sind Grundgesamtheit und Stichprobe. Individuen verhalten sich bei gleicher Behandlung unterschiedlich. Deshalb reicht es in

[Einführung](#)

[Deskriptive Statistik](#)

[Wahrscheinlichkeit](#)

[Testen](#)

[Regression](#)

[Testfragen](#)

[Index](#)

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 7 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)



Seite 8 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

der Medizin nicht, Einzelfälle zu dokumentieren. Es ist aber auch unmöglich und nicht unbedingt wünschenswert, das Verhalten **aller** Patienten auf eine Behandlung zu evaluieren. Dies ist einerseits eine Kostenfrage, andererseits geht es darum, Behandlungen mit Nebenwirkungen so sparsam wie möglich einzusetzen. Als Ausweg können wir eine **Stichprobe** aus der **Grundgesamtheit** (Population, statistische Grundgesamtheit) ziehen.

Die **Grundgesamtheit** ist die Gesamtheit aller Individuen, für welche Schlussfolgerungen gezogen werden sollen.

Beispiele:

1. Alle Hodgkin-Patienten der Welt, die die Krankheit überstanden haben.
2. Alle männlichen Einwohner der Schweiz über 65 Jahre, die weder an einer neurologischen noch an einer psychiatrischen Krankheit leiden.

Eine **Stichprobe** aus einer statistischen Grundgesamtheit ist die Menge der Individuen, die tatsächlich beobachtet wurden.

Beispiele:

1. Stichprobe von $n = 20$ Hodgkin-Patienten aus einer Klinik.
2. Repräsentative Stichprobe von $n = 1000$ Männern aus städtischen Gebieten.

[Animation](#)



[Einführung](#)

[Deskriptive Statistik](#)

[Wahrscheinlichkeit](#)

[Testen](#)

[Regression](#)

[Testfragen](#)

[Index](#)

[Home Page](#)

[Titelseite](#)



Seite 9 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

1.3. Empfohlene Literatur

Es gibt hunderte von Büchern zur Einführung in die Statistik. Ein grosser Teil davon ist brauchbar. Wir wollen nur einige der anwendungsorientierten Bücher mit ausreichendem theoretischem Anspruch erwähnen.

Altman, D. G. (1991). *Practical statistics for medical research*. Chapman and Hall.

Es bietet eine korrekte Statistik und sehr gute Beispiele. Aus diesem Buch haben wir einige der vorgestellten Beispiele. 600 S.

Bland M. (1995). *An introduction to medical statistics*. Oxford Medical Publications.

Eine sehr gute Einführung mit vielen Beispielen und Aufgaben. 396 S.

Johnson R. A. & Bhattacharyya G. K. (1992). *Statistics. Principles and methods*. 2nd ed., Wiley.

Eine leichte Lektüre zum Feierabend. 700 S.

Matthews, D. E. & Farewell, V. T. (1988). *Using and understanding medical statistics*. 2nd ed., Karger.

Im Gegensatz zu vielen anderen Einführungen bietet dieses Buch logistische Regression und Survivalanalyse. 200 S.

Sachs, L. (2002). *Angewandte Statistik: Anwendung statistischer Methoden*. 10. Auflage, Springer.

Im deutschsprachigen Raum das klassische Kochbuch (Nachschlagewerk) für alle Lebenslagen seit 1968. Achtung: Kein Lehrbuch! 889 S.



2. Deskriptive Statistik

Die Beschreibung von Daten ist die Grundlage jeder statistischen Analyse und ein wesentlicher Bestandteil jeder Publikation. Das Ziel besteht darin, die Daten einer Stichprobe kurz und prägnant zu charakterisieren. Dies erfolgt einerseits über statistische Kennwerte (z. B. Lage- und Streumasse) und andererseits durch graphische Verfahren. Dank PC und der Entwicklung guter, einfach handhabbarer Programme haben graphische Methoden an Bedeutung gewonnen.

Die deskriptive Statistik unterscheidet sich von der schliessenden Statistik dadurch, dass die Daten **ohne Signifikanz** präsentiert werden, und sie so ohne Wahrscheinlichkeitsannahmen auskommt. Trotzdem verweisen wir in diesem Kapitel gelegentlich auf Begriffe der Wahrscheinlichkeitsrechnung (Erwartungswert, Dichte, Normalverteilung). Diese Begriffe werden in Kapitel 3 erklärt.

Daten werden heute meistens mittels eines Tabellenkalkulations-Programmes oder einer Datenbank in den Computer eingegeben. Für erstere Option ist das Programm Excel eine beliebte und auch gute Wahl. Gut ist im Prinzip auch Filemaker, doch gibt es Probleme beim Export der Daten in Statistik-Programme. Ein Vorteil von Filemaker ist, dass Datenchecks leicht durchgeführt werden können. Für befristete und nicht zu komplexe Projekte ist die Benutzung einer Datenbank nicht angezeigt; die verbreitetsten Datenbanken erlauben aber den Export in die gängigen Statistik-Pakete. Zum Verständnis der Materie ist es hilfreich, die vorgestellten Analysen selbst durchzuführen. Um dies zu vereinfachen, bieten wir Ihnen die Datensätze und ein web-basiertes Statistikprogramm an. Die entsprechenden Links sehen so aus: [Datenanalyse](#). Wenn Sie diesem Link folgen, werden noch einige Tipps zur Verwendung des Programms und zum herunterladen der Daten gegeben.

Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)

«

»

◀

▶

Seite 10 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)



Seite 11 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

In der Regel werden die Daten als Tabelle eingegeben, wobei jede Zeile einen Patienten darstellt, die Kolonnen die jeweiligen Variablen. Hier sehen Sie die Daten einer Studie als Excel-Tabelle, die später immer wieder als Beispiel dient. Hier wurden immunologische Parameter (T_4 - und T_8 -Zellen) bei 20 Hodgkin und 20 non-Hodgkin-Patienten verglichen (siehe Abschnitt 2.4).

Patient	T4-cells	T8-cells	ln(T4-cells)	ln(T8-cells)	Disease	group
1	396	836	5.981	6.729	Hodgkin	1
2	568	978	6.342	6.886	Hodgkin	1
3	1212	1678	7.1	7.425	Hodgkin	1
...						
19	1283	336	7.157	5.817	Hodgkin	1
20	2415	936	7.789	6.842	Hodgkin	1
21	375	340	5.927	5.829	Non-Hodgkin	2
22	375	330	5.927	5.799	Non-Hodgkin	2
23	752	627	6.623	6.441	Non-Hodgkin	2
...						
39	377	108	5.932	4.682	Non-Hodgkin	2
40	503	163	6.221	5.094	Non-Hodgkin	2



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 12 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

2.1. Haupttypen von Daten

Am Anfang ist es wichtig, sich über die verschiedenen Typen von Daten klar zu werden. Verschiedene Datentypen bedingen nämlich verschiedene Präsentationen und Analysemethoden.

1. Qualitative oder **diskrete** Messdaten

Diskrete Daten sind dadurch gekennzeichnet, dass sie nur bestimmte Werte annehmen können, z. B.: {rot, grün, blau}, {O, A, B, AB}, {männlich, weiblich} oder {0, 1, 2, ...}, wobei diese Zahlen z. B. den Schweregrad einer Krankheit bedeuten können.

Wir unterscheiden:

- **nominal** oder kategorial (Zuordnung zu Kategorien):
Beziehung: gleich \iff ungleich
→ nur Anzahl und % sinnvoll
Beispiele: Geschlecht, Blutgruppe, Farbe
- **ordinal** (geordnet kategorial): Beziehung: grösser \iff kleiner (Rangordnung)
Beispiele: Schweregrad einer Krankheit, Items in Fragebogen

2. Quantitative oder **stetige** Messdaten

Stetige Daten können idealerweise alle Werte ohne Abstufungen annehmen. Oft ist es sinnvoll, auch ganzzahlige (also eigentlich diskrete) Werte als stetig zu behandeln (z. B. Grösse in cm oder Zähldaten).



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

Selbstkontrolle

Animation

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 13 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Es werden noch intervallskalierte und absolutskalierte Variablen unterschieden, doch hat dies für die Statistik selten eine Bedeutung. Bei absolutskalierten Größen ist ein absoluter Nullpunkt vorgegeben, und dadurch kann man auch sinnvolle Quotienten bilden. Beispiel: Temperatur in Kelvin anstatt Grad Celsius.

In einem psychologischen Sinne ist Farbe kategorial skaliert, durch den wissenschaftlichen Fortschritt kann man sie aber auch als stetig (Frequenzen von elektromagnetischen Wellen) auffassen.

2.2. Darstellung von diskreten Daten

Bei der Analyse von diskreten Daten spielt die **Wahrscheinlichkeit** eines einzelnen Ereignisses eine zentrale Rolle. Die relative Häufigkeit schätzt diese Wahrscheinlichkeit.

$$\text{relative Häufigkeit} = \frac{\text{Anzahl Beobachtungen des Ereignisses}}{\text{totale Anzahl Beobachtungen}}$$

- multipliziert mit 100 erhält man Prozentsätze

Achtung: Diese einfache Rechnung wird oft falsch gemacht:

„Local police are particularly pleased with decreases in the numbers of robbe-

ries, break-ins and car-related crimes. In fact robberies in the area are down by a staggering 100 per cent.“ *Ayrshire Leader*, 4 November 1993



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)

◀◀

▶▶

◀

▶

Seite 14 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 15 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Um relative Häufigkeit in 2 Gruppen zu vergleichen, können die Anzahlen von Ereignissen in einer Vierfeldertafel dargestellt werden.

Beispiel: Brown (1980) untersuchte an 52 Patienten, ob ein positives Röntgenbild geeignet zur Diagnose des Auftretens von Knoten ist.

Die Vierfeldertafel für den Zusammenhang von Auftreten von Knoten und Röntgenbefund sieht so aus:

		Röntgenbefund		32
		$x = 0$	$x = 1$	
kein Knoten ($y = 0$)	37	28	4	
	Knoten ($y = 1$)	9	11	20
		37	15	52

Die Vorhersagekraft des Röntgenbefundes wird mit **Sensitivität** = $11/20 = 55\%$ und **Spezifität** = $28/32 = 87\%$ bewertet. Die Sensitivität gibt bei diagnostischen Tests an, welchen Prozentsatz der Untersuchten zu recht als krank identifiziert wurde. Die Spezifität wiederum zeigt, ob Gesunde auch als Gesunde anerkannt wurden.



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

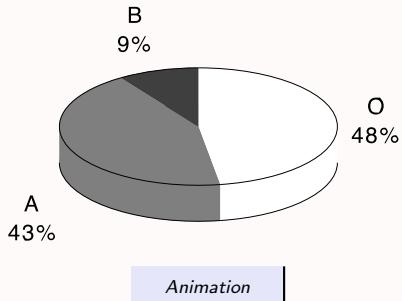
Index

Beispiel: Anteil von Blutgruppen in einer gesunden Population. So könnte die Präsentation der Daten aussehen:

Tabelle

Blutgruppe	Anzahl	relative Häufigkeit
O	2892	48%
A	2625	43%
B	570	9%
total	6087	100%

Kuchendiagramm (pie chart)



Animation

Beachten Sie, dass in der 3-dimensionalen Darstellung die Proportionen durch den perspektivischen Effekt visuell verfälscht werden.

[Home Page](#)

[Titelseite](#)



Seite 16 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)

[Selbstkontrolle](#)



Seite 17 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Balkendiagramm (bar chart)



Beachten Sie, dass die Balken immer vom Nullpunkt ausgehen sollten.



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)



Seite 18 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

2.3. Darstellung von Verläufen

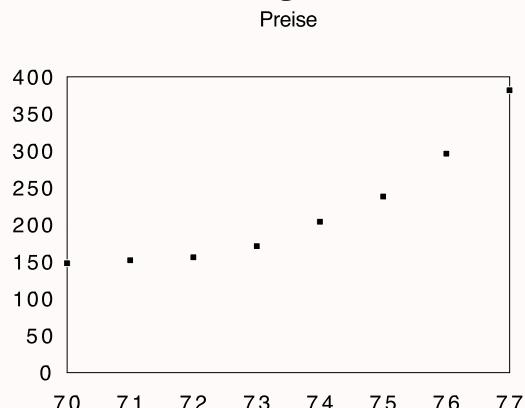
- Ordinale und stetige Daten

Die Darstellung von Verläufen ist ein heikles Thema. Hier eine extrem missbräuchliche und eine korrekte Darstellung:

Liniendiagramm



Scattergramm



[Animation](#)

Die Gründe dafür sind:

- Das Verhältnis der x - zu der y -Skala ist extrem gewählt.
- Der Nullpunkt ist nicht in der Graphik enthalten!

Es ist übrigens in gewissen Graphikprogrammen nicht einfach, den Nullpunkt in ein Scattergramm einzubeziehen.



- Benutzen Sie Ihren gesunden Menschenverstand, um zu entscheiden, ob der Nullpunkt für die Darstellung bedeutsam ist.

Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

Home Page

Titelseite

◀◀

▶▶

◀

▶

Seite 19 von 100

Zurück

Vollbild

Schließen

Beenden



[Einführung](#)

[Deskriptive Statistik](#)

[Wahrscheinlichkeit](#)

[Testen](#)

[Regression](#)

[Testfragen](#)

[Index](#)

[Home Page](#)

[Titelseite](#)



[Seite 20 von 100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

2.4. Darstellung von stetigen Daten – das Histogramm

Ziel:

Verteilung der Daten graphisch zu charakterisieren, im Sinne einer „Datendichte“ durch das Histogramm.

Vorgehen:

- Bereich der Daten in gleiche, nicht überlappende Intervalle (Zellen, Klassen) zerlegen
- Anzahl Beobachtungen pro Intervall bestimmen

$$\text{relative Häufigkeit im Intervall} = \frac{\text{Anzahl Beobachtungen im Intervall}}{\text{totale Anzahl Beobachtungen}}$$

- relative (oder absolute) Häufigkeiten über Intervalle in Balkendiagramm darstellen

Beispiel: Immunologische Variablen als Anzahl von T_4 - und T_8 -Zell-Anzahlen bei verschiedenen Krebspatienten. Dies ist ein Beispiel aus dem Am. J. Med. Sci., das wir noch mehrmals benutzen werden:



Immunologic Status of Patients in Remission from Hodgkin's Disease and Disseminated Malignancies

BY CHARLES M. SHAPIRO, MD, ENRIQUE BECKMANN, MD, PhD,
NEAL CHRISTIANSEN, MD, JACOB D. BITRAN, MD, MARK KOZLOFF, MD,
ARTHUR A. BILLINGS, MD, MARGARET C. TELFER, MD

ABSTRACT: The immunologic phenotypes of peripheral blood lymphocytes of 20 patients cured of Hodgkin's disease (Group I) and of 20 patients cured of diverse, disseminated malignancies (Group II) were compared. Eleven patients in Group I had an increase in circulating B lymphocytes, and 6 patients in Group II had a decrease in this lymphocyte subset ($0.0005 < p < 0.0025$). The T_s lymphocyte subset was increased in four patients in Group I and decreased in five patients in Group II ($0.005 < p < 0.0125$). The T_4/T_s lymphocyte ratio was less than 1.0 in eight patients in Group I and in no patients in Group II ($p < 0.001$). These statistically significant differences between the two groups were unrelated to type or duration of therapy, and thus suggest basic biologic differences between the immune status of cured Hodgkin's disease patients and cured patients with other malignancies. KEY INDEXING TERMS: Immunologic Phenotypes; T_4/T_s Lymphocyte Ratio; Hodgkin's Disease; Cancer. [Am J Med Sci 1986; 293(6):366-370.]

ease.⁷⁻⁹ Less defined are the status of the immune system in patients in remission from HD and other malignancies and the effect of different prior therapies on immune function.^{10,11}

In 1983, Lauria et al described several abnormalities in T lymphocyte subsets in disease-free HD patients, 5 years after therapy.¹² They noted a decrease in the proportion of T_4 lymphocytes. The T_s lymphocyte subset was increased, both proportionally and in absolute numbers, resulting in an abnormal T_4/T_s ratio. This study was undertaken to confirm these findings and to determine whether they are unique to patients with HD.

Materials and Methods

Forty patients with disseminated malignancies were entered into the study. They were disease-free for periods of 3–20 years. There were 20 HD patients who had been staged according to the Ann Arbor classification (Table 1).¹³ All but six patients had splenectomies during staging laparotomies. The modes of treatment consisted of standard MOPP chemotherapy (nitrogen mustard, vincristine,

Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite 21 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)



[Seite 22 von 100](#)

[Zurück](#)

[Vollbild](#)

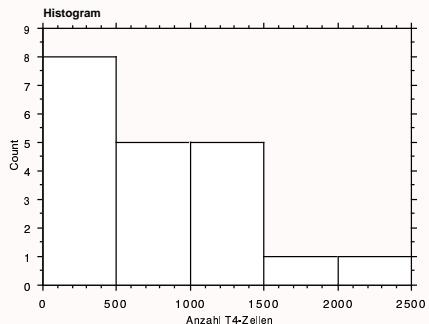
[Schließen](#)

[Beenden](#)

Für das Merkmal „Anzahl T_4 -Zellen bei Hodgkin-Patienten“ sieht die Berechnung eines Histogramms etwa so aus, wenn man Intervalle zu 500 bildet:

Intervall	Anzahl T_4 -Zellen	Anz. Beob. im Intervall	relative Häufigkeit
1–500	171 257 288 295 396 397 431 435	8	40%
501–1000	554 568 795 902 958	5	25%
1001–1500	1004 1104 1212 1283 1378	5	25%
1501–2000	1621	1	5%
2001–2500	2415	1	5%
total		20	100%

Aus der Tabelle ergibt sich das folgende Histogramm:



Bemerkung: „Count“ bei der y –Achse bedeutet Anzahl Patienten pro Intervall.



Einführung

Deskriptive Statistik

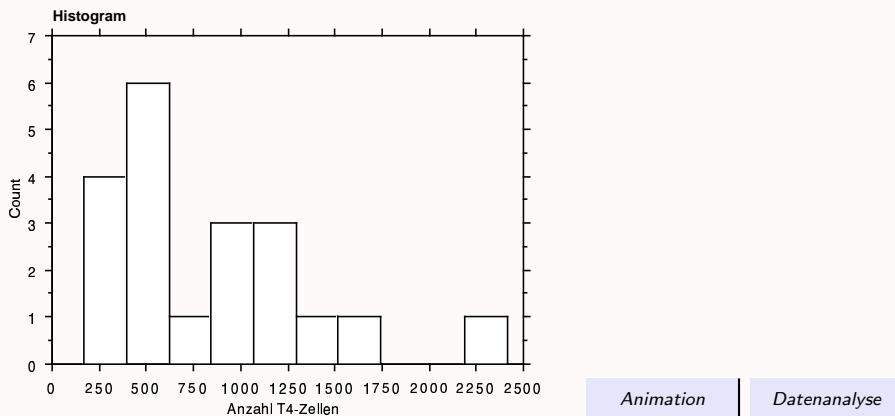
Wahrscheinlichkeit

Testen

Regression

Testfragen

Index



- Vergleichen Sie diese beiden Histogramme und Sie bemerken, dass man die Graphiken unterschiedlich interpretieren kann.
 - Steigt die relative Häufigkeit am Anfang an?
 - Gibt es Lücken?
- Eine wesentlich Erkenntnis ist, dass die Anzahl T_4 -Zellen (rechts-)schief verteilt ist, d. h. dass rechts vom Häufigkeitszentrum mehr Werte als links davon auftreten. Eine logarithmische Transformation führt zu einer eher symmetrischen Verteilung, ähnlich einer Normalverteilung (Gaussche Glockenkurve).
- Die Aussage eines Histogramms hängt wesentlich von der **Klassenbreite** und dem **Klassenzentrum** ab.

Wenn man eine deutlich zu kleine Intervalllänge wählt, erhält man z. B. ein sehr variables Histogramm.

[Home Page](#)

[Titelseite](#)



Seite 23 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Das Histogramm ist ein Schätzer der Wahrscheinlichkeitsdichte.

- Histogramme sind einfach und verbreitet.
- Es gibt bessere Dichteschätzer, die aber nur in professionelleren Statistikpaketen verfügbar sind.

Eine ähnliche, wenn auch weniger prägnante Information, vermittelt die empirische Verteilungsfunktion: Dabei steigt die Funktion treppenartig um $1/n$, wenn ein Datenpunkt dazukommt ($n = \text{Stichprobengröße}$). Man spricht von "kumulativer" Darstellung. Dass kleine Werte gegenüber grossen Werten überwiegen (d.h., dass die Verteilung rechtsschief ist) sieht man am starken Anstieg zu Beginn, aber weniger gut als im Histogramm. Die empirische Verteilungsfunktion hat eine gewisse Bedeutung bei der Analyse von Überlebenszeiten. Die Verteilungsfunktion beginnt dadurch bei $y = 0$ und steigt dann an bis $y = 1$.

[Einführung](#)

[Deskriptive Statistik](#)

[Wahrscheinlichkeit](#)

[Testen](#)

[Regression](#)

[Testfragen](#)

[Index](#)

[Home Page](#)

[Titelseite](#)

◀◀

▶▶

◀

▶

Seite 24 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)

◀◀

▶▶

◀

▶

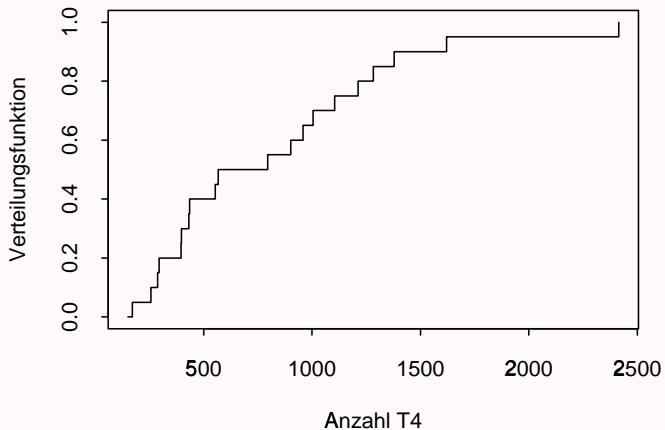
Seite 25 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Im allgemeinen kann man aus der kumulativen Darstellung weniger ersehen als aus dem Histogramm, was die grössere Verbreitung des Histogramms erklärt.

2.5. Lage- und Streumasse

2.5.1. Perzentile oder Quantile

Perzentile sind Hilfsmittel zur Beschreibung der Verteilung der Daten ohne irgendwelche Verteilungsannahmen.

Bedeutung:

1. zur Charakterisierung einer Stichprobe



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)



[Seite 26 von 100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

2. zur Konstruktion von Normwerten (biochemisch, anthropometrisch, psychometrisch)

Erklärung $\alpha\%$ – Perzentil: $\alpha \times 100\%$ der Daten sind kleiner als das $\alpha \times 100\%$ – Perzentil.

In der folgenden Graphik finden Sie eine Erklärung für Perzentile. Die Treppenlinie stellt die empirische (Stichproben-) Verteilungsfunktion der Anzahl T_4 -Zellen für $n = 20$ Hodgkin-Patienten dar. Wenn man von $y = \alpha \times 100\%$ eine Horizontale zieht und beim Schnittpunkt mit der Verteilungsfunktion den zugehörigen x -Wert abliest, erhält man das $\alpha \times 100\%$ Perzentil. Für das 77.5% Perzentil trifft man eine Stufe, und damit ist dieses Perzentil eindeutig definiert. Beim 50% Perzentil oder Median trifft man auf ein horizontales Stück und trifft damit 2 Datenpunkte. Als Median wählt man dann den Mittelwert der 2 x -Werte.



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)



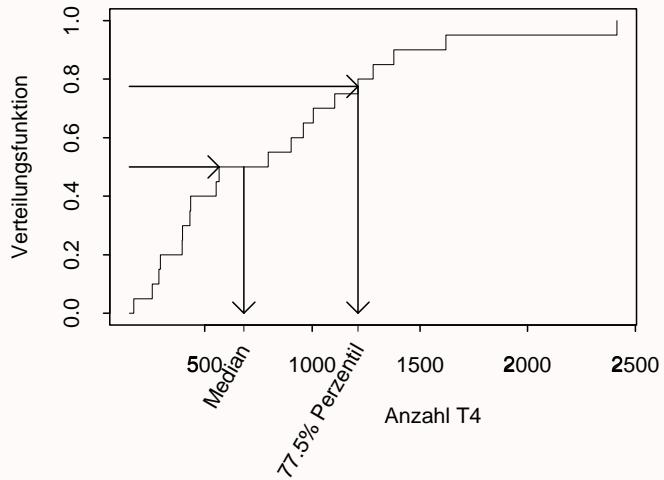
Seite 27 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Genaue **Definition:** Mindestens $\alpha\%$ der Daten sind gleich oder kleiner und mindestens $(100 - \alpha)\%$ sind gleich oder grösser als das $\alpha\%$ – Perzentil.

Quantile, in der Statistik gebräuchlich, sind das gleiche wie Perzentile bis auf den Faktor 100 (also z. B. α – Quantil).

Wichtige Perzentile:

- **Median** = 50% Perzentil

Der Median ist ein Lagemaß, der das Zentrum der Daten charakterisiert.



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

Home Page

Titelseite



Seite 28 von 100

Zurück

Vollbild

Schließen

Beenden

- **Quartile** = 25% und 75% Perzentile

Der Abstand zwischen den Quartilen (Interquartilabstand oder englisch „interquartile range“) ist ein Streumass. Dieser Bereich enthält die zentralen 50% der Daten.

Achtung! Die Werte sind nicht eindeutig! Verschiedene Programme können also etwas verschiedene Werte liefern. Dies hat damit zu tun, dass man auf verschiedene Arten Anteile von Daten verrechnen kann, wenn die Prozentrechnung nicht eine eindeutige Beobachtung ergibt.

Der Median von 9 Beobachtungen berechnet sich z. B. so, dass man diese ordnet und die 5. Beobachtung als Median nimmt. Bei 10 Werten liegt der Median zwischen dem 5. und 6. geordneten Wert, und man nimmt dann üblicherweise das Mittel beider Werte.

So sieht das Vorgehen am Beispiel der Anzahl T_4 -Zellen aus:



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)



[Seite 29 von 100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Nr.	Anzahl T_4 -Zellen	Percentil
1	171	
2	257	← 10% Perzentil = 272.5
3	288	
4	295	
5	396	← unteres Quartil = 396.5
6	397	
7	431	
8	435	
9	554	
10	568	← Median = 681.5
11	795	
12	902	
13	958	
14	1004	
15	1104	← oberes Quartil = 1158
16	1212	
17	1283	
18	1378	← 90% Perzentil = 1499.5
19	1621	
20	2415	



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 30 von 100

[Zurück](#)

[Vollbild](#)

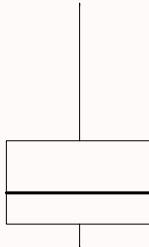
[Schließen](#)

[Beenden](#)

Boxplots

Boxplots sind ein gutes Hilfsmittel, um optisch die Verteilung der Daten zu erfassen und Ausreisser zu erkennen. Es handelt sich um eine Methode, die auf Perzentilen basiert.

Das ist die einfachste Form:



Die „Box“ gibt den Bereich vom 25. zum 75. Perzentil an, der horizontale Strich in der Box den Median. Die Stäbe (whiskers), die aus der Box hinausführen, sind nicht einheitlich definiert. In StatView geben sie die 10% und 90% Perzentile an. In anderen Statistikprogrammen gehen sie vom Minimum zum Maximum oder sie charakterisieren die Grenzen für Ausreisser und extreme Werte. In SPSS z. B. ist ein Ausreisser eine Beobachtung, die weiter als 1.5 Interquartilabstände (Box-Längen) von der Box entfernt ist.



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)

[◀◀](#)

[▶▶](#)

[◀](#)

[▶](#)

Seite 31 von 100

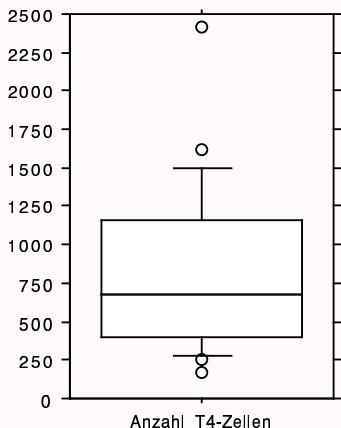
[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

So sieht der Boxplot in StatView aus:



Die unteren und oberen Querlinien (whiskers) geben die 10% und 90% Perzentile an, „o“ sind extreme Werte. Die Asymmetrie des Boxplots deutet auf eine rechtsschiefe Verteilung hin: Perzentile, die symmetrisch um den Median liegen (25. und 75. Perzentil, 10. und 90. Perzentil), haben gegen oben einen größeren Abstand als gegen unten. Sehr klar tritt der Extremwert bei $2415\text{ T}_4\text{-Zellen/mm}^3$ hervor.

