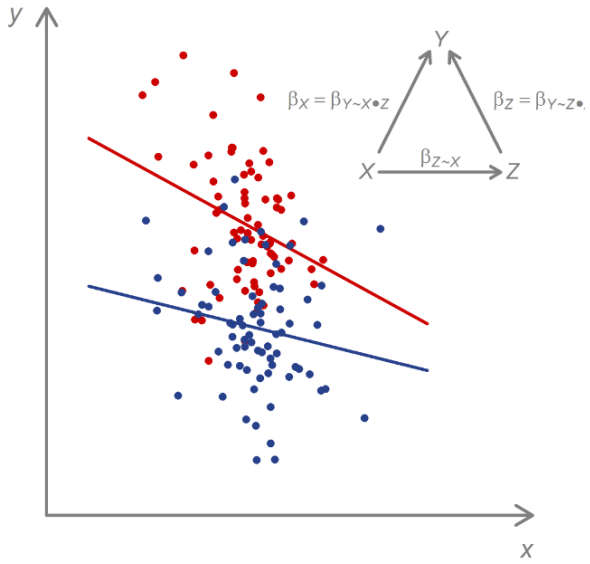
STATISTIK

GRUNDLAGEN UND METHODIK



Vorlesungsfolien zum Lehrbuch „Statistik - Grundlagen und Methodik“
von Stocker, T. C. und Steinke I. (2022)

Inhaltsverzeichnis

1 Einführung in die Statistik	1
Teil 1: Deskriptive Statistik – Empirischer Kalkül	22
2 Einführung in die deskriptive Statistik	23
3 Gesamtbeschreibung empirischer Verteilungen	30
4 Spezifizierende Beschreibung empirischer Verteilungen	61
5 Beschreibung und Analyse empirischer Zusammenhänge	145
Teil 2: Wahrscheinlichkeitsrechnung – Theoretischer Kalkül	270
6 Einführung in die Wahrscheinlichkeitsrechnung	271
7 Theoretische Verteilungen und Abhängigkeiten	330
8 Ergänzungen und Verallgemeinerungen	560

Inhaltsverzeichnis

Teil 3: Induktive Statistik – Verbindung von Empirie und Theorie	612
9 Einführung in die induktive Statistik	613
10 Statistisches Schätzen	633
11 Statistisches Testen	731
12 Das lineare Regressionsmodell	845

1 Einführung in die Statistik

1 Einführung in die Statistik

1.1 Eingrenzungen des Begriffs „Statistik“	4
1.1.1 Komprimierende Kennwerte	4
1.1.2 Staatswissenschaftliche Disziplin	5
1.1.3 Wahrscheinlichkeitstheoretisch fundierte Datenwissenschaft	6
1.2 Grundzüge statistischer Methoden	8
1.2.1 System und Zufall	8
1.2.2 Irrtums- und Sicherheitswahrscheinlichkeiten	10
1.2.3 Deskription und Induktion	14
1.2.4 Empirischer und theoretischer Kalkül	16
1.3 Teilbereiche, Spezialgebiete und Grundlagen	18
1.3.1 Datengewinnung und Aufbereitung	18
1.3.2 Spezielle Analysemethoden und Instrumente	18
1.3.3 Wissenschaftsspezifische Ausrichtungen	21

1.3.4 Mathematische und technische Grundlagen 21

1.1 Eingrenzungen des Begriffs „Statistik“

1.1.1 Komprimierende Kennwerte

> **Statistik**: „Zusammenfassender“ Kennwert einer größeren Menge von Zahlen (**Daten**)

> Singular- und Pluralform möglich (die **Statistik**, die **Statistiken**)

> Beispiele: Mittelwert (Durchschnittswert), prozentuale Anteile, Minimal- und Maximalwerte oder Spannweite

> Zielsetzung: Komprimierung von Information, Zeit und Kosten bei Entscheidungen sparen

> Problematik: Informationsverlust

1.1 Eingrenzungen des Begriffs „Statistik“

1.1.2 Staatswissenschaftliche Disziplin

> Wortherkunft: *status* (lat. für Zustand)
=> Zustandsbeschreibung des Staates

> **Statistik**: Teilgebiet der Staatswissenschaften

> Beispiele tiefergehender (wirtschaftsstatistischer) Problemstellungen:
– Wie kann Arbeitslosigkeit (sinnvoll) definiert und gemessen werden?
– Wie kann Inflation (sinnvoll) definiert und gemessen werden?
– Wie lässt sich Armut definieren und messen?

> **Adäquationsprobleme:** Adäquate Operationalisierung von Statistiken zur Beschreibung und Analyse eines bestimmten Sachverhalts

=> Darum wird es hier nicht gehen!

1.1 Eingrenzungen des Begriffs „Statistik“

1.1.3 Wahrscheinlichkeitstheoretisch fundierte Datenwissenschaft

- **Charakterisierung und Definition** -

> **Statistik**: Datenwissenschaft:

> Befasst sich mit allen praktische und theoretischen Aspekten, die bei der Gewinnung, Aufbereitung, Analyse und Interpretation von Daten eine Rolle spielen

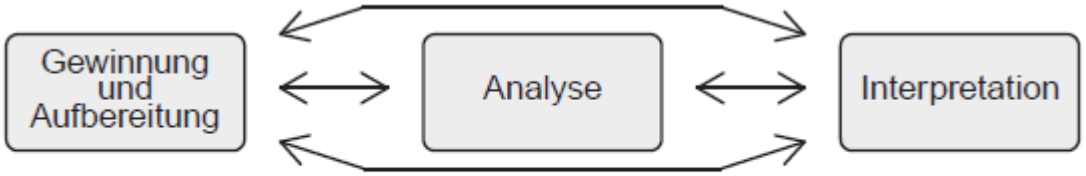


Abb. 1.1.1: Verlaufsstufen statistischer Untersuchungen von Daten

1.1 Eingrenzungen des Begriffs „Statistik“

Statistik ist die Wissenschaft von der Gewinnung, Aufbereitung, Analyse und Interpretation von Daten unter besonderer Verwendung der Wahrscheinlichkeitsrechnung.

• Historischer Abriss •

> England und USA als Ursprungsländer

> Yule (1911): *By **statistics** we mean quantitative data affected to a marked extent by a multiplicity of causes. By **statistical methods** we mean methods specifically adapted to the elucidation of quantitative data affected by a multiplicity of causes*“

> Entscheidend: Weiterentwicklung der Wahrscheinlichkeitstheorie in den 1930er Jahren (wie etwa durch Kolmogoroff)

> Fachkulturelle Entwicklung im angelsächsischen Raum anders als etwa in Deutschland...

1.2 Grundzüge statistischer Methoden

1.2.1 System und Zufall

● Zufallsvorgänge ●

> **Zufallsvorgang**: Ein Vorgang mit mehreren möglichen Ergebnissen, dessen Ausgang (konkretes Ergebnis) nicht exakt vorhersagbar (erklärbar) ist

> Beispiele: Mensabefragung auf Stichprobenbasis, Abfertigungszeiten am Gepäckschalter, Stromverbrauch einer Großstadt in einem bestimmten Zeitfenster, Ergebnisse von Bildungsstudien wie *PISA*, Arbeitslosenquote im nächsten Quartal, usw.

1.2 Grundzüge statistischer Methoden

- **Streuungszerlegung** -

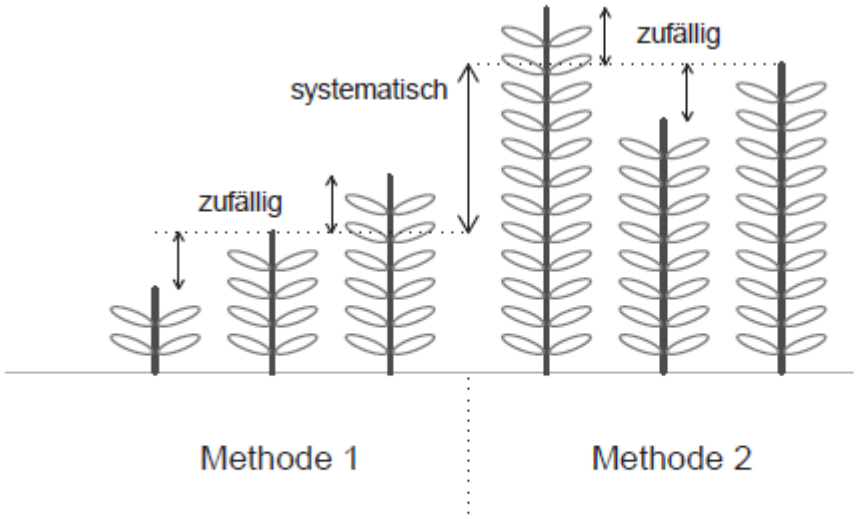


Abb. 1.2.1: Streuungszerlegung – Anbaumethoden und Ertrag von Pflanzen

1.2.2 Irrtums- und Sicherheitswahrscheinlichkeiten

● Statistische Probleme ●

Ein **statistisches Problem** liegt vor, falls ein Zufallsvorgang potenziell zu stochastischen Schwankungen von Daten und Statistiken führt und dadurch bedingt interpretatorische Unsicherheiten entstehen.

● Beispiel ●

Von 50 zufällig befragten Studierenden präferieren 30 die Mensa. Dies entspricht einem Anteil von 60%. Bedeutet dies, dass von allen 12 000 Studierenden an dieser Uni genau 7200, also exakt 60%, die Mensa präferieren?

1.2 Grundzüge statistischer Methoden

● Reale und hypothetische Stichproben ●

> **Reale Stichprobe**:
Statistisches Problem ensteht durch „echte Stichprobe“

> **Hypothetische Stichprobe**:
Statistisches Problem ensteht durch anderweitige Nichtvorhersagbarkeit (Experiment mit der Wuchshöhe von Pflanzen, Wertpapierentwicklung über die Zeit, ...)

1.2 Grundzüge statistischer Methoden

● Statistisches Modell ●

> Eine maßgeblich auf W‘rechnung basierende mathematische Beschreibung wie zufallsabhängige Daten in einer bestimmten Situation zustande kommen

> Zugewinn: Entscheidungen und Aussagen können mit **Irrtums-** und **Sicherheitswahrscheinlichkeiten** angegeben werden.

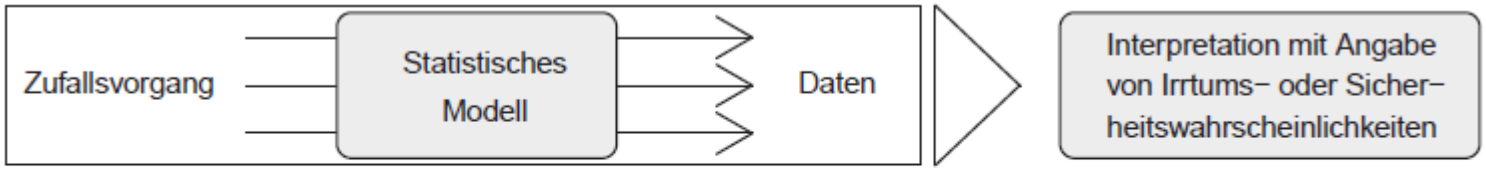


Abb. 1.2.2: Statistische Modellierung

1.2 Grundzüge statistischer Methoden

Beispiel 1.2.1: Irrtums- und Sicherheitswahrscheinlichkeiten

> Situation 1: Sind 31 von 50 wirklich eine Mehrheit?

> Situation 2: Wo liegt der wahre Anteil, wenn 40 von 50 „dafür“ stimmen?

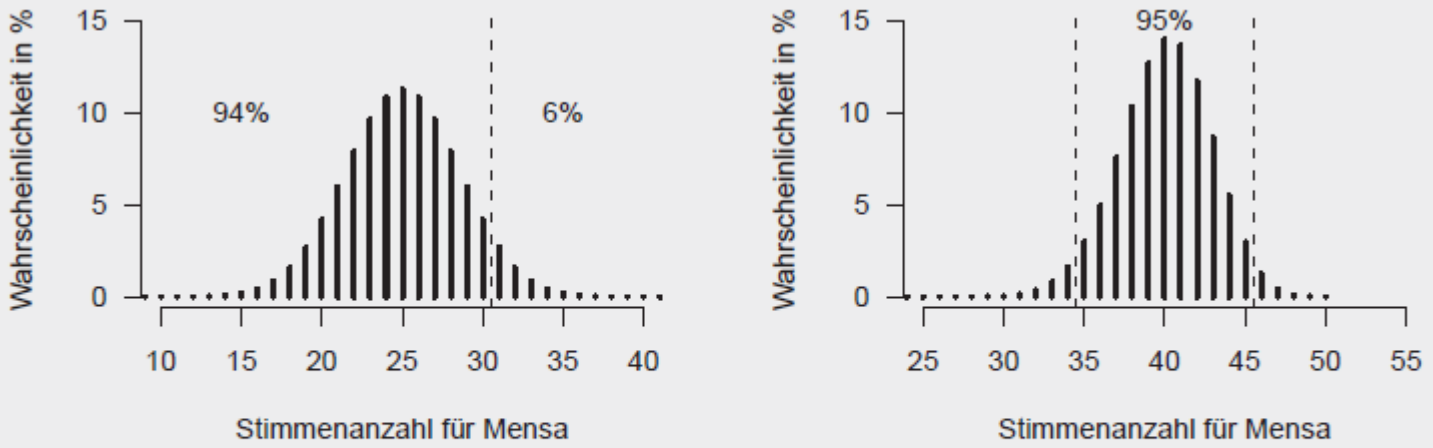


Abb. 1.2.3: 50% Zustimmung zur Mensa bzw. 80% Zustimmung zur Mensa

1.2.3 Deskription und Induktion

- **Methodik empirischer Wissenschaften** -

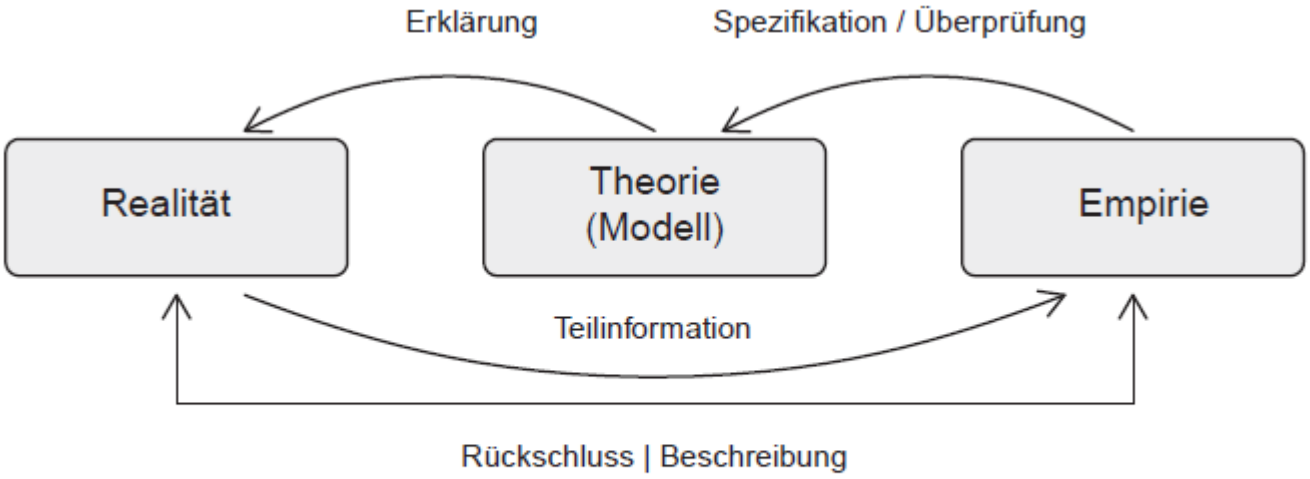


Abb. 1.2.4: Methodik empirischer Wissenschaften

1.2 Grundzüge statistischer Methoden

- **Statistische Methodik** -

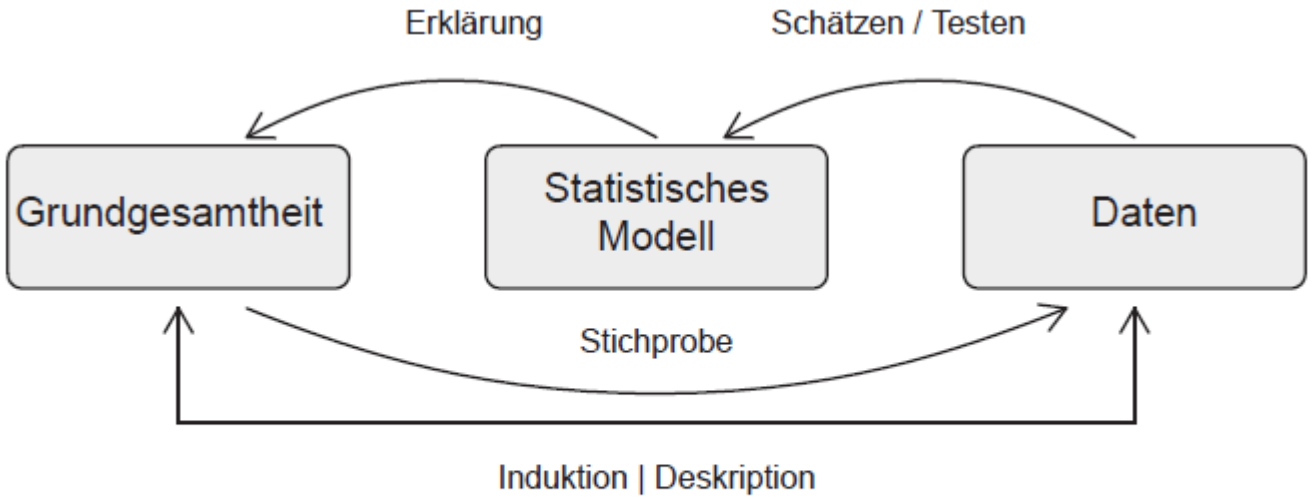


Abb. 1.2.5: Statistische Methodik

1.2.4 Empirischer und theoretischer Kalkül

● **Hintergrund** ●

> **Datengestützter (empirischer) Kalkül:** Berechnungen beruhen auf konkret vorliegenden Daten

> **Stochastischer (theoretischer) Kalkül:** Berechnungen beruhen auf einem Wahrscheinlichkeitskalkül (auf theoretischen Überlegungen)

● **Beispiel 1.2.1 fortgesetzt** ●

> *„80% der befragten Studierenden präferieren die Mensa“*

> *„Die Befragung ergab, dass Studierende mit 80% Wahrscheinlichkeit die Mensa präferieren“*

> *„Die Wahrscheinlichkeit, dass ein zufällig befragter Student die Mensa präferiert, beträgt 80%“*

> *„95%-Sicherheitsintervall [0.7, 0.9]“*?

1.3 Teilbereiche, Spezialgebiete und Grundlagen

Tab. 1.3.1: Statistische Methoden – Spezialgebiete und Grundlagen

Statistik		
Methodik	Deskriptiv	Induktiv
Einige Spezialgebiete		
	Explorative Datenanalyse	Statistische Versuchsplanung
	Data Mining	Stichprobentheorie
	Multivariate Statistik	
	Lineare Modelle	
	Zeitreihenanalyse	
	Nichtparametrische Statistik	
	Bayes–Statistik	
Wissenschaftsspezifische Ausrichtungen		
	„Metrien–Fächer" wie z.B.	
	Biometrie, Ökonometrie	
	Psychometrie oder Technometrie	
Mathematische und technische Grundlagen		
Analysis und Lineare Algebra		
Wahrscheinlichkeitstheorie	}	Stochastik
Mathematische Statistik	}	
Theorie stochastischer Prozesse	}	
Computergestützte Statistik		

1.3 Teilbereiche, Spezialgebiete und Grundlagen

1.3.1 Datengewinnung und Aufbereitung

- Statistische Versuchsplanung
- Stichprobentheorie
- Primär- und Sekundärstatistiken
- Datenaufbereitung

1.3.2 Spezielle Analysemethoden und Instrumente

- Deskriptive und induktive Methoden
- Explorative Datenanalyse und Data Mining

1.3 Teilbereiche, Spezialgebiete und Grundlagen

- **Korrelations- und Regressionsanalyse**

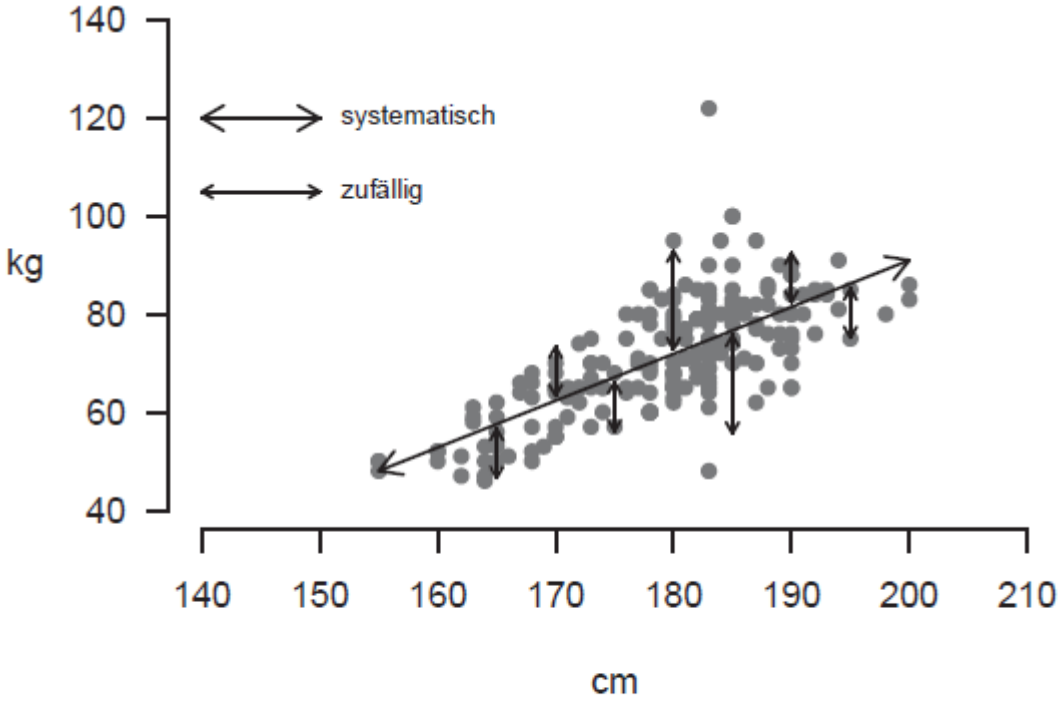
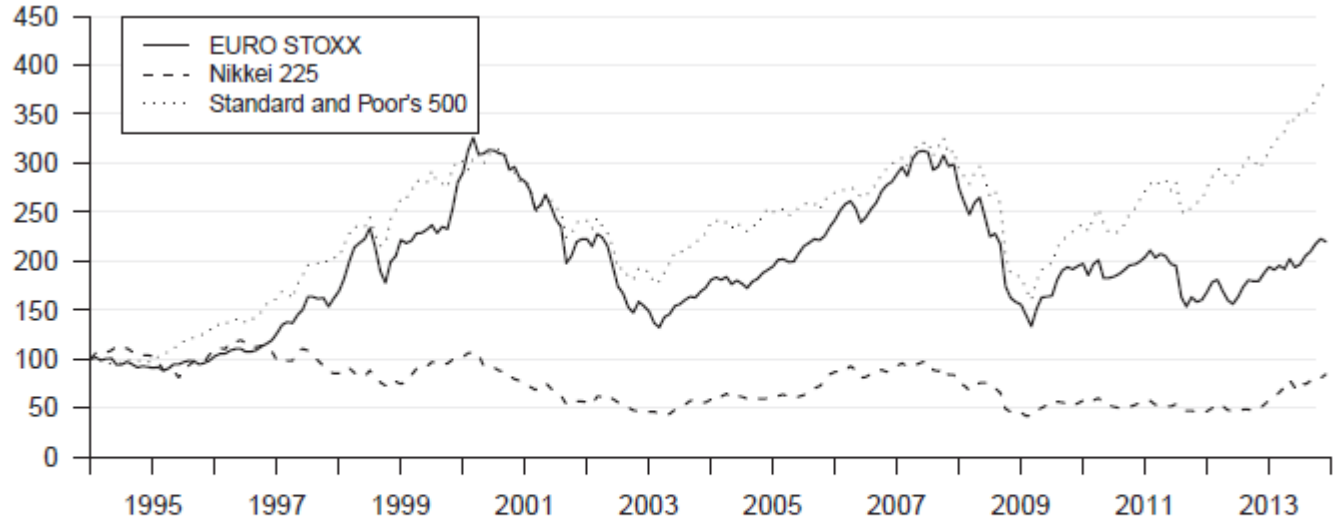


Abb. 1.3.1: Lineare Regression – Gewicht und Größe von Personen

- **Multivariate Statistik**

1.3 Teilbereiche, Spezialgebiete und Grundlagen

● Zeitreihenanalyse ●



Daten: Europäische Zentralbank [2014a]

Abb. 1.3.2: Zeitreihen – Aktienindizes 1994–2013 (monatlich, Jan. 1994 = 100) (nachgestellt)

1.3 Teilbereiche, Spezialgebiete und Grundlagen

- **Nichtparametrische Statistik**
- **Bayes-Statistik**

1.3.3 Wissenschaftsspezifische Ausrichtungen

> **Biometrie**, **Ökonometrie**, **Psychometrie**, **Technometrie**, usw.

> Angewandte Statistik vs. **Mathematische** (theoretische) **Statistik**

1.3.4 Mathematische und technische Grundlagen

> **Wahrscheinlichkeitsrechnung** (**Wahrscheinlichkeitstheorie**, **Stochastik**)

> **Analysis**, **Lineare Algebra** (Vektor- und Matrizenrechnung)

> **Numerik**, **Computergestützte Statistik**

Teil 1:
Deskriptive Statistik - Empirischer Kalkül

2 Einführung in die deskriptive Statistik

2 Einführung in die deskriptive Statistik

2.1 Daten, Datensätze und Variablen	26
2.2 Grundgesamtheit, Merkmalsträger und Skalierungsarten	28
2.3 Empirische Verteilungen und Zusammenhänge	29

2.1 Daten, Datensätze und Variablen

● Daten in der Statistik ●

... auf in einem bestimmten Kontext wesentliche Merkmale (**Variablen**) verkürzte Information über Personen, Objekte oder Vorgänge mittels Kategorisierung oder Quantifizierung

● Allgemeinere Formen von Daten ●

... etwa in der Informatik oder Nachrichtentechnik („kodierte Information“)

● Datensätze, Variablen und Dimension ●

> **Datensatz**: Rechteckanordnung von Zahlen und Kategorien

> Jedem Datensatz liegt eine Menge vergleichbarer Objekte zugrunde, (**Untersuchungseinheiten**, **Merkmalsträger**), die einem einheitlichen Bewertungssystem (**Skalierung**) unterliegen

> **Variable:** Merkmal mit verschiedenen Ausprägungsmöglichkeiten

2.1 Daten, Datensätze und Variablen

> **Dimension(alität)** von Daten: Anzahl von Variablen in einem Datensatz (teils auch mit Anzahl von Beobachtungen)

Tab. 2.1.1: Datensatz von Lehrveranstaltungen (fiktiv)

Veranstaltung	SWS	Prüfungsdauer (in Minuten)	ECTS	Sprache	Gasthörer
Statistik 2	6	120	7	Deutsch	nein
Multivariate Statistics	4	120	7	Englisch	ja
Versuchsplanung	2	NA	4	Deutsch	nein
Computergestützte Statistik	4	90	6	Deutsch	beschränkt
Econometrics I	4	90	6	Englisch	nein

- **Probleme bei der Erstellung von Datensätzen** -

Fragen in Bezug auf die Eingrenzung, Vergleichbarkeit und Eindeutigkeit von Untersuchungseinheiten sowie Messbarkeitsprobleme

● Grundgesamtheit ●

> Zielmenge von Objekten, Personen oder Vorgängen bei einer statistischen, die durch eindeutige Identifikationskriterien (sachlich, zeitlich, räumlich) abgegrenzt wird. Elemente der Menge heißen **Untersuchungseinheiten** oder **Merkmalsträger**.

> Später noch Unterscheidung zwischen **realer** und **hypothetischer Grundgesamtheit**

● Merkmalstypen ●

● Skalierungsarten ●

Merkmal		
Qualitativ	**Typ**	Quantitativ
Namen und Kategorien	**Ausprägungsform**	Zahlen (diskret oder stetig)
nominal oder ordinal	**Skalierung**	metrisch (kardinal)

Abb. 2.2.1: Merkmalstypen und Skalierungsarten

• Ein- und mehrdimensionale Verteilungen •

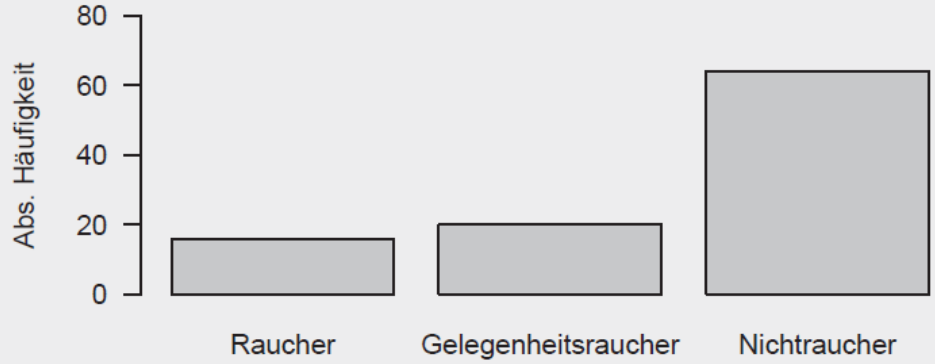


Abb. 2.3.1: Verteilung des Raucherstatus (fiktiv)

Raucherstatus / Geschlecht	Raucher	Gelegenheits–raucher	Nichtraucher
weiblich	4	8	28
männlich	12	12	36

Abb. 2.3.2: Gemeinsame Verteilung von Raucherstatuts und Geschlecht (fiktiv)

2.3 Empirische Verteilungen und Zusammenhänge

● Empirische Zusammenhänge ●

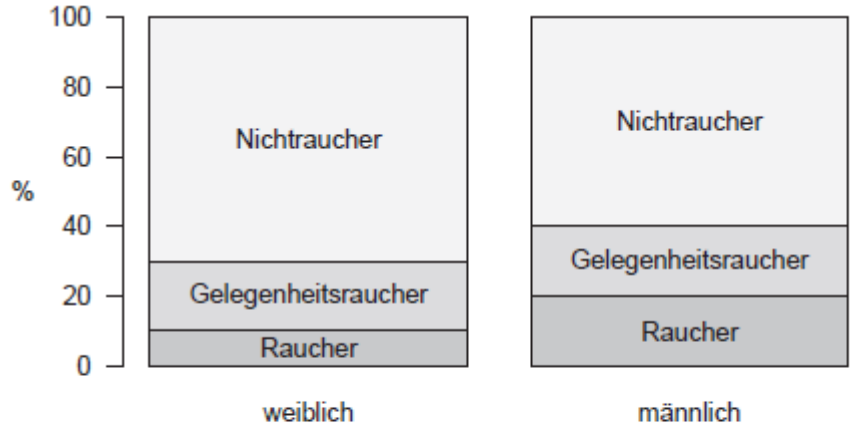


Abb. 2.3.3: Zusammenhang von Raucherstatus und Geschlecht (fiktiv)

● Wahrscheinlichkeitstheoretische Gegenstücke ●

Theoretische Verteilungen und **stochastische Zusammenhänge** (für **Zufallsvariablen**) folgen später in W‘rechnung (Kap. 7)

3 Gesamtbeschreibung empirischer Verteilungen

3 Gesamtbeschreibung empirischer Verteilungen

3.1 Tabellarische Darstellungsmöglichkeiten	32
3.1.1 Elementare Begriffe und Notation	32
3.1.2 Häufigkeitstabellen bei Urlisten	34
3.1.3 Häufigkeitstabellen bei klassierten Daten	36
3.2 Grafische Darstellungsmöglichkeiten	38
3.2.1 Kreis-, Säulen- und Balkendiagramme	38
3.2.2 Stamm-Blatt-Diagramme	41
3.2.3 Histogramme (Häufigkeitsdichten)	43
3.2.4 Boxplots	47
3.3 Empirische Verteilungsfunktion (EVF)	52
3.3.1 Häufigkeitsfunktion und EVF bei Urlisten	52
3.3.2 Häufigkeitsdichtefunktion und EVF bei klassierten Daten	56

3.1 Tabellarische Darstellungsmöglichkeiten

3.1.1 Elementare Begriffe und Notation

● Urliste und Stichprobenumfang ●

> **Urliste**: Nichtmanipulierte ursprünglich Beobachtungswerte

> **Stichprobenumfang**: Anzahl an Beobachtungswerten; notiert mit n

● Merkmale und Ausprägungsmöglichkeiten ●

> **Merkmal (Variable):** Interessierende Eigenschaft von Person, Objekt oder Vorgang. Variable betont die Variabilität hinsichtlich verschiedener **Ausprägungsmöglichkeiten**, meist notiert mit Großbuchstaben X, Y, Z, ...

> Beobachtungswerte eines Merkmals werden mit Kleinbuchstaben notiert:
$x_1, x_2, \ldots, x_n$ bzw. $y_1, y_2, \ldots, y_n$ usw.

> **Ausprägungsmöglichkeiten:** Menge aller Werte oder Kategorien, welche ein Merkmal (eine Variable) potenziell annehmen kann
Im Falle von Klausuren könnte z. B. gelten a_1= 1.0, a_2= 1.3, a_3= 1.7,...
im Falle des Merkmal „Raucherstatus: a_1= Raucher, a_2= Nichtraucher, ...

3.1 Tabellarische Darstellungsmöglichkeiten

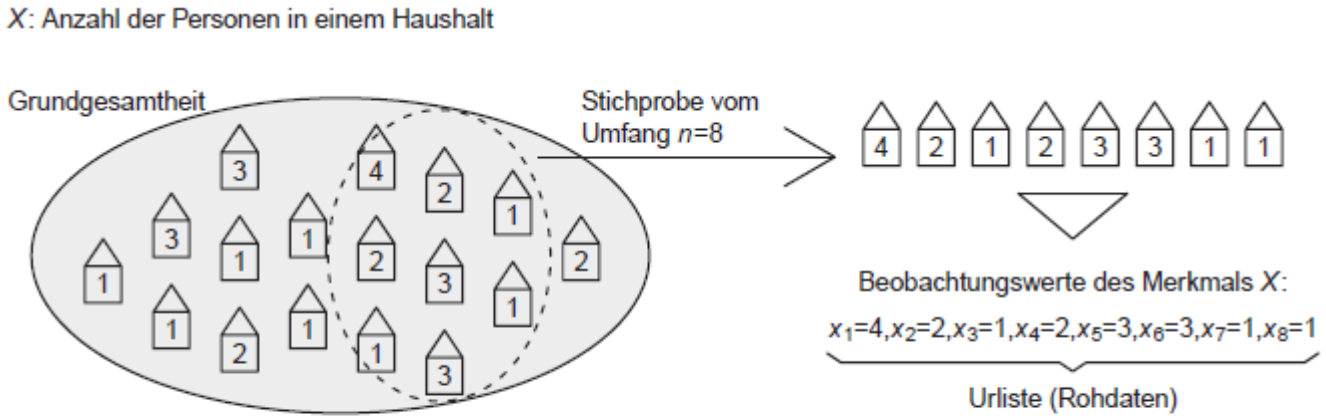


Abb. 3.1.1: Beispiel einer eindimensionalen Stichprobe

Elementare Begriffe und Notation

$X, Y, Z, \ldots$	Merkmale (Variablen)
$x_1, x_2, \ldots, x_n$	Beobachtungswerte des Merkmals X
oder	
x_i für $i = 1, 2, \ldots, n$	i heißt Indexvariable, $\{1, 2, \ldots, n\}$ heißt Indexmenge
n	Anzahl der Beobachtungen (Stichprobenumfang)
$a_1, a_2, \ldots, a_k$	Ausprägungsmöglichkeiten

3.1.2 Häufigkeitstabellen bei Urlisten

● Absolute und relative Häufigkeiten ●

> Beispiel:

$$n_1 = 3, \quad n_2 = 2, \quad n_3 = 2, \quad n_4 = 1$$

$$f_1 = n_1/n = 3/8 = 0.375, f_2 = n_2/n = 2/8 = 0.25, \ldots$$

> Allgemein gilt: $f_j = n_j/n$

● Tabellarische Darstellung der Häufigkeitsverteilung ●

Tab. 3.1.1: Formale Darstellung einer Häufigkeitsverteilung

j	a_j	n_j	f_j
1	1	3	0.375
2	2	2	0.250
3	3	2	0.250
4	4	1	0.125
Σ		8	1

3.1 Tabellarische Darstellungsmöglichkeiten

Tab. 3.1.2: Darstellung einer Häufigkeitsverteilung zu Präsentationszwecken

Personenanzahl	Häufigkeit	Prozent
1	3	37.5
2	2	25.0
3	2	25.0
4	1	12.5
Gesamt	8	100

Größe von Privathaushalten

Notation und Bezeichnungen bei Urlisten

n_j	Absolute Häufigkeit der j-ten Ausprägung
f_j	Relative Häufigkeit der j-ten Ausprägung
$n_1, n_2, \ldots, n_k$	Absolute Häufigkeitsverteilung
$f_1, f_2, \ldots, f_k$	Relative Häufigkeitsverteilung

3.1.2 Häufigkeitstabellen bei klassierten Daten

• Größenklassierung •

> Angenommen, bei einer Erhebung wurde die monatliche Absatzmenge einer bestimmten Brötchensorte in 30 Filialen eines Bäckereibetriebs erhoben:

37176,	29901,	15144,	20112,	25432,	18320,	32770,	38696,	17160,	8524,
22138,	13007,	20556,	24748,	27936,	28791,	37322,	19207,	21086,	21316,
12941,	44981,	36180,	18428,	51525,	12601,	5588,	39070,	41004,	47688.

> Einteilung in Größenklassen mit jeweils „ähnlich großen“ Werten

> **Klassengrenzen:** c_{j-1} und c_j für linke bzw. rechte Grenze der j–ten Klasse

> Beispiel: $c_0 = 0,\ \ c_1 = 10000$ Klassengrenzen der 1. Klasse

• Absolute und relative Klassenhäufigkeiten •

Beispiel: $\tilde{n}_1 = 2, \qquad \tilde{f}_1 = \tilde{n}_1/n = 2/30 \approx 0.067$

3.1 Tabellarische Darstellungsmöglichkeiten

Tab. 3.1.3: Häufigkeitsverteilung – Absatzmengen von Brötchen (in Tausend)

j	Klasse $(c_{j-1}, c_j]$	$\tilde{n}_j$	$\tilde{f}_j$
1	(0,10]	2	0.067
2	(10,20]	8	0.267
3	(20,30]	10	0.333
4	(30,40]	6	0.200
5	(40,50]	3	0.100
6	(50,60]	1	0.033
Σ		30	1

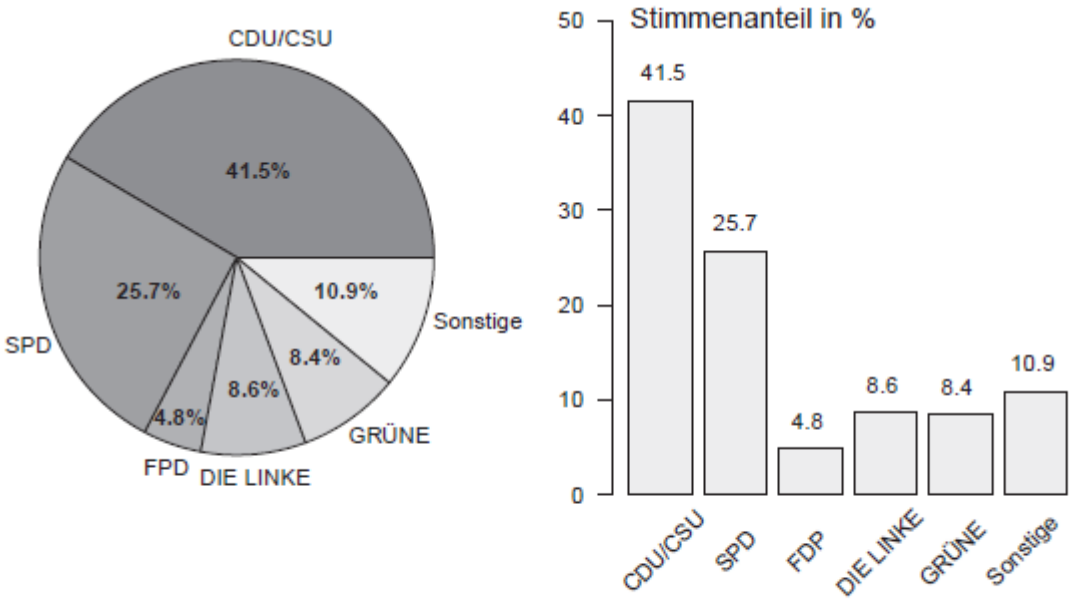
Notation und Bezeichnungen bei klassierten Daten

$(c_{j-1}, c_j]$	j-te Größenklasse als Intervall von ausschließlich c_{j-1} bis einschließlich c_j
c_{j-1} bzw. c_j	linke bzw. rechte Klassengrenze der j-ten Klasse
$\tilde{n}_j$	absolute Klassenhäufigkeit der j-ten Klasse
$\tilde{f}_j$	relative Klassenhäufigkeit der j-ten Klasse
$\tilde{n}_1, \tilde{n}_2, \ldots, \tilde{n}_k$	absolute Häufigkeitsverteilung (klassierter Fall)
$\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_k$ (klassierter Fall)	relative Häufigkeitsverteilung

3.2 Grafische Darstellungsmöglichkeiten

3.2.1 Kreis-, Säulen- und Balkendiagramme

- Kreis- und Ringdiagramme
- Säulendiagramme

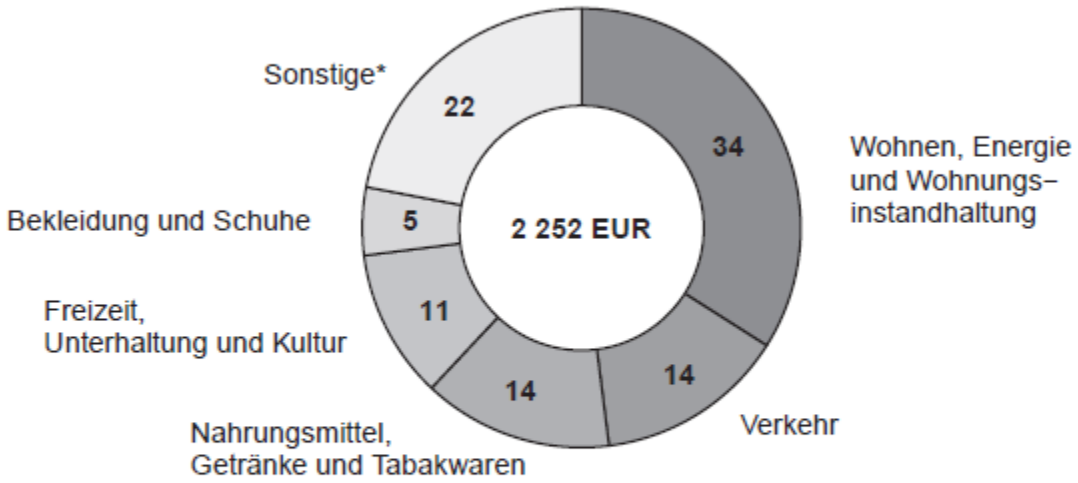


Daten: Der Bundeswahlleiter [2016]

Abb. 3.2.1: Kreis- und Säulendiagramm – Ergebnis der Bundestagswahl 2013 (Zweitstimmen)

3.2 Grafische Darstellungsmöglichkeiten

Ergebnisse der Laufenden Wirtschaftsrechnungen (LWR)



*Innenausstattung, Haushaltsgeräte,Gesundheitspflege, Nachrichtenübermittlung, Bildungswesen, u.a.

Daten: Statistisches Bundesamt [2013a, S.167]

Abb. 3.2.2: Ringdiagramm – Konsumausgaben privater Haushalte 2011 in % (nachgestellt)

3.2 Grafische Darstellungsmöglichkeiten

● Balkendiagramme ●



Daten: Statistisches Bundesamt [2013a, S. 47]

Abb. 3.2.3: Top 10-Zielländer der ausgewanderten Deutschen und Top 10-Herkunftsländer der zugewanderten Ausländer im Jahr 2011 (absolute Zahlen) (nachgestellt)

3.2.2 Stamm-Blatt-Diagramme

Hier nochmals die Urliste

37176, 29901, 15144, 20112, 25432, 18320, 32770, 38696, 17160, 8524,
22138, 13007, 20556, 24748, 27936, 28791, 37322, 19207, 21086, 21316,
12941, 44981, 36180, 18428, 51525, 12601, 5588, 39070, 41004, 47688.

The decimal point is 4 digit(s) to the right of the |

```
0 | 69
1 | 333
1 | 57889
2 | 01112
2 | 5589
3 | 03
3 | 67799
4 | 1
4 | 58
5 | 2
```

Abb. 3.2.4: Stamm-Blatt Diagramm – Monatliche Absatzmengen von Brötchen

3.2 Grafische Darstellungsmöglichkeiten

```
The decimal point is at the |

   1 | 4
   2 | 1111122334456667777788888888899999999
   3 | 000011111222222223333334555555556666667777778888889999
   4 | 00000000111111111122223333333444444555556667777788899999999999
   5 | 0000000001111233344444455555566677788888899999
   6 | 000000111122222333344444555556666677888999
   7 | 00000111222334445556666777888899999
   8 | 000011122222244555666
   9 | 00122223344444556667789 9
  10 | 001222333344456666677779
  11 | 00001255668
  12 | 00111224457889
  13 | 0011223677
  14 | 12358
  15 | 1
  16 | 37
```

Daten: Statistische Ämter des Bundes und der Länder [2014]

Abb. 3.2.5: Arbeitslosenquote in Kreisen und kreisfreien Städten Deutschlands (2011)

3.2.3 Histogramme

● Konzept und Beispiel ●

> Säulendiagramm über Größenklassen, dessen **Flächen** den relativen Klassenhäufigkeiten entsprechen

> Es gilt: *Säulenhöhe = Relative Klassenhäufigkeit / Klassenbreite*

> Der Wert $\tilde{f}_j / d_j$ heißt auch **Häufigkeitsdichte**

Tab. 3.2.1: Arbeitstabelle zur Erstellung eines Histogramms

j	Klasse $(c_{j-1}, c_j]$	d_j	$\tilde{n}_j$	$\tilde{f}_j$	$\tilde{f}_j / d_j$
1	(0,10]	10	2	0.067	0.0067
2	(10,20]	10	8	0.267	0.0267
3	(20,30]	10	10	0.333	0.0333
4	(30,40]	10	6	0.200	0.0200
5	(40,50]	10	3	0.100	0.0100
6	(50,60]	10	1	0.033	0.0033
Σ			30	1	

3.2 Grafische Darstellungsmöglichkeiten

> Beispiel anhand von Größenklasse 3, d. h. $j = 3$

$$\tilde{f}_3 = \tilde{n}_3/n = 10/30 \approx 0.3333$$

$$\tilde{f}_3/d_3 = 0.3333/10 \approx 0.0333$$

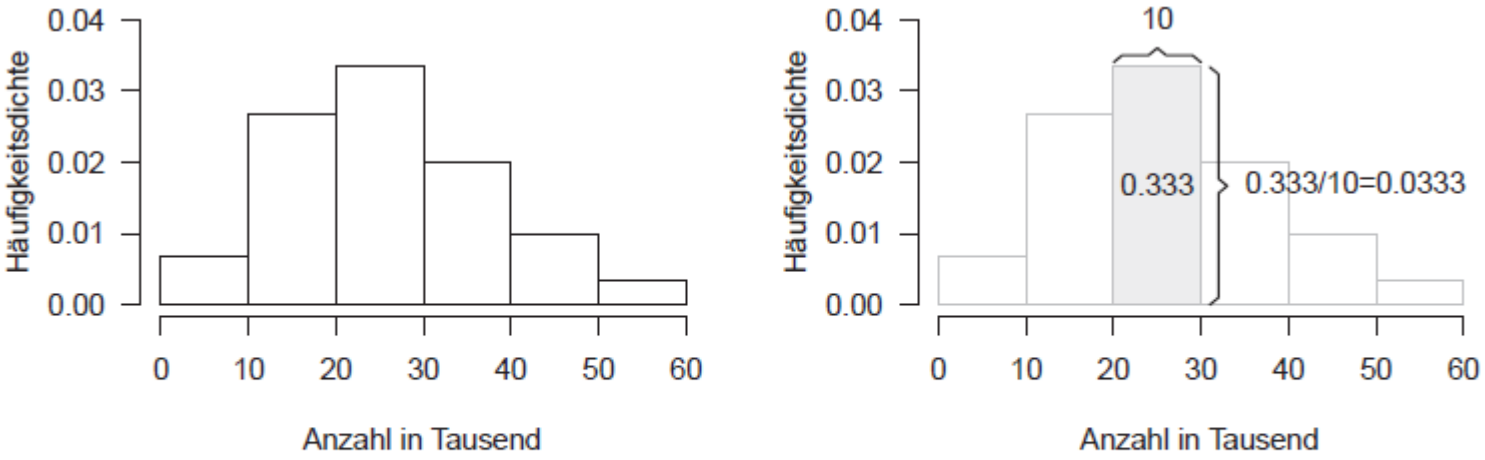


Abb. 3.2.6: Histogramm: Absatzmengen

3.2 Grafische Darstellungsmöglichkeiten

- **Problem der Klassenwahl** -

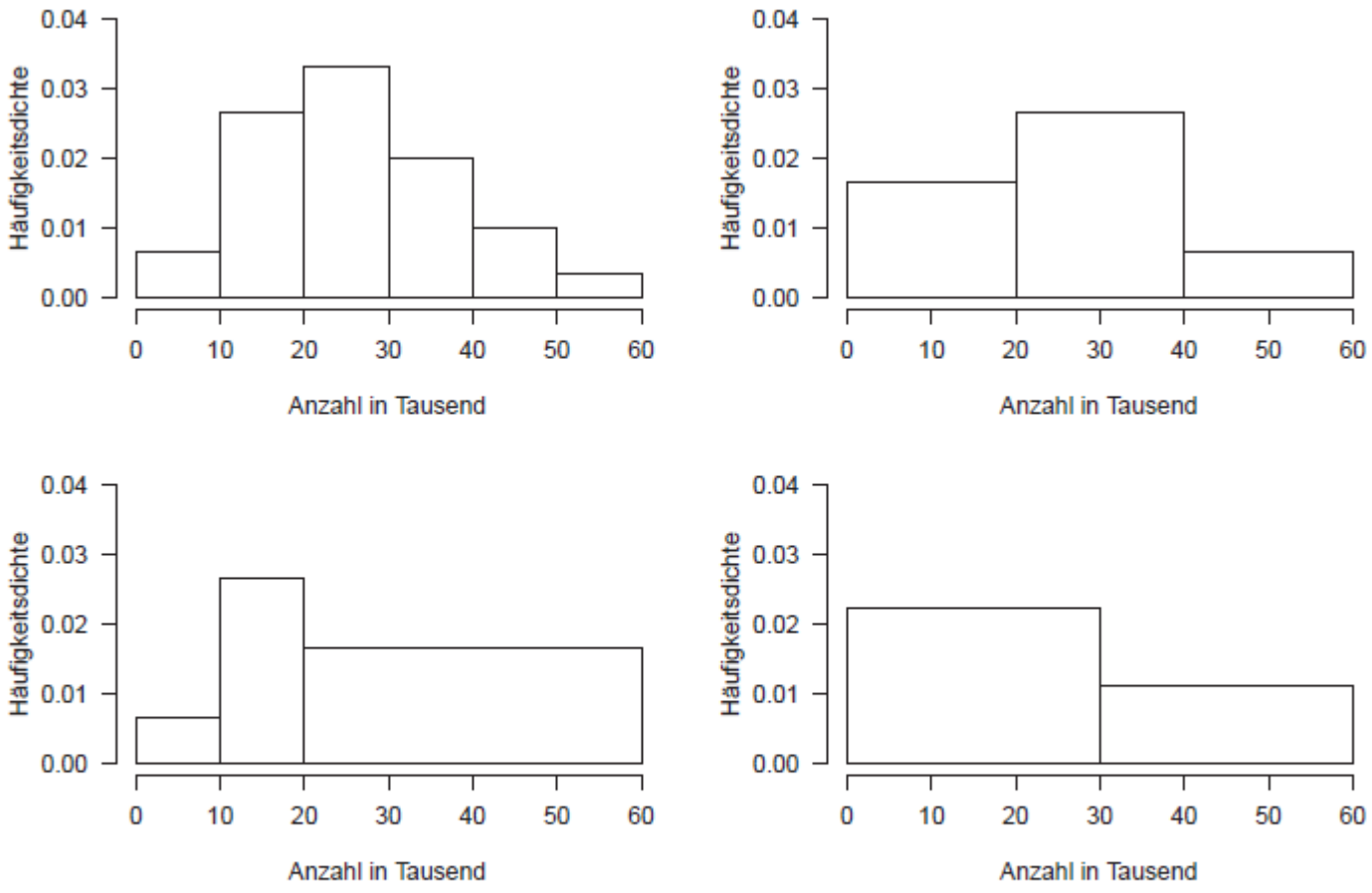


Abb. 3.2.7: Histogramme mit unterschiedlicher Klasseneinteilung

● Regeln zur Klassenwahl ●

> Nur einige sehr elementarge Regeln:

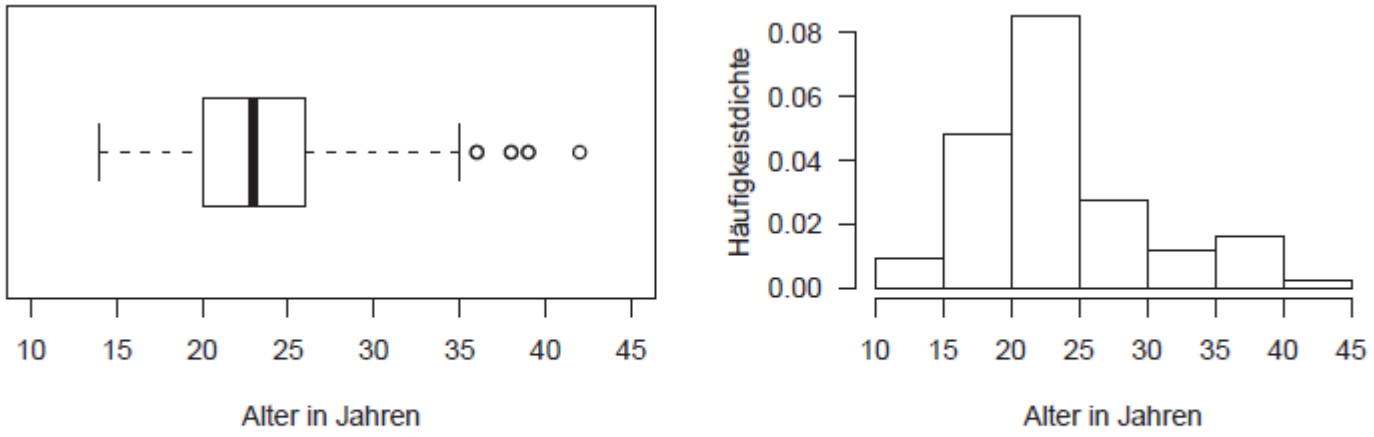
1. Nach Möglichkeit sollten Klassen ohne Werte („leere Klassen“) vermieden werden.
2. Die Werte sollten innerhalb der Klassen möglichst gleichmäßig verteilt sein.
3. Nach Möglichkeit sollten gleich breite Klassen gewählt werden (erleichtert die Interpretation).