Besonders hilfreich sind Boxplots, wenn mehrere Merkmale bzw. Gruppen verglichen werden:



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)



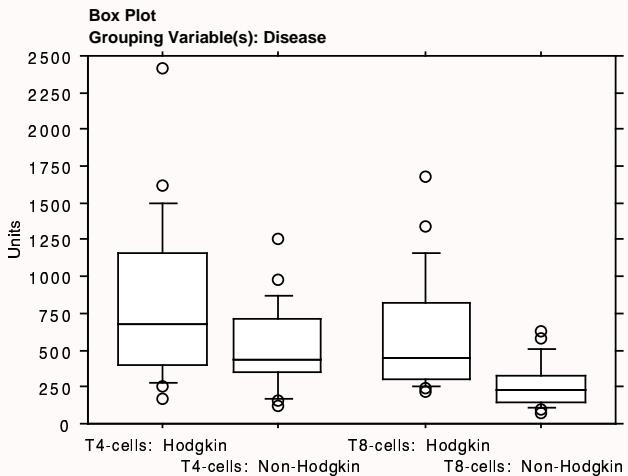
Seite 32 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Was sieht man auf einen Blick?

- Die immunologischen Variablen sind schief verteilt, denn der Median teilt den Interquartilsabstand nicht gleichmässig und Extremwerte sind ebenfalls asymmetrisch.
- Offensichtlich liegen die Anzahlen von T_4 - und T_8 -Zellen bei non-Hodgkin-Patienten tiefer und streuen weniger als bei Hodgkin-Patienten. Dies entspricht der wissenschaftlichen These der Hodgkin-Studie.
- Die Anzahlen von T_8 -Zellen liegen tiefer als die von T_4 -Zellen und streuen weniger, und zwar bei beiden Gruppen.



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)



Seite 33 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

2.5.2. Charakterisierung des Zentrums der Daten

- Was ist ein typischer, mittlerer Wert ?

Möglichkeiten:

- graphisch: Histogramme und Boxplots vermitteln einen Eindruck, wo sich die Daten befinden. Um eine objektive Zusammenfassung der Daten zu erhalten, benötigen wir aber klar definierte Masszahlen für das Zentrum der Daten.
- quantitativ durch den **Mittelwert**: Der Mittelwert (mean, average) beschreibt das Verhalten der Daten „im Mittel“ (\sum = Summenzeichen).

$$\bar{x} = (x_1 + x_2 + \dots + x_n) / n = \frac{1}{n} \sum_{i=1}^n x_i$$

Bei **Normalverteilung** ist der Mittelwert optimaler Schätzer des **Erwartungswertes**.

- quantitativ durch den **Median**: Der Median (median) beschreibt die „Mitte“ der Daten. Er ist der 50% – Punkt, d. h. die Hälfte der Stichprobe liegt über dem Median, und die andere Hälfte liegt darunter (siehe Perzentile).

Ein anschauliches Beispiel für die Unterschiede zwischen Mittelwert und Median gibt die Einkommensverteilung. Wenn z. B. in einer Gemeinde ein extrem gutverdienender Einwohner wohnt, so ist der Mittelwert relativ hoch, auch wenn die übrigen



[Einführung](#)

[Deskriptive Statistik](#)

[Wahrscheinlichkeit](#)

[Testen](#)

[Regression](#)

[Testfragen](#)

[Index](#)

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite 34 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Einwohner weniger verdienen. Der Median charakterisiert dann besser das typische Einkommen der Bürger, während das Mittel für die Steuerkraft ein aussagekräftiger Index ist.

Bei Verteilungen, die rechtschief sind, d. h. wo grössere Abweichungen nach oben als nach unten auftreten, ist der Mittelwert grösser als der Median (Beispiel: Einkommen und viele biochemische Variablen). Bei symmetrischer Verteilung (im besonderen Normalverteilung) sind Median und Mittelwert identisch (Beispiel: Körpergrösse).

Vorteile der verschiedenen Methoden:

Die Wahl zwischen Mittelwert und Median ist

- abhängig davon, ob ein typischer oder mittlerer Wert gesucht wird,
- abhängig von der Verteilung (normal, schief, gibt es Ausreisser?),
- abhängig davon, ob statistische Präzision oder Robustheit im Vordergrund stehen.

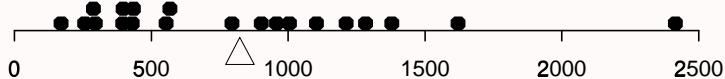
Der folgende Vergleich soll die Unterschiede zwischen Mittelwert und Median illustrieren.

Der **Mittelwert** ist derjenige Wert, der die Daten auf einer „Waage“ ausbalanciert.

Wir gehen von einer Waage ohne Gewicht und gleich schweren Beobachtungen aus.



$$\text{Mittelwert: } \sum_{i=1}^n (x_i - \bar{x}) = 0$$



- Entfernte Beobachtungen haben eine starke „Hebelkraft“.

Der Mittelwert ist **empfindlich** gegen Ausreißer.

Median: Beim Median spielt der Abstand der Beobachtungen vom Zentrum keine Rolle.



- Der Median ist **robust** gegen Ausreißer.

[Animation](#)

[Einführung](#)

[Deskriptive Statistik](#)

[Wahrscheinlichkeit](#)

[Testen](#)

[Regression](#)

[Testfragen](#)

[Index](#)

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 35 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



2.5.3. Streuung oder Variabilität einer Stichprobe

- Wie stark variieren die Daten um die mittlere Lage ?

Jetzt sind also nicht die x_i , sondern die $(x_i - \bar{x})$ relevant.

Möglichkeiten:

- graphisch: Histogramme und Boxplots vermitteln einen visuellen Eindruck der Variabilität der Daten.
- quantitativ durch die **Varianz** s^2 : Die Variabilität ist durch die Grösse der Abweichungen $(x_1 - \bar{x}), \dots, (x_n - \bar{x})$ vom Mittelwert charakterisiert.

Aufsummieren dieser Abweichungen bringt nichts, da sich positive und negative Abweichungen aufheben (Siehe dazu das Beispiel mit der Waage in diesem Kapitel.) Die absolute Grösse der Abweichungen zu ermitteln wäre eine mögliche Strategie. Aus theoretischen Gründen ist aber das Mittel der Quadrate dieser Abweichungen sinnvoller. Dies wird Varianz genannt:

$$\begin{aligned}s^2 &= \{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2\} / (n - 1) \\ &= \frac{1}{n - 1} \sum_{i=1}^n (x_i - \bar{x})^2\end{aligned}$$

Der Nenner $(n - 1)$ (anstatt n) ist mathematisch begründbar, da mit n die Werte systematisch etwas zu hoch liegen. Für grössere n sind die Differenzen aber klein.



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 37 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

3. quantitativ durch die **Standardabweichung**: (standard deviation, SD)

- Störend an der Varianz ist, dass wir quadrierte Einheiten erhalten (und damit oft grosse Masszahlen). So kann es passieren, dass bei einem mittleren Monatseinkommen von Fr. 6900.- sich eine Varianz von 144'000 Fr² ergibt. Deshalb ist es gebräuchlich die Standardabweichung s als Wurzel der Varianz anzugeben: $s = \sqrt{\text{Varianz}}$ (z. B. in Fr.)

Die Standardabweichung des monatlichen Einkommens ist dann Fr. 1200.-.

Die Standardabweichung und die Varianz sind empfindlich gegen Ausreisser.

Bei der **Normalverteilung** liegen 68% der Daten im Bereich Mittelwert ± 1 SD und 95% im Bereich Mittelwert ± 2 SD.

4. quantitativ durch die **Spannweite**

= Maximum – Minimum

- Die Spannweite gibt den Bereich (range) aller Daten an. Sie ist stark durch Extremwerte beeinflusst und hängt zudem von n ab.

5. quantitativ durch den **Interquartilabstand** (interquartile range)

= 75% Perzentil – 25% Perzentil

= oberes Quartil – unteres Quartil

= 3. Quartil – 1. Quartil

⇒ umfasst zentrale 50% der Daten

Der Interquartilabstand ist wie die Standardabweichung ein Mass für die Grösse des Bereichs der zentralen Daten.



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite 38 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Bei der **Normalverteilung** ist der halbe Interquartilabstand 0.67 SD.

Es gibt keine Masszahl, die unter allen Umständen optimal ist. Weitaus am verbreitesten ist die Standardabweichung.

Mean \pm SD

- Daten werden oft als Mittelwert plus–minus Standardabweichung summarisiert.

Mean \pm SEM

- Der **Standardfehler** des Mittelwertes (standard error of mean, $SE(\bar{x})$, SEM) ist die Standardabweichung des Mittelwertes, beschreibt also die Streuung der Masszahl „Mittelwert“:

$$SE(\bar{x}) = s/\sqrt{n}$$

Der Standardfehler beschreibt **nicht** die Daten, sondern die Genauigkeit einer Schätzung. Er hat in diesem Kapitel eigentlich nichts zu suchen. Der SEM wird trotzdem häufig verwendet, da eine kleinere Masszahl für die Variabilität besser wirkt (Vorsicht Falle!).

Der folgende Plot zeigt die verschiedenen Masszahlen des Zentrums und der Variabilität am Beispiel der Anzahl von T_4 -Zellen. Links: Boxplot (whiskers: Minimum–Maximum), Mitte: $mean \pm SD$, rechts: $mean \pm SE(\bar{x})$. In der rechten Hälfte wurde die maximale Beobachtung gestrichen, um einen Eindruck von der Abhängigkeit der Masszahlen von Einzelwerten zu geben.



Einführung

Deskriptive Statistik

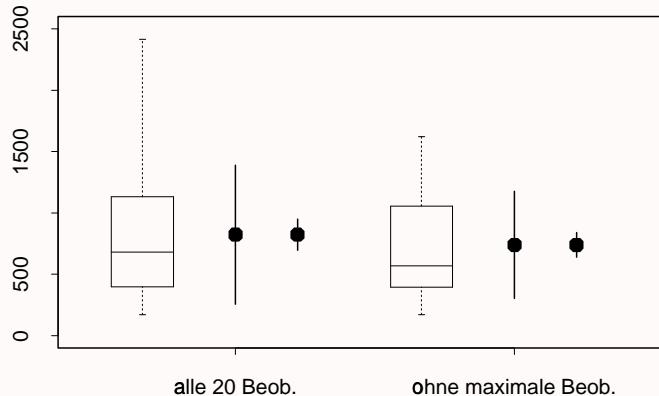
Wahrscheinlichkeit

Testen

Regression

Testfragen

Index



[Home Page](#)

[Titelseite](#)



[Seite 39 von 100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Beispiel: Statistische Masszahlen (aus StatView) für Anzahl T_4 - und T_8 -Zellen von $n = 20$ Hodgkin- und $n = 20$ non-Hodgkin-Patienten.

Bemerkung: Die Gruppe „Total“ (alle Patienten zusammen) macht hier wenig Sinn, wird aber in StatView (leider) automatisch mitanalysiert.

Klicken Sie hier, um den Datensatz 'T-Zellen' selbst zu analysieren.

[Datenanalyse](#)



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)



Seite 40 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Descriptive Statistics

Split By: Disease

	Mean	Std. Dev.	Std. Error	Count	Variance	Median
T4-cells, Total	672.625	470.487	74.391	40	221358.240	528.500
T4-cells, Hodgkin	823.200	566.385	126.648	20	320792.274	681.500
T4-cells, Non-Hodgkin	522.050	292.989	65.514	20	85842.366	433.000
T8-cells, Total	436.950	347.680	54.973	40	120881.177	319.000
T8-cells, Hodgkin	613.900	397.854	88.963	20	158288.200	447.500
T8-cells, Non-Hodgkin	260.000	154.654	34.582	20	23917.789	231.500
In(T4-cells), Total	6.288	.692	.109	40	.479	6.269
In(T4-cells), Hodgkin	6.487	.708	.158	20	.502	6.510
In(T4-cells), Non-Hodgkin	6.089	.632	.141	20	.399	6.071
In(T8-cells), Total	5.816	.735	.116	40	.541	5.765
In(T8-cells), Hodgkin	6.237	.613	.137	20	.376	6.104
In(T8-cells), Non-Hodgkin	5.394	.600	.134	20	.360	5.435

Percentiles

Split By: Disease

	T4-cells: Total	T4-cells: Hodgkin	T4-cells: Non-Hodgkin
10	200.000	272.500	171.500
25	375.000	396.500	345.000
50	528.500	681.500	433.000
75	930.000	1158.000	718.000
90	1267.500	1499.500	875.000

Percentiles

Split By: Disease

	In(T4-cells): Total	In(T4-cells): Hodgkin	In(T4-cells): Non-Hodgkin
10	5.298	5.606	5.137
25	5.927	5.983	5.840
50	6.269	6.510	6.071
75	6.835	7.053	6.576
90	7.145	7.310	6.767

Selbstkontrolle

[Beenden](#)



3. Wahrscheinlichkeitsrechnung und Versuchsplanung

3.1. Ereignisse und ihre Wahrscheinlichkeiten

Wozu benötigen wir Wahrscheinlichkeitsrechnung? Oft sind wir nicht sicher, ob ein gewisses Ereignis eintreffen wird oder nicht. Beispiele sind eine 6 bei einmal würfeln oder schönes Wetter am nächsten Sonntag oder ob ein Patient tatsächlich Krebs hat, wenn ein verdächtiges Röntgenbild vorliegt. Die Wahrscheinlichkeitsrechnung ist eine mathematische Methode, um diese Unsicherheit quantitativ einzugrenzen. In der Statistik benötigen wir die Wahrscheinlichkeitsrechnung vor allem beim Hypothesentesten und für Konfidenzintervalle (s. Kap 4). Dabei möchten wir z. B. nachweisen, dass ein empirischer Mittelwertsunterschied nur mit kleiner Wahrscheinlichkeit per Zufall auftreten kann (Signifikanztest). Weiter möchten wir um einen empirischen Wert herum Intervalle angeben, in denen der unbekannte wahre Wert mit 95% Wahrscheinlichkeit liegt (95%-Konfidenzintervalle). Insofern erlaubt uns die Wahrscheinlichkeitsrechnung Aussagen, die auf einer Stichprobe beruhen, auf die Grundgesamtheit zu verallgemeinern. Die Wahrscheinlichkeitsrechnung hat zudem Bedeutung für die stochastische Modellierung, z. B. die Ausbreitung von Epidemien oder in der Populationsgenetik.

Wahrscheinlichkeit = relative Häufigkeit in der Grundgesamtheit

Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

◀◀

▶▶

◀

▶

Seite 41 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 42 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Die Wahrscheinlichkeit wird mit P bezeichnet (für “probability”). Wenn die Wahrscheinlichkeit P für ein bestimmtes Ereignis E $P(E) = 0.4$ ist, so tritt in der Grundgesamtheit das Ereignis E im Mittel in 40 von 100 Fällen auf. $P(E) \times 100$ ist der Prozentsatz, $P(E)$ die wahre relative Häufigkeit. Beim simplen Experiment eines Münzwurfes sind die möglichen Ereignisse die Augenzahlen 1, 2, 3, 4, 5, 6, und bei einem gleichmässigen Würfel gilt $P(E = I) = 1/6$, wo $I = 1, 2, \dots, 6$. Wenn ein Patient die Praxis betritt, so hat er mit einer gewissen Wahrscheinlichkeit eine der vier Blutgruppen, es gibt vier mögliche Ereignisse. Als Konvention gilt, dass $P(E) = 0$ bedeutet, dass das Ereignis E unmöglich eintreten kann, bei $P(E) = 1$ tritt das Ereignis sicher auf (z. B. tritt sicher eine Augenzahl von 1–6 bei einmal würfeln auf). Weiter gilt für das Ereignis „Nicht-E“, (abgekürzt E_c) $P(E_c) = 1 - P(E)$, da ja entweder E oder E_c eintreten muss.

Beispiele:

- $E_1 = \text{Körpergrösse} \leq 180 \text{ cm}$
- $E_2 = 170 \text{ cm} \leq \text{Körpergrösse} \leq 180 \text{ cm}$

Solche Ereignisse sind natürlicherweise bei stetigen Variablen (siehe Kapitel 2) von Interesse, und man kann ihnen ebenfalls Wahrscheinlichkeiten zuordnen. Z. B. ist qualitativ klar, dass für einen erwachsenen männlichen Europäer die Wahrscheinlichkeit höher ist, zwischen 175 und 180 cm gross zu sein, als zwischen 160 und 165 cm gross zu sein.



3.2. Bedingte Wahrscheinlichkeit und Unabhängigkeit

Von zentraler Bedeutung sind die Begriffe der bedingten Wahrscheinlichkeit und der stochastischen Unabhängigkeit.

Definition **bedingte Wahrscheinlichkeit** = Wahrscheinlichkeit, dass das Ereignis E_1 eintritt, gegeben, dass das Ereignis E_2 eingetreten ist.

Beispiele:

- Wahrscheinlichkeit, dass ein Patient einen Tumor hat, wenn ein auffälliges Röntgenbild vorliegt.
- Wahrscheinlichkeit, dass ein Kind grösser als 1.50 m ist, wenn bekannt ist, dass es 10.5 Jahre alt ist.

Definition **Unabhängigkeit von Ereignissen**: E_1 und E_2 sind dann (stochastisch) unabhängig, wenn die bedingte Wahrscheinlichkeit von E_1 gegeben E_2 gleich der Wahrscheinlichkeit von E_1 ist (d. h. E_2 hat keinen Einfluss auf E_1).

Aus der Unabhängigkeit der Ereignisse E_1 und E_2 folgt das Produktgesetz für deren Wahrscheinlichkeiten, das Sie vielleicht aus der Mittelschule kennen:

$$P[E_1 \text{ trifft ein und } E_2 \text{ trifft ein}] = P[E_1 \cap E_2] = P[E_1] \times P[E_2].$$

Körpergrösse und Körpergewicht sind z. B. keine unabhängigen Variablen, während der Gesundheitszustand nacheinander aufgenommener Patienten meist unabhängig ist (siehe auch Abschnitt 3.8). Die Unabhängigkeit von Ereignissen hat wichtige Implikationen für die Gesetze der grossen Zahlen (siehe Kapitel 3.5), und sie hilft, viele Formeln zu vereinfachen.

[Home Page](#)

[Titelseite](#)

[«](#) [»](#)

[◀](#) [▶](#)

Seite 43 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Beispiele:

Die Häufigkeit einer Krankheit bei einem Patientenkollektiv sei $p = 0.2$ (20%).

- Wahrscheinlichkeit, dass 3 aufeinanderfolgende Patienten an dieser Krankheit leiden:

$$P[\text{Patient 1 und Patient 2 und Patient 3 krank}] = 0.2 \times 0.2 \times 0.2 = 0.008$$

Das heisst, diese Wahrscheinlichkeit ist nur noch 0.8%, also sehr klein.

- Wahrscheinlichkeit, dass von 3 aufeinanderfolgenden Patienten mindestens einer krank ist:

$$P[\text{mind. 1 } \times \text{ krank}] = 1 - P[\text{dreimal gesund}] = 1 - (0.8 \times 0.8 \times 0.8) = 0.49,$$

da $P(E_c) = 1 - P(E)$. Das heisst, diese Wahrscheinlichkeit ist fast 50%.

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite 44 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



[Einführung](#)
[Deskriptive Statistik](#)
Wahrscheinlichkeit
[Testen](#)
[Regression](#)
[Testfragen](#)
[Index](#)

3.3. Wahrscheinlichkeitsverteilungen

Für die Statistik sind Zufallsvariablen wichtiger als Ereignisse:

Den Ausgang einer Messung oder einer Beobachtung können wir als Zufallsvariable X bezeichnen (die Bezeichnung X ist eine Konvention). Beim Würfeln kann X per Zufall die ganzzahligen Werte 1 bis 6 annehmen, beim Messen einer Körpergrösse sind es positive reelle Werte. Die Körpergrösse ist insofern eine Zufallsvariable, als wir nicht wissen, wie gross der nächste Patient sein wird.

Der erhaltene Wert ($X = x$) nach einer Messung heisst **Realisierung** der Zufallsvariablen X .

Definition **Stichprobe**: n Realisierungen einer Zufallsvariablen X , die uns interessiert: x_1, \dots, x_n .

Beispiel:

Wenn die interessierende Zufallsvariable die Körpergrösse 10-jähriger Knaben ist, kann eine Stichprobe aus allen 10-jährigen Knaben eines Schulhauses bestehen.

Die Wahrscheinlichkeitsrechnung beruht auf Ereignissen, und man muss deshalb für Zufallsvariablen passende Ereignisse konstruieren. Ereignisse der Form $X \leq z$ sind wichtig, da alle anderen interessierenden Ereignisse daraus konstruierbar sind. Dabei ist z eine von uns vorgegebene Zahl.

[Home Page](#)

[Titelseite](#)

«

»

◀

▶

Seite 45 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Definition **Verteilungsfunktion** F einer Zufallsvariablen X :

$$F(z) = \text{P}[X \leq z]$$

Die empirische (kumulative) Verteilungsfunktion, die bei den graphischen Methoden eingeführt wurde (siehe 2.4), ist das Datenäquivalent zum theoretischen Begriff der Verteilungsfunktion F (ein „Schätzer“ für die Verteilungsfunktion F). Verteilungsfunktionen sind monoton steigend von 0 auf 1. Beispiele, die wir nachher kennenlernen werden sind die Normalverteilung, die χ^2 -Verteilung und die Binomialverteilung.

Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)



Seite 46 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

[◀◀](#)

[▶▶](#)

[◀](#)

[▶](#)

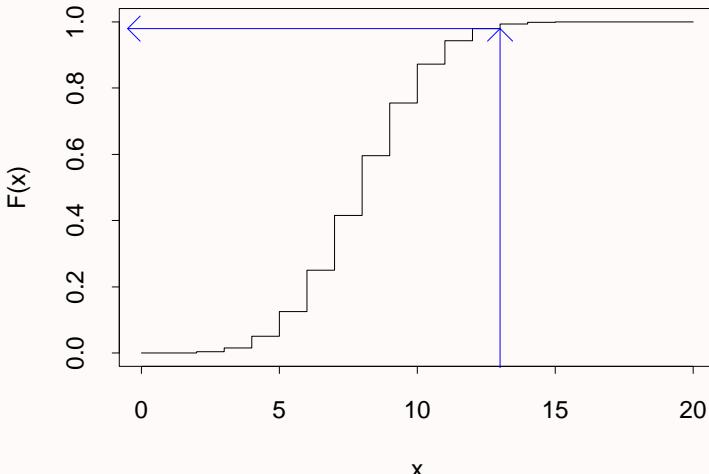
Seite 47 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Für diskrete Zufallsvariable wird F eine Treppenfunktion; im Beispiel nimmt die Zufallsvariable X ganzzahlige Werte ab 0 an (F entspricht einer Binomialverteilung), definiert in 3.4:

Werte kleiner als 0 kommen mit Wahrscheinlichkeit 0 vor. Werte grösser als 12 kommen mit Wahrscheinlichkeit 2.1% vor, denn $P[k < 13] = 97.9\%$ (siehe Grafik). Die Wahrscheinlichkeit für das Auftreten einer 2 ist 0.3% – es ist die Höhe der Stufe der Treppenfunktion bei $x = 2$.



Definition **Wahrscheinlichkeitsdichte** f :

- a) diskrete Variable $X : f(z_i) = P[X = z_i]$
- b) stetige Variable $X : f(z) = F'(Z)$ (Ableitung von F)

Das Histogramm ist ein Schätzer für die Wahrscheinlichkeitsdichte. Die Dichtefunktion ist visuell informativer als die Verteilungsfunktion, was z. B. die Form und Breite der Verteilung angeht, enthält aber grundsätzlich dieselbe Information wie die Verteilungsfunktion.

Dichtefunktionen sind nichtnegativ, und die Wahrscheinlichkeit, dass die Zufallsvariable X in einem Intervall $[a, b]$ liegt, ist die **Fläche unter der Kurve** über dem Intervall $[a, b]$.

Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)



Seite 48 von 100

[Zurück](#)

[Vollbild](#)

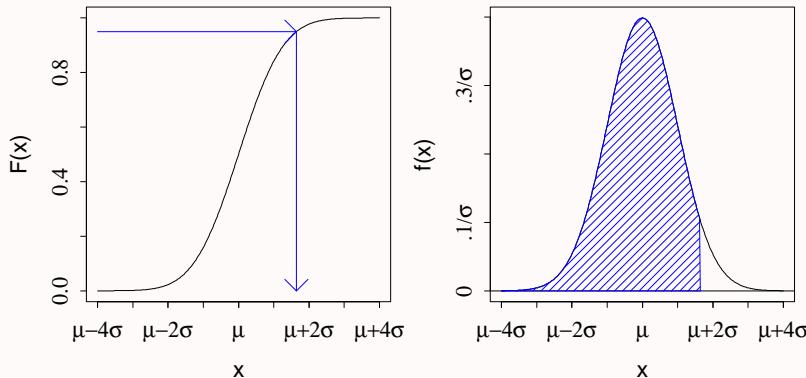
[Schließen](#)

[Beenden](#)



Sie sehen nachstehend eine Graphik der Verteilungsfunktion der Normalverteilung (mit theoretischem Mittelwert μ und theoretischer Standardabweichung σ).

Die Körpergrösse ist approximativ normalverteilt (da multifunktionell genetisch) und Männer haben einen Mittelwert $\mu = 178$ cm und eine Standardabweichung $\sigma = 7.0$ cm. Informativer als die Verteilungsfunktion ist die Wahrscheinlichkeitsdichte $f = F'$ (rechts in der Graphik, „Glockenkurve“).



Bei normalverteilten Grössen ist die Wahrscheinlichkeit, Messungen im Intervall $(\mu \pm \sigma)$ zu erhalten, 68% (siehe Abschnitt 2.5.1). Die Wahrscheinlichkeit für Messungen ausserhalb des Intervalls $(\mu \pm 3\sigma)$ ist sehr klein, nämlich ca. 1%. Die Wahrscheinlichkeit, dass eine Zufallsvariable x in einem Intervall a bis b liegt, ist gleich der Fläche der Wahrscheinlichkeitsdichte über diesem Intervall (oder gleich der Differenz $F(b) - F(a)$). Verteilungsfunktion F und Wahrscheinlichkeitsdichte f sind



[Einführung](#)
[Deskriptive Statistik](#)
Wahrscheinlichkeit
[Testen](#)
[Regression](#)
[Testfragen](#)
[Index](#)

demnach geeignet, um mit Wahrscheinlichkeiten zu rechnen.

Im Abschnitt 2.5.1 haben Sie empirische Perzentile kennen gelernt. Bei den statistischen Testen werden wir theoretische Perzentile benötigen, und zwar die extremen. Typischerweise sind dies das 2.5% und das 5% Perzentil bzw. das 97.5% oder 95% Perzentil, sodass darunter bzw. darüber nur mit kleiner Wahrscheinlichkeit Resultate zu erwarten sind. In der Graphik der Normalverteilungs-Funktion können Sie sehen, wie man das 95% Perzentil findet: man startet bei $F(x) = 0.95$, geht horizontal zum Kreuzpunkte mit F und dann vertikal hinunter. Jetzt hat man das gesuchte 95. Perzentil, in Kapitel 4 jeweils mit $z_{0.95}$ abgekürzt. In der rechten Graphik können Sie sehen, wie Wahrscheinlichkeiten mit Hilfe der Dichte berechnet werden: Die Fläche unter der Dichte bis $z_{0.95}$ ist gleich 0.95.

[Home Page](#)

[Titelseite](#)

◀◀

▶▶

◀

▶

Seite 50 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Sie haben statistische Kennzahlen wie Mittelwert \bar{x} und Standardabweichung s kennengelernt, um Daten einer Stichprobe zu charakterisieren (exakter: um deren empirische Verteilung zu charakterisieren). Es gibt auch die entsprechenden theoretischen Größen, welche die wahre Verteilung in der Grundgesamtheit charakterisieren. Zu \bar{x} gehört μ und zu s^2 gehört σ^2 . In der Statistik sagt man, dass \bar{x} ein Schätzer für μ ist, und s ein Schätzer für σ .

Theoretischer Mittelwert („Erwartungswert“) $\mu = \int_{-\infty}^{\infty} xf(x)dx$

Varianz

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$$

Standardabweichung

$$\sigma = \sqrt{\sigma^2}$$

[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
[Testen](#)
[Regression](#)
[Testfragen](#)
[Index](#)

[Home Page](#)

[Titelseite](#)



Seite 51 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung
Deskriptive Statistik
Wahrscheinlichkeit

Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite 52 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

3.4. Einige wichtige Verteilungen

Es gibt eine sehr grosse Zahl von Verteilungen, die bedeutsam geworden sind. Hier seien 3 für die Statistik wichtige Verteilungen vorgestellt.

1. Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ (auch Gauss–Verteilung genannt).

\mathcal{N} ist das Symbol der Normalverteilung, die folgende Dichte hat:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Dabei ist „exp“ die Exponentialfunktion. Für $\mu = 0$ und $\sigma^2 = 1$ heisst die Verteilung Standardnormalverteilung. Die $\alpha \times 100\%$ –Perzentile der Standardnormalverteilung werden konventionell mit z_α bezeichnet.



Die Normalverteilung ist omnipräsent wegen des zentralen Grenzwertsatzes (siehe [3.5](#)).

Eigenschaften:

- symmetrische Verteilung (siehe auch Plot der Dichte), Median = theoretischer Mittelwert
- charakterisiert durch 2 Parameter μ (= Populationsmittelwert) und σ (= Standardabweichung), die intuitiv besonders einfach sind. Wenn man diese kennt, kennt man das Wahrscheinlichkeitsgesetz.
- Die Wahrscheinlichkeiten für grosse Abweichungen vom Erwartungswert sind klein. Die Wahrscheinlichkeit von Abweichungen grösser als 2σ vom Erwartungswert μ beträgt etwa 5% („dünne Schwänze“ der Verteilung).

So sehen Stichproben aus einer standardnormalverteilten Grundgesamtheit aus:

[Animation](#)

[Home Page](#)

[Titelseite](#)



[Seite 53 von 100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung
Deskriptive Statistik
Wahrscheinlichkeit

Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite 54 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

2. χ^2 -Verteilung

Die χ^2 -Verteilung wird häufig gebraucht, wenn man in Kreuztabelchen auf Unterschiede für kategoriale Daten testen will (siehe Kapitel 4). Dort werden Unterschiede zwischen beobachteten und hypothetischen Häufigkeiten quadriert und aufsummiert. Seien Z_1, \dots, Z_ν unabhängige standardnormalverteilte Zufallsvariablen $\mathcal{N}(0, 1)$ (d. h. Erwartungswert $\mu = 0$ und Varianz $\sigma^2 = 1$). Dann ist die χ^2 -Verteilung definiert als Summe der quadrierten Z_i :

$$\chi_\nu^2 = \sum_{i=1}^{\nu} Z_i^2 \quad \chi^2\text{-verteilt mit } \nu \text{ Freiheitsgraden } (\nu = \text{Anzahl Summanden})$$

Die Freiheitsgrade werden in Programmen oft mit “df” (*degrees of freedom*) bezeichnet. Eigenschaften:

- asymmetrisch (rechts-schief)
- Erwartungswert und Varianz werden durch einen gemeinsamen Parameter ν beschrieben $\mu = \nu$, $\sigma^2 = 2\nu$.

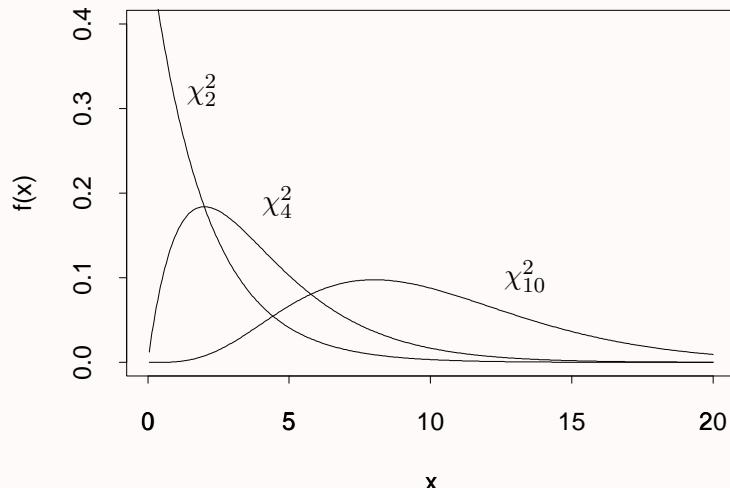
Auch hier empfiehlt sich unser Verteilungssimulator:

[Animation](#)

[Selbstkontrolle](#)



Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index



Dichten von χ^2 -Verteilungen mit 2, 4 und 10 Freiheitsgraden ($\chi^2_2, \chi^2_4, \chi^2_{10}$).

[Home Page](#)

[Titelseite](#)

[◀◀](#)

[▶▶](#)

[◀](#)

[▶](#)

Seite 55 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

[Seite 56 von 100](#)

[Zurück](#)

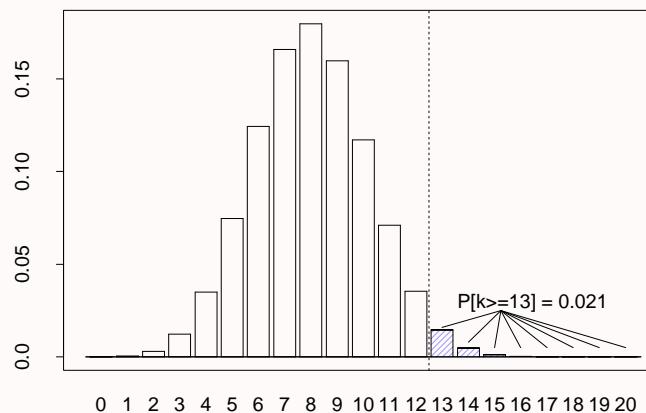
[Vollbild](#)

[Schließen](#)

[Beenden](#)

3. Binomialverteilung

Bei der Einführung des Testens von Hypothesen (siehe Kapitel 4) wird an $n = 20$ Patienten geprüft, ob ein neues Medikament eine höhere Heilungswahrscheinlichkeit als $p = 0.4$ (d. h. 40%) hat. Die Anzahl k der geheilten Patienten ($k = 0$ bis 20 möglich) folgt einer Binomialverteilung. Falls man eine Wahrscheinlichkeit von $p = 0.4$ annimmt, ergibt sich nachfolgende Wahrscheinlichkeitsverteilung für die Anzahl Heilungen:





Dies bedeutet, dass 13 oder mehr Heilungen nur mit einer Wahrscheinlichkeit von 2.1% zu erwarten sind.

Allgemein hat die diskrete Dichtefunktion folgende Form:

$$P[X = k] = \binom{n}{k} p^k (1-p)^{n-k} \quad 0 \leq k \leq n$$

- $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ $n! = 1 \cdot 2 \cdot 3 \dots n$ (Binomialkoeffizient)
- Erwartungswert einer binomialverteilten Variablen = $n p$, Varianz = $n p (1-p)$

[Selbstkontrolle](#)

[Home Page](#)

[Titelseite](#)



Seite 57 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



[Einführung](#)

[Descriptive Statistik](#)

[Wahrscheinlichkeit](#)

[Testen](#)

[Regression](#)

[Testfragen](#)

[Index](#)

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 58 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

3.5. Gesetze der grossen Zahlen

Eine Stichprobe von n unabhängigen Zufallsvariablen liefert Masszahlen, zum Beispiel den Mittelwert \bar{x} , die diese Stichprobe beschreiben. Eigentlich interessiert uns der zugrundeliegende Populationswert μ .

Fragestellung: Streben die empirischen Werte („Schätzer“, \bar{x}) gegen die wahren (theoretischen) Werte (μ), wenn die Stichprobe immer grösser wird ($n \rightarrow \infty$)?

Die **Gesetze der grossen Zahlen** zeigen mathematisch, dass diese erwünschte Eigenschaft gilt. Der **zentrale Grenzwertsatz** besagt, dass die empirischen Kennzahlen approximativ normalverteilt sind, falls n gross ist.

Der zentrale Grenzwertsatz erklärt aber auch, weshalb die Normalverteilung so oft in der Natur gilt und für die Statistik wichtig ist:

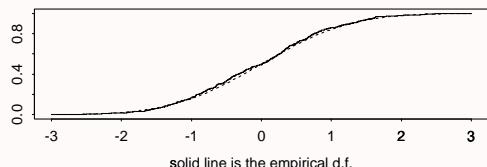
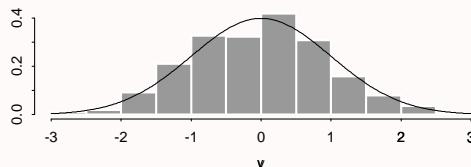
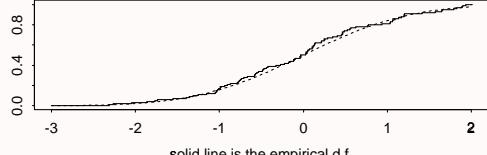
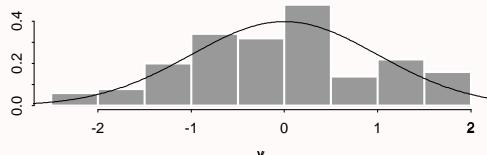
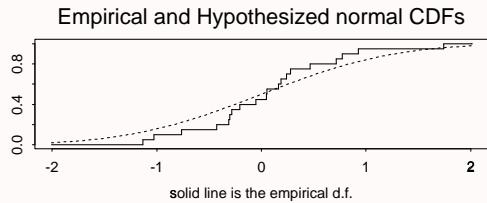
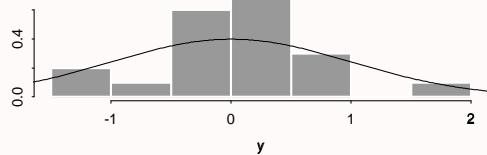
Viele Phänomene der Natur entstehen durch Überlagerung vieler kleiner Effekte, so dass das Resultat etwa normalverteilt ist. Multiplikative Effekte führen hingegen auf eine schiefe Verteilung, z. B. die χ^2 -Verteilung oder die lognormale Verteilung. Multiplikative Regeln haben wir z. B. bei Messungen, die durch die Zellteilung beeinflusst sind, weshalb viele biomedizinische Messgrössen nicht normalverteilt sind sondern schief verteilt, (z. B. T_4 und T_8 Zellen in Abschnitt 2.4).

Hier ein klassisches Experiment dazu (Galtonbrett):

[Animation](#)



Nachstehend wird illustriert, wie das Histogramm und die empirische Verteilungsfunktion gegen die wahre Grösse (im Beispiel die Normalverteilung bzw. deren Dichte) konvergieren. Das Beispiel basiert auf simulierten Pseudo-Zufallsvariablen, $n = 20, 100, 500$.

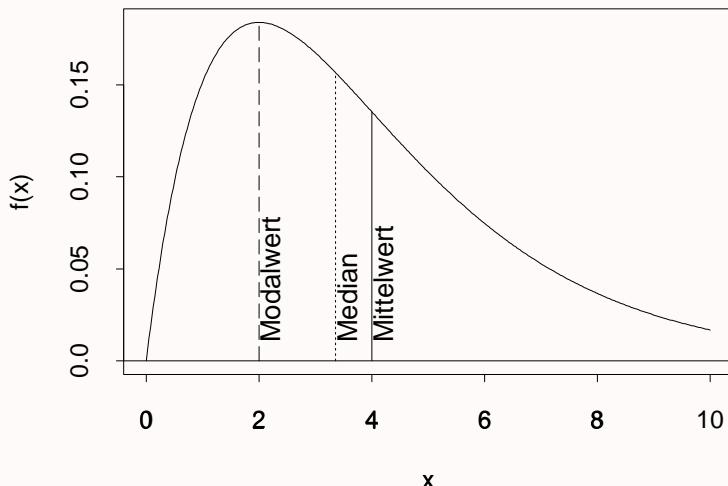


Man sieht, dass das geschätzte Histogramm für kleine und mittlere Stichproben von der wahren Wahrscheinlichkeitsdichte abweichen kann. Entsprechend vorsichtig muss man bei der Interpretation sein.



3.6. Transformationen von Daten und Verteilungen

Bei vielen biochemischen Daten, aber auch bei anderen Variablen wie Körpergewicht oder Vermögen, haben wir bei weitem keine Normalverteilung. Vielmehr erhalten wir oft eine Dichte etwa der folgenden Form (Modalwert = Wert mit maximaler Wahrscheinlichkeitsdichte):



Die Verteilung ist **rechtsschief** (wie z. B. die χ^2 -Verteilung) und nicht wie die Normalverteilung symmetrisch. Mittelwert, Median und Modalwert repräsentieren unterschiedliche Aspekte eines typischen Wertes der Verteilung. Ein Nachteil einer schießen Verteilung ist diese Uneindeutigkeit; mindestens so gravierend ist, dass viele statistische Verfahren — die auf der Normalverteilung basieren — nicht mehr ohne weiteres anwendbar sind.



Konventionellerweise werden solche Variablen X in Medizin und Ingenieurwissenschaften oft logarithmiert, um eine approximativ symmetrische Verteilung zu erhalten: $Y = \log(X)$

Damit erreicht man, dass Standardverfahren der Statistik anwendbar sind, und zumeist wird auch die Interpretation der statistischen Analyse einfacher. Neben der logarithmischen Transformation gibt es auch andere Transformationen.

Eine Alternative zur Transformation von Daten ist die Anwendung von Rangverfahren. Diese stehen aber nicht immer zur Verfügung. (Z. B. in Kapitel 4 der Mann-Whitney-Test anstatt der t-Test.)

3.7. Schätzverfahren für statistische Kennwerte

Studien kosten Geld und Zeit, und Patienten sind nicht beliebig verfügbar. Man möchte deshalb die Daten statistisch effizient nutzen, „**gute Schätzer**“ für interessierende wahre Kennwerte erhalten.

Angenommen, man möchte einen unbekannten Parameter θ schätzen, z. B. $\theta = \mu$, den theoretischen Mittelwert. Für einen Schätzer $\hat{\theta}$ eines Populationsparameters θ fordern wir mathematisch als Minimum, dass für wachsende Stichprobengröße n sich der Schätzer $\hat{\theta}$ dem wahren Wert θ immer mehr annähert:

Wenn man einen interessierenden wahren Parameter θ durch den empirischen Wert („Schätzer“) $\hat{\theta}$ bestimmt, so stellt man folgende Minimalforderungen:

- $\hat{\theta}$ soll mit wachsender Stichprobengröße n den wahren Wert θ immer besser approximieren.
- $\hat{\theta}$ sollte für grosse n etwa normalverteilt sein.



Beides ist i. A. erfüllt, und man möchte zusätzlich – quantitativ – dass die Abweichung des Schätzwertes $\hat{\theta}$ vom wahren Wert θ möglichst klein ist.

Wenn der Schätzwert $\hat{\theta}$ systematisch zu hoch oder zu niedrig liegt, so nennt man diese Abweichung Bias (systematischer Fehler). Idealerweise ist der Bias null, und solche Schätzer sind „erwartungstreu“, d. h. in diesem Sinne optimal. Viele — aber nicht alle — Verfahren der Statistik erfüllen diese Voraussetzung. Zusätzlich möchte man, dass ein Schätzwert $\hat{\theta}$ nicht zu sehr variiert, von Stichprobe zu Stichprobe möglichst stabil bleibt. Optimal ist es, wenn die Varianz von $\hat{\theta}$ möglichst klein ist („Minimum–Varianzschätzer“).

Bei vielen praktisch wichtigen Problemen ist es nicht möglich, einen Schätzwert $\hat{\theta}$ zu berechnen, der zugleich keinen Bias (d. h. keine systematische Abweichung) und minimale Varianz erzielt. Man hat deshalb nach einem weiteren allgemeinen Berechnungsprinzip gesucht und es in der **Maximum–Likelihood–Schätzung** gefunden.

Für diejenigen unter Ihnen, die es gern etwas genauer wüssten, geben wir noch eine mehr mathematisch orientierte Erklärung (die Anderen können zu Abschnitt 3.8 weitergehen).

[Einführung](#)
[Deskriptive Statistik](#)
Wahrscheinlichkeit
[Testen](#)
[Regression](#)
[Testfragen](#)
[Index](#)

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 62 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

[◀◀](#)

[▶▶](#)

[◀](#)

[▶](#)

[Seite 63 von 100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Ein Teil des Schätzfehlers kommt vom systematischen Fehler her, und dieser wird durch die Forderung der Erwartungstreue ausgeschlossen:

Kriterium 1: **Erwartungstreue** von $\hat{\theta}$ („im Mittel richtig“)

$$E[\hat{\theta} - \theta] = 0 \text{ oder } E[\hat{\theta}] = \theta$$

$E[\hat{\theta} - \theta]$ heisst **Bias** von $\hat{\theta}$ („systematischer Fehler“).

Beispiele:

- Für die Ausfallstatistik eines medizinischen Gerätes wird jeden Tag festgehalten

$x_i = 0$: Gerät nicht ausgefallen

$x_i = 1$: Gerät ausgefallen

Ein natürlicher Schätzwert \hat{p} für die Ausfallwahrscheinlichkeit p ist

$$\hat{p} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Es gilt:

$$E[\hat{p}] = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = p$$

Also ist $\hat{p} = \bar{x}$ erwartungstreu.



- Wenn man zwei Geräte zugleich betreibt, so interessiert man sich für die Wahrscheinlichkeit, dass beide zugleich kaputtgehen. Naiv würde man nehmen $\hat{p}^2 = \bar{x}^2$. Aber wegen

$$E[\bar{x}^2] = \text{Var}(\bar{x}) + E[\bar{x}]^2 = \frac{p(1-p)}{n} + p^2$$

resultiert ein Bias $= \frac{p(1-p)}{n} \neq 0$, der für grosse n aber unbedeutend wird.

Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite 64 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Kriterium 2:
Minimum–Varianz–Schätzung

Konstruiere $\hat{\theta}$ so, dass

$$\text{Var}(\hat{\theta}) = \text{minimal}$$

Denn es ist naheliegend, dass ein Schätzwert nicht zu variabel sein sollte.

Erwartungstreue Schätzer kleinster Varianz sind meistens gut.

Leider sind solche Schätzer nicht immer konstruierbar und auch nicht immer optimal, deshalb wurde in der mathematischen Statistik nach einem weiteren allgemeinen Prinzip gesucht. Ein solches — und auch das am weitesten verbreitete — stellt die **Maximum–Likelihood–Schätzung** dar. Diese ist so definiert, dass die Übereinstimmung des Schätzers mit den Beobachtungen optimal wird. Dabei muss für die Daten eine Wahrscheinlichkeitsverteilung angenommen werden, z. B. die Normalverteilung.

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 65 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



[Einführung](#)
[Deskriptive Statistik](#)
Wahrscheinlichkeit
[Testen](#)
[Regression](#)
[Testfragen](#)
[Index](#)

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 66 von 100

[Zurück](#)

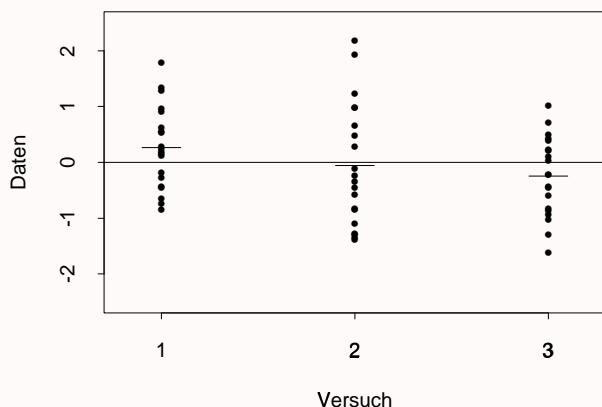
[Vollbild](#)

[Schließen](#)

[Beenden](#)

Variabilität zwischen Datensätzen

Sie sehen nachstehend 3 auf dem Computer simulierte normalverteilte Stichproben mit $\mu = 0$, $\sigma^2 = 1$ und $n = 20$.



Wie Sie an nachfolgender Tabelle sehen, schwanken die Mittelwerte \bar{x} um den Wert $\mu = 0$ und die Standardabweichungen s um den Wert $\sigma = 1$, mit zum Teil merklichen Abweichungen.

	1.	2.	3.
\bar{x}	0.27	-0.06	-0.25
s	0.74	1.10	0.71

Weitere simulierte Datensätze finden Sie hier:

[Animation](#)



Mit den Methoden der Wahrscheinlichkeitsrechnung (siehe Kapitel 3) kann man die Varianz des Mittelwertes herleiten:

$$\text{Var}(\bar{x}) = \frac{\sigma^2}{n},$$

wobei σ^2 die Populations–Varianz einer Einzelmessung ist. Durch Ersetzen von σ durch den Schätzwert s erhält man den Standardfehler (engl.: *standard error of the mean*, SEM).

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

Man muss zwischen s (= Standardabweichung = Variabilität in der Stichprobe) und $s_{\bar{x}}$ (= Standardfehler = Standardabweichung des Mittelwerts = Variabilität des Mittelwerts) klar unterscheiden. Beide haben ihren – unterschiedlichen – Platz. Für die log– T_4 –Zellanzahl bei Hodgkin–Patienten erhielten wir in Kapitel 2 $\bar{x} = 6.49$ und $s = 0.71$. Bei $n = 20$ ergibt sich ein $s_{\bar{x}} = 0.16$.

[Einführung](#)
[Deskriptive Statistik](#)
Wahrscheinlichkeit
[Testen](#)
[Regression](#)
[Testfragen](#)
[Index](#)

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite 67 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
[Testen](#)
[Regression](#)
[Testfragen](#)
[Index](#)

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

[Seite 68 von 100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Beispiel: Genauigkeit von Verhältnissen

$n = 80$ Personen wurden über Asthma befragt

$k = 7$ Asthmatiker wurden gefunden

Also ist die relative Häufigkeit von Asthma: $\hat{p} = k/n = 0.088$, in der Epidemiologie Prävalenz genannt.

Wie genau ist \hat{p} bestimmt (n ist klein)? Dies basiert auf der Binomialverteilung, da eine ja/nein-Befragung eine binäre Variable ergibt:

$$\text{Var}(\hat{p}) = \frac{p(1-p)}{n}$$

$$\implies s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Prävalenz–Beispiel: $s_{\hat{p}} = 0.032$

falls $\hat{p} = 0.5$: $s_{\hat{p}} = 0.056$

Die Formel für $s_{\hat{p}}$ zeigt, dass mittlere Häufigkeiten mehr streuen als Extreme.

Wie wir nachher sehen werden, liegt die wahre Prävalenz etwa im Intervall $(\hat{p} \pm 2s_{\hat{p}})$,

d. h. (0.088 ± 0.064) , sie ist also schlecht bestimmt.

[Datenanalyse](#)



Bei beiden Beispielen verbessert sich die Genauigkeit des Kennwerts mit $1/\sqrt{n}$ und nicht etwa mit $1/n$ ("Wurzel- n -Gesetz"). Dies gilt für fast alle statistischen Fragestellungen und hat grosse praktische Konsequenzen: Um eine Verdoppelung der Genauigkeit der späteren Datennanlayse zu erhalten, müssen viermal soviele Patienten in die Studie einbezogen werden. Umgekehrt steigt auch der Aufwand für grosse Studien stärker als proportional zu n , so dass eine seriöse Versuchsplanung angezeigt ist, gerade was die Wahl der Stichprobengrösse angeht. (Siehe auch Kapitel 4.)

Die Streuung eines Schätzers — den Standardfehler — erhalten wir bei den meisten Programmpaketen mitgeliefert und damit auch ein Bild über die Genauigkeit der Statistik.

3.8. Versuchsplanung

3.8.1. Repräsentative Stichprobe

Bei jeder Studie wollen wir — implizit oder explizit — über eine Grundgesamtheit (Population) von Versuchseinheiten (Menschen, Tiere, Proben, Spitäler) Aussagen machen. Üblicherweise wird eine Stichprobe, also nicht eine Population, in eine Studie einbezogen, d. h. es werden n Versuchseinheiten aus der Grundgesamtheit gezogen. Die Stichprobe bedingt den Zufallseffekt. Sehr wichtig ist die Wahl der Stichprobengrösse, das „ n “. Es gibt dazu viel Literatur; einige Gesichtspunkte werden in Kapitel 5 diskutiert. Selten unternimmt man auch eine Gesamterhebung. Beispiele sind die Volkszählungen oder Krankenregister (z. B. Krebsregister des Kantons Zürich). Gesamterhebungen bedingen einen hohen Aufwand, und es können nur relativ grobe Parameter erfasst werden. Deshalb wird neben der Volkszählung ein Mikrozensus durchgeführt, um detailliertere sozialwissenschaftliche Ergebnisse zu

[Home Page](#)

[Titelseite](#)

«

»

◀

▶

Seite 69 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



erhalten.

Das Beispiel eines Abstracts aus dem „British Medical Journal“ von 1996 zeigt, dass Studien detailliert bezüglich ihres Versuchsplans („design“) umschrieben werden müssen, um akzeptiert zu werden.

Adequacy of cervical cytology sampling with the Cervex brush and the Aylesbury spatula: a population based randomised controlled trial

Paola Dey, Stuart Collins, Minaxi Desai, Ciaran Woodman

Abstract

Objective—To compare the adequacy of cervical cytology sampling with two sampling instruments commonly used in primary care—namely, the Aylesbury spatula and the Cervex brush.

Design—Pair matched, population based randomised controlled trial.

Setting—86 general practices and family planning clinics in Greater Manchester.

Subjects—15 882 cervical smears taken from women aged 20–64 years as part of the national cervical screening programme.

Interventions—Participating centres were allo-

Introduction

Inadequate cervical smears are not only a cause of needless anxiety and inconvenience to women but are also an additional cost to the NHS. Attempts to reduce the rate of inadequate smears have focused on improving the competence of smear takers and the design of sampling instruments, but these instruments have rarely been evaluated in population based settings. We compared the adequacy of cytological sampling in a primary care setting with two commonly used instruments, the Cervex brush and the Aylesbury spatula.

Wie soll man die Stichprobe auswählen („ziehen“) ?

- Sie soll repräsentativ für die Population sein (Verallgemeinerungsfähigkeit).
- Wenn mehrere Gruppen untersucht werden, sollten sie in den wesentlichen

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 70 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)

[Testen](#)
[Regression](#)
[Testfragen](#)
[Index](#)

[Home Page](#)

[Titelseite](#)



Seite 71 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Merkmale vergleichbar sein (Beispiel: Alter, Schweregrad der Krankheit).

- Unabhängigkeit der Versuchseinheiten muss gewährleistet sein; dies schliesst Familienangehörige aus, da diese Genetik und Umwelt teilen.
- Es gibt wichtige Ausnahmen von der Unabhängigkeit:
prä–post–Vergleiche, Versuche mit Messwiederholungen, Longitudinalstudien.

Die Repräsentativität für die Population wird durch ein Bündel von Massnahmen sichergestellt. Am wichtigsten ist die **Randomisierung**, was bedeutet, dass Patienten nach dem Zufallsprinzip einer oder mehreren Gruppen zugewiesen werden. Aber auch auf eine möglichst volle Teilnahme bis zum Schluss ist sehr zu achten. Wegen der mangelnden Randomisierung lagen die Wahlprognosen in den USA (1948) zwischen Truman und Dewey völlig daneben. Aufgrund (bequemer) Telefonumfragen wurde ein klarer Sieg von Dewey prognostiziert, gewonnen hat aber Truman. 1948 waren eben Privattelefone nicht gleichmäßig über alle Bevölkerungsschichten verteilt, so dass sich eine Verfälschung zugunsten der Mittel- und Oberschicht einschlich.

Merke:

- Freiwillige sind nicht repräsentativ.
- Verweigerer sind nicht repräsentativ.
- Patienten von Uni–Kliniken sind nicht repräsentativ für stationäre Patienten an sich (z. B. Kreisspitäler).



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)



Seite 72 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

3.8.2. Arten von Studien

- Beobachtungsstudien werden experimentellen Studien gegenübergestellt. Bei letzteren wird eine Zielgröße beeinflusst, es wird interveniert. Beispiele aus der Medizin sind:
 - Therapiestudien
 - Neurophysiologische Studien
 - Tierexperimente

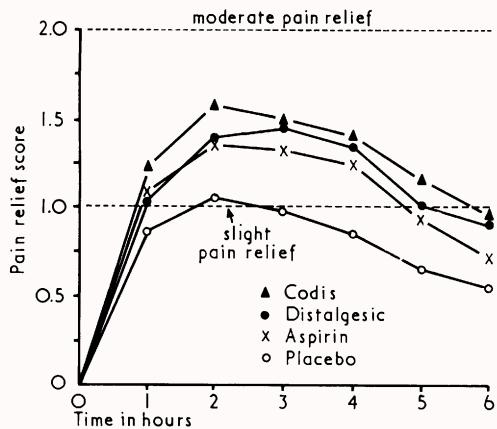
Beispiele von Beobachtungsstudien sind:

- Bestimmung der Prävalenz einer Krankheit
- Studie der kardiovaskulären Risikofaktoren
- Vergleich dementer und gesunder alter Menschen
- Des Weiteren werden prospektive und retrospektive Studientypen unterschieden. Retrospektive Studien aus Krankenblatt-Archiven werden wegen ihrer beschränkten Aussagekraft nicht mehr so häufig durchgeführt. In der Epidemiologie sind prospektive Studien durch Kohortenstudien, retrospektive durch Fall-Kontroll-Studien („case-control studies“) repräsentiert.
- Eine wichtige Unterscheidung besteht noch zwischen Querschnitts- und Längsschnitt-Studien. Letztere werden immer wichtiger, da z. B. aus dem Verlauf einer Krankheit wichtige Erkenntnisse gewonnen werden können. Es steigt dabei nicht nur der empirische, sondern auch der statistische Aufwand.

Klinische Therapiestudien

Ziel ist der Nachweis der Wirksamkeit bzw. Unwirksamkeit einer Therapie, oder die Überlegenheit einer neuen Therapie über eine Standardtherapie. Im ersteren Fall wird im allgemeinen eine Placebogruppe geführt, da allein die Verabreichung irgend-einer Pille einen (psychologischen) Effekt haben kann. Zielgröße ist der Therapieerfolg, der qualitativ („geheilt“) oder quantitativ („Veränderung des Blutdrucks“) erhoben werden kann.