> Theorie zur Konstruktion „optimaler“ Histogramme in **nichtparametrischer Statistik** im Rahmen der **Dichteschätzung**

> Implementierung von Algorithmen zur Konstruktion „optimaler“ Histogramme in Software-Pakete

3.2.4 Boxplots

- **Konzept und Beispiel**



Daten: R-Packet „UsingR“ (Datensatz „firstchi“)

Abb. 3.2.8: Boxplot und Histogramm – Alter von 87 Frauen bei der Geburt des ersten Kindes

3.2 Grafische Darstellungsmöglichkeiten

> **Unterer und oberer Hinge**: $h_L \approx \tilde{x}_{0.25}$ bzw. $h_U \approx \tilde{x}_{0.75}$

> **Interquartilsabstand**: $d_Q = \tilde{x}_{0.75} - \tilde{x}_{0.25}$

> **Oberer** (**unterer**) **Anrainer**: nicht weiter als das 1.5-fache des Interquartilsabstands vom oberen (unteren) Boxende entfernt

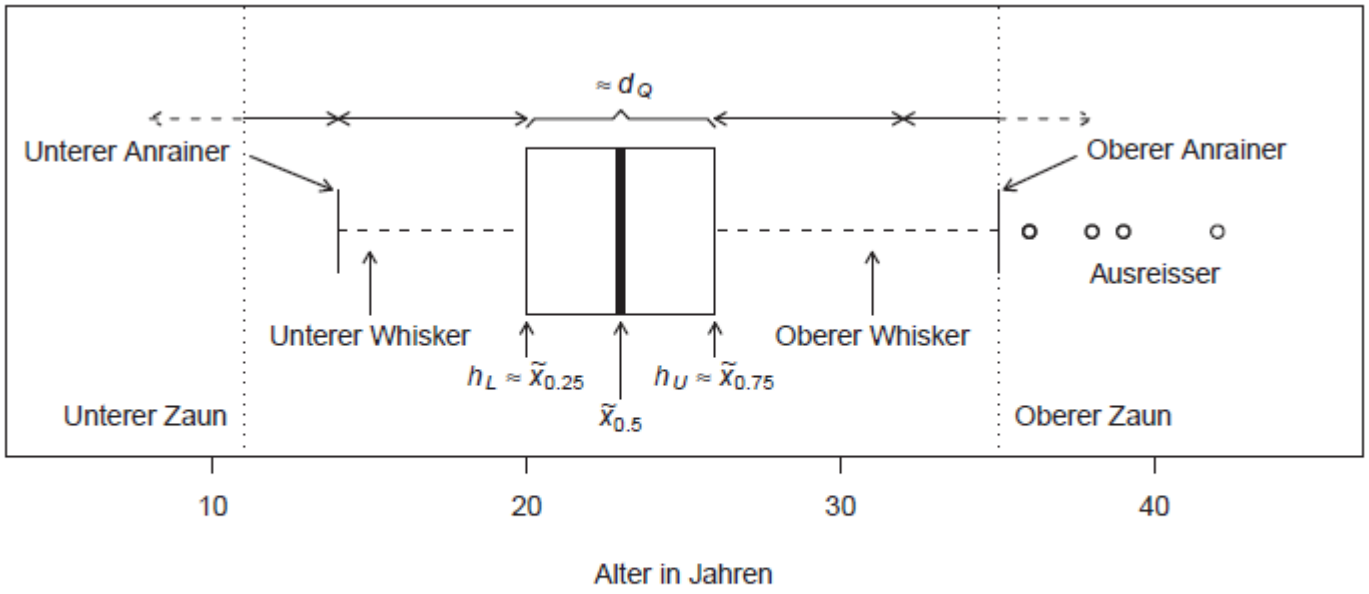
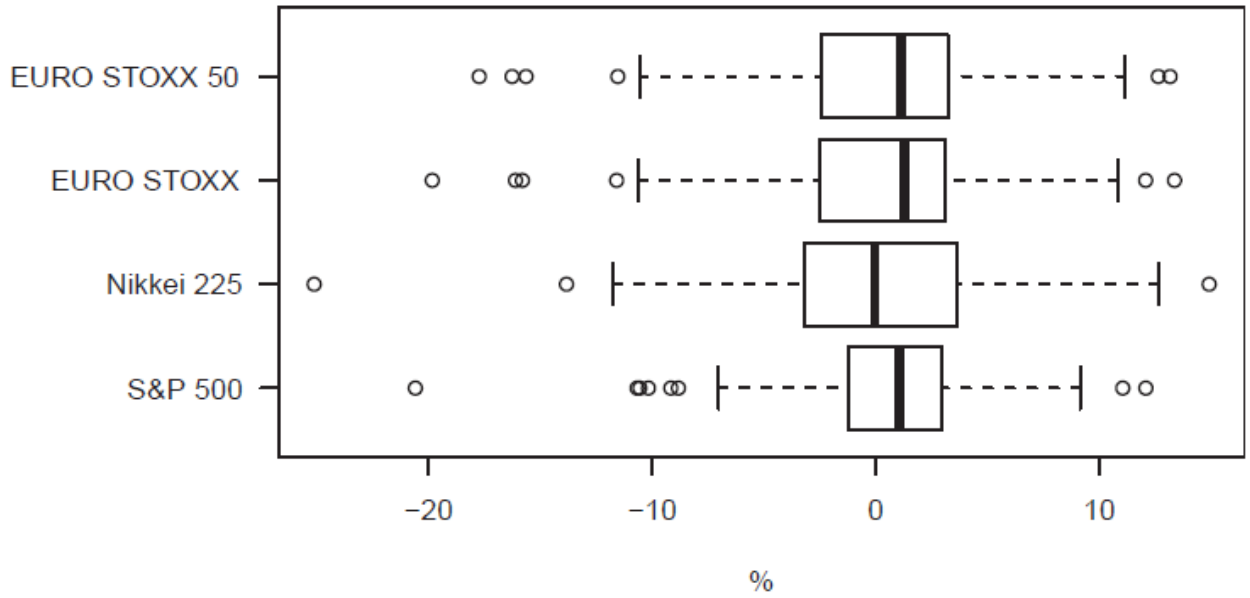


Abb. 3.2.9: Boxplot – Konstruktionsweise und Bezeichnungen

3.2 Grafische Darstellungsmöglichkeiten

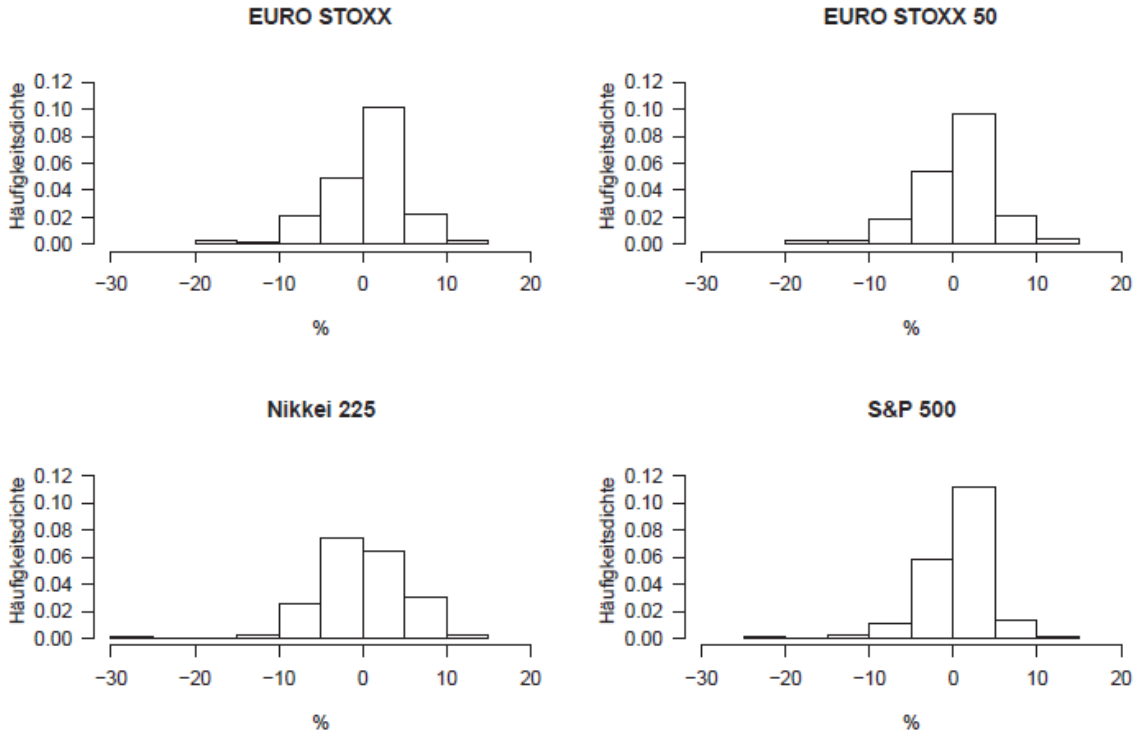
- **Deutung**
- **Für Vergleichszwecke geeignet**



Daten: Europäische Zentralbank [2014a]

Abb. 3.2.10: Boxplots – Monatsrenditen (in %) von Aktienindizes Jan. 1994–Dez. 2013

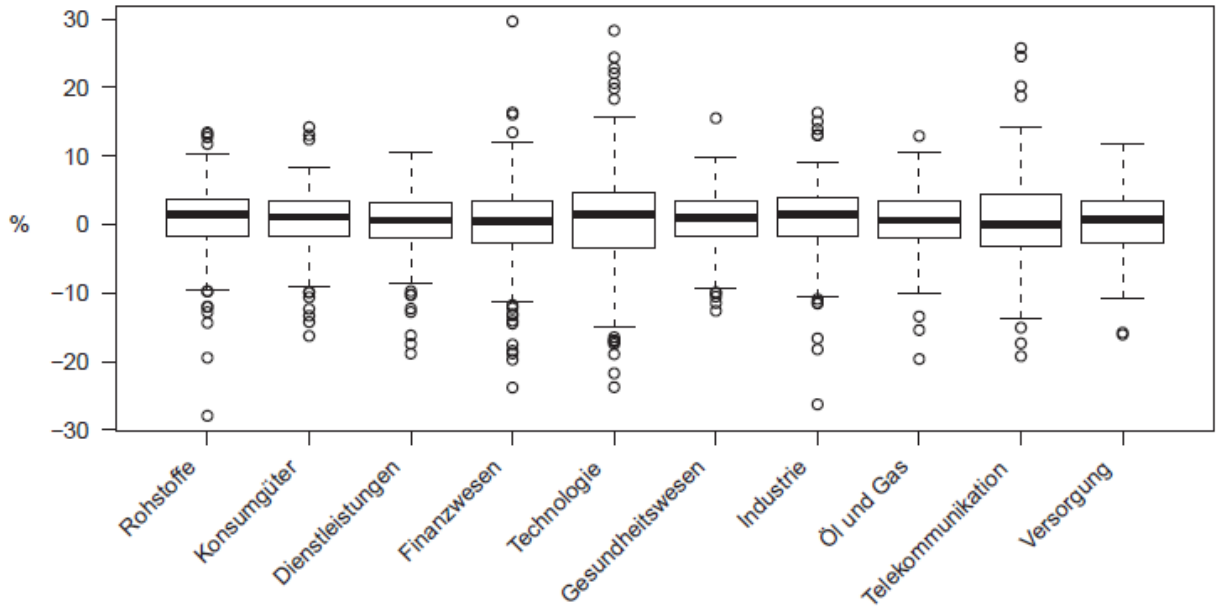
3.2 Grafische Darstellungsmöglichkeiten



Daten: Europäische Zentralbank [2014a]

Abb. 3.2.11: Histogramme – Monatsrenditen (in %) von Aktienindizes Jan. 1994–Dez. 2013

3.2 Grafische Darstellungsmöglichkeiten



Daten: Europäische Zentralbank [2014a]

Abb. 3.2.12: EURO STOXX sektorspezifische Indizes – Monatsrenditen 1994–2013

3.3 Empirische Verteilungsfunktion

3.3.1 Häufigkeitsfunktion und EVF bei Urlisten

● Häufigkeitsfunktion ●

Tab. 3.3.1: Häufigkeitsverteilung für Urliste mit kumulierten Anteilen

j	a_j	n_j	f_j	$F_n(a_j)$
1	1	3	0.375	0.375
2	2	2	0.250	0.625
3	3	2	0.250	0.875
4	4	1	0.125	1
Σ		8	1	

Die **Häufigkeitsfunktion** ordnet jedem beliebigen Wert x die relative Häufigkeit seines Auftretens in der Urliste zu, z. B.

$$f_n(-0.4) = 0,\ f_n(1) = 0.375,\quad f_n(3) = 0.250,$$
$$f_n(1.15) = 0,\quad f_n(4) = 0.125, \text{ usw.}$$

3.3 Empirische Verteilungsfunktion

● Empirische Verteilungsfunktion ●

> Kumulatives Aufsummieren der Häufigkeitsfunktion

> Gibt für jeden beliebigen Wert *x* an, wie groß der relative Anteil der Werte ist, die kleiner oder gleich *x* sind

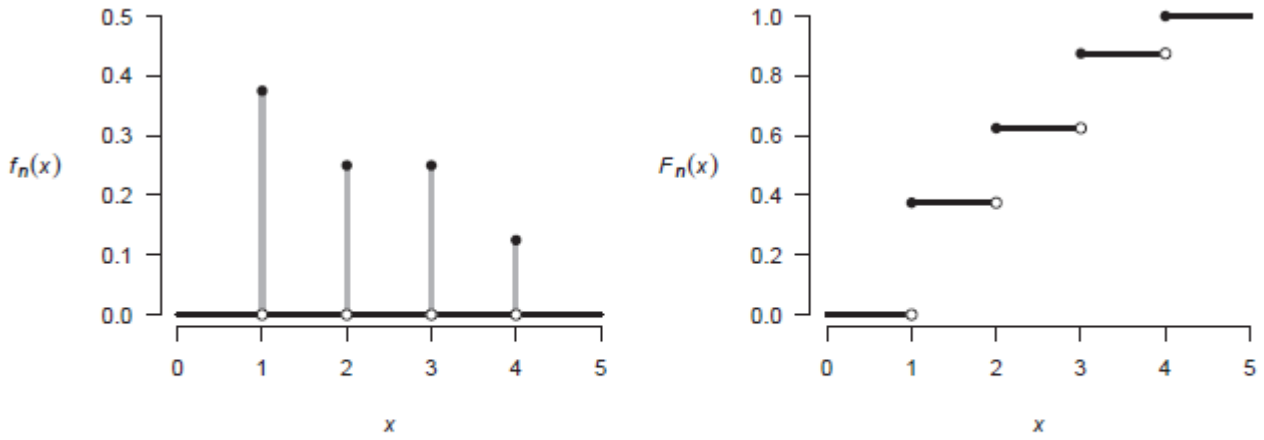


Abb. 3.3.1: Häufigkeitsfunktion und EVF bei Urliste

3.3 Empirische Verteilungsfunktion

> Beispiele:

$F_n(1) = f_n(1) = 0.375,$

$F_n(2) = f_n(1) + f_n(2) = 0.375 + 0.250 = 0.625,$

$F_n(3) = f_n(1) + f_n(2) + f_n(3) = 0.375 + 0.250 + 0.250 = 0.875$

$F_n(-12.3) = 0,$

$F_n(0.98) = 0,$

$F_n(1.99) = f_n(1) = 0.375$

$F_n(5.565) = f_n(1) + f_n(2) + f_n(3) + f_n(4)=1$

3.3 Empirische Verteilungsfunktion

Häufigkeitsfunktion und EVF bei Urlisten

Gegeben seien Beobachtungswerte eines metrischen Merkmals X mit zugehörigen Ausprägungen $a_1, a_2, \ldots, a_k$. Weiter sei $f_1, f_2, \ldots, f_k$ die zugehörige relative Häufigkeitsverteilung von X.
Dann ist die (relative) ***Häufigkeitsfunktion*** für diese Werte gegeben durch

$$f_n(x) = \begin{cases} f_j, & x = a_j,\ j = 1, \ldots, k, \\ 0, & \text{sonst.} \end{cases}$$

Die ***empirische Verteilungsfunktion*** (***EVF***) für die Urliste lautet

$$F_n(x) = \sum_{j:a_j \le x} f_n(a_j) \quad \text{für } x \in \mathbb{R}.$$

3.3.2 Häufigkeitsdichtefunktion und EVF bei klassierten Daten

• Häufigkeitsdichtefunktion •

Tab. 3.3.2: Häufigkeitsverteilung bei Klassenbildung mit kumulierten Anteilen

j	Klasse $(c_{j-1}, c_j]$	d_j	$\tilde{n}_j$	$\tilde{f}_j$	$\tilde{f}_j/d_j$	$\tilde{F}_n(c_j)$
1	(0,10]	10	2	0.067	0.0067	0.067
2	(10,20]	10	8	0.267	0.0267	0.333
3	(20,30]	10	10	0.333	0.0333	0.667
4	(30,40]	10	6	0.200	0.0200	0.867
5	(40,50]	10	3	0.100	0.0100	0.967
6	(50,60]	10	1	0.033	0.0033	1
Σ			30	1		

Die **Häufigkeitsdichtefunktion** ordnet jedem beliebigen Wert x die Häufigkeitsdichte an der Stelle x zu (klassierte Daten), z. B.

$$\tilde{f}_n(-0.5) = 0,\ \tilde{f}_n(0) = 0,\ \tilde{f}_n(8.5) = 0.0067,$$

$$\tilde{f}_n(31.2) = 0.0200,\ \tilde{f}_n(60) = 0.0033,\ \tilde{f}_n(60.01) = 0, \text{ usw.}$$

● EVF an den Klassengrenzen ●

> EVF für klassierte Daten soll kumulativen Anteil von Werten bis zu einem bestimmten Wert angeben.

> Kumulatives Integrieren der Häufigkeitsdichtefunktion

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

> Sie ordnet einem beliebigen Wert *x* die Fläche unterhalb der Häufigkeitsdichtefunktion bis zu diesem Wert zu

> An einer Klassengrenze entspricht der Wert der EVF genau der kumulierten Summe der relativen Klassenhäufigkeiten bis zu dieser Klassengrenze

3.3 Empirische Verteilungsfunktion

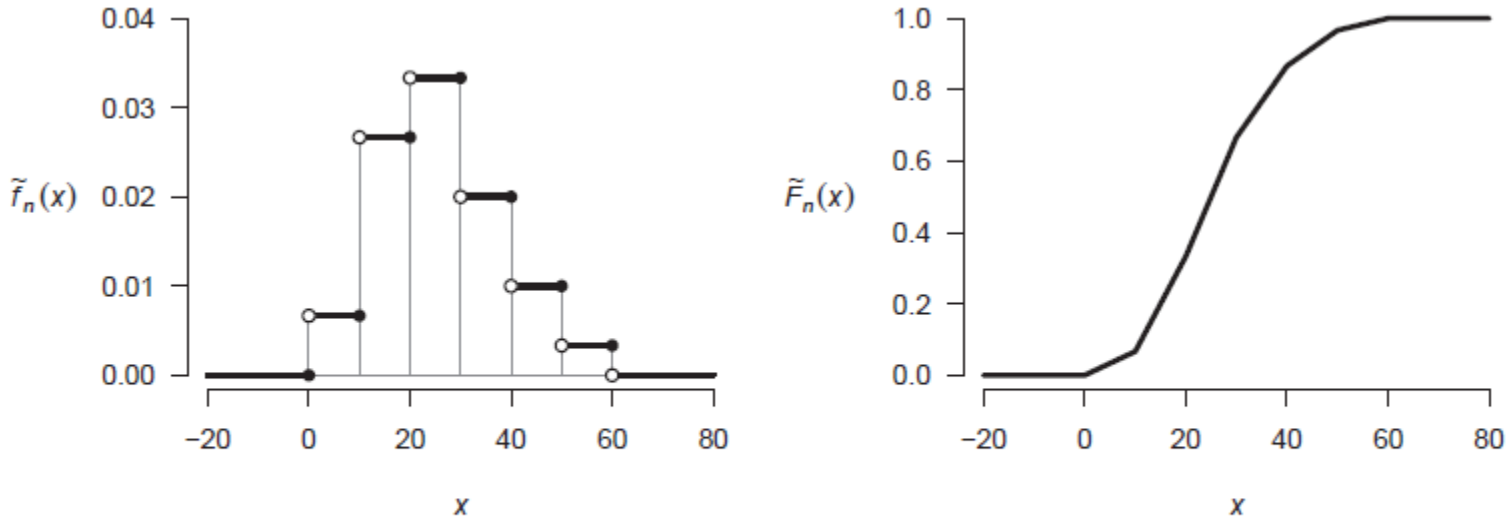
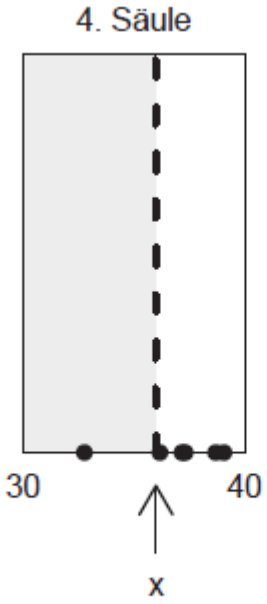


Abb. 3.3.2: Häufigkeitsdichtefunktion und EVF bei klassierten Daten

• Approximation innerhalb der Klassen •

Für einen Wert x innerhalb der j-ten Klasse gilt (als Approximation):

$$\tilde{F}_n(x) = \tilde{F}_n(c_{j-1}) + (x - c_{j-1})\frac{\tilde{f}_j}{d_j}$$



Säulenhöhe $= \frac{\tilde{f}_j}{d_j}$

Gesamte Säulenfläche $= \tilde{f}_j = 0.2$

Schattierte Fläche $= (x - 30) \times \frac{\tilde{f}_j}{d_j}$

Abb. 3.3.3: EVF: Lineare Approximation innerhalb von Klassen

3.3 Empirische Verteilungsfunktion

• 1. Ableitung der EVF •

Die Steigung der *EVF* für $x \in (c_{j-1}, c_j)$ entspricht der Häufigkeitsdichte

$$\frac{d\tilde{F}_n(x)}{dx} = \frac{\tilde{f}_j}{d_j}$$

Häufigkeitsdichtefunktion und EVF bei klassierten Daten

Gegeben seien Beobachtungswerte eines metrischen Merkmals X, die klassiert in k Größenklassen $(c_{j-1}, c_j]$, für $j = 1, \ldots, k$, vorliegen.
Dann ist die ***Häufigkeitsdichtefunktion*** für diese Werte gegeben durch

$$\tilde{f}_n(x) = \begin{cases} \tilde{f}_j/d_j, & x \in (c_{j-1}, c_j],\ j = 1, 2, \ldots, k, \\ 0, & \text{sonst.} \end{cases}$$

Die (approximative) ***empirische Verteilungsfunktion*** (***EVF***) lautet

$$\tilde{F}_n(x) = \int_{-\infty}^{x} \tilde{f}_n(t)\,dt = \begin{cases} 0, & x \le c_0, \\ \tilde{F}_n(c_{j-1}) + (x - c_{j-1})\frac{\tilde{f}_j}{d_j}, & x \in (c_{j-1}, c_j], \\ 1, & x > c_k. \end{cases}$$

4 Spezifizierende Beschreibung empirischer Verteilungen

4 Spezifizierende Beschreibung empirischer Verteilungen

4.1 Spezifika empirischer Verteilungen	65
4.2 Lagekennwerte	69
4.2.1 Arithmetisches Mittel	69
4.2.2 Median	73
4.2.3 Modalwert	77
4.2.4 Fechner‘sche Lageregeln	78
4.3 Spezielle Lagekennwerte	80
4.3.1 Arithmetisches Mittel bei gruppierten Daten	80
4.3.2 Quantile	82
4.3.3 Geometrisches Mittel	87
4.4 Streuungskennwerte	92
4.4.1 Spannweite	92
4.4.2 Mittlere absolute Abweichungen	93