Die Resultate einer Studie im British Medical Journal 1974 über die Wirkung von Schmerzmitteln bei Arthritis (siehe Graphiken unten) zeigen, dass Placebo einen nicht sehr viel schwächeren Effekt hat als z. B. Aspirin. Wenn man die Farbe von Placebo variiert, erhält man verblüffende Unterschiede, die über die therapeutischen Differenzen hinausgehen.



[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

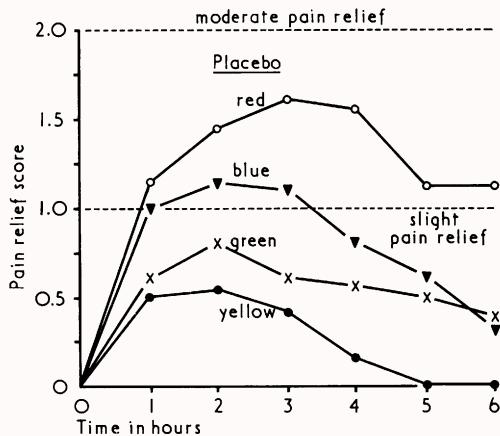
Seite 73 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Die verschiedenen Gruppen („Therapiearme“) sollten in allen wichtigen Punkten (Alter, Geschlecht, Schweregrad der Krankheit) vergleichbar sein. Die Randomisierung sollte dies approximativ garantieren; raffinierte Versuchspläne randomisieren „geschichtet“, z. B. nach Altersklasse. Trotzdem sollte man mögliche „Störgrößen“ (eben z. B. den Alterseffekt) miterheben. Man kann diese nämlich auch später, als sogenannte Kovariaten, in der statistischen Analyse quantitativ berücksichtigen.

Zusammengefasst einige Prinzipien der Versuchsplanung:

Randomisierung:

- Gleiche Chance für alle (einer Population), in eine Stichprobe zu kommen (Repräsentativität)
- gleiche Chance für alle (einer Stichprobe), in eine Gruppe zu kommen (Vergleichbarkeit)

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 74 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

[◀](#) [▶](#)

[◀](#) [▶](#)

Seite 75 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Standardisiertes Vorgehen:

- klare Einschluss-/Ausschlusskriterien
- klare diagnostische und experimentelle Bedingungen

Beispiel: In der Neurophysiologie ist genau festgelegt, wie und wo Elektroden auf die Kopfhaut zu kleben sind.

Doppelverblindung:

- Verfälschung durch Subjektivität vermeiden.

Beispiel: Bei einer Therapiestudie sollte der Patient (einfach-blind) und wenn möglich auch der Arzt (doppel-blind) nicht wissen, welches Medikament gegeben wird.

Kontrolle:

- neue Methode mit Placebo oder Standardtherapie vergleichen

Unabhängigkeit der Versuchseinheiten:

- Beine eines Versuchstieres sind nicht unabhängig.

Adäquate Stichprobengröße:

- Zu kleine Studien können keine klaren Ergebnisse liefern; zu grosse Stichproben sind unter Umständen unethisch, wenn Versuchspersonen z. B. weiterhin Placebo erhalten, obwohl die Wirksamkeit des neuen Medikaments längst feststeht.

Einfache Versuche:

- nur **zwei** Gruppen oder **zwei** Zeitpunkte vergleichen

„informed consent“:

- Ohne das Einverständnis einer Ethikkommission sind Versuchsergebnisse nicht mehr publizierbar.



[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite 76 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

3.8.3. Wo kommt der Zufall her?

Gründe für unterschiedliche Ergebnisse können sein:

1. systematisch („Bias“) oder
2. zufällig („Variabilität“)

Beispiele für systematische Diskrepanzen

- Messinstrument verstellt
- in Gruppe A mehr alte Patienten als in Gruppe B
- zwei behandelnde Ärzte mit unterschiedlichen („nicht operationalisierten“) Kriterien

Vermeidung systematischer Abweichungen durch:

- angemessene(n) Versuchsplan / Versuchsdurchführung
- statistische post-hoc Prüfungen (eventuell Kovariaten berücksichtigen)
- Trendanalysen
- Analyse der Untersuchereffekte

Zufällige Schwankungen

Zufällige Schwankungen enthalten neben dem Messfehler eine Vielzahl biologischer Schwankungen. Wenn man alle möglichen Einflussfaktoren unter Kontrolle hätte — wie angenähert in physikalischen Experimenten, wo man z. B. Druck, Temperatur und Magnetfeld konstant halten kann — gäbe es nur noch den Messfehler. Die



[Einführung](#)
[Deskriptive Statistik](#)
Wahrscheinlichkeit
[Testen](#)
[Regression](#)
[Testfragen](#)
[Index](#)

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite **77 von 100**

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

möglichen Schwankungen zwischen Patienten sind genetischer Art (dies macht oft einen grossen Teil aus), können aber auch der Umwelt zugeschrieben werden (Essen, Schlaf) oder der Lebensgeschichte (früherer Stress, Übergewicht).

Beispiel: Das Geburtsgewicht eines Kindes ist teilweise genetisch bestimmt, hängt aber auch von der Grösse der Mutter ab (kleine Mütter haben statistisch häufiger kleine Babies). Es hängt z. B. weiter vom Essen der Mutter ab, ob sie raucht, oder eine chronische Erkrankung hat.

Man versucht die zufälligen Schwankungen möglichst klein zu halten, um damit schärfere Aussagen zu erhalten. Einige Möglichkeiten sind:

- Standardisierung der Messmethode
- Kontrolle der potentiellen Einflussfaktoren
- Homogenisierung der Grundgesamtheit durch Ein- und Ausschlusskriterien (z. B. Alter, Schweregrad der Krankheit)

Die stochastische Modellierung im biomedizinischen Bereich ist nicht primär die Modellierung des Zufalls, sondern die Modellierung von Komplexität. Die vielen Einflussfaktoren, die man nicht alle erfassen kann, auch gar nicht alle kennt, werden — mit Erfolg — als stochastisch und nicht als deterministisch aufgefasst.



4. Prüfung von Hypothesen

4.1. Was ist ein statistischer Test ?

Bei der Analyse einer Stichprobe erhalten wir mittels statistischer Kennwerte nie Sicherheit über einen Sachverhalt, denn Kennwerte schwanken von Stichprobe zu Stichprobe. Statistische Tests schränken diese Unsicherheit quantitativ ein: Es sind Entscheidungsregeln, ob eine wissenschaftliche Hypothese mit grosser Wahrscheinlichkeit zutrifft. Man möchte *objektiv und quantitativ* beurteilen, ob eine Differenz oder ein Kennwert zufällig so herausgekommen ist, oder ob durch eine experimentelle Bedingung (z. B. eine Therapie) ein systematischer Effekt vorliegt. *Subjektive* Beurteilungen sind anfechtbar, weil die selbe Datenlage durch einen Forscher eher optimistisch, durch einen anderen eher pessimistisch eingeschätzt wird.

Statistische Tests werden heute in der medizinischen Literatur standardmäßig angewandt. Im Kapitel „Subjects and Methods“ einer wissenschaftlichen Arbeit gibt es üblicherweise einen Abschnitt, der darlegt, welche statistischen Methoden zum Einsatz kommen. Nachstehend ein Beispiel aus dem „British Medical Journal“ von 1996. Untersucht wird die Frage, ob der mittlere Cholesterinwert oder eine Veränderung des Cholesterinwerts das Suizidrisiko beeinflussen. Bei solch nicht naheliegenden Fragestellungen ist das Bedürfnis nach einer statistischen Prüfung besonders evident. Die Spalte „P value“ gibt an, mit welcher Wahrscheinlichkeit die gefundenen Differenzen im relativen Risiko durch Zufall zustande kommen könnten.

Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

«

»

◀

▶

Seite 78 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 79 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

STATISTICAL ANALYSIS

We used Student's *t* test and Cox's proportional hazards model for analysis. All analyses were stratified by the number of measurements of serum cholesterol for each subject (three to five).

Table 1—Relative risks (95% confidence interval) of suicide among 6393 men by average serum cholesterol concentration and change in cholesterol concentration

	No of subjects	No of suicides	Adjusted relative risk (95% confidence interval)*	P value
Average serum cholesterol concentration (mmol/l)†				
<4.78	827	10	3.16 (1.38 to 7.22)	0.007
4.78-6.21	3600	13	1.00	
>6.21	1966	9	1.28 (0.55 to 3.01)	0.56
Change in serum cholesterol concentration (mmol/l a year)‡				
Decline >0.13	1143	11	2.17 (0.97 to 4.84)	0.056
Change ≤0.13	2795	13	1.00	
Increase >0.13	2455	8	0.72 (0.30 to 1.72)	0.46

*Relative risks for average cholesterol concentration were adjusted, using Cox's proportional hazards model, for age, smoking habits (never, former, or current), and mean corpuscular volume at first examination. Relative risks for change in cholesterol concentration were adjusted as above and for average serum cholesterol concentration.

†Mean of serum cholesterol concentrations from all examinations.

‡Estimated using within person linear regression method (0.13 mmol/l equivalent to 5 mg/dl).



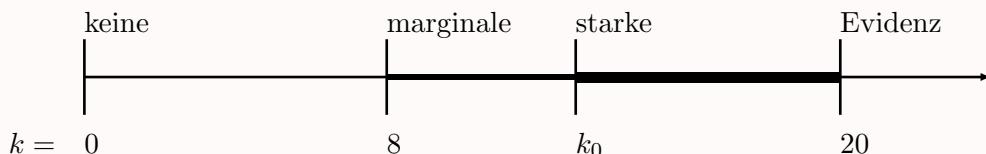
Einführendes Beispiel: Ein Standardmedikament wirkt in 40% aller Fälle. Ist ein neues Medikament besser?

In einer Studie wurden $n = 20$ Patienten mit dem neuen Medikament behandelt. Falls alt und neu gleich gut wären, würden im Mittel $k = 8$ Patienten geheilt. Falls weniger als 8 Patienten geheilt werden, spricht statistisch nichts für das neue Medikament. Falls es unwesentlich mehr als 8 Patienten sind, bleiben wir unsicher, ob dies nicht zufällig so ist. Ab einem gewissen k_0 , das deutlich grösser als 8 ist, glaubt man subjektiv an die Überlegenheit des neuen Medikaments. Kann man dies wahrscheinlichkeitsmässig erhärteten?

An einem anderen Beispiel wird diese Problematik aufgrund eines virtuellen Experimentes erläutert:

[Animation](#)

Die Evidenz, dass für die Erfolgswahrscheinlichkeit $p_{neu} > 0.4$ gilt, ist für verschiedene mögliche Resultate von k = „Anzahl geheilter Patienten“ etwa wie folgt:



Frage: Wie wahrscheinlich ist ein $k \geq k_0$, falls doch $p_{neu} = 0.4$ gilt? Anders: Mit welcher Wahrscheinlichkeit schliesst man auf Verbesserung, obwohl keine da ist? Falls $p_{neu} = 0.4$ gilt, ist k binomialverteilt mit $p = 0.4$. Daraus ergeben sich folgende

[Home Page](#)

[Titelseite](#)

«

»

◀

▶

Seite 80 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Wahrscheinlichkeiten (aus Tabellen der Binomialverteilung):

$$P[k \geq 11] = 0.128$$

$$P[k \geq 12] = 0.057$$

$$P[k \geq 13] = 0.021$$

$$P[k \geq 14] = 0.006$$

Was schliessen wir daraus? Falls das neue Medikament nicht besser ist (d. h. $p_{neu} = 0.4$), dann ist es sehr unwahrscheinlich, dass mehr als 12 Patienten geheilt werden. Falls wir also 13, 14 oder gar mehr Heilungen feststellen, dann ist das neue Medikament mit grosser Wahrscheinlichkeit besser (d. h. $p_{neu} > 0.4$). Angenommen, man führt die Studie wiederholt durch: Dann würden nur in ca. 2% der Studien mehr als 12 Patienten geheilt werden, wenn das neue Medikament tatsächlich nicht besser ist.

Allgemeine Formulierung:

Ein Studienplan startet mit einer wissenschaftlichen Hypothese. Diese kann auf klinischen oder wissenschaftlichen Erfahrungen basieren oder der Literatur entnommen werden.

H_1 : Wissenschaftliche Hypothese oder Alternativhypothese

Beispiel: $H_1: p_{neu} > 0.4$

Eine Alternativhypothese der Form $p_{neu} > 0.4$ heisst einseitig, weil man sich für Abweichungen in nur einer Richtung interessiert (hier: Verbesserung). Falls man die wissenschaftliche Hypothese $p_{neu} \neq 0.4$ (Verbesserung oder Verschlechterung) prüfen möchte, so nennt man dies eine zweiseitige Alternative.



Sowohl die wissenschaftliche Hypothese als auch die anschliessend zu formulierende statistische Hypothese (oder Nullhypothese) beziehen sich auf die unbekannten wahren Werte, nicht auf statistische Kennwerte (diese sind ja bekannte Zahlen!).

H_0 : Statistische Hypothese oder Nullhypothese

Beispiel: $H_0: p_{neu} = 0.4$

Ein statistischer Test prüft wahrscheinlichkeitsmässig die Nullhypothese, obwohl man bei der Planung einer Studie von der wissenschaftlichen Hypothese ausgeht. Die Schlussweise ist indirekt: Wenn die Nullhypothese aufgrund der Daten mit hoher Wahrscheinlichkeit abgelehnt werden kann, entscheidet man sich aufgrund dieser Evidenz für die wissenschaftliche Hypothese.

Man hat demnach zwei mögliche Entscheidungen zur Auswahl:

- verwerfe H_0 und bejahe H_1 oder
- verwerfe H_0 nicht und betrachte H_1 als nicht nachgewiesen.

Man hat dabei gemäss nachstehender Tabelle zwei Möglichkeiten, richtig zu entscheiden, und zwei Fehlermöglichkeiten.

Entscheidung	Wahrheit	
	H_0 stimmt	H_0 stimmt nicht
H_0 nicht verworfen	richtig	Fehler 2. Art: „ β “
H_0 verworfen	Fehler 1. Art: „ α “	richtig

Der α -Fehler bedeutet, eine neue Therapie oder wissenschaftliche Theorie zu befürworten, obwohl sie nicht besser ist als die alte. Der β -Fehler impliziert, dass

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 82 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



man eine neue Therapie oder Theorie nicht als besser erkennt, obwohl das in Wahrheit der Fall ist.

Es gibt eine Analogie zum Gerichtsverfahren eines nicht geständigen Angeklagten, wo auch eine Entscheidung bei Unsicherheit zu fällen ist.

	Gerichtsverfahren	Hypothesentesten
braucht starke Evidenz	Schuldspruch	neue Hypothese bejahen
Nullhypothese H_0	nicht schuldig	alte Theorie stimmt
Alternativhypothese H_1	schuldig	neue Theorie stimmt
Haltung	plädiere nicht schuldig ohne starke Evidenz für Schuld	behalte Nullhypothese, ausser sie sei mit den Daten sehr unverträglich

Wir führen jetzt wichtige statistische Begriffe ein:

Definition **Irrtumswahrscheinlichkeit** oder **Signifikanzniveau** eines Tests = α
= (maximale) Wahrscheinlichkeit eines Fehlers 1. Art
= Wahrscheinlichkeit, neue Therapie oder Theorie als besser zu betrachten,
obwohl dies nicht der Fall ist.

Konventionell wird ein $\alpha = 5\%$ oder selten 1% vorgegeben. Wenn die Daten beim Testen eine kleinere Wahrscheinlichkeit als α ergeben, lehnt man die Nullhypothese ab und nimmt die wissenschaftliche Hypothese an.

Meist geben Programm pakete nicht an, ob ein Resultat signifikant oder nicht signifikant ist, sondern geben einen p -Wert.



[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
Testen
[Regression](#)
[Testfragen](#)
[Index](#)

Definition ***p*-Wert** eines Tests

1. korrekt: $p = \text{kleinstes } \alpha$, das noch signifikant würde.
2. lax, aber verständlicher: $p = \text{Wahrscheinlichkeit, gegeben } H_0 \text{ stimme, für die Testgrösse einen so grossen oder grösseren als den aus den Daten berechneten Wert zu erhalten.}$

Falls $p \leq \alpha$, wird die Null-Hypothese abgelehnt (und die wissenschaftliche Hypothese angenommen); falls $p > \alpha$, wird die Null-Hypothese nicht abgelehnt. In der Praxis wird aber ein ganz kleiner p -Wert (z. B. < 0.0001) als „starke“ Ablehnung von H_0 gewertet und als „hochsignifikant“ eingestuft. Auch wenn dies nicht ganz korrekt ist, ist eine solche Interpretation tolerierbar.

Wir illustrieren die Begriffe mit dem Medikamentenbeispiel, indem wir ein Signifikanzniveau von $\alpha = 0.05$ vorgeben:

- Falls das Ergebnis $k = 13$ Heilungen ist, erhalten wir $p = P[k \geq 13] = 0.021$ („ p -Wert“). Da demnach $p \leq \alpha = 0.05$ gilt, können wir H_0 verwerfen und mit 5% Irrtumswahrscheinlichkeit schliessen, dass das neue Medikament besser ist.
- Falls aber nur $k = 12$ Patienten geheilt werden, ergibt sich $p = P[k \geq 12] = 0.057$, so dass $p > \alpha = 0.05$. Man kann also H_0 nicht verwerfen, der statistische Nachweis der Überlegenheit des neuen Medikaments ist nicht gelungen.

[Home Page](#)

[Titelseite](#)

«

»

◀

▶

Seite 84 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
Testen
[Regression](#)
[Testfragen](#)
[Index](#)

Selbstkontrolle

Eventuell war die Stichprobe in letzterem Fall nicht gross genug gewählt, um eine relevante Verbesserung mit genügender Wahrscheinlichkeit festzustellen (zu kleine Trennschärfe). Deswegen darf man in einem solchen Fall auch nicht sagen, dass statistisch bewiesen wird, dass das neue Medikament nicht besser ist. Richtig ist zu sagen, dass man eine Verbesserung nicht nachweisen konnte.

Definition **Trennschärfe** oder **Macht** („power“) eines Tests $= 1 - \beta$
 $= 1 - \text{Wahrscheinlichkeit eines Fehlers 2. Art}$
 $= \text{Wahrscheinlichkeit, neue Therapie oder Theorie als besser nachzuweisen, wenn sie tatsächlich besser ist.}$

Die Trennschärfe eines Tests hängt von der Stichprobengrösse n und der **Effektgrösse** ab.

Im vorher diskutierten Medikamenten–Beispiel hängt die Trennschärfe von der wahren Erfolgswahrscheinlichkeit des neuen Medikaments ab. Wenn $p_{neu} = 0.8$ ist (grosser Effekt), dann ist die Trennschärfe $1 - \beta = P[k \geq 13] = 0.97$, und nichtsignifikante Ergebnisse sind selten ($\beta = 3\%$). Wenn das neue Medikament nur unwesentlich besser ist als das alte ($p_{neu} = 0.5$), ist $1 - \beta = P[k \geq 13] = 0.13$. Die Stichprobe ist also viel zu klein, um derartige Effekte nachweisen zu können ($\beta = 87\%$).

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 85 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
Testen
[Regression](#)
[Testfragen](#)
[Index](#)

Beispiel für die Konstruktion eines Tests

Man möchte prüfen, ob herzkranke Babies später zu laufen beginnen als gesunde Babies. Dafür wird eine empirische Studie mit $n = 20$ herzkranken Kindern durchgeführt. Deren Werte sollen mit Normwerten aus der Literatur verglichen werden, die man als fest annimmt. Für den Beginn des Laufens liefert die Normpopulation ein Mittel von $\mu_0 = 12$ Monaten mit einer Populationsstreuung von $\sigma_0 = 1.8$ Monaten. Die Studie mit den herzkranken Babies ergibt ein Mittel von $\bar{x} = 12.8$ Monaten (das zugehörige, unbekannte Populationsmittel sei μ). Der Einfachheit halber nehmen wir an, dass die Standardabweichung der Norm $\sigma_0 = 1.8$ auch gültig ist.

- Wissenschaftliche Hypothese: Kinder mit angeborenem Herzleiden laufen später.

$$H_1: \mu > \mu_0$$

- Statistische (Null-) Hypothese:

$$H_0: \mu = \mu_0$$

[Home Page](#)

[Titelseite](#)



Seite 86 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Die statistische Prüfung will nun quantitativ nachweisen, dass die Grösse der Abweichung $(\bar{x} - \mu_0)$ nicht durch den Zufall erklärt werden kann:

- Berechne Differenz $(\bar{x} - \mu_0)$.
- Die Grösse von $(\bar{x} - \mu_0)$ wird auf den Standardfehler σ_0/\sqrt{n} der Differenz bezogen. Dies Führt zur Teststatistik (Testgrösse) z :

$$z = \frac{\bar{x} - \mu_0}{\sigma_0/\sqrt{n}} = \frac{0.8}{1.8/\sqrt{20}} = 1.99$$

Die Wahrscheinlichkeit, einen so grossen oder noch grösseren Wert für die Testgrösse per Zufall zu erhalten — obwohl die Null-Hypothese gilt — bezeichnen wir mit p („ p -Wert“).

Um die Wahrscheinlichkeit p berechnen zu können, müssen wir eine Annahme über die Verteilung der Daten machen. Der Einfachheit halber nehmen wir Normalverteilung an. In der Praxis muss dies geprüft werden.

Die folgende Tabelle zeigt die Resultate nicht nur für $n = 20$, sondern auch für $n = 10, 40, 80$ bei identischem \bar{x} .

n	10	20	40	80
z	1.41	1.99	2.81	3.98
p	0.079	0.023	0.0025	0.0003

Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 87 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
Testen
[Regression](#)
[Testfragen](#)
[Index](#)

[Selbstkontrolle](#)

Allgemeines Prozedere für einen statistischen Test

- Formuliere wissenschaftliche Hypothese H_1 und Nullhypothese H_0 (bezogen auf Populationswerte).
- Setze Irrtumswahrscheinlichkeit α fest.
- Es werden Daten x_1, \dots, x_n gesammelt.
- Definiere Test-(Prüf-)Statistik $T(x_1, \dots, x_n)$.
Erwünscht ist:
 - T soll empfindlich auf H_1 reagieren.
 - Die Verteilung von T unter Annahme der Nullhypothese H_0 soll mathematisch berechenbar sein.

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite 88 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Typische Form für T :

- bei einseitiger Alternative:

$$T = \frac{\text{beobachteter Wert} - \text{hypothetischer Wert}}{\text{Standardfehler des beobachteten Wertes}}$$

- bei zweiseitiger Alternative:

$$T = \left| \frac{\text{beobachteter Wert} - \text{hypothetischer Wert}}{\text{Standardfehler des beobachteten Wertes}} \right|$$

- Berechne Teststatistik für gegebene Daten $x_1, \dots, x_n \longrightarrow T_0$.
- Die Verteilung $F_T(x)$ zu T sei unter der Annahme der Hypothese H_0 bekannt.
- Ermittle p -Wert zu beobachtetem T_0 :
= Wahrscheinlichkeit, per Zufall (bei der Nullhypothese) einen so extremen Wert wie T_0 zu erhalten:

$$p = 1 - F_T(T_0)$$

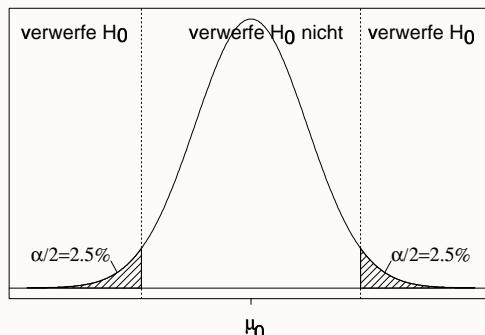
- Entscheide: Falls $p \leq \alpha \implies$ verwirfe H_0 .
Falls $p > \alpha \implies$ verwirfe H_0 nicht.

Für die verschiedenen wissenschaftlichen Fragestellungen gibt es eine Vielzahl von Tests. In diesem Kapitel behandeln wir einige von grosser praktischer Bedeutung.

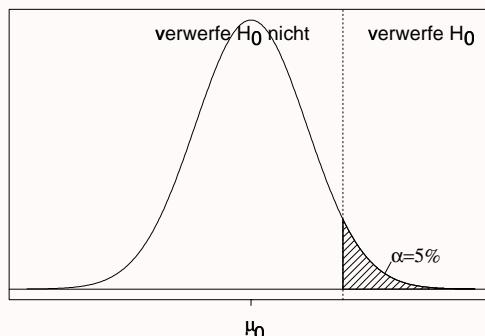
Nachfolgend werden graphisch die Bereiche dargestellt, in denen die Nullhypothese verworfen bzw. nicht verworfen wird, sowie die Wahrscheinlichkeiten α ($= 0.05$) und $1 - \beta$. Dies erfolgt für den Nachweis eines Mittelwertsunterschiedes, d. h. $H_0: \mu = \mu_0$ und $H_1: \mu = \mu_1$. Für die Daten wird eine Normalverteilung bekannter Varianz angenommen.

Falls die Nullhypothese stimmt, d. h. $\mu = \mu_0$, stellen sich die Entscheidungsbereiche wie folgt dar:

a) Alternative zweiseitig ($\mu \neq \mu_0$):



b) Alternative einseitig ($\mu > \mu_0$ oder je nach Fragestellung auch $\mu < \mu_0$):

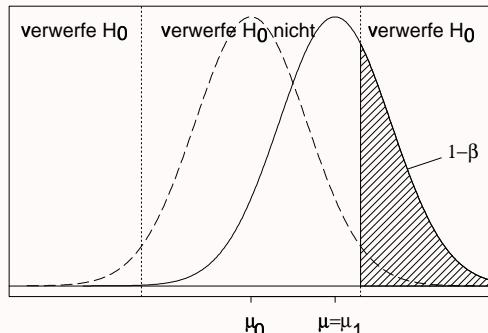


Die schraffierten Flächen geben die Wahrscheinlichkeiten an, dass die Nullhypothese fälschlicherweise verworfen wird.

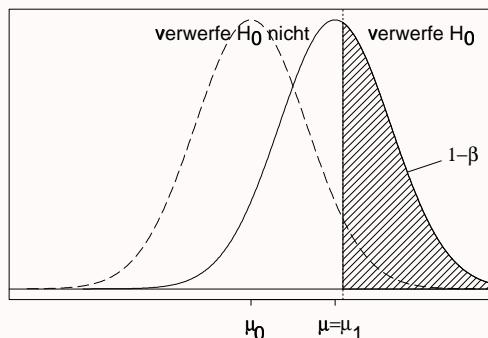


Falls die Nullhypothese nicht stimmt und $\mu = \mu_1 > \mu_0$ gilt, führt dies zu folgenden Bereichen (gestrichelt: Wahrscheinlichkeitsdichte unter H_0 , durchgezogen unter H_1)

a) Alternative zweiseitig ($\mu \neq \mu_0$):



b) Alternative einseitig (in der Graphik $\mu > \mu_0$):



[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite 91 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 92 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Man sieht, dass die Trennschärfe ($1 - \beta$) bei gleicher Differenz ($\mu - \mu_0$) grösser wird, wenn man einseitig testet (die Verwerfungsgrenze „rutscht hinunter“). Da man manchmal so bei einseitigem Testen Signifikanz erhält, wenn dies zweiseitig nicht möglich ist, und da Signifikanz die Publizierbarkeit verbessert, sind viele Zeitschriften einseitigen Tests gegenüber skeptisch (Gefahr, post-hoc eine Richtung zu postulieren).

Macht (Trennschärfe) eines Tests

- Optimale Tests sind so definiert, dass sie bei vorgegebenem α maximale Trennschärfe haben (Beispiel: t -Tests, falls Normalverteilung vorliegt).
- Die Macht sinkt, wenn α kleiner wird („Unschärferelation“: wenn man den einen Fehler kleiner macht, wird der andere grösser).
- Die Macht steigt, wenn die Variabilität kleiner wird. Dies bedeutet, dass homogenere Gruppen oder bessere Messmethodik von Vorteil sind.
- Die Macht ist bei einseitigen Tests besser.
- Die Stichprobengrösse n kann mit dem Versuchsplan so gewählt werden, dass zum Beispiel $\beta = 0.10$ oder 0.05 erreicht wird (d. h. vorgegebene Macht 90% oder 95%). Damit ist ein klarer Entscheid bezüglich der Nullhypothese und der Alternative möglich.

Wir werden die Bestimmung der Stichprobengrösse jetzt näher erläutern („Power Analyse“). Wem das zu speziell ist, kann direkt bei Abschnitt [4.2](#) weiterlesen.



Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

Bestimmung der Stichprobengrösse

Die Bestimmung der notwendigen Stichprobengrösse soll am Beispiel des Nachweises eines Mittelwertsunterschiedes von einem vorgegebenen Wert μ_0 bei bekannter Varianz σ_0^2 und unabhängigen normalverteilten Daten x_1, \dots, x_n erläutert werden. Die Teststatistik ist

$$z = \sqrt{n} \frac{\bar{x} - \mu_0}{\sigma_0}.$$

Unter der Nullhypothese $H_0 : \mu = \mu_0$ ist z standardnormalverteilt. Zur Erinnerung: Mit z_α bezeichnen wir das $\alpha \times 100\%$ -Perzentil der Standardnormalverteilung. H_0 wird verworfen, wenn $|z| > z_{1-\alpha/2}$, da

$$P_0[|z| > z_{1-\alpha/2}] = \alpha$$

(siehe auch die Graphik zum Fehler erster Art bei zweiseitiger Alternative, oben). Wenn $\mu = \mu_1 > \mu_0$ gilt, dann ist

$$z = \sqrt{n} \frac{\bar{x} - \mu_1}{\sigma_0} + \sqrt{n} \frac{\mu_1 - \mu_0}{\sigma_0} \sim \mathcal{N}(\sqrt{n}\delta, 1).$$

Die Grösse $\delta = \frac{\mu_1 - \mu_0}{\sigma_0}$ ist hier die **Effektgrösse**.