4.4.3 Median absoluter Abweichungen	97
4.4.4 Varianz, Standardabweichung und Schwankungsintervalle	100
4.5 Spezielle Streuungskennwerte	104
4.5.1 Varianz bei gruppierten Daten	104
4.5.2 Quantilsabstände	107
4.5.3 Variationskoeffizient	109
4.6 Standardisierung mittels Lage und Streuung	113
4.7 Messung von Schiefe	117
4.8 Darstellung und Messung von Konzentration	120
4.8.1 Lorenzkurve	120
4.8.2 Gini-Koeffizient	123

4 Spezifizierende Beschreibung empirischer Verteilungen

4.9 Spezifische Eigenschaften empirischer Kennwerte	128
4.9.1 Minimumeigenschaft des arithmetischen Mittels	128
4.9.2 Minimumeigenschaft des Medians	132
4.9.3 Transformationseigenschaften	136
4.9.4 Robustheit	144

4.1 Spezifika empirischer Verteilungen

- **Unimodalität und Multimodalität**
- **Symmetrie und Schiefe**
- **Lage und Streuung**

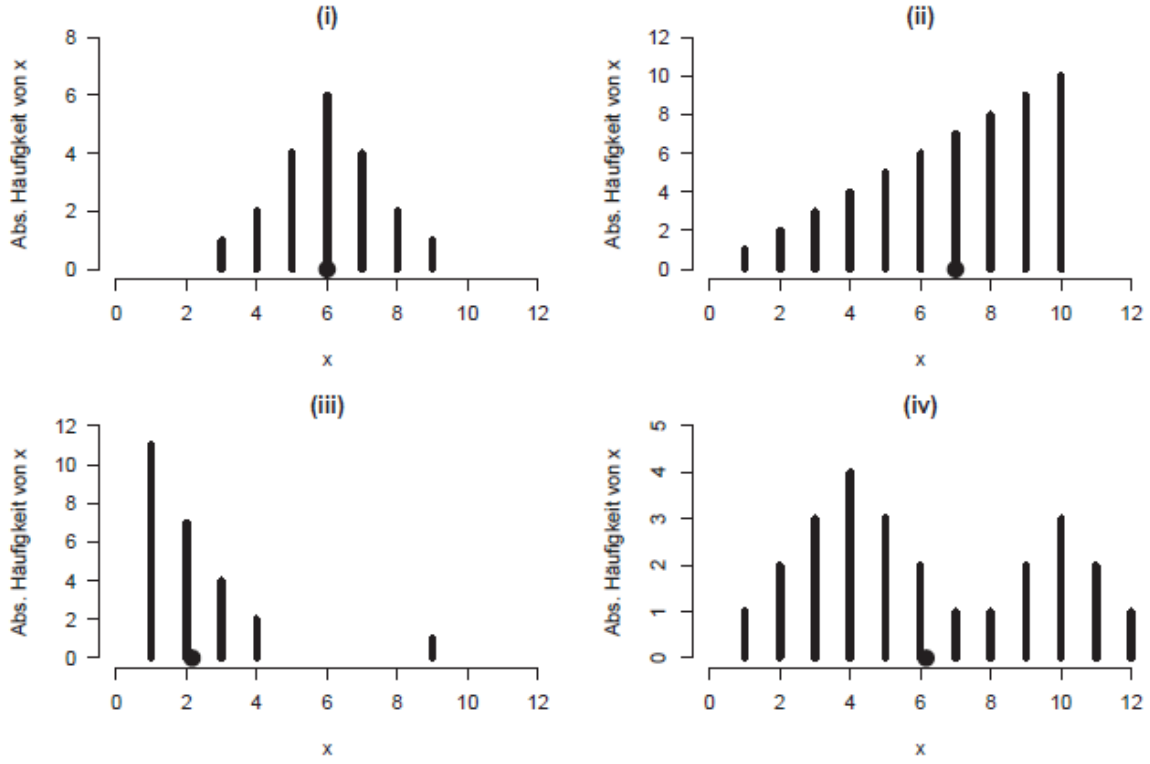
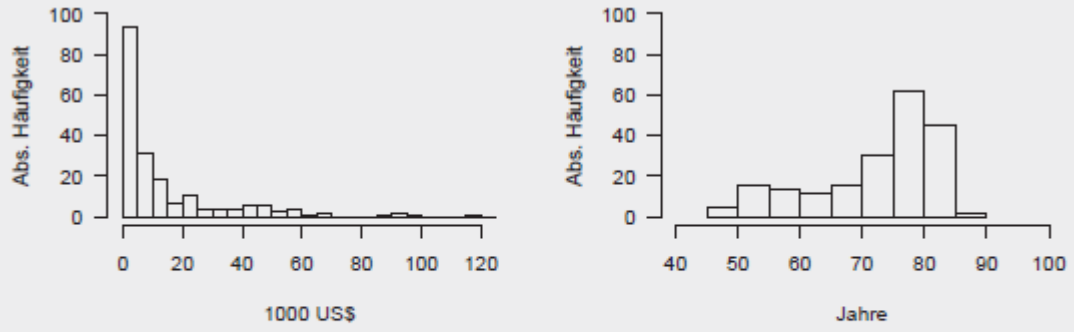


Abb. 4.1.1: Verteilungseigenschaften – Stilisierte Beispiele

4.1 Spezifika empirischer Verteilungen

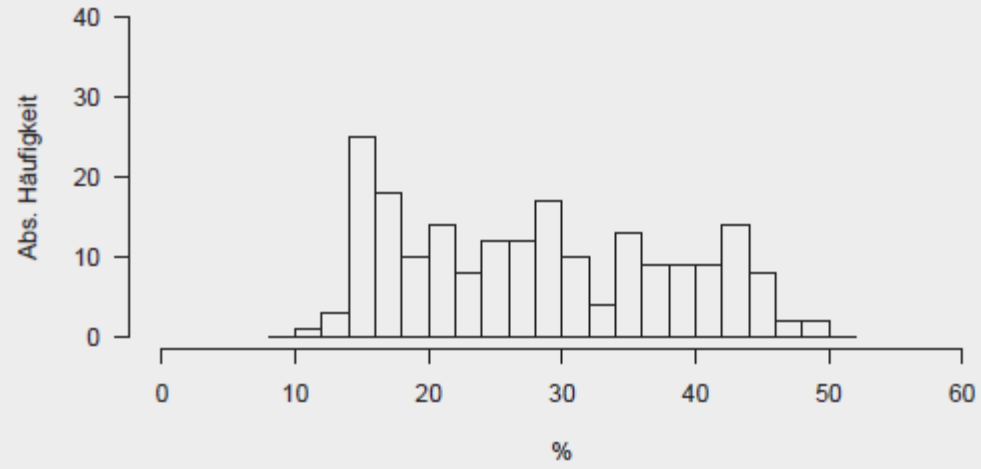
Beispiel 4.1.1: Schiefe und Multimodalität



Daten: United Nations Statistics Division [2014]

Abb. 4.1.2: Weltweites Pro-Kopf-BIP und weltweite Lebenserwartung

4.1 Spezifika empirischer Verteilungen



Daten: United Nations Statistics Division [2014]

Abb. 4.1.3: Weltweiter Anteil der unter 15-Jährigen an der Gesamtbevölkerung

4.1 Spezifika empirischer Verteilungen



Daten: United Nations Statistics Division [2014]

Abb. 4.1.4: Weltkarte: Anteil der unter 15-Jährigen im Jahr 2012

4.2.1 Arithmetisches Mittel

● Definition und Berechnung ●

> Gegeben metrisch skalierte Beobachtungswerte $x_1, x_2, \ldots x_n$:

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

> **Merkmalssumme**

$$\sum_{i=1}^{n} x_i = n \times \bar{x}$$

● Interpretation ●

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

4.2 Lagekennwerte

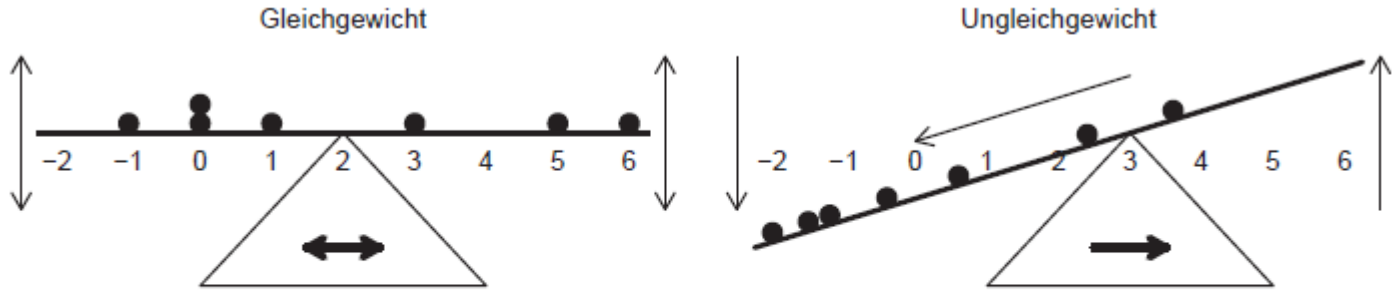


Abb. 4.2.1: Arithmetisches Mittel als physikalischer Schwerpunkt

• Berechnung bei klassierten Daten •

Tab. 4.2.1: Häufigkeitsverteilung mit Klassenmitten

j	Klasse $(c_{j-1}, c_j]$	$\tilde{n}_j$	$\tilde{f}_j$	m_j
1	(0,10]	2	0.067	5
2	(10,20]	8	0.267	15
3	(20,30]	10	0.333	25
4	(30,40]	6	0.200	35
5	(40,50]	3	0.100	45
6	(50,60]	1	0.033	55
Σ		30	1	

4.2 Lagekennwerte

> Feststellung: Bei gleichmäßiger Verteilung aller Werte innerhalb der Klassen stimmen **Klassenmittelwert** und **Klassenmitte** ungefähr überein

> Im Beispiel (Tab. 4.2.1) gilt deshalb mit $k = 6$ Klassen:

$$\bar{x} \approx \frac{1}{n}\sum_{j=1}^{6} m_j \tilde{n}_j$$

$$= \frac{1}{30}(5 \times 2 + 15 \times 8 + 25 \times 10 + 35 \times 6 + 45 \times 3 + \underbrace{55}_{m_6} \times \overbrace{1}^{\tilde{n}_6})$$

$$= 26$$

m_j ... Klassenmitte der j-ten Klasse

● **Kein robuster Kennwert** ●

Arithmetisches Mittel

Für metrische Beobachtungswerte $x_1, \ldots, x_n$ berechnet sich das arithmetische Mittel als

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Unter Vorgabe absoluter oder relativen Häufigkeiten aller vorkommenden Ausprägungen a_j, für $j = 1, \ldots, k$, gilt entsprechend

$$\overline{x} = \frac{1}{n} \sum_{j=1}^{k} a_j n_j = \sum_{j=1}^{k} a_j f_j.$$

Liegen die Werte in k Größenklassen mit Klassenmitten m_j vor, gilt unter Vorgabe absoluter oder relativer Klassenhäufigkeiten näherungsweise

$$\overline{x} \approx \frac{1}{n} \sum_{j=1}^{k} m_j \tilde{n}_j = \sum_{j=1}^{k} m_j \tilde{f}_j.$$

4.2.2 Median

- **Definition und Interpretation**

$-1, 0, 0, 1, 3, 5, 6$ $\quad\quad$ $-1, 0, \boxed{1, 3,} 5, 6$

$\uparrow$ $\tilde{x}_{0.5} = 1$ $\quad\quad$ $\uparrow$ $\tilde{x}_{0.5} = 2$

- **Berechnung bei Urlisten**

$x_1 = 0,\ x_2 = 1,\ \ x_3 = 5,\ \ x_4 = 6,\ \ x_5 = 3,\ \ x_6 = 0,\ \ x_7 = -1$

Geordnete Werte:

$x_{(1)} = -1,\ x_{(2)} = 0,\ \ x_{(3)} = 0,\ \ x_{(4)} = 1, x_{(5)} = 3,\ x_{(6)} = 5,\ \ x_{(7)} = 6$

$n = 7$ ungerade:

$$\tilde{x}_{0.5} = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{7+1}{2}\right)} = x_{(4)} = 1$$

$$\tilde{x}_{0.5} = 0.5\left(x_{(n/2)} + x_{(n/2+1)}\right)$$ falls n gerade wäre

• Berechnung bei klassierten Daten •

Tab. 4.2.2: Berechnung des Medians bei klassierten Daten

j	Klasse $(c_{j-1}, c_j]$	10	$\tilde{f}_j$	$\tilde{f}_j/d_j$	$\tilde{F}_n(c_j)$
1	(0,10]	10	0.067	0.0067	0.067
2	(10,20]	10	0.267	0.0267	0.333
3	**(20,30]**	**10**	**0.333**	**0.0333**	**0.667**
4	(30,40]	10	0.200	0.0200	0.867
5	(40,50]	10	0.100	0.0100	0.967
6	(50,60]	10	0.033	0.0033	1
Σ			1		

Einfallsklasse des Medians ist hier $j = 3$

$c_{j-1} = c_2 = 20, \qquad \tilde{F}_n(c_2) = 0.333, \qquad \tilde{f}_3/d_3 = 0.0333$

4.2 Lagekennwerte

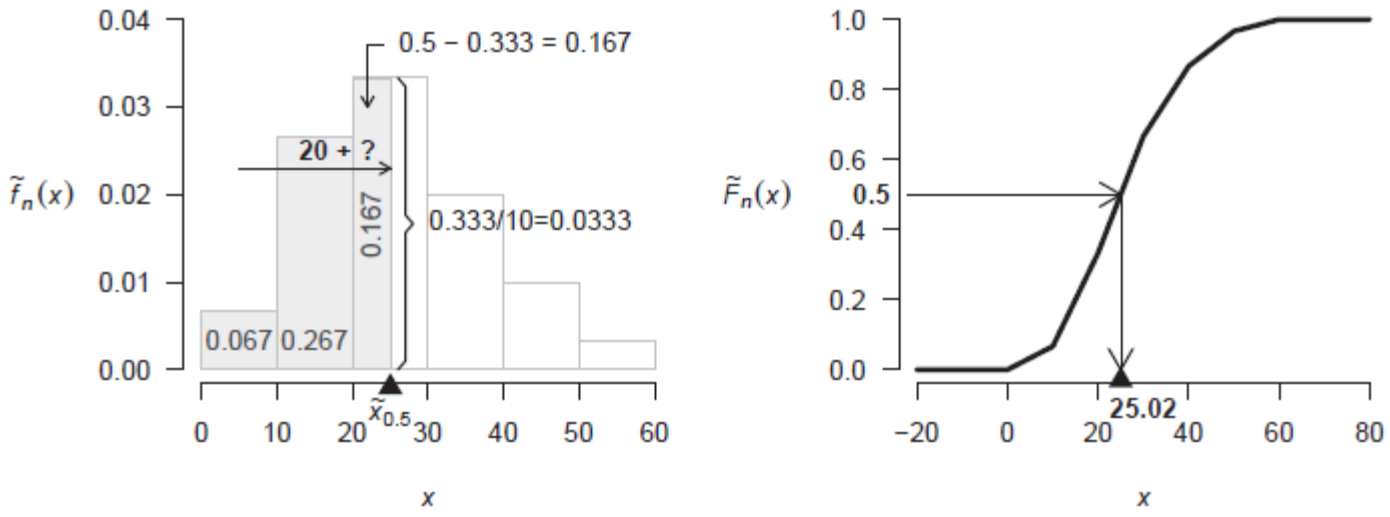


Abb. 4.2.2: Berechnung des Medians und Bestimmung aus der EVF

$$\tilde{x}_{0.5} \approx 20 + \frac{0.5 - 0.333}{0.0333} \approx 25.02$$

Geordnete Werte, Minimum, Maximum und Median

Gegeben seien metrische Beobachtungswerte $x_1, \ldots, x_n$. Die **geordneten Werte** werden mit $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ notiert. Es gilt:

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq \cdots \leq x_{(n)}.$$

Dabei bezeichnet $x_{(1)}$ das **Minimum** (den kleinsten Wert) und $x_{(n)}$ das **Maximum** (den größten Wert).

Der Median ist definiert als

$$\tilde{x}_{0.5} = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{für } n \text{ ungerade,} \\ \frac{1}{2}\left(x_{(n/2)} + x_{(n/2+1)}\right), & \text{für } n \text{ gerade.} \end{cases}$$

Liegen die Werte klassiert in k Größenklassen $(c_0, c_1], (c_1, c_2], \ldots, (c_{k-1}, c_k]$ vor, wird der Median näherungsweise bestimmt über

$$\tilde{x}_{0.5} = c_{j-1} + \frac{(0.5 - \tilde{F}_n(c_{j-1}))}{\tilde{f}_j / d_j},$$

falls er in die j-te Klasse hineinfällt.

4.2.3 Modalwert

Modalwert

Der **Modalwert** oder **Modus**, notiert mit x_{mod}, ist der am häufigsten vorkommende Wert einer Urliste. Im Falle klassierter Daten lässt er sich als Klassenmitte der Klasse mit der höchsten Häufigkeitsdichte definieren.

4.2 Lagekennwerte

4.2.4 Fechner‘sche Lageregeln

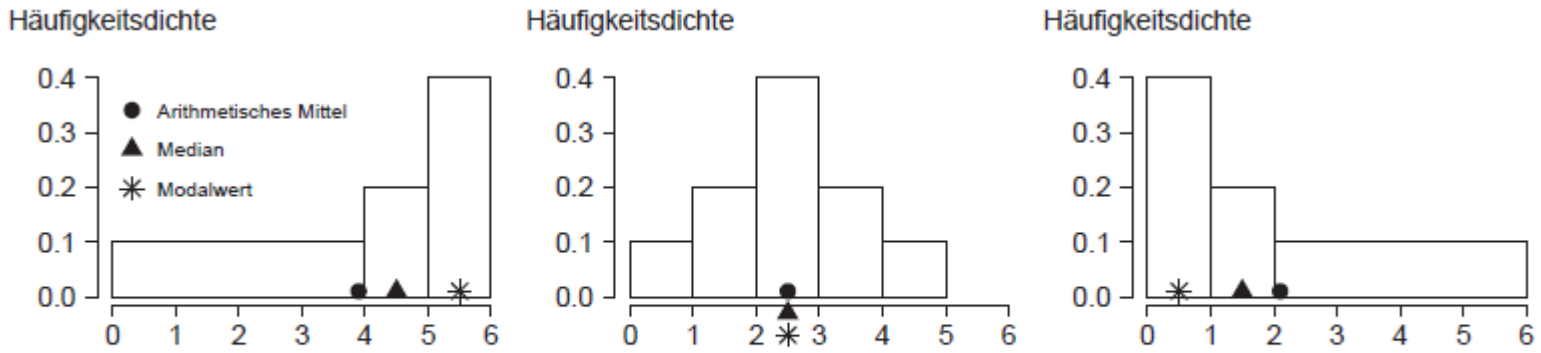


Abb. 4.2.3: Lageregeln: Arithmetisches Mittel, Median und Modalwert

Fechner’sche Lageregeln

Bei einer unimodalen Verteilung eines metrischen Merkmals gilt:

$$\overline{x} = \tilde{x}_{0.5} = x_{\text{mod}} \qquad \text{bei (perfekter) Symmetrie,}$$

$$x_{\text{mod}} < \tilde{x}_{0.5} < \overline{x} \qquad \text{bei Rechtsschiefe,}$$

$$\overline{x} < \tilde{x}_{0.5} < x_{\text{mod}} \qquad \text{bei Linksschiefe.}$$

Beispiel 4.2.1: Durchschnittseinkommen, Bundesmedian und Armutsgefährdung

Tab. 4.2.3: Durchschnitt und Median: Einkommen in Deutschland 2012 (in Euro)

Bezugsgröße/Einkommensart	Durchschnitt	Median
Haushalte/Bruttoeinkommen	43 078	34 375
Haushalte/Verfügbares Einkommen	31 644	26 263
Nettoäquivalenzeinkommen	22 022	19 595
Personen/Bruttoeinkommen	24 323	19 200

Quelle: Statistisches Bundesamt [2014a]

4.3 Spezielle Lagekennwerte

4.3.1 Arithmetisches Mittel bei gruppierten Daten

• Hintergrund •

Tab. 4.3.1: Haushaltsgröße in 3 Stadtbezirken

Stadtbezirk	Durchschnittliche Haushaltsgröße	Anzahl von Haushalten
Bezirk 1	1.5	282
Bezirk 2	2.4	585
Bezirk 3	1.6	250

$\bar{x}_3 = 1.6$ $\tilde{n}_3 = 250$

$k = 3$ Gruppen

$n = 282 + 585 + 250$
$= 1117$

• Berechnung •

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{k} \bar{x}_j \tilde{n}_j = \frac{1}{1117}(1.5 \times 282 + 2.4 \times 585 + 1.6 \times 250)$$
$$\approx 1.99$$

4.3 Spezielle Lagekennwerte

● Klassierung als Spezialfall ●

> Gruppen sind gegeben als Größenklassen

> Problem: Klassenmittelwerte häufig unbekannt

> Klassenmitten ersetzen approximativ die Klassenmittel $m_j \approx \bar{x}_j$

Arithmetisches Mittel bei gruppierten Daten

Gegeben seien n metrische Beobachtungswerte, gruppiert in k Gruppen mit jeweiligen Gruppenmittelwerten $\bar{x}_j$ und **Gruppenumfängen** $\tilde{n}_j$, für $j = 1, \ldots, k$. Dann gilt:

$$\bar{x} = \frac{1}{n} \sum_{j=1}^{k} \bar{x}_j \tilde{n}_j = \sum_{j=1}^{k} \bar{x}_j \tilde{f}_j,$$

wobei $\tilde{f}_j = \tilde{n}_j / n$ das **Gruppengewicht** der j-ten Gruppe bezeichnet. In diesem Zusammenhang spricht man auch von einem **gewichteten Mittel.**

4.3.2 Quantile

● Definition und Interpretation ●

> Ein ***x*%-Quantil** wird (grob) gesagt von *x*% der Werte unterschritten und von (100–*x*)% überschritten

> Speziell: **50%-Quantil** oder **0.5-Quantil** = **Median**

● Berechnung bei Urlisten ●

Werte geordnet!

5588, 8524, 12601, 12941, 13007, 15144, 17160, **18320,** 18428, 19207,
20112, 20556, 21086, 21316, 22138, 24748, 25432, 27936, 28791, 29901,
32770, 36180, 37176, 37322, 38696, 39070, **41004, 44981,** 47688, 51525.

z. B. 0.25- und 0.9-Quantil:

$0.25 \times 30 = 7.5 \notin \mathbb{N}$ $\qquad$ $0.9 \times 30 = 27 \in \mathbb{N}$

$\tilde{x}_{0.25} = x_{([7.5]+1)} = x_{(8)} = 18\,320$

$\tilde{x}_{0.9} = 0.5\left(x_{(27)} + x_{(28)}\right) = 0.5(41004 + 44981) = 42992.5$

• Berechnung bei klassierten Daten •

Tab. 4.3.2: Berechnung von Quantilen bei klassierten Daten

j	Klasse $(c_{j-1}, c_j]$	d_j	$\tilde{f}_j$	$\tilde{f}_j/d_j$	$\tilde{F}_n(c_j)$
1	(0,10]	10	0.067	0.0067	0.067
2	**(10,20]**	**10**	**0.267**	**0.0267**	**0.333**
3	(20,30]	10	0.333	0.0333	0.667
4	(30,40]	10	0.200	0.0200	0.867
5	**(40,50]**	**10**	**0.100**	**0.0100**	**0.967**
6	(50,60]	10	0.033	0.0033	1
Σ			1		

Einfallsklasse des 0.25-Quantils ist hier $j = 2$

$c_{j-1} = c_1 = 10, \qquad \tilde{F}_n(c_1) = 0.067, \qquad \tilde{f}_2/d_2 = 0.0267$

4.3 Spezielle Lagekennwerte

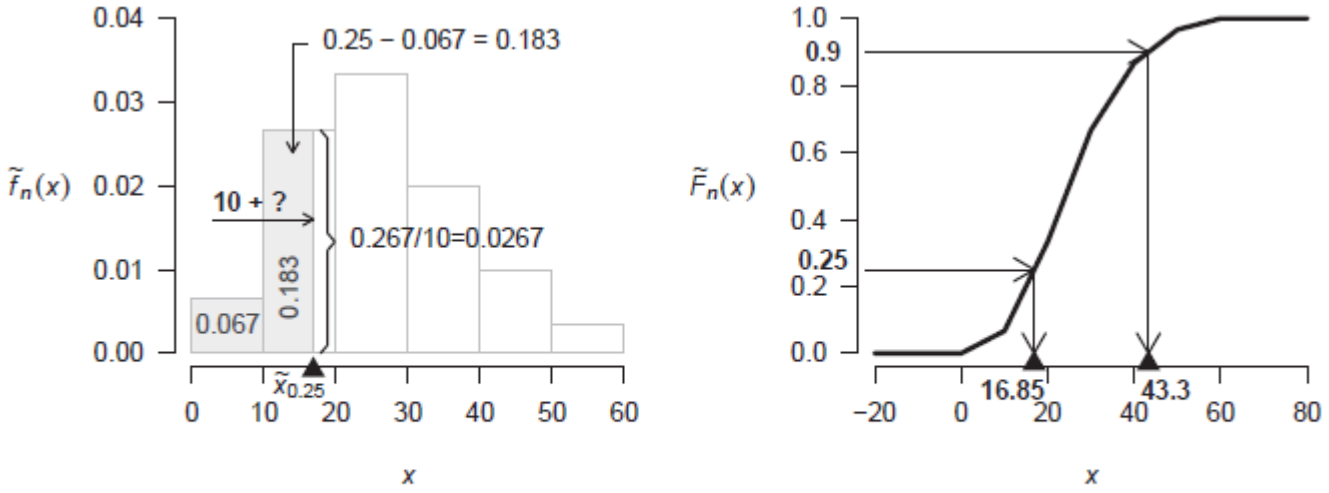


Abb. 4.3.1: Berechnung des 0.25-Quantils und Bestimmung von Quantilen aus der EVF

$$\tilde{x}_{0.25} \approx 10 + \frac{0.25 - 0.067}{0.0267} \approx 16.85$$

$$\tilde{x}_{0.9} \approx 40 + \frac{0.9 - 0.867}{0.01} \approx 43.3$$

Quantile

Gegeben seien metrische Beobachtungswerte $x_1, \ldots, x_n$. Dann ist für $\alpha \in (0, 1)$, das α-**Quantil** gegeben durch

$$\tilde{x}_\alpha = \begin{cases} x_{([n\alpha]+1)}, & \text{falls } n\alpha \text{ keine natürliche Zahl ist,} \\ \frac{1}{2}\left(x_{(n\alpha)} + x_{(n\alpha+1)}\right), & \text{falls } n\alpha \text{ eine natürliche Zahl ist,} \end{cases}$$

wobei „[...]“ die Gauß-Klammer bezeichnet.

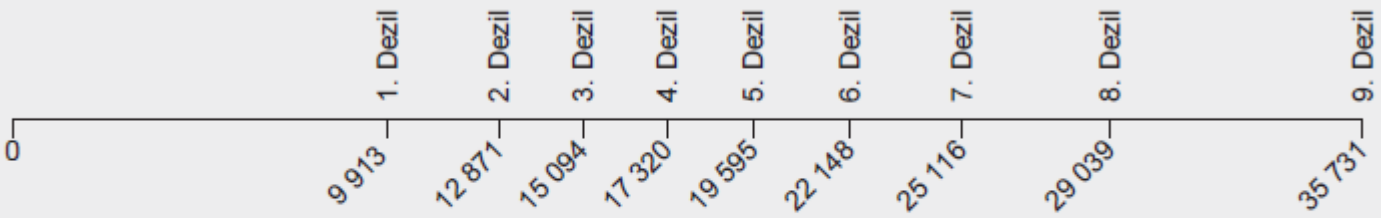
Liegen die Werte klassiert in k Größenklassen $(c_0, c_1], (c_1, c_2], \ldots, (c_{k-1}, c_k]$ vor, wird das α-Quantil näherungsweise bestimmt über

$$\tilde{x}_\alpha \approx c_{j-1} + \frac{(\alpha - \tilde{F}_n(c_{j-1}))}{\tilde{f}_j / d_j} \quad \text{für } j = 1, \ldots, k,$$

falls es in die j-te Klasse hineinfällt.

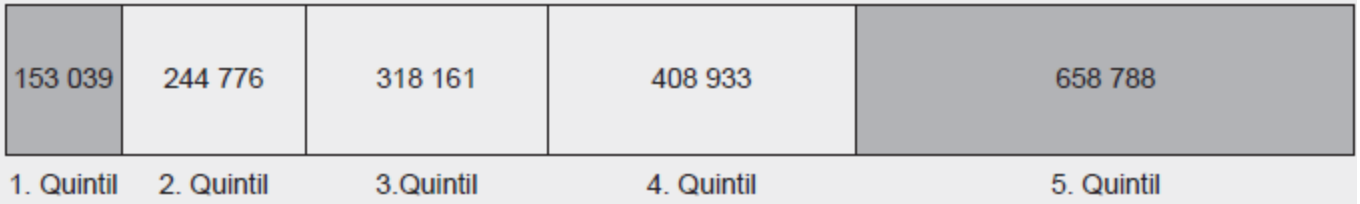
4.3 Spezielle Lagekennwerte

Beispiel 4.3.1: Dezile und Quintilsverhältnis der Einkommensverteilung



Daten: Statistisches Bundesamt [2014a]

Abb. 4.3.2: Dezile des Nettoäquivalenzeinkommen in Euro (2012)



Daten: Statistisches Bundesamt [2014a]

Abb. 4.3.3: Einkommensvolumina nach Quintilen in Mio. Euro

4.3.3 Geometrisches Mittel

• Hintergrund •

Umsätze eines Unternehmens

Jahr	2010	2011	2012	2013
Umsatz	1000	1100	990	1287

$+10\%\ \ -10\%\ \ +30\%$ (Wachstumsraten)

Problem: $\frac{1}{3}(0.1 - 0.1 + 0.3) = 0.1$ aber



$1000 \times 1.1 \times 1.1 \times 1.1 = 1331 \neq 1287$

Angabe des durchschnittlichen Wachstums mit 10% nicht sinnvoll

4.3 Spezielle Lagekennwerte

● Wachstumsfaktoren und Wachstumsraten ●

> **Wachstumsfaktoren** im Beispiel: 1.1, 0.9, 1.3

z.B. $1.3 = \frac{1287}{990}$ („späterer Zeitpunkt/früherer Zeitpunkt“)

> **Wachstumsraten** im Beispiel: +10%, –10%, +30%

Wachstumrate =(Wachstumsfaktor – 1) × 100%

● Definition und Berechnung ●

$$\bar{x}_{geom} = (1.1 \times 0.9 \times 1.3)^{1/3} \approx 1.0877$$

Alternativ: $\bar{x}_{geom} = (1287/1000)^{1/3} \approx 1.0877$

Damit gilt tatsächlich:

$1000 \times \bar{x}_{geom} \times \bar{x}_{geom} \times \bar{x}_{geom} = 1000 \times 1.0877^3 = 1287$

4.3 Spezielle Lagekennwerte

Geometrisches Mittel

Seien $x_0, x_1, \ldots, x_n$ zeitlich aufeinanderfolgende metrische Werte einer Zeitreihe korrespondierend zu den Zeitpunkten $0, 1, 2, \ldots, n$. Dann heißen

$$W_i = \frac{x_i}{x_{i-1}} \quad \text{bzw.} \quad R_i = \frac{x_i - x_{i-1}}{x_{i-1}} = W_i - 1$$

Wachstumsfaktor bzw. **Wachstumsrate** zum Zeitpunkt i für $i = 1, \ldots, n$. Das **geometrische Mittel** der Wachstumsfaktoren ist definiert als

$$\overline{x}_{geom} = (W_1 \times W_2 \times \cdots \times W_n)^{1/n}.$$

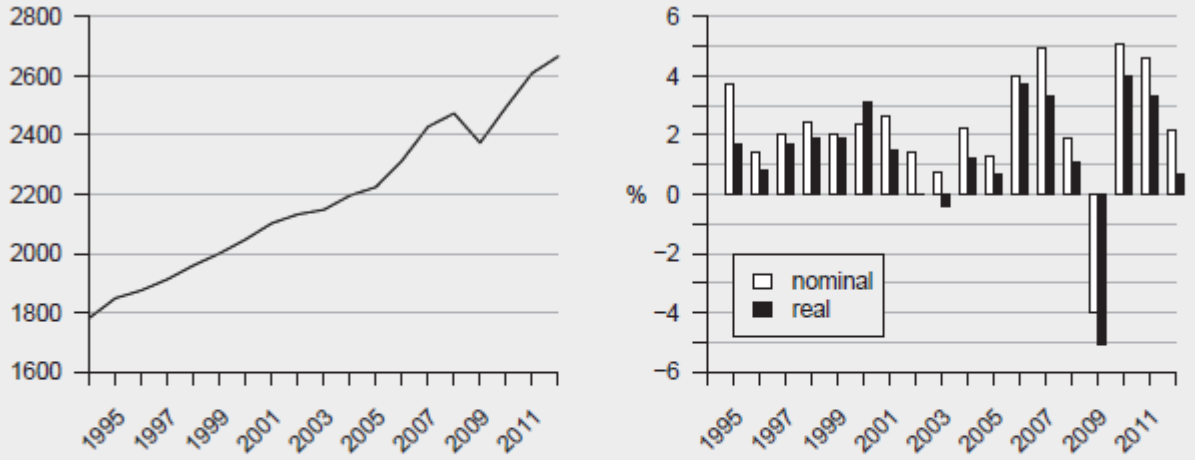
Dabei gilt:

$$x_0 \times \overline{x}_{geom}^n = x_n \quad \text{bzw.} \quad \overline{x}_{geom} = \left(\frac{x_n}{x_0}\right)^{1/n}.$$

Die durchschnittliche prozentuale Wachstumsrate beträgt $100 \times (\overline{x}_{geom} - 1)\%$.

4.3 Spezielle Lagekennwerte

Beispiel 4.3.2: Wirtschaftswachstum in Deutschland



Daten: Statistisches Bundesamt [2014b]

Abb. 4.3.4: Bruttoinlandsprodukt für Deutschland 1994–2012 (nominal, in Mrd. Euro) und jährliche Wachstumsraten (nominal und real)

4.3 Spezielle Lagekennwerte

Tab. 4.3.3: Reales BIP-Wachstum (prozentuale Änderung zum Vorjahr)

Jahr	1995	1996	1997	1998	1999	2000	2001	2002	2003
%	1.7	0.8	1.7	1.9	1.9	3.1	1.5	0.0	−0.4
Jahr	2004	2005	2006	2007	2008	2009	2010	2011	2012
%	1.2	0.7	3.7	3.3	1.1	−5.1	4.0	3.3	0.7

Daten: Statistisches Bundesamt [2014b]

> Durchschnittliches Wachstum des realen BIP von 1994 bis 2007

$$\bar{x}_{geom} = (1.017 \times 1.008 \times 1.017 \times \cdots \times 1.033)^{1/13} \approx 1.016$$

> Durchschnittliches Wachstum des realen BIP von 2008 bis 2012

$$\bar{x}_{geom} = (1.011 \times 0.949 \times 1.040 \times 1.033 \times 1.007)^{1/5} \approx 0.007$$

> Durchschnittliches Wachstum des nominalen BIP von 1994 bis 2012

$$\bar{x}_{geom} = (2666.4/1782.2)^{1/18} \approx 1.023$$

4.4 Streuungskennwerte

4.4.1 Spannweite

$-1, 0, 0, 1, 3, 5, 6$ (geordnet!)

$$\mathrm{R} = x_{(7)} - x_{(1)} = 6 - (-1) = 7$$

Spannweite

Gegeben seien metrische Beobachtungswerte $x_1, \ldots, x_n$. Dann ist die Spannweite gegeben durch

$$R = x_{(n)} - x_{(1)},$$

wobei $x_{(n)}$ der größte (**Maximum**) und $x_{(1)}$ der kleinste Wert (**Minimum**) ist.

4.4.2 Mittlere absolute Abweichungen

● Definition und Berechnung ●

$$-1, 0, 0, 1, 3, 5, 6$$

$$\Longrightarrow \quad \bar{x} = 2:$$

$$d^* = \frac{1}{7}(|-1-2| + |0-2| + |0-2| + |1-2| + |3-2| + |5-2| + |6-2|)$$

$$= \frac{1}{7}(3 + 2 + 2 + 1 + 1 + 3 + 4) \approx 2.29$$

$$\Longrightarrow \quad \tilde{x}_{0.5} = 1:$$

$$d = \frac{1}{7}(|-1-1| + |0-1| + |0-1| + |1-1| + |3-1| + |5-1| + |6-1|)$$

$$= \frac{1}{7}(2 + 1 + 1 + 0 + 2 + 4 + 5) \approx 2.14$$

4.4 Streuungskennwerte

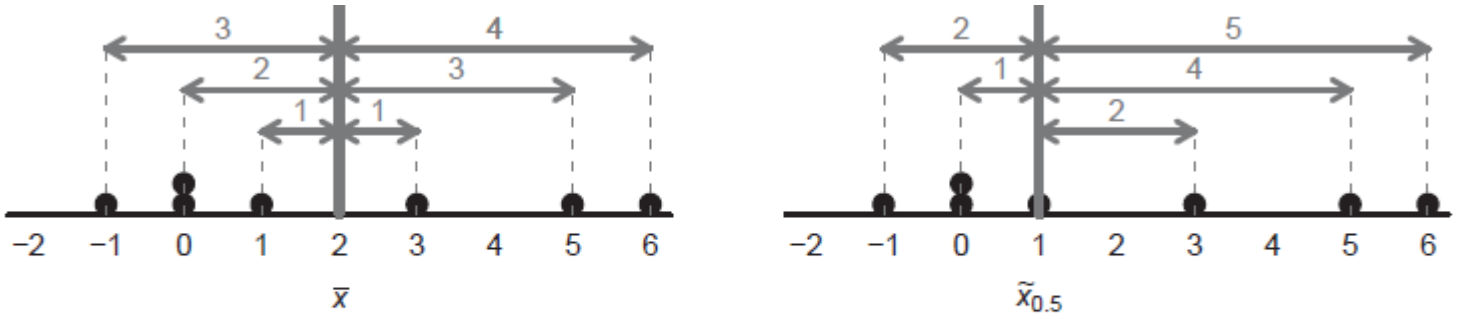


Abb. 4.4.1: Absolute Abweichungen

• Interpretation •

> d^* ... Durchschnittsabstand aller Werte zum arithmetischen Mittel
(**Mittlere absolute Abweichung vom arithmetischen Mittel**)

> d ... Durchschnittsabstand aller Werte zum Median
(**Mittlere absolute Abweichung vom Median**)

4.4 Streuungskennwerte

• Berechnung bei klassierten Daten •

Tab. 4.2.1: Häufigkeitsverteilung mit Klassenmitten

j	Klasse $(c_{j-1}, c_j]$	$\tilde{n}_j$	$\tilde{f}_j$	m_j
1	(0,10]	2	0.067	5
2	(10,20]	8	0.267	15
3	(20,30]	10	0.333	25
4	(30,40]	6	0.200	35
5	(40,50]	3	0.100	45
6	(50,60]	1	0.033	55
Σ		30	1	

$$\Rightarrow \quad \bar{x} = 26 \quad \text{(Folie 71)}$$

$$\Rightarrow \quad \tilde{x}_{0.5} = 25.02 \quad \text{(Folie 75)}$$

$$d^* \approx \frac{1}{30}(2 \times |5 - 26| + 8 \times |15 - 26| + \cdots + 1 \times |55 - 26|) \approx 9.33$$

$$d \approx \frac{1}{30}(2 \times |5 - 25.02| + 8 \times |15 - 25.02| + \cdots + 1 \times |55 - 25.02|) \approx 9.00$$

4.4 Streuungskennwerte

- **Median als präferierter Bezugswert** -

> ... d theoretisch fundiert über **Minimumeigenschaft des Medians**

> ... d^* eher unüblich

Mittlere absolute Abweichungen

Gegeben seien metrische Beobachtungswerte $x_1, \ldots, x_n$. Dann heißen die Kennwerte

$$d^* = \frac{1}{n}\sum_{i=1}^{n}|x_i - \overline{x}| \quad \text{und} \quad d = \frac{1}{n}\sum_{i=1}^{n}|x_i - \tilde{x}_{0.5}|$$

mittlere absolute Abweichung vom arithmetischen Mittel bzw. **mittlere absolute Abweichung vom Median.** Liegen die Werte klassiert in k Größenklassen mit Klassenmitten m_j vor, so gelten die approximativen Berechnungsformeln

$$d^* \approx \frac{1}{n}\sum_{j=1}^{k}\tilde{n}_j|m_j - \overline{x}| = \sum_{j=1}^{k}\tilde{f}_j|m_j - \overline{x}| \quad \text{bzw.}$$

$$d \approx \frac{1}{n}\sum_{j=1}^{k}\tilde{n}_j|m_j - \tilde{x}_{0.5}| = \sum_{j=1}^{k}\tilde{f}_j|m_j - \tilde{x}_{0.5}|,$$

wobei arithmetisches Mittel bzw. Median approximativ bestimmt werden.

4.4.3 Median absoluter Abweichungen

● Hintergrund ●

> Problem: Mittelwerte sind nicht **robust**, so auch nicht mittlere absolute Abweichungen (auch nicht wenn Bezugspunkt Median ist)

> Situation ohne Ausreißer

Beobachtungswerte: 1, 2, 3, 4, 5 $\Longrightarrow$ $\bar{x} = 3$ $\tilde{x}_{0.5} = 3$

Absolute Abweichungen: 2, 1, 0, 1, 2 $\Longrightarrow$ $d^* = 1.2$ $d = 1.2$

... vom arithmetischen Mittel bzw. Median

> Situation mit Ausreißer: 1, 2, 3, 4, 500 $\Longrightarrow$ $\bar{x} = 102$ $\tilde{x}_{0.5} = 3$

Absolute Abweichungen: 101, 100, 99,98,398 $\Longrightarrow$ $d^* = 159.2$

... vom arithmetischen Mittel

2, 1, 0, 1, 497 $\Longrightarrow$ $d = 100.2$

... vom Median

4.4 Streuungskennwerte

● Definition und Berechnung ●

Situation mit Ausreißer: 1, 2, 3, 4, 500 $\Longrightarrow$ $\tilde{x}_{0.5} = 3$

Absolute Abweichungen vom Median: 2, 1, 0, 1, 497

Geordnet: 0, 1, 1, 2, 497

Davon der Median: *MAD* = 1

MAD

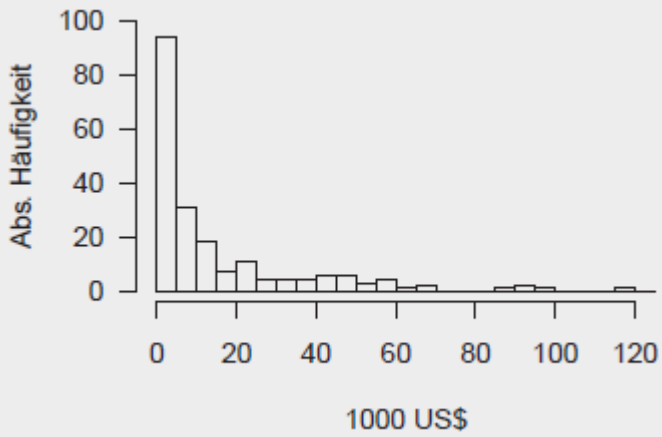
Gegeben seien metrische Beobachtungswerte $x_1, \ldots, x_n$. Die absoluten Abweichungen vom Median sind definiert als

$$|x_1 - \tilde{x}_{0.5}|,\ |x_2 - \tilde{x}_{0.5}|, \ldots, |x_n - \tilde{x}_{0.5}|.$$

Dann heißt der Median dieser Abweichungen **Median der absoluten Abweichungen vom Median** oder auch **MAD (Median Absolute Deviation)**.

4.4 Streuungskennwerte

Beispiel 4.4.1: Streuung des weltweiten Pro-Kopf-BIP (vgl. Abb. 4.1.2)



$\bar{x} = 14\,936$

$\tilde{x}_{0.5} = 5\,476$

$d = 12\,744$

$MAD = 4\,722$

(alle Einheiten in US-Dollar)

4.4.4 Varianz, Standardabweichung und Schwankungsintervalle

● Definition und Berechnung ●

> **Empirische Varianz** in „originärer Formel“

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

> **Empirische Standardabweichung**

$$\tilde{s} = \sqrt{\tilde{s}^2}$$

> Rechenbeispiel:

$$-1, 0, 0, 1, 3, 5, 6 \quad \Longrightarrow \quad \bar{x} = 2$$

$$\Longrightarrow \quad \tilde{s}^2 = \frac{1}{7}\left((-1-2)^2 + (0-2)^2 + (0-2)^2 + \cdots (6-2)^2\right) \approx 6.2857$$

$$\Longrightarrow \quad \tilde{s} = \sqrt{6.28571} \approx 2.51$$

● Verschiebungsformel für die empirische Varianz ●

> **Allgemeine Verschiebungsformel**

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

> Für $c = 0$ folgt daraus die **Verschiebungsformel für die emirische Varianz**

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

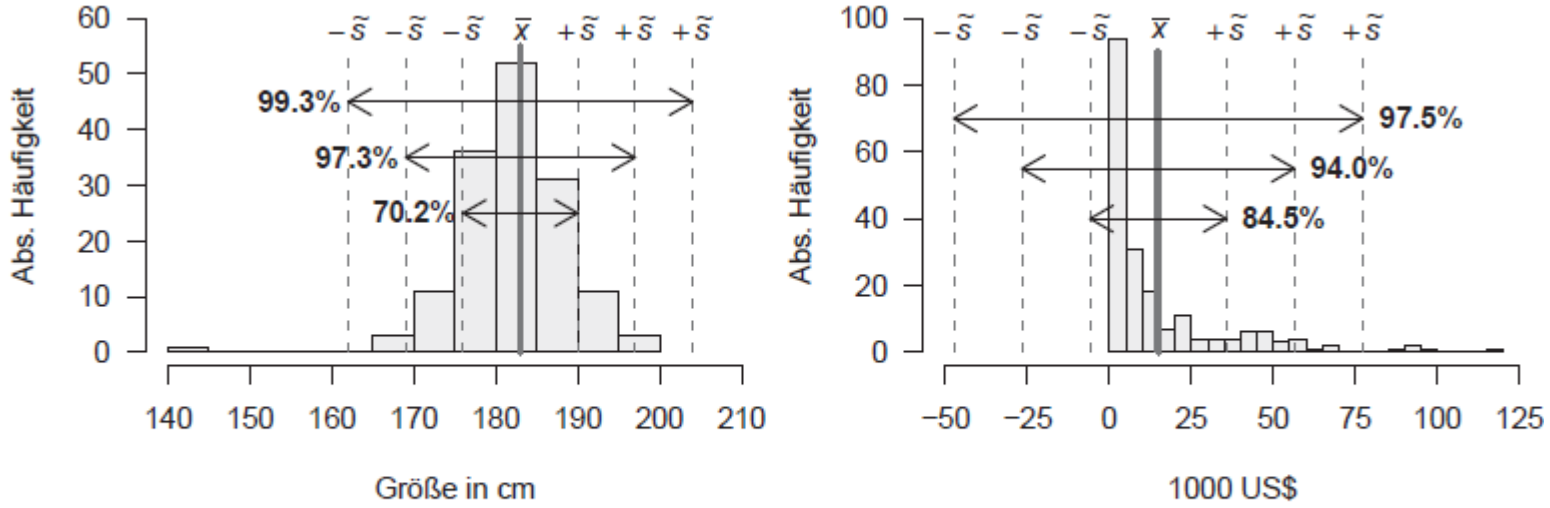
> Rechenbeispiel

$-1, 0, 0, 1, 3, 5, 6 \quad \Longrightarrow \quad \bar{x}=2$

$$\Longrightarrow \quad \tilde{s}^2=\frac{1}{7}\left((-1)^2+0^2+0^2+1^2+3^2+5^2+6^2\right)-2^2\approx 6.2857$$

4.4 Streuungskennwerte

• Standardabweichung und Interpretation •



Quelle: Eigenerhebung bzw. Vereinte Nationen

Abb. 4.4.2: Empirische Schwankungsintervalle mit empirischen Abdeckungsraten

Es liegen in den **empirischen Schwankungsintervallen**

$$[\bar{x} - \tilde{s}, \bar{x} + \tilde{s}], \quad [\bar{x} - 2\tilde{s}, \bar{x} + 2\tilde{s}] \text{ und } \quad [\bar{x} - 3\tilde{s}, \bar{x} + 3\tilde{s}]$$

ca. 68% 95% bzw. 99%

aller Beobachtungswerte.

4.4 Streuungskennwerte

• Hintergründe •

Bedeutendstes Streuungsmaß in der Statistik aus verschiedenen Gründen...

Varianz und Standardabweichung

Gegeben seien metrische Beobachtungswerte $x_1, \ldots, x_n$. Dann ist durch

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

die (***empirische***) ***Varianz*** und durch

$$\tilde{s} = \sqrt{\tilde{s}^2}$$

die (***empirische***) ***Standardabweichung*** gegeben.

Liegen die Werte klassiert in k Größenklassen mit Klassenmitten m_j vor, so wird die Varianz näherungsweise über

$$\tilde{s}^2 \approx \frac{1}{n} \sum_{j=1}^{k} \tilde{n}_j (m_j - \overline{x})^2 = \sum_{j=1}^{k} \tilde{f}_j (m_j - \overline{x})^2$$

berechnet, wobei das arithmetische Mittel approximativ bestimmt wird.

4.5 Spezielle Streuungskennwerte

4.5.1 Varianz bei gruppierten Daten

• Berechnung •

Streuungszerlegungsformel

Gegeben seien n metrische Beobachtungswerte gruppiert in k Gruppen mit jeweiligen Gruppenmittelwerten $\overline{x}_j$, Gruppenvarianzen $\tilde{s}_j^2$ und Gruppenumfängen $\tilde{n}_j$, für $j = 1, \ldots, k$. Dann gilt:

$$\tilde{s}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum_{j=1}^{k}\tilde{n}_j\tilde{s}_j^2 + \frac{1}{n}\sum_{j=1}^{k}\tilde{n}_j(\overline{x}_j - \overline{x})^2$$

$$= \sum_{j=1}^{k}\tilde{f}_j\tilde{s}_j^2 + \sum_{j=1}^{k}\tilde{f}_j(\overline{x}_j - \overline{x})^2,$$

wobei $\tilde{f}_j = \tilde{n}_j/n$ die Gruppenanteile (Gruppengewichte) bezeichnen und

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n}x_i = \frac{1}{n}\sum_{j=1}^{k}\tilde{n}_j\overline{x}_j = \sum_{j=1}^{k}\tilde{f}_j\overline{x}_j$$

das Gesamtmittel ist. Die beiden Streuungsbestandteile

$$\sum_{j=1}^{k}\tilde{f}_j\tilde{s}_j^2 \quad \text{und} \quad \sum_{j=1}^{k}\tilde{f}_j(\overline{x}_j - \overline{x})^2$$

werden als ***interne Streuung*** bzw. ***externe Streuung*** bezeichnet. Es gilt:

Gesamtstreuung = Interne Streuung + Externe Streuung.