[Home Page](#)

[Titelseite](#)



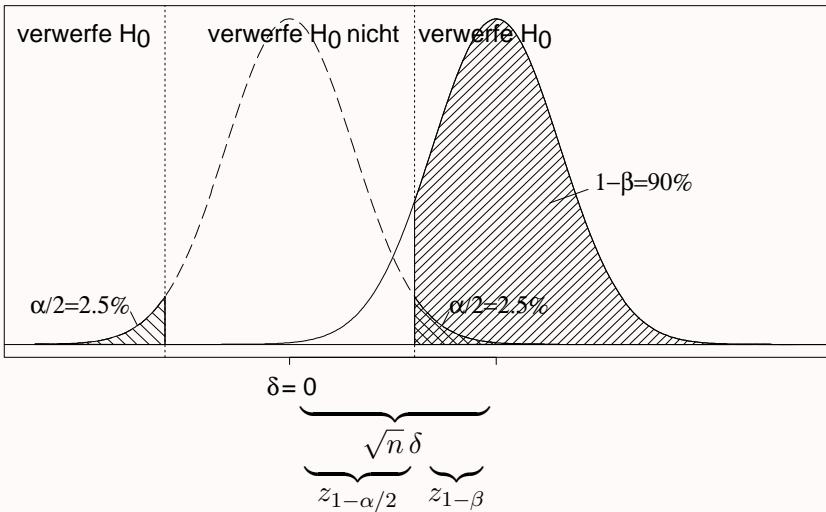
[Seite 93 von 100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Die Trennschärfe $1 - \beta$ erhält man aus der obigen Graphik. Für eine vorgegebene Effektgrösse δ und eine vorgegebene Trennschärfe $1 - \beta$ berechnet sich die notwendige Stichprobengrösse damit aus

$$\sqrt{n}\delta = z_{1-\alpha/2} + z_{1-\beta}$$

und folglich

$$n = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{\delta^2}.$$

Die notwendige Stichprobengrösse n ist proportional zur Varianz σ_0^2 und indirekt proportional zum quadrierten Mittelwertsabstand $(\mu_1 - \mu_0)^2$. Das ist analog zur Genauigkeit von Schätzungen: Um einen halb so grossen Effekt nachweisen zu können,

[Home Page](#)

[Titelseite](#)

«

»

◀

▶

Seite 94 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite 95 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

brauchen wir 4 mal so viele Beobachtungen. Die Stichprobengröße n wächst ausserdem bei einer Reduktion des Signifikanzniveaus α und bei einer Erhöhung der Trennschärfe $1 - \beta$. Diese Zusammenhänge sind aber nichtlinear und deshalb nicht sofort zu überblicken.

4.2. Tests auf Mittelwertsunterschiede

Wenn man eine experimentelle Bedingung ändert oder eine neue Therapie erprobt, wird man sich zuerst für Veränderungen im Mittelwert interessieren. Beim Vergleich von 2 Gruppen gibt es im wesentlichen 3 Situationen:

- Vergleich eines Mittelwertes mit einem bekannten festen Wert (Einstichprobenproblem),
- Vergleich der Mittelwerte zweier unabhängiger Stichproben (Zweistichprobenproblem),
- Vergleich der Mittelwerte zweier verbundener Stichproben (gepaartes Testproblem).

Wenn wir für die Beobachtungen eine Normalverteilung voraussetzen, werden die Hypothesen mit t -Tests geprüft, sonst mit entsprechenden Rangverfahren. Im folgenden wollen wir die verschiedenen Testverfahren vorstellen.

Dabei zeigen wir auch, wie die entsprechenden Testgrössen konstruiert werden. Dadurch kann man ein Gefühl dafür bekommen, wie das soeben beschriebene allgemeine Procedere bei diesen Problemen funktioniert. Für die praktische Anwendung ist die Kenntnis der Konstruktion der Testgrössen nicht notwendig, da die Verfahren in allen gängigen Statistikprogrammen verfügbar sind.



[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
Testen
[Regression](#)
[Testfragen](#)
[Index](#)

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 96 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

4.2.1. Einstichproben– t –Test

Beim einführenden Beispiel des Laufens von Babies benutzten wir einen Einstichprobentest, da wir mit einer bekannten Norm — als wahr vorausgesetzt — und nicht mit einer Gruppe von gesunden Kindern verglichen haben. Letzteres führt auf einen Zweistichproben–Test (siehe unten). Da wir normalverteilte Daten und eine bekannte Standardabweichung σ_0 angenommen haben, resultiert eine normalverteilte Teststatistik z .

Beispiel: Laufen von Babies \Rightarrow Einstichproben– z –Test

$x_1, \dots, x_n \sim \mathcal{N}(\mu, \sigma_0^2)$, wo σ_0^2 bekannt ist

$$H_0: \mu = \mu_0$$

$$\Rightarrow z = \frac{\bar{x} - \mu_0}{\sigma_0 / \sqrt{n}} \quad \text{unter der Nullhypothese normalverteilt } \mathcal{N}(0, 1)$$

Falls aber — was im allgemeinen realistisch ist — die Varianz σ^2 nicht bekannt ist, muss sie durch s^2 aus den Daten geschätzt werden. Da man somit durch eine zufällige Größe teilt, ist die entsprechende Statistik t nicht mehr normalverteilt. Dies führt auf den t –Test:

Teststatistik	$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$	Einstichproben– t –Test
---------------	--	---------------------------

Die Teststatistik t ist t –verteilt mit $(n - 1)$ Freiheitsgraden.

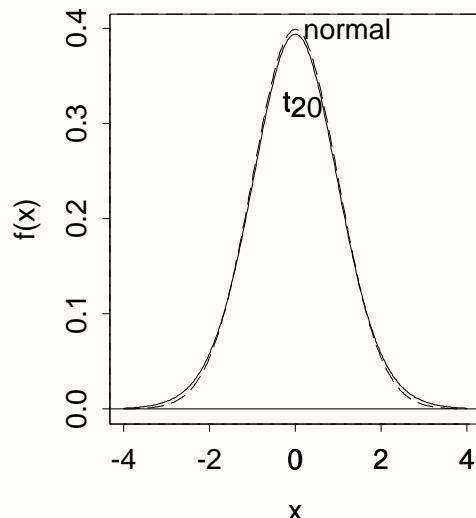
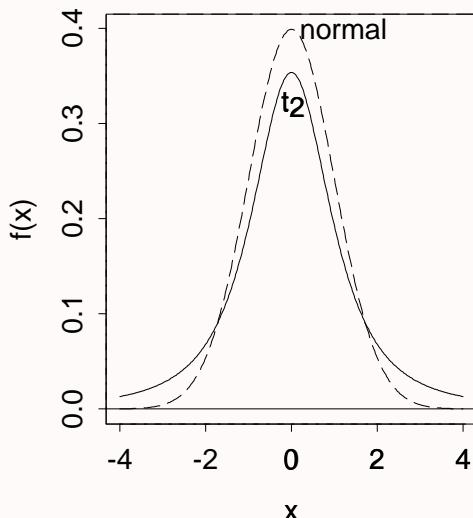
Definition der **t –Verteilung**:

X_1, \dots, X_n seien unabhängig verteilt nach $\mathcal{N}(0, 1)$.



Dann gilt: $t = \frac{\bar{x}}{s/\sqrt{n}}$ ist t -verteilt mit $(n - 1)$ Freiheitsgraden.

Auf den folgenden Abbildungen sehen Sie, dass die t -Verteilung mit wenig Freiheitsgraden (d. h. n klein) beträchtlich von der Normalverteilung abweicht, für viele Freiheitsgrade wenig (für $n \rightarrow \infty$ wird die Approximation dann exakt).



Vergleich der t - und der Standardnormalverteilung: links t mit 2, rechts mit 20 Freiheitsgraden.

[Home Page](#)

[Titelseite](#)



Seite 97 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite 98 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Unter der Annahme, dass $s = 1.8$ gilt, erhalten wir beim vorangehenden Beispiel mit $\bar{x} = 12.8$:

n	10	20	40	80
t	1.41	1.99	2.81	3.98
p	0.0961	0.0306	0.0039	0.0008

Die p -Werte sind im Vergleich zur z -Teststatistik (mit σ_0 anstatt s) etwas grösser geworden.

4.2.2. Zweistichproben- t -Test für unabhängige Stichproben

Man möchte die Mittelwerte von zwei Stichproben (von zwei verschiedenen Gruppen) statistisch vergleichen. In der Praxis ist dieses Problem viel häufiger als das Einstichproben-Problem.

Beispiel: Vergleich der log- T_4 -Zellanzahl von Hodgkin- und non-Hodgkin-Patienten

Gruppe 1 (Hodgkin): $n = 20, \bar{x} = 6.49, s_x = 0.71$

Gruppe 2 (non-Hodgkin): $m = 20, \bar{y} = 6.09, s_y = 0.63$

Die wissenschaftliche Hypothese ist: Die T_4 -Zellzahl bei Hodgkin-Patienten ist auch nach der Remission erhöht verglichen mit non-Hodgkin-Patienten.

$$H_1: \mu_x > \mu_y \text{ (einseitige Alternative)}$$

$$H_0: \mu_x = \mu_y \implies \mu_x - \mu_y = 0$$

Die Konstruktion einer Zweistichproben-Teststatistik erfolgt nach dem bewährten Prinzip:



Beobachtet – Erwartet bei $H_0 = (\bar{x} - \bar{y}) - 0 = 0.4$

Liegt 0.4 genügend weit von 0 weg, so dass die Nullhypothese verworfen werden kann? Dividiere durch den Standardfehler von $(\bar{x} - \bar{y})$!

Machen wir die Annahme, dass $\sigma_x^2 = \sigma_y^2 = \sigma^2$ gilt, d. h. gleiche Varianz in beiden Stichproben. Dann erhalten wir für die Standardabweichung von $(\bar{x} - \bar{y})$:

$$\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}$$

Die gemeinsame Standardabweichung σ ist nicht bekannt und muss durch s geschätzt werden. Die Formel für die aus beiden Stichproben kombinierte (gepoolte) Standardabweichung s lautet:

$$s = \sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}}$$

Damit ergibt sich als Testgrösse:

Teststatistik	$t = \frac{\bar{x} - \bar{y}}{s \sqrt{\frac{1}{n} + \frac{1}{m}}}$	Zweistichproben- t -Test
---------------	--	----------------------------

Annahme: Seien $x_1, \dots, x_n, y_1, \dots, y_m$ unabhängige, **normalverteilte** Grössen mit gleicher Varianz.

Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 99 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
Testen
[Regression](#)
[Testfragen](#)
[Index](#)

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite **100** von **100**

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Dann gilt: Die Teststatistik t hat eine t -Verteilung mit $n + m - 2$ Freiheitsgraden. Bei einer Datenanalyse sollten die Annahmen der Normalverteilung und gleicher Varianz wenigstens grob geprüft werden, z.B. mittels Box-Plots.

Beispiel: Wir kehren zurück zum Vergleich von log- T_4 -Zellanzahl für $n = 20$ Hodgkin- und $m = 20$ non-Hodgkin-Patienten.

Nach der obigen Formel erhalten wir ein $s = 0.67$ und damit als Testgrösse:

$$t = \frac{0.4}{0.67\sqrt{\frac{1}{20} + \frac{1}{20}}} = 1.88$$

$P[t \geq 1.88] = \text{einseitiges } p = 0.034$,

also ist $p \leq \alpha = 0.05$: Hodgkin-Patienten haben signifikant mehr T_4 -Zellen.

Wie bereits erwähnt, haben einseitige Tests eine höhere Trennschärfe als zweiseitige, was zu Misbrauch führen kann. Der zweiseitige Test hat den p -Wert

$$p = P[t \leq -1.88 \text{ oder } t \geq 1.88] = 0.068 : p > \alpha = 0.05$$

Es ergibt sich also kein signifikanter Unterschied zwischen den beiden Gruppen.

Wenn man mehr als zwei Gruppen miteinander vergleichen will, sollte eine Varianzanalyse als Verallgemeinerung des Zweistichproben- t -Tests durchgeführt werden
(im Internet-Kurs nicht behandelt).

[Selbstkontrolle](#)



[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
Testen
[Regression](#)
[Testfragen](#)
[Index](#)

4.2.3. Der gepaarte t -Test

(t -Test für „gepaarte“ oder „verbundene“ Stichproben)

Ein relativ häufiger Fehler ist, dass der eben besprochene t -Test auch für den Vergleich von Daten benutzt wird, die am selben Menschen gewonnen werden, also nicht unabhängig sind. Dafür ist der gepaarte t -Test konzipiert.

Beispiele:

- prä–post–Vergleiche bei Therapiestudien
- Mehrfachuntersuchungen an denselben Patienten
- Vergleich EEG linke und rechte Hemisphäre

Beispiel: Es werden kardiologische Funktionen bei Typ I Diabetikern untersucht.

Es ist bekannt, dass Diabetiker schlechtere kardiovaskuläre Werte im Vergleich zu Gesunden aufweisen. Die Frage, ob sich die Funktionen bei guter Glukose–Einstellung verbessern, wurde bei $n = 8$ Patienten untersucht. Hier wird die Herzrate analysiert.

x_1, \dots, x_n — Daten zum Zeitpunkt 1, bei schlechter Einstellung

y_1, \dots, y_n — Daten zum Zeitpunkt 2, bei guter Einstellung

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 101 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)

[◀◀](#)

[▶▶](#)

[◀](#)

[▶](#)

Seite [102](#) von [100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Man erwartet eine Erniedrigung der Herzrate bei verbesserter Einstellung (Alternativ-Hypothese H_1), was zu folgendem Hypothesenpaar führt:

$$H_0: \mu_x = \mu_y$$

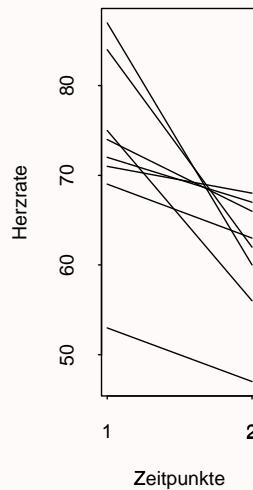
$$H_1: \mu_x > \mu_y$$

Es wurden dazu die individuellen Veränderungen (post-prä Werte) $d_i = y_i - x_i$ in der Herzrate analysiert. Die entsprechende Formulierung der Hypothese ist dann mit $\delta = \mu_y - \mu_x$ wie folgt:

$$H_0: \delta = 0$$

$$H_1: \delta < 0$$

Nr	x	y	d
1	74	66	-8
2	72	67	-5
3	84	62	-22
4	53	47	-6
5	75	56	-19
6	87	60	-27
7	69	63	-6
8	71	68	-3
Mittelwert	73	61	-12
s	10	7	9.2





[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
Testen
[Regression](#)
[Testfragen](#)
[Index](#)

Wenn man bei den d_i angelangt ist, reduziert sich der gepaarte t –Test formal auf den Einstichproben– t –Test, d. h. auf folgende Teststatistik:

Teststatistik	$t = \frac{\bar{d}}{s_d / \sqrt{n}}$	Gepaarter t –Test
---------------	--------------------------------------	---------------------

Unter der Annahme, dass die d_i normalverteilt sind, folgt die Teststatistik t unter der Nullhypothese einer t –Verteilung mit $(n - 1)$ Freiheitsgraden.

$$t = \frac{-12}{9.2 / \sqrt{8}} = -3.7$$

Die Wahrscheinlichkeit ($t \leq -3.7$) zu erhalten ist 0.004. Man erhält demnach eine signifikante Verbesserung bei guter therapeutischer „Compliance“. Der Effekt ist deutlich und deshalb bereits für $n = 8$ Patienten nachweisbar.

[Datenanalyse](#)

Falsch wäre die Anwendung des (normalen) Zweistichproben– t –Tests.

[Selbstkontrolle](#)

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 103 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



4.2.4. Rangtests: Mann–Whitney– und Wilcoxon–Test

In der Praxis sind die t –Tests oft problematisch, weil normalverteilte Daten vorausgesetzt werden, bzw. weil andernfalls eine Transformation zur Normalverteilung gesucht werden muss (z. B. logarithmieren der Anzahl T_4 –Zellen). Wir wollen hier Tests vorstellen, die eine gute Trennschärfe haben, aber die Normalverteilungsannahme nicht erfordern. Die Idee ist dabei, nur die Rangordnung der Daten und nicht die Daten selbst zu benützen. Es besteht eine gewisse Ähnlichkeit zur Benutzung von Median und Perzentilen als Kennwerte; diese beruhen auch auf der Rangordnung. Daraus ergibt sich auch die Unempfindlichkeit gegen Ausreisser und extreme Werte: Wie gross ein Wert numerisch auch sei, in der Rangordnung einer Stichprobe der Grösse n erhält er höchstens den Rang n .

Der Test zum Vergleich der Zentren von zwei unabhängigen Gruppen und damit das Analogon zum Zweistichproben– t –Test heisst **Mann–Whitney Test**. Er wird auch Wilcoxon–Test für unabhängige Stichproben oder Wilcoxon–Rangsummen–Test genannt. Das Analogon zum gepaarten t –Test ist der **Wilcoxon–Test für Paardifferenzen** (Wilcoxon signed rank test).

Wie bei den t –Tests setzt man voraus, dass die Beobachtungen bzw. Paare unabhängige Zufallsgrössen sind.

Die Trennschärfe dieser Rangtests ist für ziemlich beliebige Situationen gut, auch dann, wenn die entsprechenden t –Tests nicht gültig sind. Die Rangtests haben aber auch bei Vorliegen einer Normalverteilung eine Trennschärfe, die nahe den optimalen t –Tests ist. Im Zweifelsfall sollte man also immer Rangtests anstelle von t –Tests verwenden.

Zur Illustration des Vorgehens beim Mann–Whitney Test vergleichen wir zwei unabhängige Stichproben x_1, \dots, x_6 und y_1, \dots, y_6 , indem wir sie in eine **gemeinsa**

[Home Page](#)

[Titelseite](#)

«

»

◀

▶

Seite 104 von 100

[Zurück](#)

[Vollbild](#)

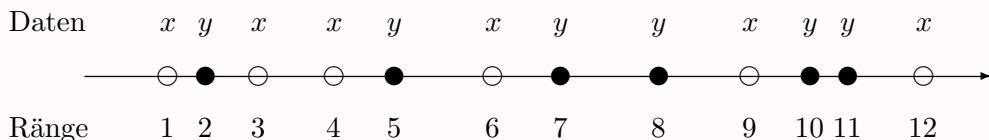
[Schließen](#)

[Beenden](#)



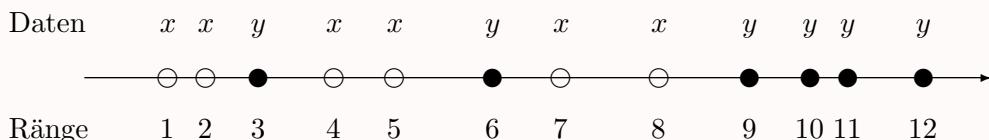
me Rangordnung bringen.

Situation 1: $\mu_x \approx \mu_y$



Der mittlere Rang der x_i -Werte ist 5.8, derjenige der y_i ist 7.2, so dass sie annähernd gleich sind.

Situation 2: $\mu_y > \mu_x$



Der mittlere Rang der y_i ist mit 8.5 deutlich grösser als derjenige der x_i mit 4.5.

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 105 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 106 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Voraussetzung: Seien x_1, \dots, x_n und y_1, \dots, y_m unabhängige, zufällige Größen.

Das Vorgehen ist nun wie folgt:

1. Erstelle **gemeinsame** Rangordnung.
2. Berechne getrennt mittlere Ränge der x_i und der y_i .
3. Die mittleren Ränge dienen dem Programm dazu, p -Werte zu berechnen.

Beispiel: Wir wollen wieder prüfen, ob die T_4 -Zellanzahl bei Hodgkin-Patienten erhöht ist, verglichen mit non-Hodgkin-Patienten.

Die T_4 -Zellanzahl ist nicht normalverteilt, sondern rechtsschief, weshalb wir früher logarithmiert haben, um den t -Test anwenden zu können.

Hier ein Ausschnitt der geordneten Werte:

Gruppe	nH	nH	H	nH	nH	H	...
T_4 -Zellen	116	151	171	192	208	257	...
Rang	1	2	3	4	5	6	...

Es fällt auf, dass die non-Hodgkin-Patienten bei kleinen Werten stark vertreten sind. StatView liefert einen p -Wert von $p = 0.04$ bei einseitigem Testen, und damit eine signifikante Erhöhung der T_4 -Zellanzahl bei remittierten Hodgkin-Patienten. Man erhält somit ein vergleichbares Ergebnis wie wenn man logarithmiert und den t -Test (einseitiges $p = 0.034$) anwendet.

[Selbstkontrolle](#)



4.3. Tests für Proportionen oder Wahrscheinlichkeiten

4.3.1. Einstichprobenfall

Man möchte prüfen, ob eine berechnete Proportion oder relative Häufigkeit von einem vorgegebenen festen Wert signifikant abweicht.

$$H_0: p = p_0, \quad H_1: p > p_0 \text{ respektive } p < p_0 \text{ (einseitig)}$$
$$p \neq p_0 \text{ (zweiseitig)}$$

Der Test erfolgt mittels Binomialverteilung. Beispiele von früher sind:

1. Das Standardmedikament heilt 40% der Patienten ($p_0 = 0.4$). Ist das neue Medikament besser, d. h. $H_1: p_{neu} > p_0$?
2. Ist die Häufigkeit von Knabengeburten tatsächlich grösser als 0.5?

$$H_0: p = 0.5 = p_0, \quad H_1: p > 0.5 = p_0$$

Die Antwort ist „ja“, siehe Konfidenzintervalle im Abschnitt 4.6.3.

Datenanalyse

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite 107 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

4.3.2. Zweistichprobenfall

Statistisch verglichen werden nun zwei empirische Proportionen oder Häufigkeiten.

Beispiel: Bei 34 von 113 Knaben und bei 54 von 139 getesteten Mädchen erfolgt der Nachweis eines Grippevirus-Antikörpers. Gibt es einen Geschlechtsunterschied in der Häufigkeit?

Man kann — ähnlich zum Zweistichproben- t -Test — die Differenz der beiden Häufigkeiten geeignet normieren und die Testgrösse mit den Perzentilen der ap-



[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite [108](#) von [100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

proximativ gültigen Standardnormalverteilung vergleichen. Für das Beispiel führt dies auf einen zweiseitigen p-Wert von $p = 0.14$, d. h. der Geschlechtsunterschied ist nicht signifikant. Natürlich ist es aber, die Resultate als Vierfelder-Tafel darzustellen und mit einem χ^2 -Homogenitätstest statistisch zu prüfen (siehe Abschnitt 4.4.2).

Achtung: Auch hier muss man bei gepaarten Studien — z. B. wenn die Häufigkeit von Schmerzen vor und nach einer Behandlung bei denselben Patienten untersucht wird — anders vorgehen (McNemar-Test).

4.4. Der χ^2 -Test

Der χ^2 -Test — oder besser die χ^2 -Statistik — eignet sich zur Beantwortung qualitativ unterschiedlicher Fragen. Gemeinsam ist ihnen, dass wir es mit kategorialen Daten zu tun haben.

4.4.1. χ^2 -Anpassungstest

Es geht hier darum, die empirische Verteilung kategorialer Daten mit einer vorgegebenen Verteilung statistisch zu vergleichen.

Beispiel: Die Genotypen A, B und C kommen nach einem Vererbungsmodell in den Häufigkeiten $1/4$, $1/2$, $1/4$ vor. Hier ist die Verteilung durch ein Modell vorgegeben.

Um das Modell zu überprüfen, werden 100 Pflanzen gezüchtet, mit folgendem Ergebnis.



3 Zellen:

A	B	C
18	55	27

Die Übereinstimmung der Daten mit dem Modell soll nun statistisch geprüft werden.

$$H_0: p_A = 1/4, \quad p_B = 1/2, \quad p_C = 1/4$$

Unter der Nullhypothese erwartet man Zellhäufigkeiten 25, 50, 25.

Diese Fragestellung entspricht für kategoriale Daten in etwa derjenigen eines Einstichprobentests bei kontinuierlichen Daten.

Idee: Vergleiche beobachtete („Obs“ für „observed“) und erwartete („Exp“ für „expected“) Zellhäufigkeiten:

$$(18 - 25)^2, \quad (55 - 50)^2, \quad (27 - 25)^2$$

Der Standardfehler dieser quadratischen Abweichungen ist jeweils die erwartete Zellhäufigkeit, im Beispiel 25, 50, 25. Dies folgt aus einem mathematischen Argument, dessen Details hier zu weit führen würden. Damit erhalten wir folgende Teststatistik X^2 :

$$X^2 = \frac{(18 - 25)^2}{25} + \frac{(55 - 50)^2}{50} + \frac{(27 - 25)^2}{25} = 2.62$$

[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
Testen
[Regression](#)
[Testfragen](#)
[Index](#)

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite **109** von **100**

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

Teststatistik	$X^2 = \sum_{\text{Zellen}} \frac{(\text{Obs} - \text{Exp})^2}{\text{Exp}}$	χ^2 -Anpassungstest
---------------	---	--------------------------

Die Prüfverteilung ist eine χ^2 -Verteilung mit $(k - 1)$ Freiheitsgraden (siehe Abschnitt 3.4). Der p -Wert ist aber nur approximativ gültig (für grosse n wird er exakt).

Beispiel: $X^2 = 2.62 \implies \chi^2$ verteilt mit 2 Freiheitsgraden
oberes 5% Perzentil von $\chi^2_2 = 5.99$
 $2.62 < 5.99 \implies$ nicht signifikant

Der Versuch spricht also nicht gegen die Verteilung, die aus den genetischen Modellen abgeleitet wurde. Beachten Sie, dass es beim χ^2 -Test keine einseitige Alternative gibt, da die Vorzeichen der Abweichungen durch das Quadrieren verschwinden.

4.4.2. Testen auf Homogenität in Kontingenztafeln

Im Unterschied zu Abschnitt 4.4.1 vergleicht man nicht empirische Häufigkeiten mit theoretischen, sondern empirische Häufigkeiten von 2 oder mehr unabhängigen Gruppen miteinander, was in der Praxis viel häufiger ist.

Beispiel: Vergleich von Medikament A mit Medikament B an $n = 150$ Patienten.

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite 110 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite [111](#) von [100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Die klinische Beurteilung des Gesundheitszustandes ist dreistufig als sehr gut, gut bzw. schlecht gegeben.

Nach Randomisierung erhalten 80 Patienten Medikament *A* und 70 Patienten Medikament *B*. Die Daten werden dann in einer Kontingenztafel (Kreuztabelle) angeordnet:

	sehr gut	gut	schlecht	<i>n</i>
<i>A</i>	37 (<i>A</i> ₁)	24 (<i>A</i> ₂)	19 (<i>A</i> ₃)	80
<i>B</i>	17 (<i>B</i> ₁)	33 (<i>B</i> ₂)	20 (<i>B</i> ₃)	70
Total	54	57	39	150

H_0 : *A* und *B* sind gleich gut, d. h., die Wahrscheinlichkeiten für die Wirkung sind identisch.

$$p_{A_1} = p_{B_1} = p_1, \quad p_{A_2} = p_{B_2} = p_2, \quad p_{A_3} = p_{B_3} = p_3$$

Das Problem ist analog zu einem Zweistichproben–Problem bei kontinuierlichen Daten.

Das Testprinzip besteht wieder darin, in jeder Zelle die beobachtete (Obs) und erwartete (Exp) Anzahl zu vergleichen:

Teststatistik	$X^2 = \sum_{\text{Zellen}} \frac{(\text{Obs} - \text{Exp})^2}{\text{Exp}}$	χ^2 –Test auf Homogenität
---------------	---	--------------------------------



Die Prüfverteilung ist eine χ^2 -Verteilung mit $(r - 1) \times (c - 1)$ Freiheitsgraden.

r = Anzahl Zeilen in der Kreuztabelle (r für „rows“)

c = Anzahl Kolonnen in der Kreuztabelle (c für „columns“)

Im Medikamenten-Beispiel ist $r = 2$ (Medikament A, B) und $c = 3$ (Rating: sehr gut, gut, schlecht).

	sehr gut	gut	schlecht	n
A	37 (28.8)	24 (30.4)	19 (20.8)	80
B	17 (25.2)	33 (26.6)	20 (18.2)	70
total	54	57	39	150

In Klammern stehen die erwarteten Häufigkeiten, falls beide Medikamente gleich gut sind. Sie werden berechnet, indem die Ergebnisse aus der gesamten Stichprobe im Verhältnis 80:70 aufgeteilt werden. $\Rightarrow X^2 = 8.22$, $p = P[\chi_2^2 \geq 8.22] = 0.016 < 0.05$. Demnach sind die Medikamente A und B signifikant verschieden (mit Irrtumswahrscheinlichkeit $\alpha = 0.05$).