4.5 Spezielle Streuungskennwerte

• Beispiel 4.5.1 •

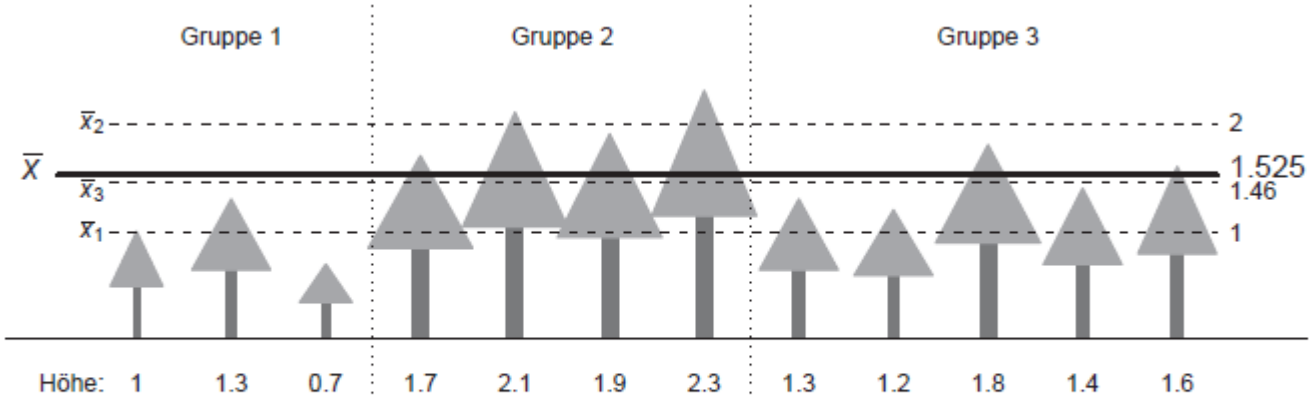


Abb. 4.5.1: Varianz bei gruppierten Daten – Streuungszerlegung

$\bar{x}_1 = 1.0$	$\bar{x}_2 = 2.0$	$\bar{x}_3 = 1.46$
$\tilde{s}_1^2 = 0.06$	$\tilde{s}_2^2 = 0.05$	$\tilde{s}_3^2 = 0.0464$
$\tilde{n}_1 = 3$	$\tilde{n}_2 = 4$	$\tilde{n}_3 = 5$

$$\Rightarrow \quad \bar{x} = \sum_{j=1}^{3} \tilde{f}_j \bar{x}_j = \frac{3}{12} \times 1.0 + \frac{4}{12} \times 2.0 + \frac{5}{12} \times 1.46 = 1.525$$

$$\Rightarrow \sum_{j=1}^{3} \tilde{f}_j \tilde{s}_j^2 = \frac{3}{12} \times 0.06 + \frac{4}{12} \times 0.05 + \frac{5}{12} \times 0.0464 = 0.051$$

$$\Rightarrow \sum_{j=1}^{3} \tilde{f}_j \left(\bar{x}_j - \bar{x}\right)^2 = \frac{3}{12} \times (1 - 1.525)^2 + \frac{4}{12} \times (2 - 1.525)^2 + \frac{5}{12} \times (1.46 - 1.525)^2$$

$$= 0.145875$$

$$\Rightarrow \tilde{s}^2 = 0.051 + 0.145875 = 0.196875$$

Intern Extern Gesamt

Tatsächlich erhält man diesen Wert auch auf „konventionellem Wege“:

$$\tilde{s}^2 = \frac{1}{12} \sum_{i=1}^{12} x_i^2 - \bar{x}^2 \quad = \frac{1}{12} (1^2 + 1.3^2 + \cdots + 1.6^2) - 1.525^2 = 0.196875$$

- **Interpretation**
- **Hintergründe**

4.5.2 Quantilsabstände

Quantilsabstände

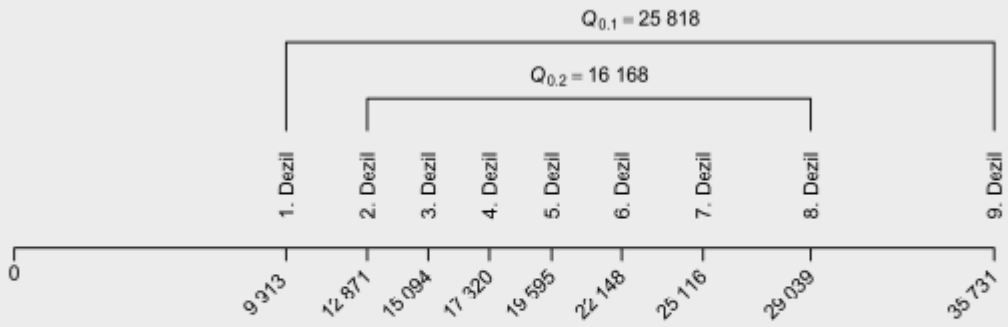
Für metrische Beobachtungswerte ist der α-**Quantilsabstand** definiert als

$$Q_\alpha = \tilde{x}_{1-\alpha} - \tilde{x}_\alpha \quad \text{für } \alpha \in (0, 0.5).$$

Er entspricht der Spannweite der mittleren $100\times(1-2\alpha)\%$ aller Werte. Im Speziellen heißt $Q_{0.25}$ auch **Interquartilsabstand.**

4.5 Spezielle Streuungskennwerte

Beispiel 4.5.2: Quantilsabstände der Einkommensverteilung



Daten: Statistisches Bundesamt [2014a]

Abb. 4.5.2: Quantilsabstände des Nettoäquivalenzeinkommens in Euro (2012)

$$Q_{0.1} = \tilde{x}_{0.9} - \tilde{x}_{0.1} = 35731 - 9913 = 25818$$

$$Q_{0.2} = \tilde{x}_{0.8} - \tilde{x}_{0.2} = 29039 - 12871 = 16168$$

4.5.3 Variationskoeffizient

● Hintergrund ●

> **Verschiebungsinvarianz** der Varianz ist manchmal störend:

1, 2, 3, 4, 5 ← identische Varianz → 15, 16, 17, 18, 19

> Beispiel:

Angenommen, im Rahmen einer Marktstudie wird die Streuung von Preisen für bestimmte Produkte bei 3 verschiedenen Lebensmitteldiscountern verglichen. Dabei ergeben sich folgende Preise für

100 g Speisezalz: 0.29, 0.39, 0.49 Euro
10 kg Waschmittel: 19.79, 19.89, 19.99 Euro

Somit erhalten wir:

$$\bar{x}_{Salz} = 0.39 \qquad \tilde{s}_{Salz} = 0.0816$$
$$\bar{x}_{Waschen} = 19.89 \qquad \tilde{s}_{Waschen} = 0.0816$$

Identisch!

Bezogen auf das Preisniveau variiert der Preis des Waschmittels relativ betrachtet weniger als der Preis des Salzes

4.5 Spezielle Streuungskennwerte

● Definition und Berechnung ●

> **Variationskoeffizient** als Maß der relativen Streuung:

$$v = \frac{\tilde{s}}{\bar{x}} \quad \text{nichtnegative Werte mit } \bar{x} > 0 \text{ vorausgesetzt}$$

> Beispiel von zuvor:

$$v_{Salz} = \frac{\tilde{s}_{Salz}}{\bar{x}_{Salz}} = \frac{0.0816}{0.39} \approx 0.209$$

$$v_{Waschen} = \frac{\tilde{s}_{Waschen}}{\bar{x}_{Waschen}} = \frac{0.0816}{19.89} \approx 0.004$$

● Interpretation ●

> Variationskoeffizient entspricht der Standardabweichung der prozentualen Abweichungen vom arithmetischen Mittel

4.5 Spezielle Streuungskennwerte

> Im Beispiel des Salzes mit $\bar{x}_{Salz} = 0.39$

Preise:	0.29	0.39	0.49
Abweichung in %:	-25.64%	0%	25.64%

z. B. $\frac{0.49-0.39}{0.39} \times 100\% = 25.64\%$

Standardabweichung :

$$\tilde{s} = \sqrt{\tilde{s}^2} = \sqrt{\frac{1}{3}\left((-25.64-0)^2 + (0-0)^2 + (25.64-0)^2\right)} \approx 20.9$$

Variationskoeffizient

Seien $x_1, \ldots, x_n$ nichtnegative metrische Beobachtungswerte mit $\bar{x} > 0$. Dann ist der **Variationskoeffizient** definiert als

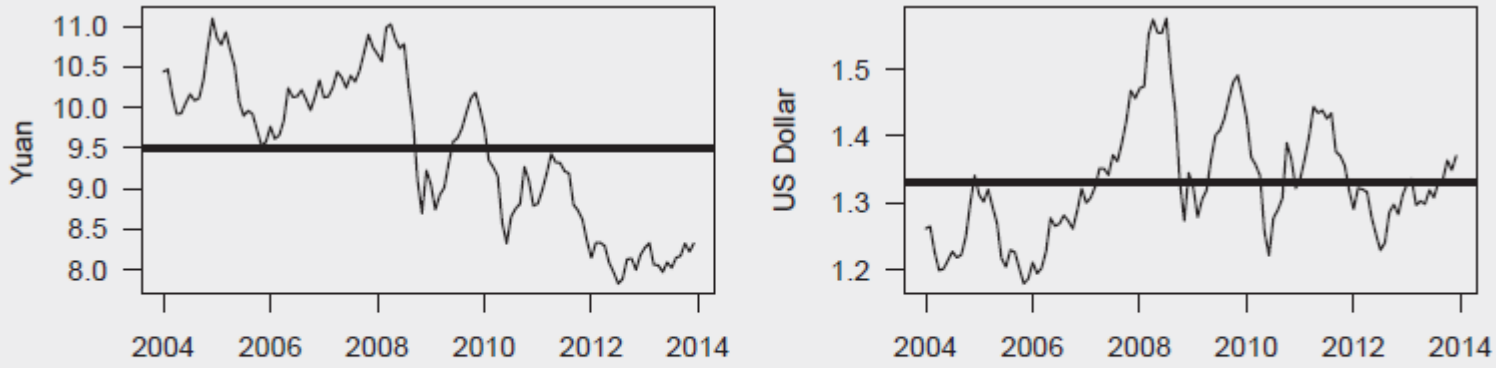
$$v = \tilde{s}/\bar{x}.$$

Der Wert $100 \times v$ entspricht der Standardabweichung der prozentualen Abweichungen vom arithmetischen Mittel, die definiert sind als

$$y_i = 100 \times \left(\frac{x_i}{\bar{x}} - 1\right) \quad \text{für } i = 1, \ldots, n.$$

4.5 Spezielle Streuungskennwerte

Beispiel 4.5.3: Variationsvergleich von Wechselkursen



Daten: Europäische Zentralbank [2014b]

Abb. 4.5.3: Wechselkurse des Euro; Jan. 2004–Dez. 2013

$$\bar{x}_{Yuan} = 9.49 \quad \tilde{s}_{Yuan} = 0.93 \quad v_{Yuan} = 0.10$$

$$\bar{x}_{Dollar} = 1.33 \quad \tilde{s}_{Dollar} = 0.09 \quad v_{Dollar} = 0.07$$

4.6 Standardisierung mittels Lage und Streuung

● Hintergrund ●

> Beobachtungswerte aus unterschiedlichen Grundgesamtheiten lassen sich nicht immer sinnvoll vergleichen

> Beispiel: Einkommensvergleich in zwei unterschiedlichen Ländern
Deutscher: 2800 Euro bei einem Durchschnitt von 2500 Euro und einer Standardabweichung von 150 Euro
Schweizer: 5500 Franken bei einem Durchschnitt von 5000 Franken und einer Standardabweichung von 400 Franken

● Berechnung und Interpretation ●

> Allgemeine Form:

Standardisierter Wert = (Originalwert – Lagewert) / Streuungswert.

> Spezialfall *Z*-Standardisierung:

$$z_i = \frac{x_i - \bar{x}}{\tilde{s}_X}$$

4.6 Standardisierung mittels Lage und Streuung

> Im Beispiel von zuvor

$$\text{Deutscher: } \frac{2800 - 2500}{150} = 2 \qquad \text{Schweizer: } \frac{5500 - 5000}{400} = 1.25$$

Fazit: Deutscher (Schweizer) liegt 2 (1.25) Standardabweichungen über dem Durchschnitt

● Eigenschaften *z*-standardisierter Werte ●

> Mittelwert *z*-standardisierter Werte ist stets 0:

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i = \frac{1}{n}\sum_{i=1}^{n} \frac{x_i - \bar{x}}{\tilde{s}_X}$$

$$= \frac{1}{n\tilde{s}_X}\sum_{i=1}^{n}(x_i - \bar{x}) = 0 \qquad \text{(vgl. Folie 69)}$$

4.6 Standardisierung mittels Lage und Streuung

> Varianz (Standardabweichung) z-standardisierter Werte ist stets 1:

$$\tilde{s}_Z^2 = \frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})^2 = \frac{1}{n}\sum_{i=1}^{n} z_i^2 - \bar{z}^2 = \frac{1}{n}\sum_{i=1}^{n} z_i^2$$

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

z-Standardisierung

Gegeben seien metrische Beobachtungswerte $x_1, \ldots, x_n$. Dann sind die z-standardisierten Werte gegeben durch

$$z_i = \frac{x_i - \overline{x}}{\tilde{s}_X} \quad \text{für } i = 1, \ldots, n.$$

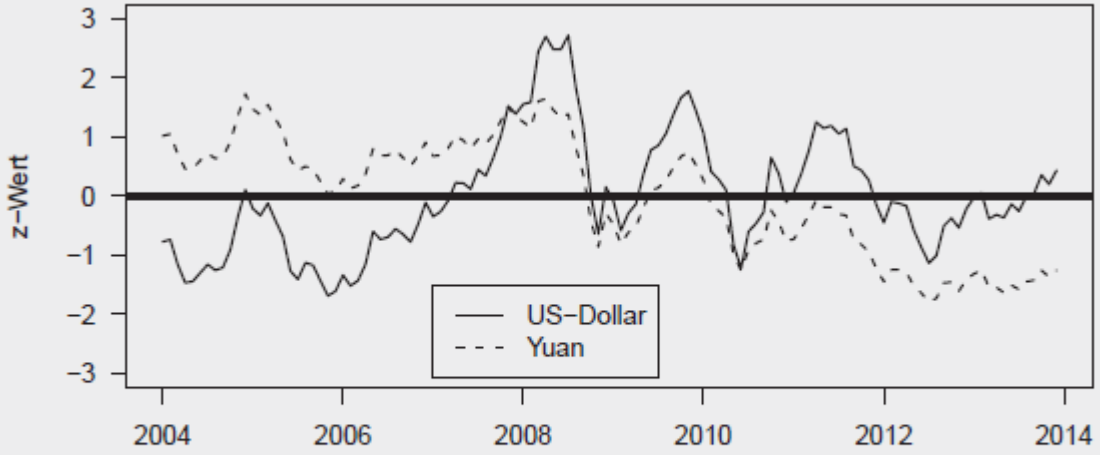
Es gilt:

$$\overline{z} = 0 \quad \text{und} \quad \tilde{s}_Z^2 = 1.$$

Z-standardisierte Werte sind verschiebungs- und skaleninvariant.

4.6 Standardisierung mittels Lage und Streuung

Beispiel 4.6.1: Standardisierte Zeitreihen



Daten: Europäische Zentralbank [2014b]

Abb. 4.6.1: Standardisierte Wechselkurse des Euro; Jan. 2004 – Dez. 2013

4.7 Messung von Schiefe

● Konzept und Definition ●

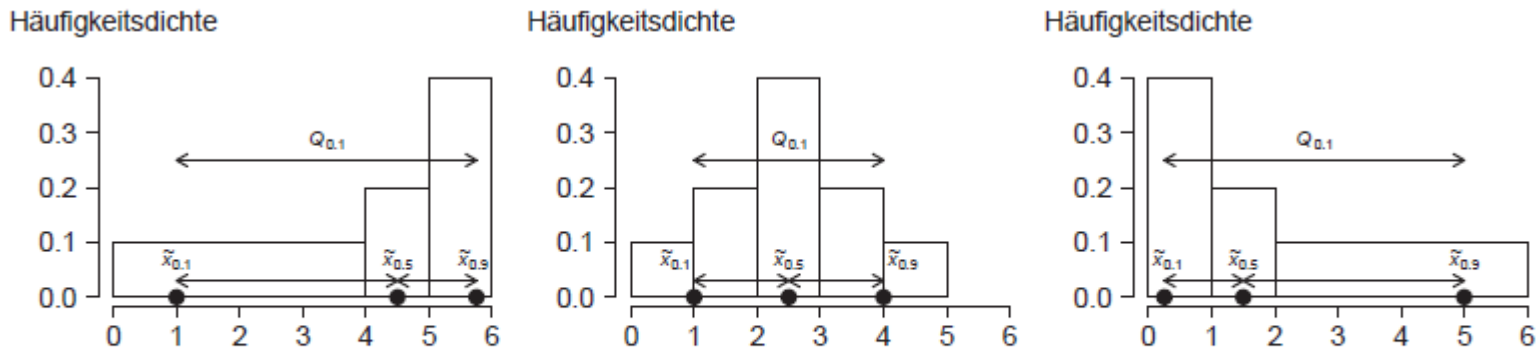


Abb. 4.7.1: Zerlegung des Quantilsabstands

> Feststellung: Bei schiefen Verteilungen liegen obere und untere Quantile gewöhnlich unterschiedlich weit vom Median entfernt

> Erster Vorschritt: Zerlegung des **Quantilsabstands** im Sinne von

$$\underbrace{\tilde{x}_{1-\alpha} - \tilde{x}_{\alpha}}_{Q_\alpha} = (\tilde{x}_{0.5} - \tilde{x}_{\alpha}) + (\tilde{x}_{1-\alpha} - \tilde{x}_{0.5})$$

4.7 Messung von Schiefe

> **Quantilskoeffizient der Schiefe** setzt die Differenz der beiden Abstände ins Verhältnis zum Quantilsabstand (siehe Kasten)

● Interpretation ●

... siehe Kasten

Quantilskoeffizient der Schiefe

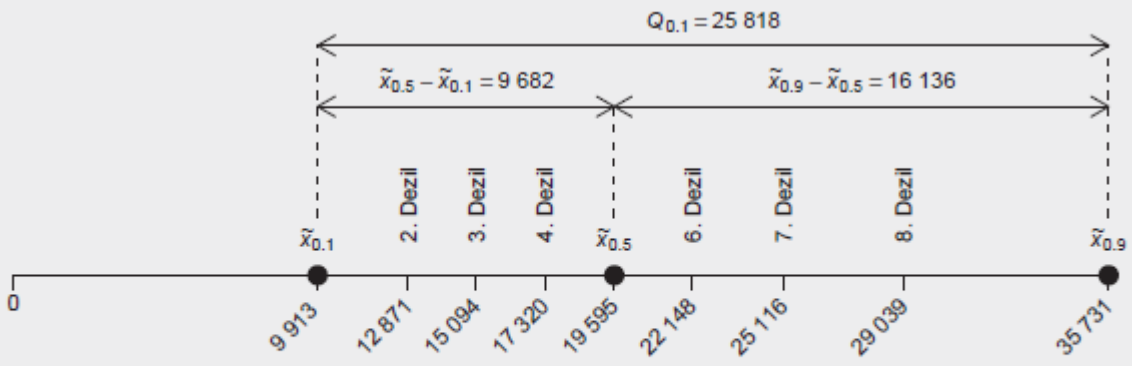
Gegeben seien metrische Beobachtungswerte $x_1, \ldots, x_n$. Dann ist durch

$$QS_\alpha = \frac{(\tilde{x}_{1-\alpha} - \tilde{x}_{0.5}) - (\tilde{x}_{0.5} - \tilde{x}_\alpha)}{\tilde{x}_{1-\alpha} - \tilde{x}_\alpha} \quad \text{für } \alpha \in (0, 0.5),$$

der α-**Quantilskoeffzient** der Schiefe gegeben.
Speziell wird $QS_{0.25}$ als **Quartilskoeffizient der Schiefe** bezeichnet.
Für $QS_\alpha > 0$ sind die mittleren $(1 - 2\alpha) \times 100\%$ der Werte rechtsschief, für $QS_\alpha < 0$ linksschief und für $QS_\alpha = 0$ symmetrisch verteilt.

4.7 Messung von Schiefe

Beispiel 4.7.1: Schiefe der Einkommensverteilung



Daten: Statistisches Bundesamt [2014]

Abb. 4.7.2: 0.1-Quantilskoeffizient der Schiefe für die Einkommensverteilung

$$QS_{0.1} = 0.25$$

$$QS_{0.2} = 0.17$$

$$QS_{0.3} = 0.10$$

$$QS_{0.1} = \frac{(\tilde{x}_{0.9} - \tilde{x}_{0.5}) - (\tilde{x}_{0.5} - \tilde{x}_{0.1})}{\tilde{x}_{0.9} - \tilde{x}_{0.1}} = \frac{(35731 - 19595) - (19595 - 9913)}{35731 - 9913}$$

$$\approx 0.25$$

4.8 Darstellung und Messung von Konzentration

4.8.1 Lorenz-Kurve

• Was versteht man unter Konzentration? •

Man vergleiche die folgenden beiden Datensätze

Datensatz 1: 4, 4, 4, 4, 4,

Datensatz 2: 2, 2, 2, 2, 12

Merkmalssumme jeweils = 20

Wie verteilt sich diese auf die einzelnen Merkmalsträger?
Wie stark ist die Ungleichverteilung (**Konzentration**) ausgeprägt?

• Beispiel 4.8.1 •

Man stelle sich 3 Branchen mit jeweils 5 Firmen vor:

Branche 1:	1, 100, 1, 1, 1,	$\bar{x}_1 = 20.8,\ \tilde{s}_1 = 39.6,\ \nu_1 = 1.90,$
Branche 2:	100, 100, 1, 100, 100,	$\bar{x}_2 = 80.2,\ \tilde{s}_2 = 39.6,\ \nu_2 = 0.49,$
Branche 3:	100, 60, 20, 40, 80.	$\bar{x}_3 = 60,\ \tilde{s}_3 = 28.28,\ \nu_3 = 0.47.$

Feststellung: Phänomen der **Konzentration** wird durch herkömmliche Maße nicht adäquat gemessen

4.8 Darstellung und Messung von Konzentration

• Konstruktion einer Lorenzkurve •

Tab. 4.8.1: Arbeitstabelle zur Erstellung einer Lorenzkurve für Branche 3

i	$x_{(i)}$	Kum. Anteil der Beobachtungen	Kum. Anteil an der Merkmalssumme
1	20	1/5=0.2	20/300≈ 0.067
2	40	1/5+1/5=0.4	(20+40)/300=0.2
3	60	1/5+1/5+1/5=0.6	(20+40+60)/300=0.4
4	80	1/5+1/5+1/5+1/5=0.8	(20+40+60+80)/300≈ 0.667
5	100	1/5+1/5+1/5+1/5+1/5=1.0	(20+40+60+80+100)/300=1

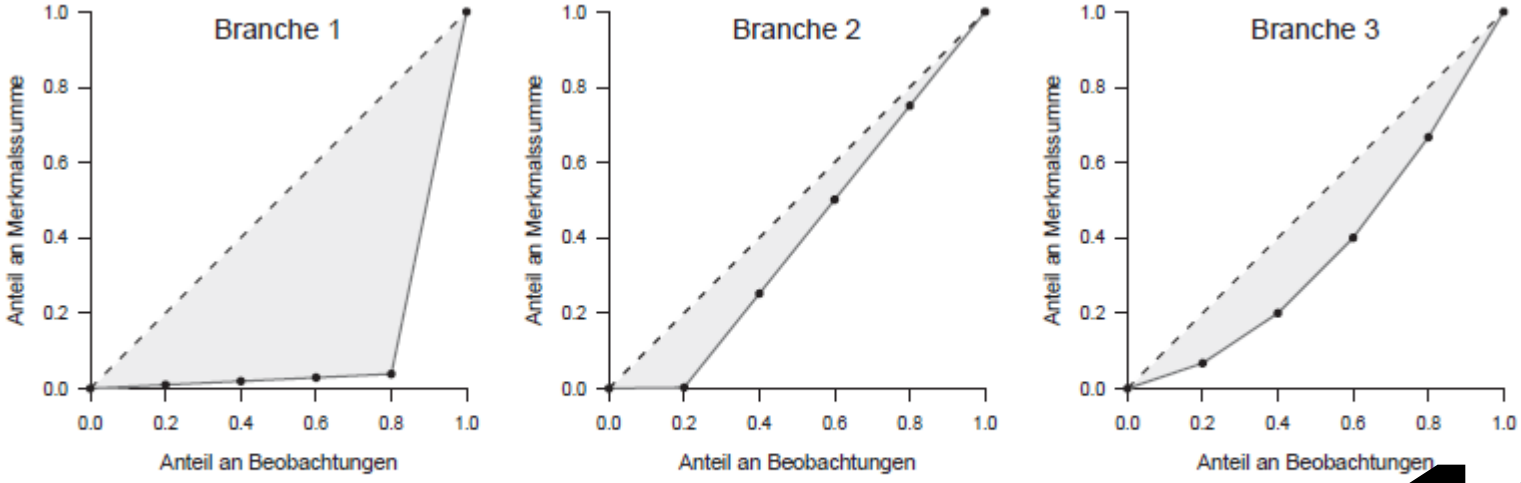


Abb. 4.8.1: Lorenzkurven der einzelnen Branchen

4.8 Darstellung und Messung von Konzentration

● Interpretation ●

> Die Fläche zwischen Lorenzkurve und Winkelhalbierender wird immer dann groß, wenn

wenige Merkmalsträger (relativ) viel von der Merkmalssumme auf sich vereinigen.

> Keine Konzentration liegt vor, wenn die Lorenzkurve mit der Winkelhalbierenden zusammenfällt. Dann

sind alle Werte gleich. Damit verteilt sich die Merkmalssumme gleichmäßig auf die Merkmalsträger.

> Beispiel: Koordinatenpunkte der Lorenzkurve, falls Branche mit 5 Unternehmen keinerlei Konzentration aufweist:

(0, 0), (0.2, 0.2), (0.4, 0.4), (0.6, 0.6), (0.8, 0.8), (1, 1)

4.8.2 Gini-Koeffizient

● Definition und Interpretation ●

> (Nichtnormierter) **Gini-Koeffizient**:
Zweifaches der Fläche zwischen Lorenzkurve und Winkelhalbierender

> Gini-Koeffizient = 0, falls alle Werte gleich sind

> Gini-Koeffizient ≈ 1 bei maximaler Konzentration, d. h. wenn ein einzelner Merkmalsträger die gesamte Merkmalssumme auf sich vereint

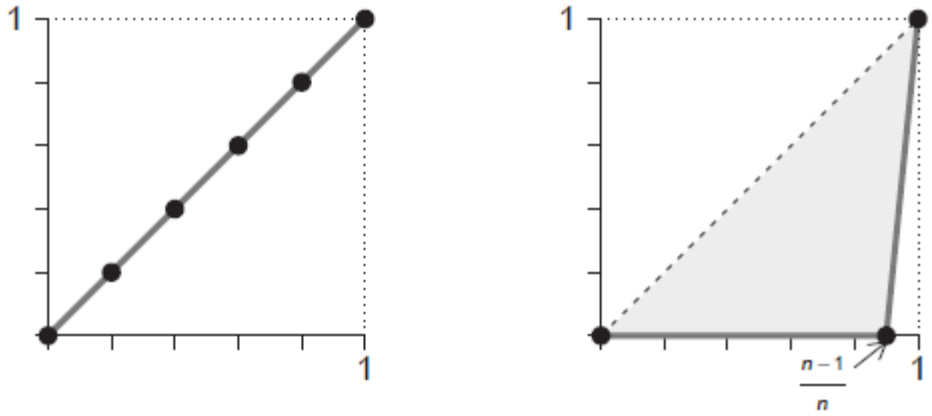


Abb. 4.8.2: Minimale und maximale Konzentration

● Berechnung ●

Formel: $$G = \frac{2\sum i x_{(i)}}{n\sum x_i} - \frac{n+1}{n}$$

Beispielrechnung für Branche 3:

20, 40, 60, 80, 100 (geordnet!)

$$\Rightarrow \quad G_3 = \frac{2(1\times 20 + 2\times 40 + 3\times 60 + 4\times 80 + 5\times 100)}{5\times 300} - \frac{6}{5}$$

$$\approx 0.2667$$

Analog: $G_1 = 0.76,$ $G_2 = 0.1975$

4.8 Darstellung und Messung von Konzentration

• Wertebereich und Normierung •

G ist maximal $\frac{n-1}{n}$; soll der Maximalwert = 1 sein:

$$G^* = \frac{n}{n-1} G$$

Normierter Gini-Koeffizient

Gini-Koeffizient

Gegeben seien metrische, nichtnegative Beobachtungswerte $x_1, \ldots, x_n$, wobei mindestens ein Wert positiv ist. Dann ist der **Gini-Koeffizient** gegeben durch

$$G = \frac{2 \sum_{i=1}^{n} i x_{(i)}}{n \sum_{i=1}^{n} x_i} - \frac{n+1}{n},$$

wobei $x_{(i)}$ die geordneten Werte sind. Sein Wert entspricht dem Doppelten der Fläche, welche die Lorenzkurve zur Winkelhalbierenden einschließt. Dabei gilt:

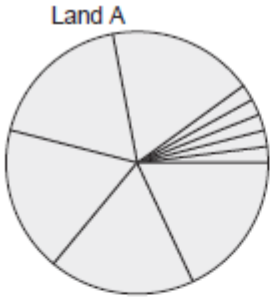
$$G \in [0, 1 - 1/n].$$

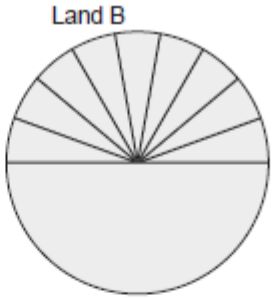
Der Gini-Koeffizient ist gleich 0, falls $x_1 = x_2 = \cdots = x_n$ und ist gleich $1 - 1/n$, falls $x_{(1)} = x_{(2)} = \cdots = x_{(n-1)} = 0$ und $x_{(n)} > 0$.

● Vorsicht bei der Interpretation ●

vgl. Fahrmeir et al. (2010)

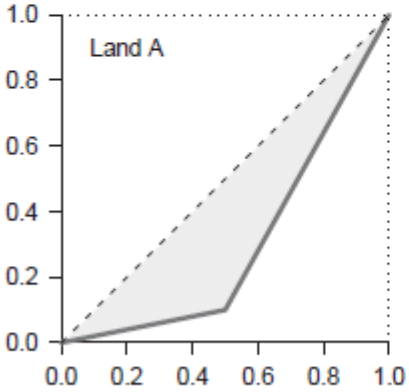
$G_A = 0.4$





$G_B = 0.4$

Fazit: Gini-Koeffizient allein verrät nicht alles!



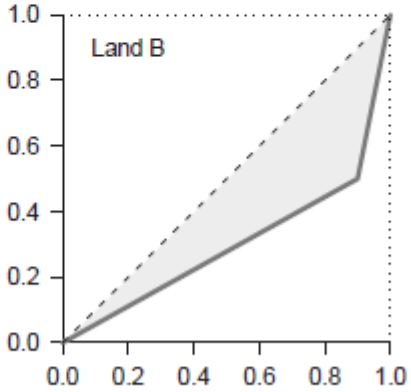


Abb. 4.8.3: Aufteilung des Grundbesitzes und korrespondierende Lorenzkurven

4.8 Darstellung und Messung von Konzentration

Beispiel 4.8.2: Konzentration von Einkommen in Deutschland

Tab. 4.8.2: Konzentration der Einkommensverteilung gemäß Gini-Koeffizient

	Markteinkommen			Nettoeinkommen		
Jahr	West-	Ost	Deutschland	West-	Ost-	Deutschland
1991	0.406	0.375	0.411	0.245	0.205	0.247
1996	0.439	0.476	0.450	0.254	0.212	0.251
2001	0.449	0.506	0.462	0.265	0.219	0.261
2006	0.479	0.540	0.494	0.290	0.239	0.286
2011	0.472	0.529	0.485	0.291	0.257	0.288

Daten: Sachverständigenrat [2013, Schaubild 91]

4.9.1 Minimumeigenschaft des arithmetischen Mittels

● Hintergrund ●

> Bekannt: $\tilde{s}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ verwendet $\bar{x}$ als Referenzpunkt in der Frage „*Wie stark streuen die Werte um ...?*“

> Frage: Welcher Referenzpunkt c minimiert den Ausdruck $\frac{1}{n}\sum_{i=1}^{n}(x_i - c)^2$?

> Antwort: Die Lösung lautet tatsächlich $c = \bar{x}$

● Analytischer Nachweis ●

> Minimierungsproblem: $\min_c f(c)$ mit $f(c) = \frac{1}{n}\sum_{i=1}^{n}(x_i - c)^2$

$$\frac{df(c)}{dc} = -\frac{2}{n}\sum_{i=1}^{n}(x_i - c) = -\frac{2}{n}\sum_{i=1}^{n}x_i + \frac{2}{n}\sum_{i=1}^{n}c$$

$$= -2\bar{x} + 2c = 2(c - \bar{x})$$

$$\Rightarrow \quad f \text{ minimal für } c = \bar{x}$$

> Alternativ: Betrachte allgemeine Verschiebungsformel (Folie 101)

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

$$\Leftrightarrow \quad \frac{1}{n}\sum_{i=1}^{n}(x_i - c)^2 = \tilde{s}^2 + (\bar{x} - c)^2$$

Ist minimal, falls $c = \bar{x}$

4.9 Spezifische Eigenschaften empirischer Kennwerte

● **Beispiel 4.9.1** ●

Fall 1 : $-1, 0, 0, 1, 3, 5, 6$

$\Rightarrow \quad \bar{x} = 2, \quad \tilde{s}^2 = 6.28571$

$\Rightarrow \quad f_1(c) = (2 - c)^2 + 6.28571$

Fall 2 : $-1, 0, 0, 1, 3, 4, 5, 6$

$\Rightarrow \quad \bar{x} = 2.25, \quad \tilde{s}^2 = 5.9375$

$\Rightarrow \quad f_2(c) = (2.25 - c)^2 + 5.9375$

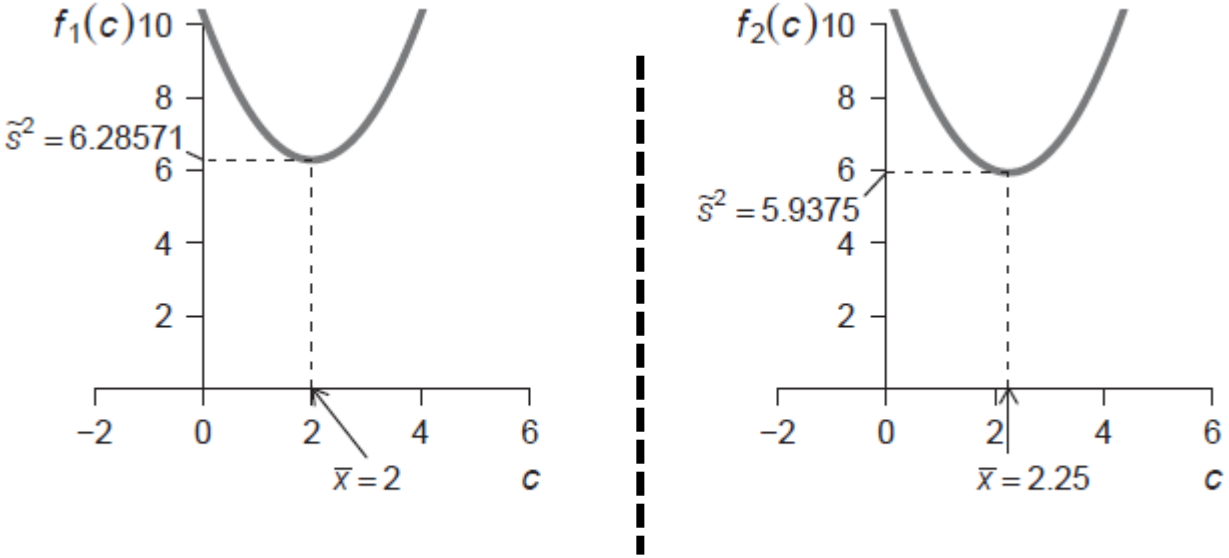


Abb. 4.9.1: Minimumeigenschaft des arithmetischen Mittels

4.9 Spezifische Eigenschaften empirischer Kennwerte

Satz 4.9.1: Minimumeigenschaft des arithmetischen Mittels

Gegeben seien metrische Beobachtungswerte $x_1, \ldots, x_n$. Dann gilt für jedes $c \in \mathbb{R}$:

$$\tilde{s}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \leq \frac{1}{n}\sum_{i=1}^{n}(x_i - c)^2.$$

4.9.2 Minimumeigenschaft des Medians

● Hintergrund ●

> Bekannt: $d = \frac{1}{n}\sum_{i=1}^{n}|x_i - \tilde{x}_{0.5}|$ verwendet $\tilde{x}_{0.5}$ als Referenzpunkt in der Frage „*Wie stark streuen die Werte um ...?*“

> Frage: Welcher Referenzpunkt c minimiert den Ausdruck

$$\frac{1}{n}\sum_{i=1}^{n}|x_i - c|\,?$$

> Antwort: Die Lösung lautet tatsächlich $c = \tilde{x}_{0.5}$

● Analytischer Nachweis ●

> Minimierungsproblem: $\min_c \tilde{f}(c)$ mit $\tilde{f}(c) = \frac{1}{n}\sum_{i=1}^{n}|x_i - c|$

> Jedoch: $\tilde{f}$ analytisch nicht gut handhabbar (teils nicht differenzierbar)
auch keine Streuungzerlegungsformel

> Minimumeigenschaft jedoch einsichtig („informaler Beweis“)

4.9 Spezifische Eigenschaften empirischer Kennwerte

• Beispiel 4.9.2 •

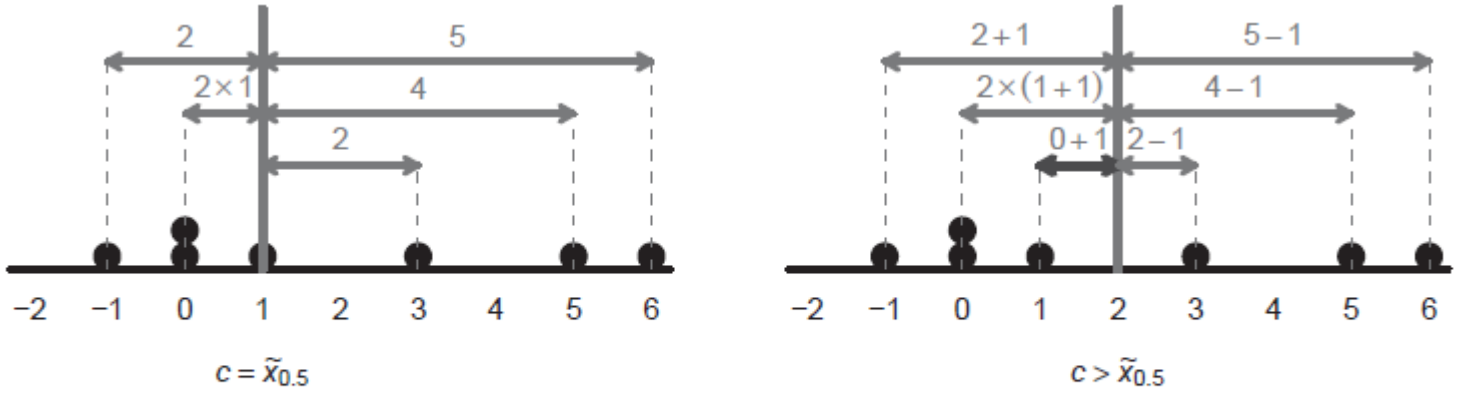


Abb. 4.9.2: Absolute Abweichungen bei ungerader Anzahl von Beobachtungen

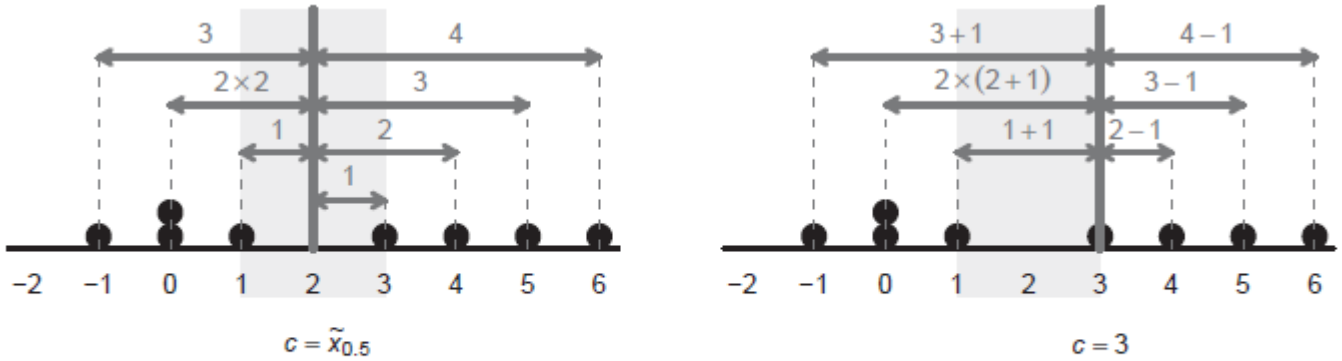


Abb. 4.9.3: Absolute Abweichungen bei gerader Anzahl von Beobachtungen

4.9 Spezifische Eigenschaften empirischer Kennwerte

Fall 1 : –1, 0, 0, 1, 3, 5, 6

$\Rightarrow \quad \tilde{x}_{0.5} = 1$

$$\Rightarrow \quad \tilde{f}_1(c) = \frac{1}{7}[|-1-c| + |0-c| + \cdots + |6-c|]$$

Fall 2 : –1, 0, 0, 1, 3, 4, 5, 6

$\Rightarrow \quad \tilde{x}_{0.5} = 2$

$$\Rightarrow \quad \tilde{f}_2(c) = \frac{1}{8}[|-1-c| + |0-c| + \cdots + |6-c|]$$

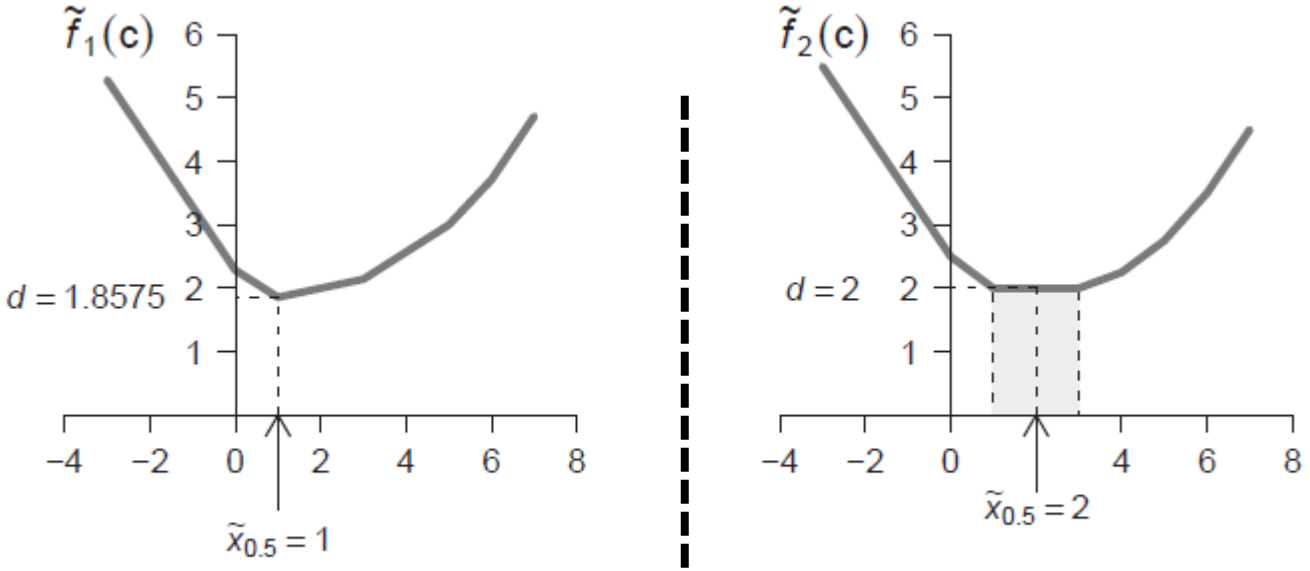


Abb. 4.9.4: Minimumeigenschaft des Medians

4.9 Spezifische Eigenschaften empirischer Kennwerte

Satz 4.9.2: Minimumeigenschaft des Medians

Gegeben seien metrische Beobachtungswerte $x_1, \ldots, x_n$. Dann gilt für jedes $c \in \mathbb{R}$:

$$d = \frac{1}{n} \sum_{i=1}^{n} |x_i - \tilde{x}_{0.5}| \leq \frac{1}{n} \sum_{i=1}^{n} |x_i - c|.$$

4.9.3 Transformationseigenschaften

● **Arten von Transformationen** ●

> Ursprungswerte $x_1, x_2, \ldots, x_n$ sollen einheitlich transformiert werden

> **Verschiebung**: $u_i = x_i + c$ für $i = 1, \ldots, n$

> **Umskalierung**: $u_i = cx_i$ für $i = 1, \ldots, n$ und $c > 0$

> Beispiel:

Ursprungswerte:	–1,	0,	0,	1,	3,	5,	6
Verschiebung um $c = 9$:	8,	9,	9,	10,	12,	14,	15
Umskalierung mit $c = 2$:	–2,	0,	0,	2,	6,	10,	12

4.9 Spezifische Eigenschaften empirischer Kennwerte

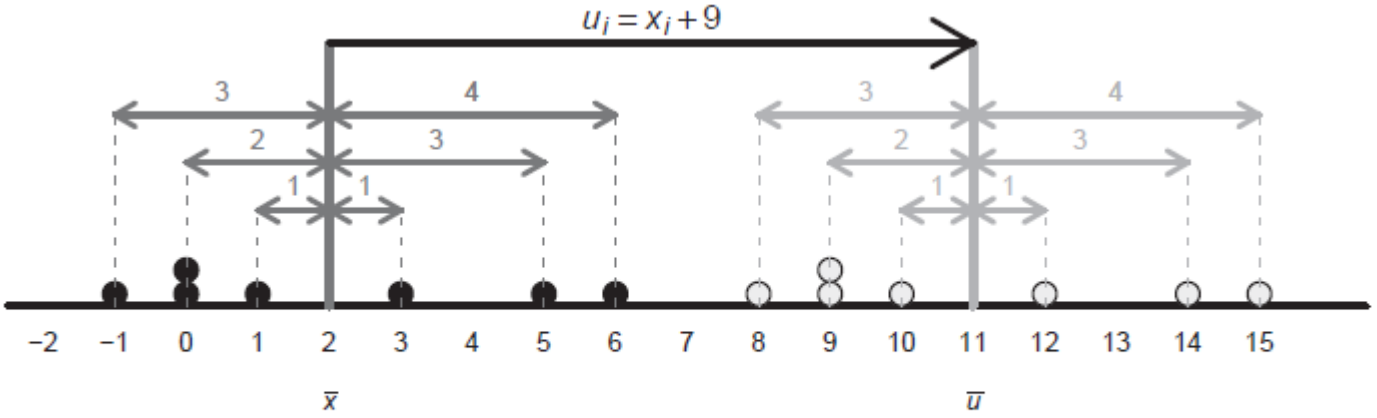


Abb. 4.9.5: Beobachtungswerte und absolute Abweichungen bei Verschiebung

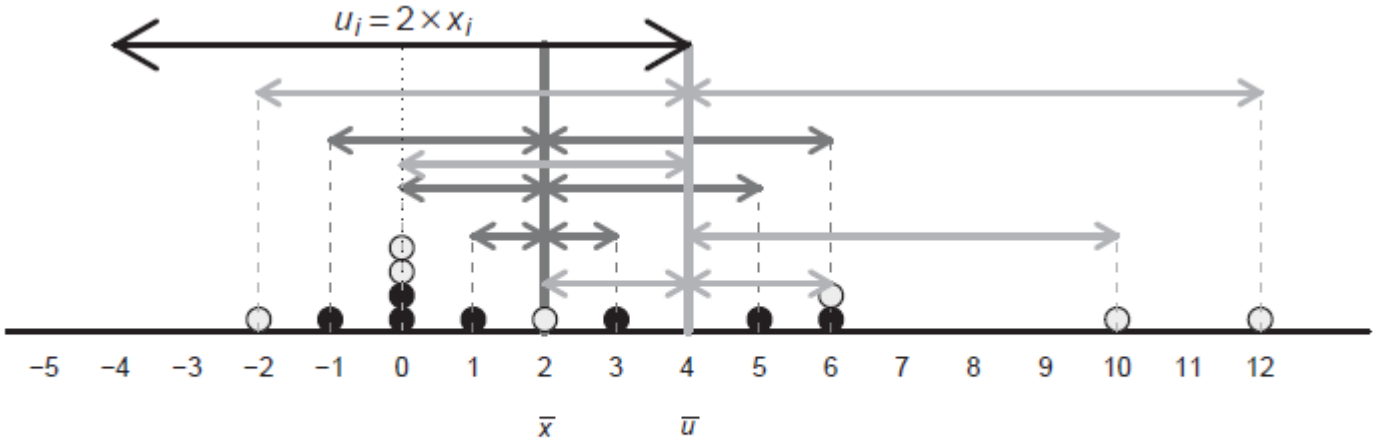


Abb. 4.9.6: Beobachtungswerte und absolute Abweichungen bei Umskalierung

4.9 Spezifische Eigenschaften empirischer Kennwerte

- **Verschiebungsäquivarianz und Verschiebungsinvarianz** -

> **Äquivarianz** ... „Gleichartige Veränderlichkeit“ bei Transformation

> **Invarianz** ... „Unveränderlichkeit “ bei Transformation

> Beispiel 1: Verschiebungsäquivarianz des arithmetischen Mittels:

Transformation: $u_i = x_i + c$

$$\bar{u} = \frac{1}{n}\sum_{i=1}^{n} u_i = \frac{1}{n}\sum_{i=1}^{n}(x_i + c) =$$

$$= \bar{x} + \frac{1}{n} \times nc = \bar{x} + c$$

$$= \bar{x} + c$$

> Beispiel 2: Verschiebungsinvarianz der (empirischen) Varianz

$$\tilde{s}_U^2 = \frac{1}{n}\sum_{i=1}^{n}(u_i - \bar{u})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i + c - \bar{x} - c)^2$$

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

$$= \tilde{s}_X^2$$

● Skalenäquivarianz und Skaleninvarianz ●

> Beispiel 3: Skalenäquivarianz des arithmetischen Mittels

Transformation: $u_i = cx_i$

$$\bar{u} = \frac{1}{n}\sum_{i=1}^{n} u_i = \frac{1}{n}\sum_{i=1}^{n} cx_i = c \times \frac{1}{n}\sum_{i=1}^{n} x_i = c\bar{x}$$

> Beispiel 4: Skalenäquivarianz der Standardabweichung

Transformation: $u_i = cx_i$ mit $c > 0$

$$\tilde{s}_U^2 = \frac{1}{n}\sum_{i=1}^{n}(u_i - \bar{u})^2 = \frac{1}{n}\sum_{i=1}^{n}(cx_i - c\bar{x})^2$$

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

$$= c^2 \tilde{s}_X^2$$

=> Fazit: Varianz ist weder äquivariant noch invariant, aber

$$\tilde{s}_U = \sqrt{\tilde{s}_U^2} = \sqrt{c^2 \tilde{s}_X^2} = c\tilde{s}_X$$ Beachte: $c > 0$

=> Fazit: Standardabweichung ist skalenäquivariant

4.9 Spezifische Eigenschaften empirischer Kennwerte

- **Eigenschaften weiterer Kennwerte** -

Tab. 4.9.1: Transformationseigenschaften empirischer Kennwerte

Empirischer Kennwert	VÄ	VI	SÄ	SI
Arithmetisches Mittel	✓	✗	✓	✗
Median (Quantile)	✓	✗	✓	✗
Modalwert	✓	✗	✓	✗
Spannweite	✗	✓	✓	✗
Quantilsabstände	✗	✓	✓	✗
Mittlere absolute Abweichungen	✗	✓	✓	✗
MAD	✗	✓	✓	✗
Varianz	✗	✓	✗	✗
Standardabweichung	✗	✓	✓	✗
Variationskoeffizient	✗	✗	✗	✓
Quantilskoeffizient der Schiefe	✗	✓	✗	✓
Gini-Koeffizient	✗	✗	✗	✓

VÄ = verschiebungsäquivariant, VI = verschiebungsinvariant
SÄ = skalenäquivariant, SI = skaleninvariant

4.9 Spezifische Eigenschaften empirischer Kennwerte

● Eigenschaften standardisierter Werte ●

> Bei Verschiebung $u_i = x_i + c$ gilt: $\bar{u} = \bar{x} + c$ und $\tilde{s}_U = \tilde{s}_X$

Daraus folgt bei einer Standardisierung verschobener Werte:

$$z_i = \frac{u_i - \bar{u}}{\tilde{s}_U} = \frac{x_i + c - \bar{x} - c}{\tilde{s}_X} = \frac{x_i - \bar{x}}{\tilde{s}_X}$$

> Bei Umskalierung $u_i = cx_i$ gilt: $\bar{u} = c\bar{x}$ und $\tilde{s}_U = c\tilde{s}_X$

Daraus folgt bei einer Standardisierung verschobener Werte:

$$z_i = \frac{u_i - \bar{u}}{\tilde{s}_U} = \frac{cx_i - c\bar{x}}{c\tilde{s}_X} = \frac{x_i - \bar{x}}{\tilde{s}_X}$$

> Fazit: Verschiebungen und Umskalierungen wirken sich nicht auf standardisierte Werte aus

4.9 Spezifische Eigenschaften empirischer Kennwerte

Definition 4.9.1: Transformationseigenschaften empirischer Kennwerte

Ein empirischer Kennwert M heißt

(i) **verschiebungsäquivariant**, falls für jedes $c \in \mathbb{R}$ gilt:

$$M(x_1 + c, x_2 + c, \ldots, x_n + c) = M(x_1, x_2, \ldots, x_n) + c.$$

(ii) **verschiebungsinvariant**, falls jedes $c \in \mathbb{R}$ gilt:

$$M(x_1 + c, x_2 + c, \ldots, x_n + c) = M(x_1, x_2, \ldots, x_n).$$

(iii) **skalenäquivariant**, falls für jedes $c > 0$ gilt:

$$M(cx_1, cx_2, \ldots, cx_n) = cM(x_1, x_2, \ldots, x_n).$$

(iv) **skaleninvariant**, falls für jedes $c > 0$ gilt:

$$M(cx_1, cx_2, \ldots, cx_n) = M(x_1, x_2, \ldots, x_n).$$

4.9 Spezifische Eigenschaften empirischer Kennwerte

4.9.4 Robustheit

- Zum Begriff
- Robuste und nicht robuste Kennwerte

Tab. 4.9.2: Robustheitseigenschaften empirischer Lage- und Streuungskennwerte

Lagekennwerte		Streuungskennwerte	
nicht robust	robust	nicht robust	robust
Arithmetisches Mittel	Median	Spannweite	Quantilsabstände
	Quantile	Mittlere absolute Abweichungen	MAD
	Modalwert	Varianz	
		Standardabweichung	
		Variationskoeffizient	