Datenanalyse

Man sollte beachten, dass im Falle eines gepaarten Versuchsplans (Medikament A und B an derselben Stichprobe erprobt) eine andere Statistik anzuwenden ist (McNemar-Test).

Der p -Wert des χ^2 -Tests ist nur für grosse n gültig. Allgemein sagt man, dass die Approximation für 2×2 Kontingenztafeln gut ist, wenn die erwarteten Häufigkeiten (Exp) in allen Zellen ≥ 5 sind. Ist diese Voraussetzung nicht gegeben, kann man **Fisher's exakten Test** anwenden. Für allgemeine $r \times c$ Tafeln genügt es für die Gültigkeit des χ^2 -Tests, dass alle erwarteten Häufigkeiten ≥ 3 sind.

[Home Page](#)

[Titelseite](#)

«

»

◀

▶

Seite 112 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



4.4.3. Test für Unabhängigkeit zweier Variablen

Im nächsten Kapitel 5 wird der Test auf Unabhängigkeit für kontinuierliche Daten (mittels Korrelation) behandelt. Hier wollen wir dasselbe für diskrete Variable untersuchen.

Problemstellung: An einer Stichprobe der Grösse n werden 2 diskrete Variable erhoben und in einer Kontingenztafel angeordnet. Sind die Merkmale unabhängig?

$$H_0: p_{ij} = p_i \times p_j \text{ für alle } i, j$$

Sie erinnern sich: Stochastische (statistische) Unabhängigkeit ist so definiert, dass Wahrscheinlichkeiten multipliziert werden können.

Beispiel: An 400 Kindern wird deren Händigkeit geprüft, und es wird geprüft, ob Vater und Mutter links- oder rechtshändig sind. Frage: Vererbt sich die Händigkeit?

Vater × Mutter	Kind		total
	rechts	links	
rechts, rechts	303 (295.8)	37 (44.2)	340
rechts, links	29 (33.1)	9 (4.9)	38
links, links	16 (19.1)	6 (2.9)	22
total	348	52	400

() = erwartet, falls unabhängig

H_0 : kein Zusammenhang („Händigkeit nicht genetisch bedingt“)

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 113 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
Testen
[Regression](#)
[Testfragen](#)
[Index](#)

Teststatistik: Obwohl wir eine andere Problemstellung haben, geht alles gleich wie beim Test zum Vergleich zweier kategorialer Stichproben (siehe Abschnitt [4.4.2](#)). Man vergleicht beobachtete und erwartete Zellhäufigkeiten und bildet die Summe der normierten Quadrate ($= X^2$). Diese ist dann wieder approximativ $\chi^2_{(r-1) \times (c-1)}$ -verteilt.

Im Beispiel ist $X^2 = 9.15$, $p = P[\chi^2_2 \geq 9.15] = 0.010 < \alpha = 0.05$. Wir gehen daher von einer Vererblichkeit der Händigkeit aus.

Datenanalyse

Da die erwartete Häufigkeit in 2 Zellen kleiner als 5 ist, sollte besser Fisher's exakter Test benutzt werden. Damit erhält man ebenfalls einen p -Wert von 0.010, das Ergebnis wird in diesem Beispiel also bestätigt.

Selbstkontrolle

4.5. Multiples Testen

Bisher sind wir davon ausgegangen, dass wir *einen* Test durchführen, *um eine Hypothese zu testen*. Für diesen Fall wollen wir die Irrtumswahrscheinlichkeit α – meist 5% – einhalten. Oft werden in der Praxis mehrere Tests durchgeführt. Dabei werden aber die Irrtumswahrscheinlichkeiten nicht mehr eingehalten, wie das folgende Beispiel zeigt:

Wenn ich 20 mal unabhängig voneinander etwas teste, wird im Mittel bei $\alpha = 5\%$ ein Resultat signifikant, auch wenn in allen 20 Fällen die Nullhypothese richtig ist.

Beispiel: Studie mit 4 diagnostischen Gruppen, 20 Variablen erhoben

Naiv vergleicht man die 4 Gruppen paarweise untereinander, eine Variable nach der anderen.

[Home Page](#)

[Titelseite](#)

« ►

◀ ►

Seite 114 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
Testen
[Regression](#)
[Testfragen](#)
[Index](#)

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite [115](#) von [100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

$\Rightarrow 120 (= 6 \times 20)$ paarweise Vergleiche möglich

$\Rightarrow 120$ statistische Tests möglich

H_0 : Kein einziger Unterschied zwischen den Gruppen

H_1 : Unterschied in mindestens einer Variablen

Sei $\alpha = 0.05$ die Irrtumswahrscheinlichkeit für jeden einzelnen Test der paarweisen Vergleiche. Falls H_0 gilt, d. h., kein einziger Unterschied, erhalten wir trotzdem im Mittel per Zufall $0.05 \times 120 = 6$ Ablehnungen von H_0 . Das bedeutet, dass wegen der vielen Tests die Irrtumswahrscheinlichkeit α nicht für das multiple Testproblem gilt.

Allgemein: Wir führen k unabhängige Tests auf nominellem 5% Niveau durch. Wie gross ist dann die Wahrscheinlichkeit, bei Gültigkeit von H_0 mindestens einen p -Wert $p < \alpha$ zu erhalten („effektives α “)?

k	nominelles α	effektives α
1	0.05	0.05
2	0.05	0.10
3	0.05	0.14
5	0.05	0.23
10	0.05	0.40
20	0.05	0.64
50	0.05	0.92

Das viele Testen führt also auf eine „ α -Inflation“.



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

Lösungen:

1. Multivariate statistische Verfahren wie Varianzanalyse.
2. Bonferroni–Korrektur (für kleine k !)

$$\Rightarrow \alpha[\text{Einzeltest}] = \frac{\alpha}{k}$$

Die Bonferroni–Korrektur ist konservativ, d. h. der Fehler 1. Art ist deutlich kleiner als α , womit die Trennschärfe sinkt.

3. In der Versuchsplanung werden wenige strikte Hypothesen zum Testen aufgestellt. Die Daten werden ansonsten deskriptiv ausgewertet.

[Home Page](#)

[Titelseite](#)



Seite 116 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
Testen
[Regression](#)
[Testfragen](#)
[Index](#)

4.6. Konfidenzintervalle (Vertrauensbereiche)

Bei Wiederholung einer Studie erhalten wir andere statistische Kennzahlen (vgl Abschnitt 3.7). Dies ist erklärbar durch die unterschiedlichen Stichproben, was notwendigerweise einen Zufallseffekt mit sich bringt. Man möchte diese Zufallsschwankungen in den statistischen Kennzahlen quantitativ fassen.

Da der wahre Kennwert θ (zum Beispiel $\theta = \mu, p$) nicht bekannt ist, und die Schätzung mit statistischer Ungenauigkeit behaftet ist: Gibt es ein Intervall, in dem θ mit hoher Wahrscheinlichkeit liegt? („Quantifizierung der Ungenauigkeit“)

Definition: Ein **95%-Konfidenzintervall** $[\hat{\theta}_u, \hat{\theta}_o]$ ist ein zufälliges Intervall, das den unbekannten, wahren Wert θ mit Wahrscheinlichkeit 95% enthält.

In Formeln heisst dies: $P[\hat{\theta}_u \leq \theta \leq \hat{\theta}_o] \geq 0.95$

Man kann auch allgemein $(1 - \alpha) \times 100\%$ Konfidenzintervalle definieren, konventionell wird aber $\alpha = 0.05$ gesetzt.

Bei wiederholten Experimenten liegt man demnach in $\alpha \times 100\%$ der Fälle falsch. Es wird offensichtlich, dass Konfidenzintervalle etwas mit dem Konzept des Signifikanztests zu tun haben, weshalb wir sie hier einführen.

Nachfolgend ein Beispiel für die Wichtigkeit von Konfidenzbereichen für die biomedizinische Literatur. Publiziert wurde in Lancet eine multizentrische Studie zum akuten Herzinfarkt. Die Ergebnisse stützen sich stark auf Konfidenzintervalle ab, so dass diese bereits in der Zusammenfassung erscheinen.

[Home Page](#)

[Titelseite](#)

«

»

◀

▶

Seite 117 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



[Einführung](#)

[Deskriptive Statistik](#)

[Wahrscheinlichkeit](#)

[Testen](#)

[Regression](#)

[Testfragen](#)

[Index](#)

[Home Page](#)

[Titelseite](#)

◀◀

▶▶

◀

▶

Seite **118** von **100**

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Summary

In 1985 an overview of clinical trials confirmed that patients treated within 6 h of the onset of symptoms of myocardial infarction benefited from thrombolytic therapy. Doubt remained about treatment later than this and this uncertainty prompted further randomised studies. The South American multicentre trial EMERAS is one of these.

4534 patients entering hospital up to 24 h after the onset of suspected acute myocardial infarction were randomised between intravenous streptokinase (SK) 1·5 MU and placebo, during the period January, 1988, to January, 1991. Once the results of ISIS-2 were known, only patients presenting more than 6 h after symptom onset were randomised. There was no significant difference in mortality during the hospital stay (269/2257 [11·9%] deaths among SK patients vs 282/2277 [12·4%] in controls). Among the 2080 patients presenting 7–12 h from symptom onset there was a non-significant trend towards fewer deaths with SK (11·7% SK vs 13·2% control; 14% [SD 12] reduction with 95% confidence interval [CI] of 33% reduction to 12% increase), whereas there was little difference among the 1791 patients presenting after 13–24 h (11·4% vs 10·7%; 8% [16] increase with a 95% CI of 20% reduction to 45% increase). These 95% CIs are wide and are consistent with the results of previous studies among patients presenting late after symptom onset.

The EMERAS results, though not conclusive on their own, do contribute substantially to accumulating evidence on the question of whether fibrinolytic therapy really does produce any worthwhile improvement in survival among such patients.

Lancet 1993; **342:** 767–72



Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

[◀](#) [▶](#)

[◀](#) [▶](#)

Seite [119](#) von [100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Wie man sieht, stehen Konfidenzintervalle in einem engen Zusammenhang zur Prüfung von Hypothesen. In der obigen Studie konnte kein Unterschied in der Sterblichkeit nach einem Herzinfarkt zwischen den Gruppen mit und ohne thrombolytischer Therapie nachgewiesen werden (“no significant difference”). Das heisst aber nicht, dass es keinen Unterschied gibt. Die Konfidenzintervalle zeigen, dass es möglich ist, dass die Therapie klinisch relevante Verbesserungen bis zu 33% bringt. Das muss dann allerdings erst noch durch neue Studien belegt werden, da natürlich auch die andere Grenze des Konfidenzintervalls (Verschlechterung um 12%) möglich wäre. Der Bezug zum Hypothesentesten ist genau gesagt so, dass ein Resultat mit $\alpha = 5\%$ signifikant ist, wenn der Wert der Nullhypothese ausserhalb des 95%-Konfidenzintervalls liegt.

4.6.1. Konfidenzintervall für μ bei bekanntem σ

Aus tutoriellen Gründen sei eine etwas vereinfachte Situation vorausgesetzt.

- Eine physikalische Grösse soll mit einem Messgerät mit **bekannter** Streuung $\sigma = \sigma_0$ bestimmt werden.
- Die Messungen x_1, \dots, x_n seien verteilt wie $\mathcal{N}(\mu, \sigma_0^2)$, d. h. sie sind normalverteilt mit Erwartungswert μ und Varianz σ_0^2 .
- Gesucht ist das $(1 - \alpha)$ -Konfidenzintervall für μ .

(i) Der Mittelwert \bar{x} ist verteilt als $\mathcal{N}(\mu, \frac{\sigma_0^2}{n})$, da \bar{x} die Varianz $\frac{\sigma_0^2}{n}$ hat.

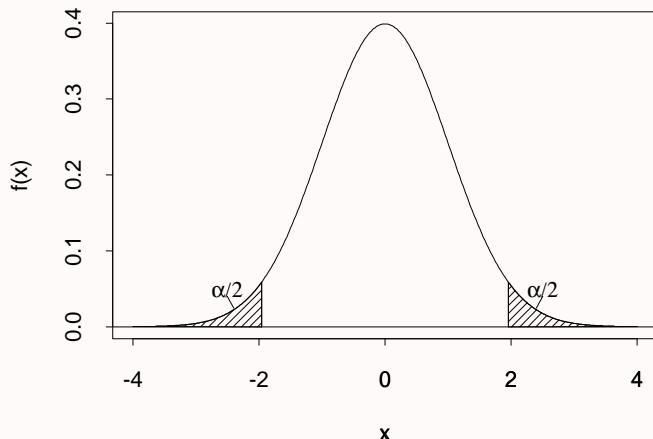
(ii) Dann ist $\frac{\bar{x} - \mu}{\sigma_0 / \sqrt{n}}$ verteilt als $\mathcal{N}(0, 1)$.



(iii) $(1 - \alpha)$ -Konfidenzintervall für Erwartungswert μ bei bekanntem σ_0 :

$$\bar{x} - z_{1-\alpha/2} \frac{\sigma_0}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{1-\alpha/2} \frac{\sigma_0}{\sqrt{n}}$$

Begründung für die Formel des Konfidenzintervalls (für Interessierte!):



In der Abbildung ist $\alpha = 0.05$ angenommen, mit $z_{\alpha/2} = -1.96$ und $z_{1-\alpha/2} = 1.96$.

- Nach Definition ist $P \left[z_{\alpha/2} \leq \frac{\bar{x} - \mu}{\sigma_0/\sqrt{n}} \leq z_{1-\alpha/2} \right] = 1 - \alpha$.
- Da die Normalverteilung symmetrisch ist, folgt $z_{\alpha/2} = -z_{1-\alpha/2}$.
- Auflösen nach μ liefert das $(1 - \alpha)$ -Konfidenzintervall:

$$-z_{1-\alpha/2} \frac{\sigma_0}{\sqrt{n}} \leq \bar{x} - \mu \leq z_{1-\alpha/2} \frac{\sigma_0}{\sqrt{n}}$$

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 120 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



$$\implies \bar{x} - z_{1-\alpha/2} \frac{\sigma_0}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{1-\alpha/2} \frac{\sigma_0}{\sqrt{n}}$$

Bemerkungen:

1. Das Konfidenzintervall für μ liegt symmetrisch um \bar{x} , was plausibel ist. Seine Länge wird kleiner (man wird „sicherer“), wenn die Standardabweichung σ_0 sinkt oder die Stichprobe grösser wird.
2. Jedes Konfidenzintervall ist zufällig, in diesem Beispiel ist es zufällig durch seine Lage bei \bar{x} .
3. Die Annahme, dass die Standardabweichung der Stichprobe $\sigma = \sigma_0$ bekannt sei, ist im allgemeinen unrealistisch, σ muss geschätzt werden (siehe unten).

Numerisches Beispiel zur Illustration der Abhängigkeit des Konfidenzintervalls für μ von α und n : $\bar{x} = 0.2$, $\sigma_0 = 0.1$

n	α		
	0.05	0.01	0.001
10	[0.14, 0.26]	[0.12, 0.28]	[0.10, 0.30]
50	[0.17, 0.23]	[0.16, 0.24]	[0.15, 0.25]
200	[0.19, 0.21]	[0.18, 0.22]	[0.18, 0.22]

Wir sehen eine „Unschärferelation“: Je sicherer wir sein wollen, dass der wahre Wert im Intervall liegt, desto länger wird dann das Konfidenzintervall.

Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 121 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



4.6.2. Konfidenzintervall für μ bei unbekanntem σ^2

Beispiel: Konfidenzintervall für den Erwartungswert μ der Anzahl T_4 -Zellen auf der Basis von $n = 20$ Hodgkin-Patienten

Problem: Die Daten sind rechtsschief, deutlich nicht normalverteilt.

Lösung: 1. Daten logarithmieren

2. logarithmierte Daten als approximativ normalverteilt betrachten

log- T_4 -Zellanzahl: $\bar{x} = 6.49$, $s = 0.71$

Die Idee ist nun, die Statistik $\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ von vorhin weiter zu verwenden, aber σ durch

s zu ersetzen: $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$

Die Statistik t wäre standardnormalverteilt, falls wir σ und nicht s im Nenner hätten. So gilt: t ist t -verteilt, (siehe 4.2.1), „wackelt draussen mehr“, da s ein Schätzer und keine feste Zahl wie σ ist („mehr Wahrscheinlichkeit in den Extrembereichen“).

Dann ergibt sich analog:

(1 - α)-Konfidenzintervall für μ bei unbekanntem σ :

$$\bar{x} - t_{1-\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{1-\alpha/2} \frac{s}{\sqrt{n}}$$



Einführung
Deskriptive Statistik
Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)

«

»

◀

▶

Seite 123 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

- Die Grösse $t_{1-\alpha/2}$ ist das $(1 - \alpha/2) \times 100\%$ -Perzentil einer t -Verteilung (mit $n - 1$ Freiheitsgraden).
- Das Intervall ist symmetrisch um \bar{x} .
- Die Länge des Intervalls ist abhängig von n und s , d. h. die Lage und die Länge des Intervalls sind jetzt zufällig.

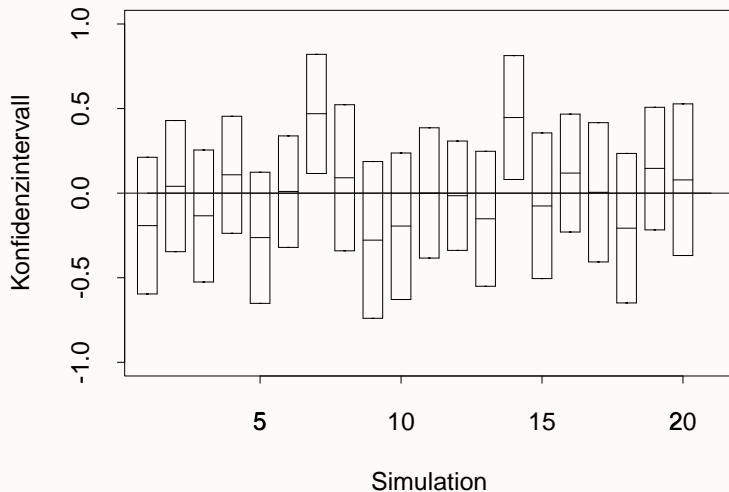
Für die $\log T_4$ -Zellanzahl erhalten wir folgende $(1 - \alpha)$ -Konfidenzintervalle:

$$\alpha = 0.05 \quad 6.14 \leq \mu \leq 6.84$$

$$\alpha = 0.01 \quad 5.90 \leq \mu \leq 6.99$$

$$\alpha = 0.001 \quad 5.76 \leq \mu \leq 7.22$$

Nach Konstruktion enthält ein Konfidenzintervall im Mittel in $(1 - \alpha) \times 100\%$ der Fälle den wahren Wert. Auf dem Computer wurden 20 Stichproben von $n = 25$ normalverteilten „Pseudo-Zufallszahlen“ simuliert ($\mu = 0$, $\sigma^2 = 1$). Für jede Stichprobe wurde das 95%-Konfidenzintervall für μ nach obiger Formel konstruiert. Nicht nur die Lage, sondern auch die Länge ändert sich beträchtlich. Im Mittel würden wir auf 20 Stichproben in einem Fall erwarten, dass der wahre Wert nicht im Konfidenzintervall liegt. Hier enthalten zufälligerweise zweimal die Konfidenzintervalle den wahren Wert nicht. Deswegen müssen wichtige Ergebnisse repliziert werden, denn nach der Produktregel für Wahrscheinlichkeiten ist es äusserst unwahrscheinlich, in 2 unabhängigen Experimenten zufällig extreme Resultate zu erhalten.



Weitere Simulationen finden Sie hier:

[Animation](#)

4.6.3. Konfidenzintervall für relative Häufigkeiten

Man kann für praktisch alle interessierenden Größen Konfidenzintervalle berechnen. Wichtig ist z. B. ein Konfidenzintervall für eine wahre relative Häufigkeit, z. B. eine Prävalenz. (Der Schätzwert für p ist $\hat{p} = k/n$, wenn z. B. bei einer Prävalenzschätzung k von n Personen eine Krankheit haben.) Die untere Grenze

[Home Page](#)

[Titelseite](#)

◀◀

▶▶

◀

▶

Seite 124 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



\hat{p}_u und die obere Grenze \hat{p}_o eines Konfidenzintervalls werden relativ kompliziert berechnet, so dass wir nur Beispiele geben:

Beispiel: Man beobachtet $n = 20$ Geburten, $k = 7$ mal wird ein Knabe geboren.

$$\hat{p} = 7/20 = 0.35$$

Welches ist der 95%-Vertrauensbereich für p ?

Datenanalyse

- Der Vertrauensbereich ist **weit**, denn n ist klein.
- Der Vertrauensbereich schliesst 0.5 mit ein (gleiche Häufigkeit der Knaben- und Mädchengeburten mit $p = 0.5$ ist möglich).

Das folgende reale Beispiel zeigt, dass die Annahme gleicher Häufigkeit von Knaben- und Mädchengeburten doch nicht plausibel ist:

1950 – 1970: 1'944'700 Geburten in der Schweiz, davon 997'600 Knaben.

Ist es Zufall, dass der Schätzwert $\hat{p} = 0.5130$ für eine Knabengeburt von $p = 0.5$ abweicht? Das 99%-Konfidenzintervall ist $(0.5121, 0.5139)$; d. h. das Konfidenzintervall ist bei diesem grossen n sehr eng. Es schliesst 0.5 (gleiche Wahrscheinlichkeit einer Knabengeburt) deutlich nicht mit ein. Damit ist mit $\alpha = 1\%$ nachgewiesen, dass Knabengeburten häufiger sind. Spekulativ könnte man einen Mechanismus der Natur postulieren, der für mehr Knabengeburten sorgt, um ihre höhere Mortalität im ersten Lebensjahr zu kompensieren.

Datenanalyse

Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

[◀](#) [▶](#)

[◀](#) [▶](#)

Seite 125 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



5. Korrelation und Regression

Bisher wurde die statistische Analyse auf der Basis **einer** Messgrösse behandelt, d. h. univariat (Ausnahme: χ^2 -Test). Jetzt geht es darum, Zusammenhänge zwischen **zwei** oder mehr stetigen Variablen (**bivariate**, **multivariate** Daten) zu untersuchen.

Mögliche Fragestellungen sind:

- Besteht eine Beziehung zwischen den Variablen ?
- Wie stark ist die Beziehung ?
- Welche Form hat die Beziehung ?
- Kann eine Variable von primärem Interesse aus der Beobachtung anderer Variabler vorhergesagt werden ?



By permission of Johnny Hart and Field Enterprises, Inc.

[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
[Testen](#)
Regression
[Testfragen](#)
[Index](#)

[Home Page](#)

[Titelseite](#)



Seite 126 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 127 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

5.1. Bivariate Daten

Man beobachtet zwei **stetige** Variable (x, y) an der selben Beobachtungseinheit und erhält **paarweise** Beobachtungen $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$.

Beispiel: Zusammenhang zwischen Gewicht und Grösse.

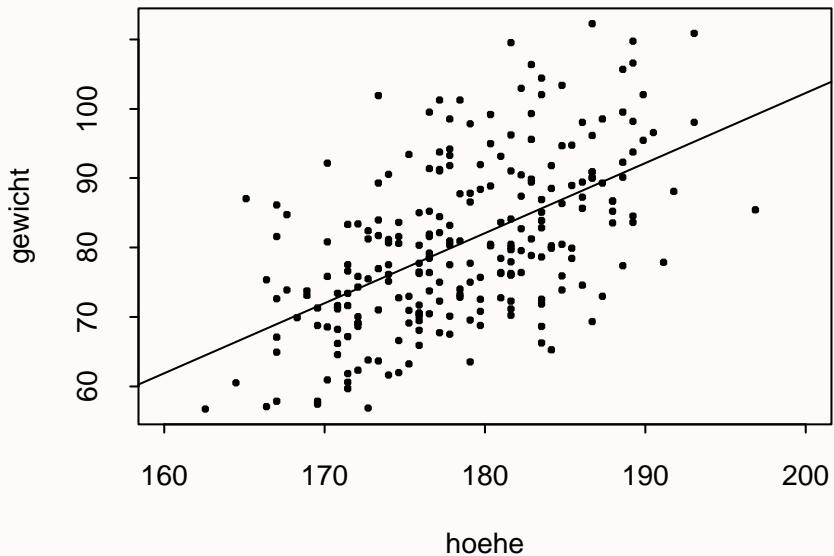
Jede Korrelations- oder Regressionsanalyse sollte mit dem Ausdruck des **Scattergramms** (Streudiagramm) begonnen werden, Beispiel: Nächste Seite.

Die folgenden Daten, die wir in diesem Kapitel analysieren wollen, wurden von R.W. Johnson (Carleton College, Northfield, MN) zur Illustration der Regressionsanalyse zur Verfügung gestellt. In dieser Stichprobe wurden bei 252 Männern der Prozentsatz an Körperfett, Alter, Gewicht und 10 Körperumfangsmasse bestimmt. Dabei wurde die Körperflechte mittels einer Wägung unter Wasser relativ aufwendig exakt gemessen und mittels gewisser Eichformeln in prozentuale Körperfett umgerechnet. Ein hoher Anteil an Körperfett stellt ein Gesundheitsrisiko dar; aus praktischen Gründen möchte man das Körperfett aus einfach zu ermittelnden Körpermassen approximativ bestimmen, wofür sich die Regressionsmethode eignet.

Es erweist sich tatsächlich, dass man das Körperfett bei Männern mittels multipler Regression bequem aus leicht zu bestimmenden Parametern schätzen kann (siehe Abschnitt 5.6).



Das folgende Scattergramm zeigt den Zusammenhang von Gewicht und Körpergrösse in der von Ausreisern befreiten Stichprobe von 241 Männern.



Man erhält so einen visuellen Eindruck vom Zusammenhang zwischen den Variablen. In diesem Fall erhärtet sich die Vermutung, dass Gewicht und Körpergrösse zusammenhängen.

Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite [128](#) von [100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
[Testen](#)
Regression
[Testfragen](#)

[Index](#)

[Home Page](#)

[Titelseite](#)

[◀](#)

[▶](#)

[◀](#)

[▶](#)

Seite [129](#) von [100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

5.2. Korrelation und ihre Eigenschaften

Eine Korrelation (auch Produkt–Moment Korrelation oder Pearson Korrelation genannt) misst, wie stark der **lineare** Zusammenhang, die lineare Übereinstimmung zwischen x und y ist.

Der Korrelationskoeffizient wird wie folgt berechnet:

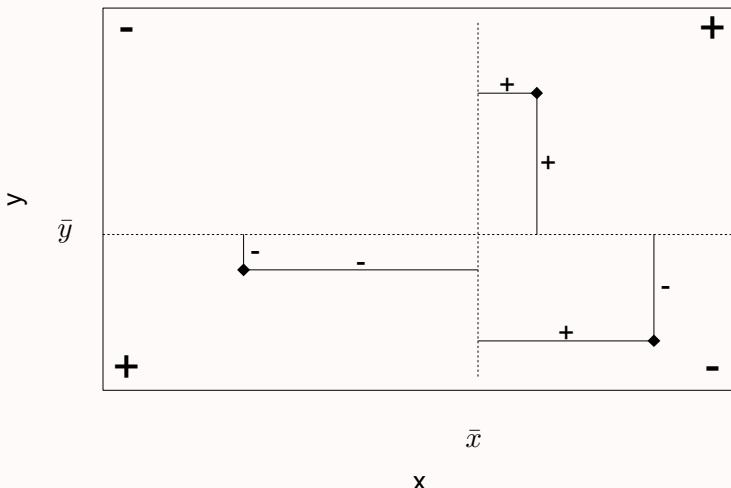
Kovarianz: $\text{Cov}(x, y) = s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$

Varianz: $s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$

Korrelation:
$$r = \frac{s_{xy}}{s_x s_y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$



Die Plausibilität des Zählers kann man anhand von 3 Datenpunkten graphisch verstehen:



Rechts oben und links unten von den Mittelwerten der Variablen geben Beobachtungen einen positiven Beitrag zur Kovarianz und damit zur Korrelation; links oben und rechts unten einen negativen Beitrag. Damit ist die Korrelation ungefähr 0, wenn sich die Beobachtungen auf alle 4 Quadranten verteilen. Sie wird deutlich positiv, wenn sich die Beobachtungen um eine Gerade mit positiver Steigung gruppieren, d. h. im Quadranten I und III. Bei Messungen vorwiegend im Quadranten II und IV wird sie negativ.

[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
[Testen](#)
Regression
[Testfragen](#)
[Index](#)

[Home Page](#)

[Titelseite](#)

[«](#)

[»](#)

[◀](#)

[▶](#)

Seite 130 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Plausibilität des Nenners:

Die Korrelation r wird durch diese Normierung mit den Standardabweichungen von den Masseinheiten unabhängig und ist damit besser interpretierbar.

Eigenschaften:

$$-1 \leq r \leq 1$$

$r = 1$ \implies deterministisch positiver linearer Zusammenhang zwischen x und y

$r = -1$ \implies deterministisch negativer linearer Zusammenhang zwischen x und y

$r = 0$ \implies kein linearer Zusammenhang

Allgemein gilt:

- Das Vorzeichen gibt die Richtung des Zusammenhangs an.
- Die Grösse gibt die Intensität des Zusammenhangs wieder.

[Home Page](#)

[Titelseite](#)

[◀◀](#)

[▶▶](#)

[◀](#)

[▶](#)

Seite **131** von **100**

[Zurück](#)

[Vollbild](#)

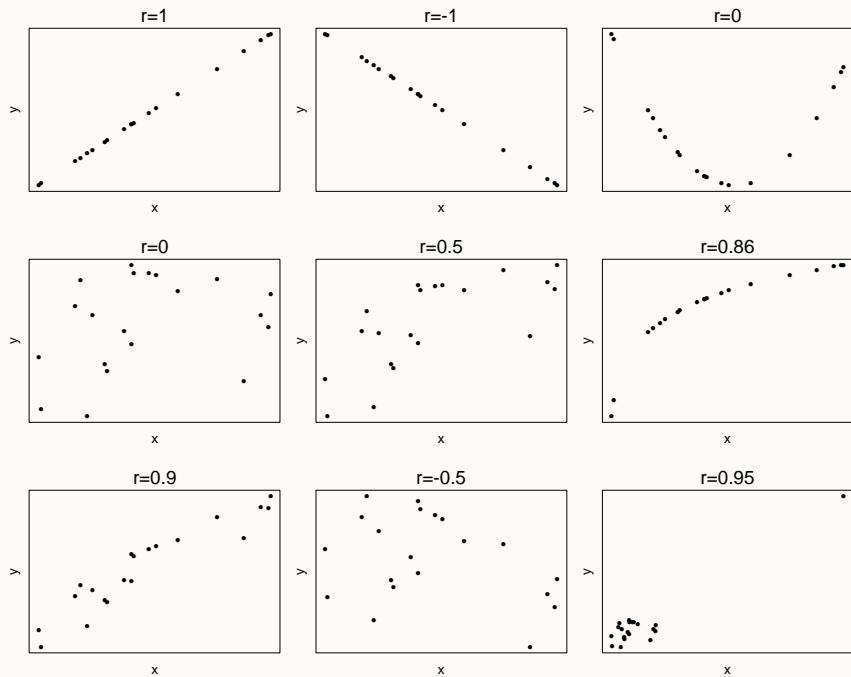
[Schließen](#)

[Beenden](#)



Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

Die folgenden Illustrationen zeigen Daten mit unterschiedlich starkem Zusammenhang. Man beachte, dass die Korrelation nur den **linearen** Zusammenhang misst, unter Umständen kann die Korrelation für gewisse deterministische nichtlineare Zusammenhänge Null werden. Werte, die in der x - und y -Richtung extrem liegen, können eine starke Korrelation vortäuschen.



[Home Page](#)

[Titelseite](#)

[◀◀](#)

[▶▶](#)

[◀](#)

[▶](#)

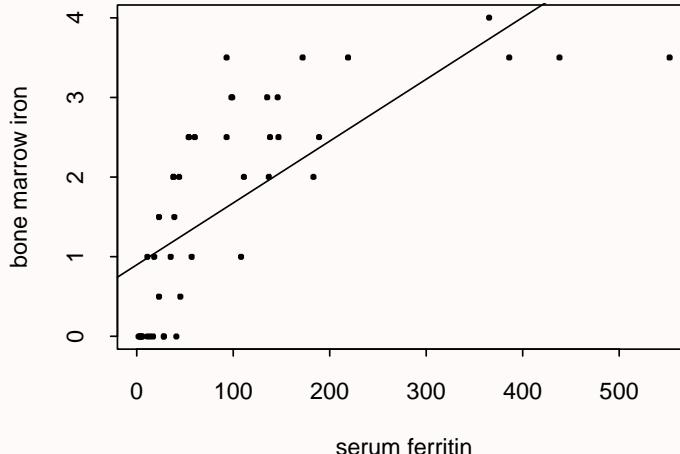
[Seite 132 von 100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Beispiel: In einer Studie an 45 anämischen Patienten (Baumann Kurer et al., British J. Haematology, 1995) wurde untersucht, ob die invasive Messung des Eisengehaltes im Knochenmark durch eine einfache Blutprobe (Ferritingehalt im Blutserum) ersetzt werden kann.

Die Stichprobenkorrelation ist $r = 0.72$. Man sieht aber, dass die Werte nicht gleichmässig um eine Gerade streuen, eine Gerade nicht zu diesen Daten passt. Die Korrelation wird stark durch weit aussen liegende Werte oben rechts bestimmt.

- Da lineare Zusammenhänge am einfachsten zu behandeln sind, sollte man versuchen, durch Transformation auf einen linearen Zusammenhang zu kommen.

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 133 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite 134 von 100

[Zurück](#)

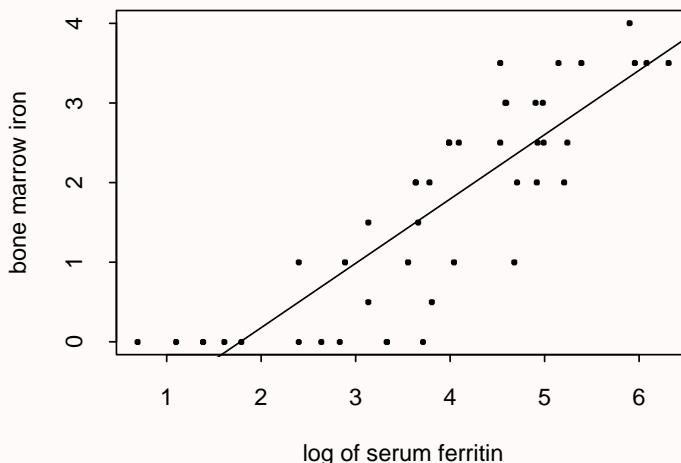
[Vollbild](#)

[Schließen](#)

[Beenden](#)

- Da weit vom Mittelwert entfernte Beobachtungen die Korrelation stark beeinflussen, sollte man versuchen, die beiden Variablen so zu transformieren, dass sie annähernd normalverteilt sind.

Häufig erfüllen Transformationen die beiden Forderungen gleichzeitig.



Im Beispiel ist die Stichprobenkorrelation nach der log–Transformation des Serum Ferritins $r = 0.85$, die Daten streuen gleichmässig um eine Gerade (diese heisst Regressionsgerade, siehe unten).



Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

- Da lineare Zusammenhänge am einfachsten zu behandeln sind, sollte man versuchen, durch Transformation auf einen linearen Zusammenhang zu kommen.
- Da weit vom Mittelwert entfernte Beobachtungen die Korrelation stark beeinflussen, sollte man versuchen, die beiden Variablen so zu transformieren, dass sie annähernd normalverteilt sind.
- Häufig erfüllen Transformationen die beiden Forderungen gleichzeitig, wie in diesem Beispiel.

Hier können Sie testen, ob Sie das Prinzip des Korrelationskoeffizienten verstanden haben:

[Animation](#)

[Home Page](#)

[Titelseite](#)



Seite 135 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



5.3. Tests auf linearen Zusammenhang und Konfidenzintervalle

Als nächstes soll geprüft werden, ob überhaupt ein linearer Zusammenhang zwischen den beiden Variablen statistisch nachzuweisen ist.

Nullhypothese: Die wahre Korrelation ρ ist gleich 0 („kein Zusammenhang“).

Annahme: (x, y) gemeinsam **normalverteilt**

Die folgende Testgrösse T ist mathematisch-statistisch begründet. Sie folgt einer t -Verteilung mit $(n - 2)$ Freiheitsgraden.

Teststatistik:

$$T = r \sqrt{\frac{n - 2}{1 - r^2}} \sim t_{n-2}$$

Sie ist insofern plausibel, als T betragsmäßig klein wird, wenn r nahe bei 0 liegt (=Nullhypothese) und immer grösser wird mit wachsender Grösse des Betrags von r (wo die Nullhypothese immer unplausibler wird).

Beispiel: Zusammenhang von Gewicht und Körpergrösse bei Männern.

$$n = 241, \quad r = 0.55$$

$$\Rightarrow T = 7.9 > t_{239;0.975} = 1.97, \quad p < 0.0001$$

Es besteht also ein signifikanter Zusammenhang zwischen Körpergrösse und Gewicht. Die Angabe eines Konfidenzintervalls ist für die Korrelation noch wichtiger als für die bisherigen statistischen Kennwerte. Mit etwas Erfahrung vermittelt die Angabe von n und (\bar{x}, s) auch ein Gefühl für die Variabilität dieser Kennwerte, was aber bei der Korrelation nicht der Fall ist. Das **Konfidenzintervall** liefert ein Intervall, in dem die wahre Korrelation mit grosser Wahrscheinlichkeit liegt.

[Home Page](#)

[Titelseite](#)

«

»

◀

▶

Seite 136 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



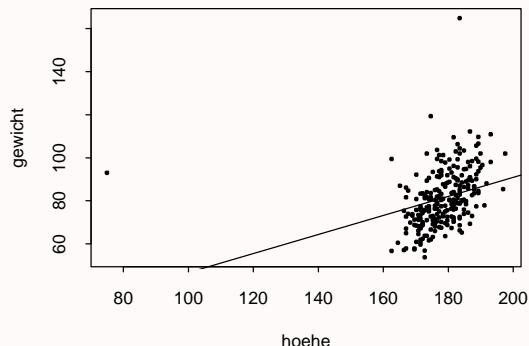
Im Beispiel ist ρ im Intervall $(0.46, 0.64)$ mit Wahrscheinlichkeit $1 - \alpha = 0.95$ (95%-Konfidenzintervall für ρ). Konfidenzintervalle für die Korrelation können ziemlich gross werden, wenn die Stichprobe mässigen Umfang hat.

5.4. Ausreisser und Gefahren der Korrelationsrechnung

- Wie wirken sich Ausreisser aus, und wie behandle ich sie ?
- Wie kann man bei Nicht-Normalverteilung die Korrelation schätzen und testen ?

Beispiel: Körpergrösse und Masse von Männern

In der vollen Stichprobe von 252 Männern sah das Scattergramm vor der Elimination von Ausreisern so aus:



Wir erhalten $r = 0.31$ und $p = 0.0001$. Die Ausreisser senken also die Korrelation von 0.55 auf 0.31, eine deutliche Verfälschung der Realität. Negativ wirkt sich

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite 137 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
[Testen](#)
Regression
[Testfragen](#)
[Index](#)

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite [138](#) von [100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

vor allem der unsinnige Wert von knapp 80 cm für die Körpergrösse bei einem Körpergewicht von über 90 kg aus. (Vermutlich handelt es sich um einen Tippfehler.)

Spearmans Rangkorrelation

Ähnlich wie wir früher auf Rangverfahren zurückgegriffen haben, um von der Normalverteilungsannahme loszukommen (Beispiel Mann–Whitney–Test), stützen wir uns auch hier auf die Rangordnung der Daten ab, um robustere Resultate zu erhalten.

Vorgehen:

1. Man bringt x_1, \dots, x_n und y_1, \dots, y_n getrennt in Rangreihen.
2. Man korreliert die Ränge miteinander anstatt die Zahlen selber.

Dadurch wird der Einfluss von Ausreisern begrenzt. Im Beispiel ergibt sich jetzt Spearmans Rangkorrelationskoeffizient $r_S = 0.52$ und $p < 0.0001$ (bei den korrekten Daten: $r_S = 0.55$, $p < 0.0001$). Rangkorrelationen können die Daten nicht „reparieren“, aber die Auswirkungen falscher oder atypischer Werte mildern.



[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite 139 von 100

[Zurück](#)

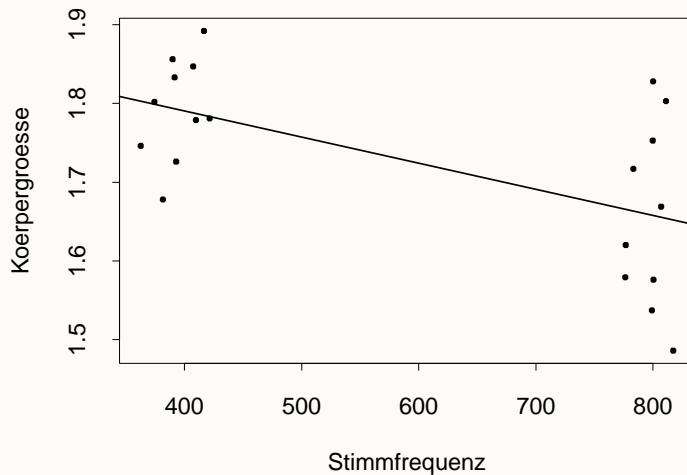
[Vollbild](#)

[Schließen](#)

[Beenden](#)

Gefahren der Korrelations-Rechnung

- Bei 10 Variablen gibt es 45 mögliche Korrelationen. Man muss sich demnach davor hüten, einzelne signifikante Korrelationen überzubewerten (siehe Problem des multiplen Testens).
- Heterogenitätskorrelation: Am Beispiel von Stimmfrequenz und Körpergrösse bei Männern und Frauen sieht man eine negative Korrelation ($r = -0.60$, $p = 0.006$), obwohl die Werte sowohl für die Gruppe der Männer als auch für die der Frauen unkorreliert sind.





3. Allgemeine Trends führen zu Scheinkorrelationen über die Zeit: Der Preis von Benzin und die Scheidungsrate korrelieren, da beide einen Zeittrend aufweisen.
4. Triviale Korrelationen: Wenn man die Körpergrösse mit 12 (= x) und mit 20 Jahren (= y) misst, müssen die Werte gut korrelieren, da ja $y = x + z$ gilt, wobei z der Zuwachs von 12 bis 20 Jahren ist.
5. Konfundierung durch 3. Variable: Die Anzahl der Störche und Geburten in einem Kanton korreliert stark („Bringt der Storch die Kinder?“). Die konfundierende Variable, die beide Grössen gleichsinnig beeinflusst, ist hier die Grösse des Kantons.

Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

◀◀

▶▶

◀

▶

Seite 140 von 100

[Zurück](#)

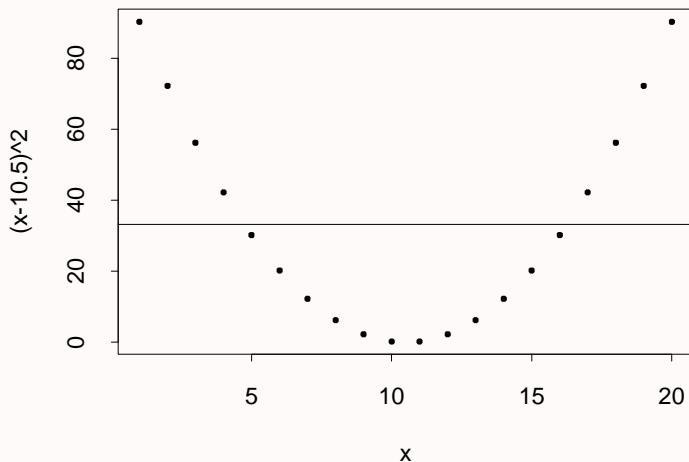
[Vollbild](#)

[Schließen](#)

[Beenden](#)



6. nichtlineare Zusammenhänge:



Hier erhält man $r = 0$, obwohl ein deterministischer aber quadratischer Zusammenhang deutlich ist.

7. Extreme Datenpunkte: Das frühere grafische Beispiel (siehe 5.1) mit einem gleichsinnig in der x - und der y -Richtung extrem verschobenem Wert zeigt, dass eine grosse Korrelation (im Beispiel 0.95) durch einzelne Werte vorgetäuscht werden kann, ohne dass ein allgemeiner Zusammenhang besteht. Aber auch extreme Datenpunkte in der x - oder y -Richtung können die Korrelation übermäßig beeinflussen.



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)



[Seite 142 von 100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

5.5. Einfache lineare Regression

- Die einfache Regressionsanalyse ist die statistische Analyse der Wirkung einer stetigen Variablen x auf eine andere Variable y . Die Beziehung ist also im Unterschied zur ungerichteten Korrelationsanalyse gerichtet.

x = unabhängige Variable, erklärende Variable, Prädiktor (oft nicht zufällig; Zeit, Alter, Messpunkt)

y = abhängige Variable, erklärte Variable, Zielvariable, Outcome, Response

Ziel: Nicht nur die Stärke und Richtung (\nearrow , \searrow) des Zusammenhangs soll bestimmt werden, sondern es soll ein quantitatives Gesetz formuliert werden: Wie ändert sich y , wenn x sich ändert ?



[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
[Testen](#)
Regression
[Testfragen](#)
[Index](#)

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 143 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Beispiel: Das Körpergewicht ist ein naheliegendes Mass für Übergewicht. Wie wir aber gesehen haben (Scattergramm, Abschnitt 5.1), hängt es von der Körpergrösse ab und ist demnach als Mass für Übergewicht nicht geeignet. Die folgende Regressionsgleichung quantifiziert dies.

Regression: $y = \text{Gewicht}$, $x = \text{Höhe}$, $n = 241$

$$\hat{y} = -99.7 + 1.01 \times x, \quad r^2 = 0.31, \quad p < 0.0001$$

Wie schwer sind Männer? Die beste Voraussage ist $\bar{y} = 80.7 \text{ kg}$. Die Standardabweichung der Messwerte um diesen Vorhersagewert beträgt $SD = s_y = 11.8 \text{ kg}$.

Wie schwer sind Männer von 175 cm? Die Zusatzinformation der Körpergrösse ändert die Vorhersage des Gewichts, indem der Wert der Regressionsgeraden für $x = 175 \text{ cm}$ genommen wird: $\hat{y} = -99.7 + 1.01 \times 175 = 77.0 \text{ kg}$. Die Standardabweichung der Messwerte um diesen Vorhersagewert beträgt $s_e = 9.8 \text{ kg}$. Wir erhalten demnach exaktere Aussagen über das Gewicht, wenn wir die Körpergrösse berücksichtigen. Allgemein gilt, dass wir exaktere Ergebnisse erhalten, wenn wir über ein Regressionsmodell wichtige Einflussgrössen berücksichtigen.



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)

[◀◀](#)

[▶▶](#)

[◀](#)

[▶](#)

Seite [144](#) von [100](#)

[Zurück](#)

[Vollbild](#)

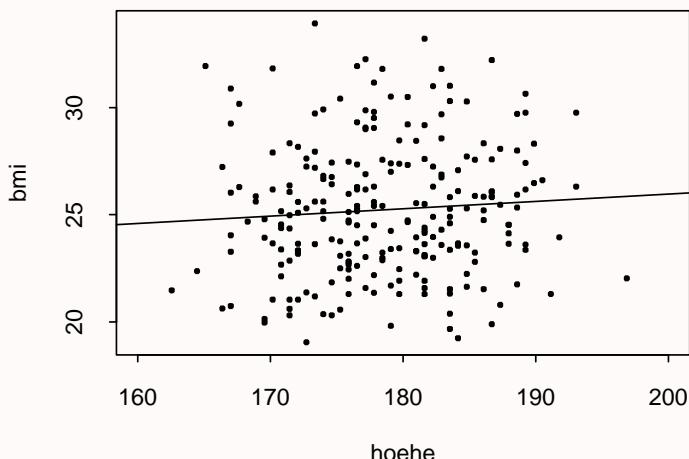
[Schließen](#)

[Beenden](#)

In der Klinik wird oft der „body mass index“ (BMI = Gewicht/Höhe²) als Mass für Übergewicht benutzt. Die Frage ist, ob der BMI tatsächlich von der Körpergrösse nicht beeinflusst ist.

Regression $y = \text{BMI} = \text{Gewicht} / \text{Höhe}^2$, $x = \text{Höhe}$

$$\hat{y} = 19.2 + 0.034 \times x, \quad r^2 = 0.005, \quad p = 0.27$$



Der BMI ist nicht oder nur wenig mit der Körpergrösse korreliert. Damit liefert er ein einfaches Mass für Übergewicht.



5.5.1. Statistisches Modell der Regression

$$y_i = f(x_i) + \varepsilon_i \quad i = 1, \dots, n$$

f = Regressionsfunktion, „wahrer Verlauf“

ε_i = unbeobachtbare, zufällige Schwankungen (Fehler oder Rauschen). Die Residuen ε_i schwanken um 0 (Mittelwert der $E_i = 0$) und haben die Varianz σ^2 .

Das Problem der Bestimmung von f vereinfacht sich sehr, wenn wir f als eine lineare Funktion annehmen („lineare Regression“).

$$f(x) = a + b x$$

Damit sind nur noch der Achsenabschnitt (intercept) a und die Steigung (slope) b der Geraden $a + b x$ unbekannt und müssen aus den Daten bestimmt („geschätzt“) werden.

Hier können Sie selbst Daten eingeben und die Regressionsanalyse rechnen:

[Animation](#)

[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
[Testen](#)
Regression
[Testfragen](#)
[Index](#)

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite **145** von **100**

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)



[Seite 146 von 100](#)

[Zurück](#)

[Vollbild](#)

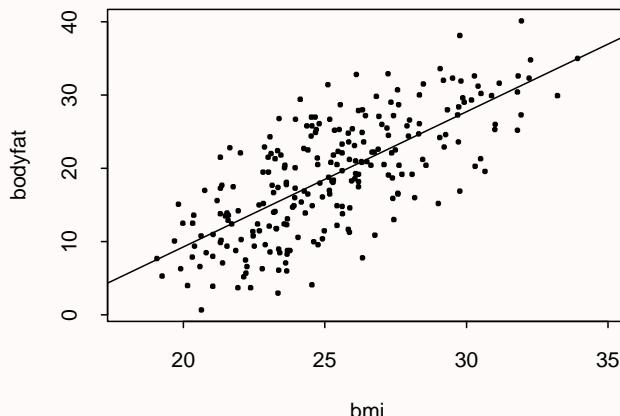
[Schließen](#)

[Beenden](#)

Beispiel: Apriori sind sowohl prozentuales Körperfett als auch der BMI als Masse für Übergewicht bei Männern von Interesse.

$x = \text{BMI} (\text{in } \text{kg}/\text{m}^2)$

$y = \text{Körperfett} (\text{in } \%)$



Geradengleichung: $\text{bodyfat} = -27.6 + 1.84 \times \text{BMI}$, $r^2 = 0.52$, $p < 0.0001$

Interpretationen:

1. Männer mit einem BMI von $25 \text{ kg}/\text{m}^2$ haben im Mittel 18 % Körperfett.
2. Wenn sich der BMI um $1 \text{ kg}/\text{m}^2$ erhöht, resultieren im Mittel 2 % mehr Körperfett.



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite [147](#) von [100](#)

[Zurück](#)

[Vollbild](#)

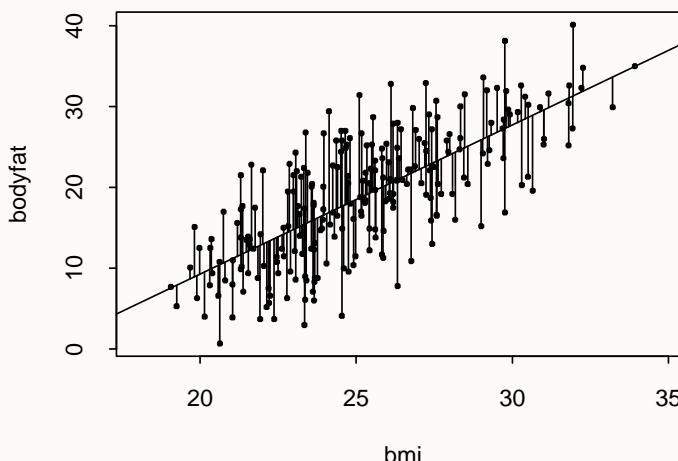
[Schließen](#)

[Beenden](#)

5.5.2. Die Methode der kleinsten Quadrate

Der Anpassung der Gerade an die Daten ist intuitiv dann am besten, wenn die Abstände der Messpunkte von der zu bestimmenden Geraden im Mittel klein sind („method of least squares“).

- Übliche Methode zur Schätzung von a und b : Es werden die vertikalen Abstände zur Geraden betrachtet, die zu bestimmen ist, und zwar — wie bei der Varianz — in der quadrierten Form.



[Animation](#)

[Selbstkontrolle](#)



Sei \hat{y}_i der Wert der geschätzten Regressionsgerade ($= \hat{a} + \hat{b} x_i$) bei einem Wert x_i .

Wähle die Schätzung der Parameter so, dass die quadratische Abweichung

$$S(\hat{a}, \hat{b}) = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

minimal wird!

Die resultierenden Schätzwerte für a und b sind:

$$\text{Steigung: } \hat{b} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = r \frac{s_y}{s_x}$$

$$\text{Achsenabschnitt: } \hat{a} = \bar{y} - \hat{b} \bar{x}$$

Man erhält diese Formeln über die Lösung eines Systems zweier linearer Gleichungen („Normalgleichungen“) in \hat{a} und \hat{b} .

[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
[Testen](#)
Regression
[Testfragen](#)
[Index](#)

[Home Page](#)

[Titelseite](#)

[◀◀](#)

[▶▶](#)

[◀](#)

[▶](#)

Seite **148** von **100**

[Zurück](#)

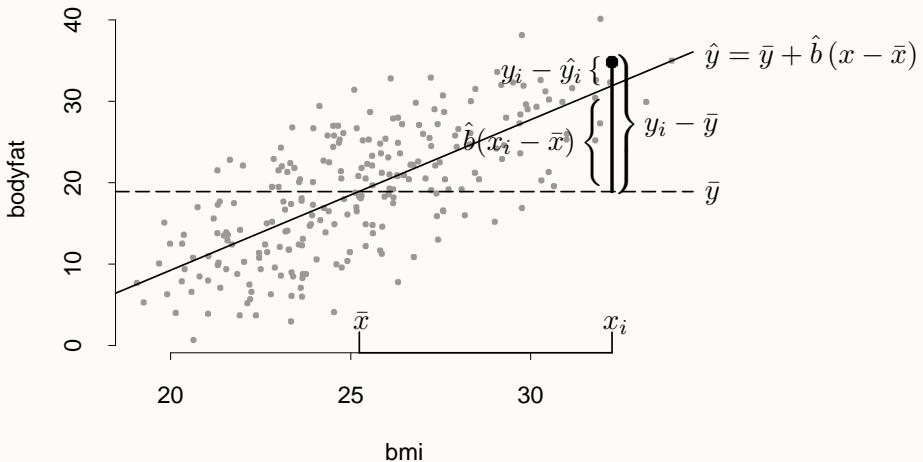
[Vollbild](#)

[Schließen](#)

[Beenden](#)



5.5.3. Durch die Regression erklärte Varianz



Wir haben gesehen, dass eine Variable y um so besser durch die Variable x erklärt werden kann, je grösser die Korrelation r zwischen beiden Variablen dem Betrage nach ist. Wir können also einen Teil der Variabilität von y durch die Regression auf x erklären. Aus mathematischen Überlegungen ergibt sich, dass diese „erklärte Varianz“ als

$$s_{reg}^2 = r^2 s_y^2$$

Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)



Seite 149 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite 150 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

berechnet werden kann. Die Grösse r^2 gibt also den Anteil der Varianz von y , der durch Kenntnis von x erklärt wird und ist in diesem Sinne bedeutsamer als r selber („Bestimmtheitsmass“). Im Spezialfall $r = \pm 1$ liegen alle Punkte auf einer Geraden, und es bleibt keine Variabilität bei y mehr übrig, wenn man x berücksichtigt. Die Varianz („Residualvarianz“)

$$s_{\text{res}}^2 = (1 - r^2)s_y^2$$

ist der Teil, der übrigbleibt, ein Schätzer der Varianz σ^2 der Residuen ε_i . In obiger Figur ist s_{res}^2 die Varianz der Abstände der Daten von der Regressionsgeraden. Die Regressionsanalyse von $y = \text{bodyfat}$ auf $x = \text{BMI}$ hat ein $r^2 = 0.52$ ergeben. Das bedeutet, dass der BMI das Körperfett – der eigentliche, aber nicht leicht zu bestimmende Risikofaktor – nur etwa zur Hälfte erklären kann. Dies liegt nicht daran, dass der BMI einen grossen Messfehler hat, sondern daran, dass der BMI z. B. auch vom Körperbau und der Muskelmasse abhängt.

Die Beobachtungen streuen um Regressionsgerade mit SD

$$s_{\text{res}} = \sqrt{1 - r^2} s_y$$

r	0.3	0.5	0.7	0.9	0.99
$\sqrt{1 - r^2}$	0.95	0.87	0.71	0.44	0.14

Dies bedeutet, dass selbst eine beträchtliche Korrelation zwischen 0.5 und 0.7 die Standardabweichung nicht sehr stark reduziert. Mit diesem Phänomen hängt zusammen, dass Vorhersagen i. A. wenig exakt sind. Ein Beispiel ist die Vorhersage der Erwachsenengrösse in der Kindheit.

Prüfen Sie, ob Sie dies verstanden haben:

[Selbstkontrolle](#)



5.5.4. Tests und Konfidenzintervalle in der linearen Regression

Hat die Variable x überhaupt einen Einfluss auf y , d.h. ist $b \neq 0$? Statistisch bedeutet diese Frage, dass man testen will, ob sich y mit x systematisch ändert. Dabei benötigen wir die folgenden mathematischen Annahmen:

- Die Fehler ε_i sind unabhängig.
- Die Fehler sind normalverteilt $\mathcal{N}(0, \sigma^2)$ mit konstanter Varianz σ^2 .

Nullhypothese: $b = 0$

Die Nullhypothese ist äquivalent dazu, dass die Korrelation Null ist (siehe 5.3). Unter der Annahme einer gemeinsamen Normalverteilung von (x, y) ergibt sich der gleiche Test wie auf Korrelation $\rho = 0$.

In der Regressionsanalyse gilt:

- Alle Analysen werden **bedingt** für gegebene Werte x_1, \dots, x_n durchgeführt.
⇒ Regressionsanalysen sind einfacher als Analysen der Korrelation.
⇒ Die Verteilung der unabhängigen Variablen x wird nebensächlich.



[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
[Testen](#)
Regression
[Testfragen](#)
[Index](#)

Beispiel: Körperfett in Abhängigkeit vom BMI bei Männern.

Mit dem Programm StatView erhält man den folgenden Ausdruck:

Regression Summary

bodyfat vs. bmi

Count	241
Num. Missing	0
R	.718
R Squared	.516
Adjusted R Squared	.514
RMS Residual	5.547

Regression Coefficients

bodyfat vs. bmi

	Coefficient	Std. Error	Std. Coeff.	t-Value	P-Value
Intercept	-27.617	2.939	-27.617	-9.398	<.0001
bmi	1.844	.116	.718	15.957	<.0001

Die Regressionsgerade (Regression Coefficients) ist

$$\widehat{\text{bodyfat}} = -27.6 + 1.84 \times \text{bmi}$$

Das Bestimmtheitsmass r^2 (R Squared) hat den Wert 0.52, d. h. 52% der Variabilität des Körperfetts können durch die Regression auf den BMI erklärt werden. Die Standardabweichung der Residuen(RMS Residual) ist $s_{res} = 5.55$. „P-value“ sind

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 152 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

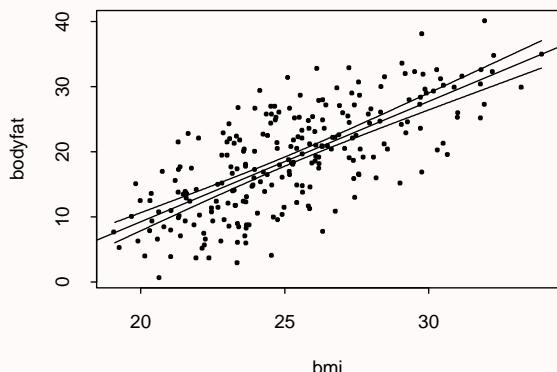
[Beenden](#)



die p -Werte der statistischen Tests. Der Regressionskoeffizient ist signifikant ($p = 0.0001$), also wird ein Zusammenhang bestätigt.

Konfidenzintervall für die Regressionsgerade

Statistikprogramme bieten $(1 - \alpha)$ -Konfidenzintervalle für den Wert der Regressionsgeraden $a + bx$ für einen gegebenen Wert von x , und tragen diese Konfidenzintervalle in den Regressionsplot ein:



Für einen BMI von z.B. $25 \text{ kg}/\text{m}^2$ können wir aus der Graphik ablesen, dass der mittlere Körperfett-Wert mit 95%iger Sicherheit zwischen 15.8 und 19.2 % liegt.

[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
[Testen](#)
[Regression](#) **Regression**
[Testfragen](#)
[Index](#)

[Home Page](#)

[Titelseite](#)

[◀◀](#)

[▶▶](#)

[◀](#)

[▶](#)

Seite 153 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite [154](#) von [100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

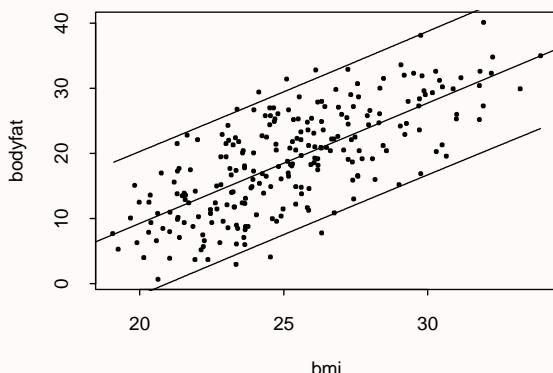
Vorhersageintervall für zukünftige Beobachtungen

Ziel ist es, vorauszusagen, in welches Intervall eine zukünftige Beobachtung für gegebenes x^* mit hoher Wahrscheinlichkeit zu liegen kommt.

Das Vorhersageintervall ist wesentlich breiter als das Konfidenzintervall, weil sich die Variabilität der geschätzten Regressionsgerade (Konfidenzintervall) und die Variabilität des Fehlers der zukünftigen Beobachtung (σ^2) addieren.

Achtung: Es besteht Verwechslungsgefahr mit dem Konfidenzintervall.

Die punktweisen Vorhersageintervalle werden üblicherweise in den Regressionsplot für alle x eingetragen:





[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
[Testen](#)
Regression
[Testfragen](#)
[Index](#)

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite **155** von **100**

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

5.6. Multiple Regression

Im Datensatz mit $y = \text{prozentuales Körperfett}$ hat man als mögliche Prädikatoren neben $x = \text{BMI}$ auch Alter und diverse leicht zugängliche Körpermasse. Es ist zu hoffen, dass mit mehreren Prädikatoren das Körperfett besser bestimmt ist. Allgemein: Wenn man anstatt einer Einflussgröße x deren k hat (x_1, \dots, x_k) , so könnte man nach dem bisherigen Stoff k einfache (univariate) Regressionsanalysen durchführen. Damit kann man aber das Zusammenspiel der k Prädiktoren nicht erfassen. Deshalb möchte man eine Regressionsanalyse mit allen k Einflussgrößen in einem Modell durchführen. Damit stellt sich das Problem, passende Modelle zu finden und statistisch zu prüfen.

Es gibt eine Reihe von Gründen, anstelle von mehreren einfachen Regressionsanalysen eine multiple Regressionsanalyse durchzuführen:

1. Man möchte mögliche Effekte von zusätzlichen „Stör“-Variablen in einer Studie eliminieren, bei der grundsätzlich nur eine Einflussgröße von Interesse ist.

Beispiel: Häufige Störgröße ist das Alter. $y = \text{Blutdruck}$, $x_1 = \text{Dosierung Hypertensivum}$, $x_2 = \text{Alter}$.

2. Man möchte mögliche Prognosefaktoren erforschen, von denen wir nicht wissen, ob sie alle wichtig oder zum Teil redundant sind.

Beispiel: $y = \text{Stenose}$, $x_1 = \text{HDL}$, $x_2 = \text{LDL}$, $x_3 = \text{BMI}$, $x_4 = \text{Rauchen}$, $x_5 = \text{Triglyceride}$.

3. Man möchte eine möglichst genaue Formel zur Vorhersage der abhängigen aus den erklärenden Variablen entwickeln.



Beispiel: y = Erwachsenengröße, x_1 = Größe als Kind, x_2 = Größe der Mutter, x_3 = Größe des Vaters.

4. Man möchte die Wirkung einer Variablen x_1 auf eine andere Variable y studieren, wobei der Einfluss weiterer Variablen x_2, \dots, x_k berücksichtigt wird.

Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite [156](#) von [100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
[Testen](#)
Regression
[Testfragen](#)
[Index](#)

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite [157](#) von [100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Neben dem BMI ist der „waist–hip–ratio“ (Taille–Hüft–Quotient) ein bekanntes Mass für Übergewicht bei Männern. Wir interessieren uns dafür, ob durch die Kenntnis beider Masse eine bessere Schätzung des Körperfetts möglich ist, als wenn wir nur eines kennen.

StatView liefert den folgenden Ausdruck:

Regression Summary
bodyfat vs. 2 Independents

Count	241
Num. Missing	0
R	.811
R Squared	.657
Adjusted R Squared	.654
RMS Residual	4.677

Regression Coefficients
bodyfat vs. 2 Independents

	Coefficient	Std. Error	Std. Coeff.	t-Value	P-Value
Intercept	-70.141	4.955	-70.141	-14.156	<.0001
weist/hip	70.367	7.100	.512	9.911	<.0001
bmi	.953	.133	.371	7.187	<.0001



In der Regressionsgleichung der multiplen Regression gibt es statt einer mehrere unabhängige Variable x_1, \dots, x_k :

$$y = a + b_1 x_1 + b_2 x_2 + \dots + b_k x_k + \varepsilon$$

In unserem Beispiel sieht die geschätzte Regressionsgleichung für $k = 2$ so aus:

$$\text{Körperfett} = -70 - 0.95 \times \text{BMI} + 70 \times \text{waist-hip-ratio}$$

Die Regressionskoeffizienten in der multiplen Regression beschreiben den Einfluss einer unabhängigen Variablen bei festgehaltenen anderen Variablen.

Versuch einer Erklärung:

Beide Übergewichtsmasse liefern zusätzlich zum jeweils anderen Mass eine signifikante Information über das Körperfett (beide $p < 0.0001$). Bei gleichem BMI haben Männer mit dickem Bauch mehr Fett, da der Bauch typischerweise nicht aus Muskelpaketen besteht. Bei gleichem „waist-hip-ratio“ führt ein hoher BMI zu weniger Fett. Dies ist überraschend, denn der BMI für sich allein würde zu mehr Fett führen. Die Erklärung ist, dass bei gleichem Bauchumfang ein hoher BMI auf viel Muskelmasse hinweist. Der Prozentsatz erklärter Varianz steigt durch den zusätzlichen Prädiktor von 52% auf 65%, was man als praktisch bedeutsam ansehen kann.

[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
[Testen](#)
Regression
[Testfragen](#)
[Index](#)

[Home Page](#)

[Titelseite](#)

«

»

◀

▶

Seite 158 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



6. Testfragen

Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

[◀◀](#)

[▶▶](#)

[◀](#)

[▶](#)

Seite **159** von **100**

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
[Testen](#)
[Regression](#)
Testfragen
[Index](#)

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite **160** von **100**

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Mit diesen Testfragen können Sie sich selber testen, ob Sie die Vorlesung in ihren Grundzügen verstanden haben.

Um die Aufgaben zu bearbeiten, müssen Sie sie zuerst durch einen Mausklick auf 'Start' aktivieren und erst dann Ihre Antwortwahl anklicken.

Start Was ist ein Abdomen?

1. Der Hinterleib des Menschen.

Ja Nein

2. Der Hinterleib des Gliederfüssers.

Ja Nein

3. $\chi^2 = 0$ belegt eindeutig, dass die Nullhypothese richtig ist.

Ja Nein

4. Die vorhergehende Antwort passt nicht zu der gestellten Frage.

Ja Nein

Auswertung

Wenn Sie auf 'Auswertung' klicken, erhalten Sie (logischerweise) eine Auswertung Ihrer Antworten. Die richtigen Antworten erhalten Sie, wenn Sie Ihren Computer mit dem Worten 'Korrigiere' dazu auffordern.

Mit dem 'Zurück'-Button springen Sie ungefähr an diejenige Textstelle, zu der die Aufgabe gehört.

Viel Spass!!

[Zurück](#)



Start Welche der folgenden Merkmale sind diskret?

1. Gewicht

Ja Nein

2. Diagnose

Ja Nein

3. Blutdruck

Ja Nein

4. Blutgruppe

Ja Nein

5. Kopfumfang

Ja Nein

6. Geschlecht

Ja Nein

Auswertung

[Zurück](#)

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite **161** von **100**

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Start Welche der folgenden Merkmale können durch ein Kuchendiagramm dargestellt werden?

1. Augenfarbe

Ja Nein

2. Blutgruppe

Ja Nein

3. Körpergrösse

Ja Nein

4. Geschlecht

Ja Nein

5. Temperatur

Ja Nein

Auswertung

[Zurück](#)

[Home Page](#)

[Titelseite](#)



[Seite 162 von 100](#)

[Zurück](#)

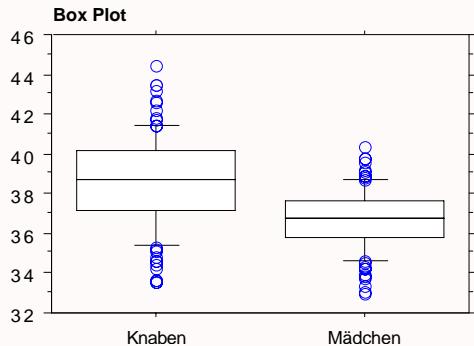
[Vollbild](#)

[Schließen](#)

[Beenden](#)



Die Schulterbreite von 15-jährigen Knaben ($n=120$) und 15-jährigen Mädchen ($n=112$) werden mit Hilfe der folgenden Boxplots verglichen. Welche Interpretationen sind stimmgig?



Start

1. Ein Kind mit einer Schulterbreite kleiner als 34cm ist immer ein Mädchen.
Ja Nein
2. Bei den Mädchen ist die Streuung kleiner als bei den Knaben.
Ja Nein
3. Ein Boxplot basiert auf Perzentilen.
Ja Nein
4. Der Strich in der Box stellt den Mittelwert dar.
Ja Nein
5. Bei den Knaben hat es prozentual mehr Beobachtungen in der Box als bei den Mädchen.
Ja Nein

Auswertung

Zurück

Schließen

Beenden

[Home Page](#)

[Titelseite](#)



Seite 163 von 100

[Zurück](#)

[Vollbild](#)



[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
[Testen](#)
[Regression](#)
Testfragen
[Index](#)

Start Ein Patient spürt nach der Einnahme eines Medikamentes Langeweile, wirft einen Fünfliber 10 mal und erhält k-mal Kopf.

1. Die Wahrscheinlichkeit, 5-mal Kopf zu erhalten ist grösser als die Wahrscheinlichkeit immer Kopf zu werfen

Ja Nein

2. Die Wahrscheinlichkeit, k-mal Kopf zu erhalten, ist für alle k von 0 bis 10 gleich ($1/11$).

Ja Nein

3. Die Anzahl der Würfe, in denen der Kopf oben liegt, ist normalverteilt mit $\mu = 5$ und $\sigma = 10$.

Ja Nein

4.

5. Die Anzahl der Würfe, in denen der Kopf oben liegt, ist binomialverteilt mit $p = 0.5$ und $n = 10$.

Ja Nein

Auswertung

[Zurück](#)

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite **164** von **100**

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

Start Die Variable x sei normalverteilt mit Mittelwert 1 und Standardabweichung 1. Welche der folgenden Antworten sind richtig?

1. Die Variable x kann nur positive Werte annehmen

Ja Nein

2. Die Fläche unter der Dichte von $x = 0$ bis $x = 2$ ist ungefähr 0.68 .

Ja Nein

3. Die Fläche unter der Dichte von $x = 0$ bis $x = 2$ ist ungefähr 0.95 .

Ja Nein

4. Der Wert der Verteilungsfunktion an der Stelle $x = 2$ ist ungefähr 0.84 .

Ja Nein

5. Der Wert der Verteilungsfunktion an der Stelle $x = 2$ ist ungefähr 0.975/,.

Ja Nein

Auswertung

[Zurück](#)

[Home Page](#)

[Titelseite](#)



[Seite 165 von 100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

[Seite 166 von 100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Start Beim Test einer Nullhypothese H_0 gegen eine Alternativhypothese H_1 bedeutet eine Wahrscheinlichkeit $\alpha = 0.05$ für den Fehler 1. Art: die Wahrscheinlichkeit ist höchstens 0.05 dafür, dass man

1. H_1 annimmt, wenn H_1 richtig ist

Ja Nein

2. H_0 beibehält, wenn H_0 richtig ist

Ja Nein

3. H_0 nicht ablehnt, wenn H_1 richtig ist

Ja Nein

4. H_0 ablehnt, obwohl H_0 richtig ist

Ja Nein

5. H_1 fälschlicherweise ablehnt

Ja Nein

Auswertung

[Zurück](#)



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 167 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Start Ein kontrollierter Versuch eines neuen Medikamentes führte zu dem Schluss, dass es signifikant besser als ein Placebo ist: $p < 0.05$. Welche der folgenden Behauptungen bevorzugen Sie?

1. Man ist jetzt sicher, dass das Medikament besser als ein Placebo ist.

Ja Nein

2. Falls das Medikament nicht effektiv ist, ist die Wahrscheinlichkeit, solche Resultate zu erhalten, kleiner als 5%.

Ja Nein

3. Der beobachtete Effekt des Medikaments ist mit 95% Sicherheit klinisch relevant.

Ja Nein

4. Das Medikament hat nur bei 5% der Patienten nicht gewirkt.

Ja Nein

Auswertung

[Zurück](#)



Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

[Seite 168 von 100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Start Die Normalverteilung spielt eine wichtige Rolle in der Statistik. Welche der folgenden Merkmale oder Eigenschaften treffen auf die Normalverteilung zu?

1. Die Normalverteilung ist symmetrisch um den Mittelwert.

Ja Nein

2. Sie lässt sich nicht auf Patienten anwenden, da diese nicht ein Normalkollektiv darstellen.

Ja Nein

3. Es liegen ca. 90% der Daten im Intervall ($\bar{x} \pm 2s$).

Ja Nein

4. Es liegen ca. 95% der Daten im Intervall ($\bar{x} \pm 2s$).

Ja Nein

5. Die Normalverteilung wird in der Medizin ausschliesslich für Normen verwendet.

Ja Nein

Auswertung

[Zurück](#)



[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
[Testen](#)
[Regression](#)
Testfragen
[Index](#)

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite [169](#) von [100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Start 22 gesunde Probanden wurden in einer Druckkammer hyperbaren Bedingungen ausgesetzt. Mit einem gepaarten t -Test ($\alpha = 5\%$) wurde geprüft, ob ein Unterschied zwischen dem Blutvolumen durchs Auge vor dem Versuch und 10 Minuten nach dem Versuch festzustellen ist. Der p -Wert betrug 0.08. Welche Aussagen sind richtig?

1. Die Nullhypothese lautet: Es gibt keinen Unterschied.

Ja Nein

2. Bei 8% der Probanden bestand kein signifikanter Unterschied zwischen den Messwerten.

Ja Nein

3. Die Nullhypothese ist zu verwerfen.

Ja Nein

4. Bei 8% der Probanden bestand ein signifikanter Unterschied zwischen den Messwerten

Ja Nein

5. Die Nullhypothese lautet: Es gibt einen Unterschied.

Ja Nein

6. Die Nullhypothese kann nicht verworfen werden.

Ja Nein

Auswertung

[Zurück](#)



Start

Welche Tests setzen normalverteilte Daten voraus?

1. gepaarter t -Test

Ja Nein

2. ungepaarter t -Test

Ja Nein

3. χ^2 -Tests

Ja Nein

4. Mann–Whitney Test

Ja Nein

5. Wilcoxon–Test für gepaarte Stichproben

Ja Nein

Auswertung

Zurück

[Home Page](#)

[Titelseite](#)



Seite 170 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

Start Gegeben seien zwei Grundgesamtheiten mit den Erwartungswerten μ_1 und μ_2 und der Varianz σ^2 . Daraus werden 2 Stichproben gezogen und deren Mittelwerte mit dem t -Test für unverbundene Stichproben überprüft. Dabei wird die Power grösser, wenn alle Grössen gleich bleiben, aber

1. der Stichprobenumfang grösser wird

Ja Nein

2. die Irrtumswahrscheinlichkeit α grösser wird

Ja Nein

3. der Betrag der Differenz $|\mu_1 - \mu_2|$ grösser wird

Ja Nein

4. die Varianz σ^2 grösser wird

Ja Nein

Auswertung

[Zurück](#)

[Home Page](#)

[Titelseite](#)



[Seite 171 von 100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 172 von 100

[Zurück](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Start

Bei einem χ^2 -Test ergibt sich für den Wert der Prüfgrösse $\chi^2 = 0$. Was besagt dieses Ergebnis?

1. Dieses Ergebnis ist unmöglich, da die Prüfgrösse nur positive Werte annehmen kann.

Ja Nein

2. Aufgrund des Testergebnisses behält man die Nullhypothese bei; ein β -Fehler ist bei dieser Entscheidung jedoch nicht auszuschliessen.

Ja Nein

3. $\chi^2 = 0$ belegt eindeutig, dass die Nullhypothese richtig ist.

Ja Nein

4. $\chi^2 = 0$ belegt eindeutig, dass die Alternativhypothese richtig ist.

Ja Nein

5. Ob man die Null- oder die Alternativhypothese annimmt, ist abhängig von der Grösse des α -Fehlers.

Ja Nein

Auswertung



Einführung

Deskriptive Statistik

Wahrscheinlichkeit

Testen

Regression

Testfragen

Index

[Home Page](#)

[Titelseite](#)



[Seite 173 von 100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Start

Zwischen der Körpergrösse von Kindern und der Grösse der beiden Eltern wird eine Korrelation von $r=0.17$ gefunden. Welche Aussagen teffen zu?

1. Die Korrelation misst, wie gross der Unterschied in den Werten zwischen Eltern und Kindern ist.

Ja Nein

2. Die Grösse stimmt Eltern und Kindern in 71% der Fälle gut überein.

Ja Nein

3. Die Kinder sind im Mittel 0.71cm grösser als ihre Eltern (säkulärer Trend).

Ja Nein

4. Die Korrelation ist ein Mass für die Übereinstimmung zwischen 2 Variablen.

Ja Nein

5. Wenn man die Körpergrösse der Eltern kennt, kann man die Variabilität bei den Kindern zu 50% erklären.

Ja Nein

6. Die Körpergrösse von Kindern und Eltern zeigt eine gewisse Übereinstimmung.

Ja Nein

Auswertung

[Zurück](#)



Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen
Index

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite **174** von **100**

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Start

In einer Untersuchung der Korrelation zwischen der Plasmakonzentration und dem Effekt eines Medikamentes erhält man: $r=+0.14$, $p<0.001$, $N=83$. Welche der folgenden Aussagen ist richtig?

- 1.** Es besteht ein starker Zusammenhang zwischen der Konzentration und dem Effekt.

Ja Nein

- 2.** Der Zusammenhang ist statistisch gesichert.

Ja Nein

- 3.** Es besteht nur ein schwacher Zusammenhang.

Ja Nein

- 4.** Je grösser die Plasmakonzentration, desto grösser der Medikamenteneffekt.

Ja Nein

- 5.** Je kleiner die Plasmakonzentration, desto grösser der Medikamenteneffekt

Ja Nein

Auswertung

[Zurück](#)



Start

Die Regressionsgerade von x auf y geht stets durch

1. ...den Nullpunkt des Koordinatensystems (0,0).

Ja Nein

2. ...den Schwerpunkt(x,y).

Ja Nein

3. ...mindestens zwei Punkte der Punktwolke.

Ja Nein

4. ...mindestens einen Punkt der Punktwolke.

Ja Nein

Auswertung

[Zurück](#)

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 175 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen

Index

[Home Page](#)

[Titelseite](#)

[◀](#)

[▶](#)

[◀](#)

[▶](#)

[Seite 176 von 100](#)

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Index

χ^2 , *siehe* chi²

Alternative, *siehe* Hypothese

Ausreisser, 140

 empfindlich gegen, 33

 robust gegen, 33

Balkendiagramm, 16

bar chart, *siehe* Balkendiagramm

Beobachtungen, *siehe* Daten

Bestimmtheitsmass, 154

Bias, 61

Binomialverteilung, 54

Bonferroni–Korrektur, 115

Boxplot, 28

chi²–Test, 107

chi²–Verteilung, 52

Daten, 12

 absolutskalierte, 13

 diskrete, 12, 13

 extreme, 144

 intervallskalierte, 13

kategoriale, 12

nominale, 12

ordinale, 12

paarweise, 129

qualitative, 12

quantitative, 12

stetige, 12, 19

Streuung der, 34

Transformation von, 58, 135, 137

Variabilität der, 34

Zentrum der, 31

Dichte, *siehe* Wahrscheinlichkeitsdichte

Effektgrösse, 83, 92

Einstichprobenproblem, 94

Ereignis, 39

 unabhängiges, 41

Erwartungswert, 31, 49

Fehler

 1. Art, 82

 2. Art, 83



Einführung
Deskriptive Statistik
Wahrscheinlichkeit
Testen
Regression
Testfragen

Index

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite 177 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

Fisher's exakter Test, 111

Gesetz der grossen Zahlen, 56

Grundgesamtheit, 7

Häufigkeit, relative, 13, 19, 39, 126

Heterogenitätskorrelation, 142

Histogramm, 19

Homogenität, 109

Test auf, 110

Hypothese

Alternativ-, 79

einseitig, 89

zweiseitig, 89

Null-, 80

Prüfung von, 76

statistische, 80

wissenschaftliche, 79

Interquartilabstand, 26, 35

interquartile range, *siehe* Interquartilabstand

Irrtumswahrscheinlichkeit, 82

Klassenbreite, 22

Klassenzentrum, 22

Konfidenzintervall, 116, 116, 138, 156

für die Regressionsgerade, 158

Konfundierung, 143

Kontingenztafel, 109

Korrelation, 49, 128, 131

Heterogenitäts-, 142

Rang-, 141

Schein-, 142

Trivial-, 143

Kovarianz, 49

Kuchendiagramm, 15

least squares, *siehe* Schätzung, Kleinste Quadrate

Liniendiagramm, 17

Macht, 83, 91

Mann-Whitney Test, 103

McNemar Test, 107, 111

Median, 26, 31, 33

Mittelwert, 31, 32, 49

Mittelwertsunterschiede, Test auf, 94

Normalgleichungen, 152

Normalverteilung, 50, 56

p-Wert, 83

Perzentil, 24



[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
[Testen](#)
[Regression](#)
[Testfragen](#)
Index

pie chart, *siehe* Kuchendiagramm
power, *siehe* Macht
Prävalenz, 65
Programmpakete, 6
Proportionen, Test für, 106

Quantil, 24
Quartil, 26

Randomisierung, 69
Rangkorrelation, 141
Rangtest, 103
Regression, 128
 einfache lineare, 145
 multiple, 160
Residualvarianz, 154

Scattergramm, 17, 129
Schätzung
 erwartungstreue, 60
 Kleinste Quadrate, 149
 Maximum–Likelihood, 63, 122
 Minimum–Varianz, 63
Scheinkorrelation, 142
Signifikanz, 82
 ohne, 10
Signifikanzniveau, 82

Spannweite, 35
Spearmans Rangkorrelation, 141
standard deviation, *siehe* Standardabweichung
standard error, *siehe* Standardfehler
Standardabweichung, 35, 49
Standardfehler, 36, 67
Stichprobe, 7, 8, 43
 repräsentative, 67
Stichprobengröße, 92

t–Test
 Einstichproben–, 94
 gepaarter, 100
 ungepaarter, 97
 Zweistichproben–, 97
t–Verteilung, 95
Test
 auf Homogenität, 110
 auf Mittelwertsunterschiede, 94
 auf Unabhängigkeit, 112
 F–, *siehe* F–Test
 für Proportionen, 106
 Fisher’s exakter, 111
 in der linearen Regression, 156
 Mann–Whitney, 103

[Home Page](#)

[Titelseite](#)

◀

▶

◀

▶

Seite 178 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)



[Einführung](#)
[Deskriptive Statistik](#)
[Wahrscheinlichkeit](#)
[Testen](#)
[Regression](#)
[Testfragen](#)
Index

[Home Page](#)

[Titelseite](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

Seite 179 von 100

[Zurück](#)

[Vollbild](#)

[Schließen](#)

[Beenden](#)

- McNemar, 107, 111
- multipler, 113
- Rang-, *siehe* Rangtest
- statistischer, 76
- t-, *siehe* t-Test
- Wilcoxon, 103
 - für Paardifferenzen, 103
 - für unabhängige Stichproben, 103
 - Rangsummen, 103
 - signed rank, 103
- Therapiestudie, 71
- Transformation
 - von Daten, 58, 135, 137
- Trennschärfe, 83, 91
- Trivialkorrelation, 143
- Unabhängigkeit, 41
 - Test auf, 112
- Varianz, 34, 49
- Verlauf, 17
- Versuchsplanung, 67
- Verteilung, 43
 - Binomial-, 54
 - chi²-, 52
 - Normal-, 50, 56
- schiefe, 58
- t-, 95
- Verteilungsfunktion, 44
 - empirische, 23
- Vertrauensbereich, *siehe* Konfidenzintervall
- Vorhersageintervall, 159
- Wahrscheinlichkeit, 13, 39
 - bedingte, 41
- Wahrscheinlichkeitsdichte, 23, 46
- Wilcoxon Test, 103
 - für Paardifferenzen, 103
 - für unabhängige Stichproben, 103
 - Rangsummen, 103
 - signed rank, 103
- Zentraler Grenzwertsatz, 56
- Zufall, 74
- Zusammenhang
 - linearer, 131, 133, 138
 - nichtlinearer, 144
 - quadratischer, 144
- Zweistichprobenproblem, 94