- Anmerkungen zur kritischen Verwendung

5 Beschreibung und Analyse empirischer Zusammenhänge

5 Beschreibung und Analyse empirischer Zusammenhänge

5.1 Zusammenhänge zwischen kategorialen Merkmalen 150

5.1.1 Kontingenztabellen 150

Verteilungen in Kontingenztabellen 150

Empirische Abhängigkeit und Unabhängigkeit 157

Konzeptionelle Vorüberlegungen für ein Zusammenhangsmaß 164

5.1.2 Zusammenhangsmaße für Kontingenztabellen 167

Chi-Quadrat-Koeffizient 167

Mittlere quadratische Kontingenz 172

Kontingenzkoeffizient nach Pearson 174

Transformationseigenschaften der Zusammenhangsmaße 177

5.1.3 Grafische Analysemöglichkeiten 180

Gestapelte und gruppierte Säulendiagramme 180

Segmentierte Säulen- und Balkendiagramme 182

Assoziationsplots 183

Mosaikplots 187

Spineplots 188

5.2 Zusammenhänge zwischen metrischen Merkmalen 189

5.2.1 Grafische Analysemöglichkeiten 189

Streudiagramme 189

Streudiagramm-Matrizen 191

Hexagonalplots 192

5.2.2 Zusammenhangsmaße für metrische Merkmale 193

Empirische Kovarianz 193

Empirischer Korrelationskoeffizient nach Pearson 199

Empirische Kovarianz- und Korrelationsmatrizen 210

Korrelationskoeffizient nach Spearman 211

5.2.3 Einfache lineare Regression	215
Zweck und allgemeine Vorgehensweise	215
KQ-Methode (L_2-Regression)	221
LAD-Methode (L_1-Regression)	231
Kritische Punkte und Alternativen	239
5.3 Ergänzende und vertiefende Themen	242
5.3.1 Zusammenhänge in anderen Skalierungsfällen	242
Grafische Analysemöglichkeiten	242
Zusammenhangsmaße	245
Spezielle Regressionsmodelle	247

5.3.2 Grafische Darstellung höherdimensionaler Daten 250

Höherdimensionale Mosaikplots 250

3D-Punktwolken 252

Biplots 253

5.3.3 Wichtige Aspekte bei der Analyse empirischer Zusammenhänge 258

Empirische Zusammenhänge und Kausalität 258

Spezielle Formen von Kausalität 260

Systematische Verzerrungen 261

Zufällige Schwankungen 267

5.1.1 Kontingenztabellen

Verteilungen in Kontingenztabellen

- Beispiel 5.1.1 •
- Allgemeines zur Notation •
- Absolute Verteilungen •
- Relative Verteilungen •

Tab. 5.1.1: Gemeinsame Verteilung von Geschlecht und Raucherstatus

	Raucher	Gelegenheits-raucher	Nichtraucher	Summe	
weiblich	n_{11} 4	n_{12} 8	n_{13} 28	40	$n_{1\bullet}$
männlich	n_{21} 12	n_{22} 12	n_{23} 36	60	$n_{2\bullet}$
Summe	$n_{\bullet 1}$ 16	$n_{\bullet 2}$ 20	$n_{\bullet 3}$ 64	100	n

5.1 Zusammenhänge zwischen kategorialen Merkmalen

Tab. 5.1.2: Allgemeine Gestalt einer Kontingenztabelle

X \ Y	b_1	b_2	...	b_j	...	b_l	Summe
a_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1l}	$n_{1\bullet}$
a_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2l}	$n_{2\bullet}$
⋮	⋮					⋮	⋮
a_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{il}	$n_{i\bullet}$
⋮	⋮					⋮	⋮
a_k	n_{k1}	n_{k2}	...	n_{kj}	...	n_{kl}	$n_{k\bullet}$
Summe	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet j}$	...	$n_{\bullet l}$	n

$$n_{i\bullet} = \sum_{j=1}^{l} n_{ij}$$

$$\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij} = n$$

$$n_{\bullet j} = \sum_{i=1}^{k} n_{ij}$$

Im Beispiel mit $k = 2$ und $l = 3$:

$$\sum_{i=1}^{2}\sum_{j=1}^{3} n_{ij} = n_{11} + n_{12} + n_{13} + n_{21} + n_{22} + n_{23} = 100$$

5.1 Zusammenhänge zwischen kategorialen Merkmalen

Tab. 5.1.3: Gemeinsame relative Verteilung von Geschlecht und Raucherstatus

	Raucher	Gelegenheits-raucher	Nichtraucher	Summe
weiblich	f_{11} 0.04	f_{12} 0.08	f_{13} 0.28	0.40 $f_{1\bullet}$
männlich	f_{21} 0.12	f_{22} 0.12	f_{23} 0.36	0.60 $f_{2\bullet}$
Summe	$f_{\bullet 1}$ 0.16	$f_{\bullet 2}$ 0.20	$f_{\bullet 3}$ 0.64	1.00

Es gilt: $f_{ij} = \frac{n_{ij}}{n}$ $f_{\bullet j} = \frac{n_{\bullet j}}{n}$ $f_{i\bullet} = \frac{n_{i\bullet}}{n}$

$$\sum_{i=1}^{k}\sum_{j=1}^{l} f_{ij} = 1$$

5.1 Zusammenhänge zwischen kategorialen Merkmalen

Gemeinsame Verteilung und Randverteilungen

(X, Y)	Merkmalsvektor
$(x_1, y_1), \ldots, (x_n, y_n)$	zweidimensionale Beobachtungswerte
$a_1, a_2, \ldots, a_k$	Ausprägungen von X
$b_1, b_2, \ldots, b_l$	Ausprägungen von Y
n_{ij}	Absolute Häufigkeit der Ausprägungskombination (a_i, b_j)
f_{ij}	Relative Häufigkeit der Ausprägungskombination (a_i, b_j)
$n_{11}, n_{12}, \ldots, n_{kl}$	Gemeinsame absolute Häufigkeitsverteilung von (X, Y)
$f_{11}, f_{12}, \ldots, f_{kl}$	Gemeinsame relative Häufigkeitsverteilung von (X, Y)
$n_{i\bullet}$ bzw. $n_{\bullet j}$	Absolute (Rand-)Häufigkeit der Ausprägung a_i bzw. b_j
$f_{i\bullet}$ bzw. $f_{\bullet j}$	Relative (Rand-)Häufigkeit der Ausprägung a_i bzw. b_j
$n_{1\bullet}, n_{2\bullet}, \ldots, n_{k\bullet}$	Absolute Randverteilung (Häufigkeitsverteilung) von X
$n_{\bullet 1}, n_{\bullet 2}, \ldots, n_{\bullet l}$	Absolute Randverteilung (Häufigkeitsverteilung) von Y
$f_{1\bullet}, f_{2\bullet}, \ldots, f_{k\bullet}$	Relative Randverteilung (Häufigkeitsverteilung) von X
$f_{\bullet 1}, f_{\bullet 2}, \ldots, f_{\bullet l}$	Relative Randverteilung (Häufigkeitsverteilung) von Y

5.1 Zusammenhänge zwischen kategorialen Merkmalen

- **Bedingte Verteilungen**
- **Notation für bedingte Verteilungen**

Tab. 5.1.4: Verteilung von Raucherstatus bedingt auf Geschlecht

X \ Y	Raucher	Gelegenheits-raucher	Nichtraucher	Summe
weiblich	0.10	0.20	0.70	1
männlich	0.20	0.20	0.60	1

Beispiele:

$$f_{11}^{Y|X} = \frac{n_{11}}{n_{1\bullet}} = \frac{4}{40} = \frac{f_{11}}{f_{1\bullet}} = \frac{0.04}{0.40} = 0.10$$

$$f_{22}^{Y|X} = \frac{n_{22}}{n_{2\bullet}} = \frac{12}{60} = \frac{f_{22}}{f_{2\bullet}} = \frac{0.12}{0.60} = 0.20$$

$$f_{23}^{Y|X} = \frac{n_{23}}{n_{2\bullet}} = \frac{36}{60} = \frac{f_{23}}{f_{2\bullet}} = \frac{0.36}{0.60} = 0.60$$

5.1 Zusammenhänge zwischen kategorialen Merkmalen

Tab. 5.1.5: Verteilung von Geschlecht bedingt auf Raucherstatus

X \ Y	Raucher	Gelegenheits-raucher	Nichtraucher
weiblich	0.25	0.40	0.4375
männlich	0.75	0.60	0.5625
Summe	1.00	1.00	1.0000

Beispiele:

$$f_{11}^{X|Y} = \frac{n_{11}}{n_{\bullet 1}} = \frac{4}{16} = \frac{f_{11}}{f_{\bullet 1}} = \frac{0.04}{0.16} = 0.25$$

$$f_{22}^{X|Y} = \frac{n_{22}}{n_{\bullet 2}} = \frac{12}{20} = \frac{f_{22}}{f_{\bullet 2}} = \frac{0.12}{0.20} = 0.60$$

$$f_{12}^{X|Y} = \frac{n_{12}}{n_{\bullet 2}} = \frac{8}{20} = \frac{f_{12}}{f_{\bullet 2}} = \frac{0.08}{0.20} = 0.40$$

5.1 Zusammenhänge zwischen kategorialen Merkmalen

Bedingte Verteilungen

$f_{ij}^{X\|Y} = \frac{n_{ij}}{n_{\bullet j}} = \frac{f_{ij}}{f_{\bullet j}}$	bedingte Häufigkeit von (a_i, b_j) bedingt auf $Y = b_j$, wobei $n_{\bullet j} > 0$ für alle j
$f_{ij}^{Y\|X} = \frac{n_{ij}}{n_{i\bullet}} = \frac{f_{ij}}{f_{i\bullet}}$	bedingte Häufigkeit von (a_i, b_j) bedingt auf $X = a_i$, wobei $n_{i\bullet} > 0$ für alle i
$f_{1j}^{X\|Y}, f_{2j}^{X\|Y}, \ldots, f_{kj}^{X\|Y}$	bedingte Verteilung von X bedingt auf $Y = b_j$
$f_{i1}^{Y\|X}, f_{i2}^{Y\|X}, \ldots, f_{il}^{Y\|X}$	bedingte Verteilung von Y bedingt auf $X = a_i$

Empirische Abhängigkeit und Unabhängigkeit

- **Empirische Abhängigkeit**

Empirische Abhängigkeit ⇔ Bedingte Verteilungen unterscheiden sich

- **Frage der Richtung der Abhängigkeit**

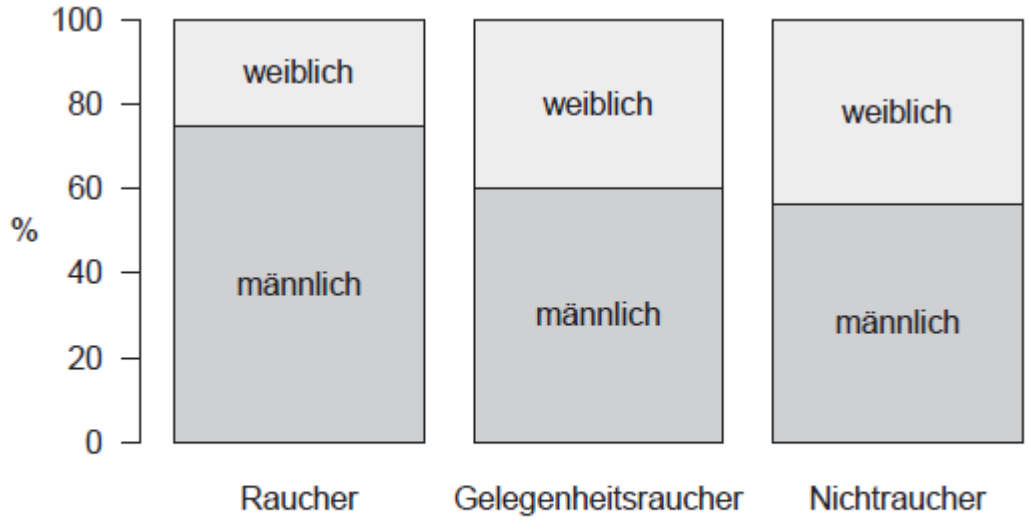


Abb. 5.1.1: Empirische Abhängigkeit von Geschlecht und Raucherstatus

5.1 Zusammenhänge zwischen kategorialen Merkmalen

● Empirische Unabhängigkeit ●

Empirische Unabhängigkeit ⇔ Bedingte Verteilungen sind identisch

Tab. 5.1.6: Gemeinsame Verteilung bei Unabhängigkeit

	Raucher	Gelegenheits-raucher	Nichtraucher	Summe
weiblich	10	10	20	40
männlich	15	15	30	60
Summe	25	25	50	100

Tab. 5.1.7: Verteilung von Raucherstatus bedingt auf Geschlecht bei Unabhängigkeit

	Raucher	Gelegenheits-raucher	Nichtraucher	Summe
weiblich	0.25	0.25	0.50	1
männlich	0.25	0.25	0.50	1

Identische „Zeilenverteilungen“ => unabhängig

5.1 Zusammenhänge zwischen kategorialen Merkmalen

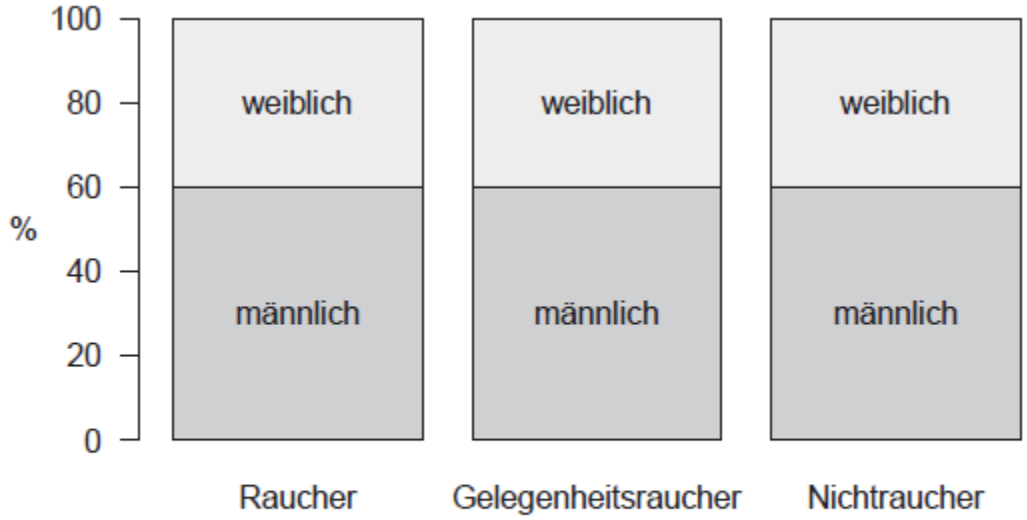


Abb. 5.1.2: Empirische Unabhängigkeit von Geschlecht und Raucherstatus

• Verhältnis zur Kausalität •

Weder impliziert empirische Abhängigkeit eine kausale Beziehung (Ursache-Wirkungs-Beziehung), noch schließt empirische Unabhängigkeit eine solche aus.

● Formale Unabhängigkeitskriterien ●

> Zwei Merkmale sind genau dann **empirisch unabhängig**, falls die bedingten Verteilungen identisch sind. Damit stimmen diese auch mit den jeweiligen Randverteilungen überein. Formal ausgedrückt gilt dann:

$$f_{ij}^{Y|X} = f_{\bullet j} \quad \text{für alle } i \text{ und } j \quad \text{bzw.}$$

$$f_{ij}^{X|Y} = f_{i\bullet} \quad \text{für alle } i \text{ und } j.$$

> Gemäß Definition bedingter Häufigkeiten resultieren daraus die Kriterien:

$$\frac{n_{ij}}{n_{i\bullet}} = f_{\bullet j} \quad \text{für alle } i \text{ und } j \quad \text{bzw.}$$

$$\frac{n_{ij}}{n_{\bullet j}} = f_{i\bullet} \quad \text{für alle } i \text{ und } j.$$

> Durch Umformung erhält man daraus

$$n_{ij} = n_{i\bullet} f_{\bullet j} \quad \text{bzw.}$$

$$n_{ij} = n_{\bullet j} f_{i\bullet}$$

> Mit $f_{\bullet j} = n_{\bullet j}/n$ und $f_{i\bullet} = n_{i\bullet}/n$ erhält man daraus schließlich

(*) $$n_{ij} = \frac{n_{i\bullet} n_{\bullet j}}{n}$$ für alle *i* und *j*

● Tatsächliche und zu erwartende Häufigkeiten ●

> n_{ij} ... tatsächliche Häufigkeit

> $\frac{n_{i\bullet} n_{\bullet j}}{n}$ ... erwartete Häufigkeit bei Unabhängigkeit

● Multiplikationskriterium ●

> Dividiert man Gleichung (*) durch n erhält man

$$f_{ij} = f_{i\bullet} f_{\bullet j} \quad \text{für alle } i \text{ und } j$$

Das Produkt der Randverteilungen ergibt die gemeinsame Verteilung

Tab. 5.1.8: Gemeinsame relative Verteilung bei Unabhängigkeit

	Raucher	Gelegenheits-raucher	Nichtraucher	Summe
weiblich	0.10	0.10	0.20	0.40
männlich	0.15	0.15	0.30	0.60
Summe	0.25	0.25	0.50	1.00

● Stochastische Unabhängigkeit ●

Analoges Konzept in W‘rechnung (Folie 314 ff. , Folie 395 ff.)

5.1 Zusammenhänge zwischen kategorialen Merkmalen

Definition 5.1.1: Empirische Abhängigkeit und Unabhängigkeit

Gegeben sei eine $(k \times l)$-Kontingenztabelle zweier Merkmale X und Y, wobei die Ausprägungen von X mit den k Zeilen und die Ausprägungen von Y mit den l Spalten der Tabelle korrespondieren. Dann sind folgende Aussagen äquivalent.

(i) X und Y sind **empirisch unabhängig**.

(ii) Die Zeilenverteilungen stimmen mit der Randverteilung von Y überein.

(iii) Die Spaltenverteilungen stimmen mit der Randverteilung von X überein.

(iv) Die absoluten Häufigkeiten stimmen mit den unter Unabhängigkeit zu erwartenden Häufigkeiten überein, d.h.

$$n_{ij} = \frac{n_{i\bullet} n_{\bullet j}}{n} \text{ für alle } i \text{ und } j.$$

(v) Das Produkt der Randverteilungen ergibt die gemeinsame Verteilung, d.h. $f_{ij} = f_{i\bullet} f_{\bullet j}$ für alle i und j.

Trifft eine der fünf Aussagen nicht zu, sind X und Y **empirisch abhängig**. Man spricht dann auch von einem **empirischen Zusammenhang**.

Konzeptionelle Vorüberlegungen für ein Zusammenhangsmaß

● Abweichungen von der Unabhängigkeit ●

Tab. 5.1.9: Tatsächliche und zu erwartende Häufigkeiten mit jeweiligen Abweichungen

	Raucher	Gelegenheits-raucher	Nichtraucher	Summe
weiblich	4 (6.4)	8 (8)	28 (25.6)	40
	−2.4	**0**	**2.4**	
männlich	12 (9.6)	12 (12)	36 (38.4)	60
	2.4	**0**	**−2.4**	
Summe	16	20	64	100

> Unter Unabhängigkeit sollte gelten

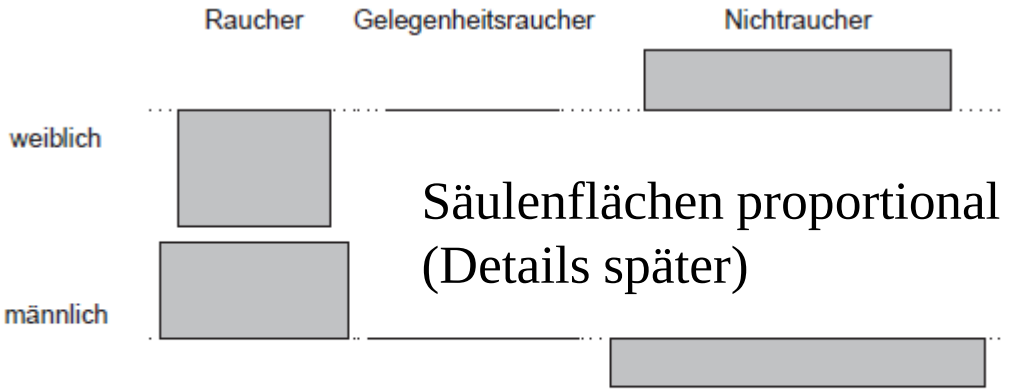
$$n_{ij} = \frac{n_{i\bullet} n_{\bullet j}}{n} \quad \text{d. h.} \quad n_{ij} - \frac{n_{i\bullet} n_{\bullet j}}{n} = 0 \quad \text{für alle } i \text{ und } j$$

> Abweichungen von der Unabhängigkeit im Beispiel (Tab. 5.1.9), z. B.

$$n_{11} - \frac{n_{1\bullet} n_{\bullet 1}}{n} = 4 - \frac{40 \times 16}{100} = 4 - 6.4 = -2.4$$

$$n_{12} - \frac{n_{1\bullet} n_{\bullet 2}}{n} = 8 - \frac{40 \times 20}{100} = 8 - 8 = 0$$

- **Assoziationsplot** -



Säulenflächen proportional zu den Abweichungen (Details später)

Abb. 5.1.3: Assoziationsplot der Abweichungen von der Unabhängigkeit

● Summe aller absoluten Abweichungen ●

> Summe aller „Abweichungen von der Unabhängigkeit“ als Maß für die Stärke der Abhängigkeit nicht geeignet, da stets 0:

$$\sum_{i=1}^{k}\sum_{j=1}^{l}\left(n_{ij}-\frac{n_{i\bullet}n_{\bullet j}}{n}\right)=\sum_{i=1}^{k}\sum_{j=1}^{l}n_{ij}-\sum_{i=1}^{k}\sum_{j=1}^{l}\frac{n_{i\bullet}n_{\bullet j}}{n}$$

$$=n-n\sum_{i=1}^{k}\sum_{j=1}^{l}\frac{n_{i\bullet}n_{\bullet j}}{n\times n}=n-n\sum_{i=1}^{k}\sum_{j=1}^{l}f_{i\bullet}f_{\bullet j}$$

$$=n-n\underbrace{\left(\sum_{i=1}^{k}f_{i\bullet}\right)}_{=1}\underbrace{\left(\sum_{j=1}^{l}f_{\bullet j}\right)}_{=1}=0$$

> Summe aller absoluten Abweichungen als Maß jedoch auch nicht üblich

5.1.2 Zusammenhangsmaße für Kontingenztabellen

Chiquadrat-Koeffizient

● **Definition und Rechenbeispiel** ●

> Definition:

$$\chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{l} \frac{\left(n_{ij} - \frac{n_{i\bullet} n_{\bullet j}}{n}\right)^2}{\frac{n_{i\bullet} n_{\bullet j}}{n}}$$

wobei $n_{i\bullet} > 0$ und $n_{\bullet j} > 0$ für alle i und j

> Rechenbeispiel zu Folie 165:

$$\chi^2 = \frac{(-2.4)^2}{6.4} + \frac{0^2}{8} + \frac{2.4^2}{25.6} + \frac{2.4^2}{9.6} + \frac{0^2}{12} + \frac{(-2.4)^2}{38.4} = 1.875$$

● Interpretation ●

> $\chi^2 = 0$ falls keine Abhängigkeit besteht, sonst > 0.

> umso größer, desto größer die Abweichungen von der Unabhängigkeit

> Abweichungen werden umso höher gewichtet, desto kleiner die Randhäufigkeiten bzw. die erwarteten Häufigkeiten sind, da Division durch

$$\frac{n_{i\bullet} n_{\bullet j}}{n}$$

● Wertebereich ●

> Allgemein: $[0, \underbrace{(M-1) \times n}_{\text{Theoretischer Maximalwert}}]$, wobei $M = \min(k, l)$

> Beispiel zu Folie 164: $M = \min(2,3) = 2, \quad n = 100$

$$\Rightarrow [0, (2-1) \times 100] = [0, 100]$$

5.1 Zusammenhänge zwischen kategorialen Merkmalen

• Maximale Abhängigkeit in nichtquadratischen Tabellen •

Tab. 5.1.10: Beispiel maximaler Abhängigkeit in nichtquadratischer Tabelle

	Raucher	Gelegenheits-raucher	Nichtraucher	Summe
weiblich	0	0	40	40
männlich	30	30	0	60
Summe	30	30	40	100

Sofern $l \geq k$ (mehr Spalten als Zeilen) mit $n_{i\bullet} > 0$ gilt:
Maximale Kontingenz $\Leftrightarrow$ In jeder Spalte genau ein positiver Zelleneintrag

• Maximale Abhängigkeit in quadratischen Tabellen •

Tab. 5.1.11: Maximale Abhängigkeit in quadratischer Tabelle

	b_1	b_2	b_3	b_4	Summe
a_1	$n_{11} > 0$	0	0	0	$n_{1\bullet} = n_{11}$
a_2	0	$n_{22} > 0$	0	0	$n_{2\bullet} = n_{22}$
a_3	0	0	$n_{33} > 0$	0	$n_{3\bullet} = n_{33}$
a_4	0	0	0	$n_{44} > 0$	$n_{4\bullet} = n_{44}$
Summe	$n_{\bullet 1} = n_{11}$	$n_{\bullet 2} = n_{22}$	$n_{\bullet 3} = n_{33}$	$n_{\bullet 4} = n_{44}$	n

Diagonalisierbarkeits-kriterium

● Berechnungsformel für (2 × 2)–Tabellen ●

> Formel: $\chi^2 = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_{\bullet 1}n_{\bullet 2}n_{1\bullet}n_{2\bullet}}$

> Beispiel :

	Raucher	Gelegenheits-raucher	Summe
weiblich	4	8	12
männlich	12	12	24
Summe	16	20	36

$$\chi^2 = \frac{36 \times (4 \times 12 - 12 \times 8)^2}{16 \times 20 \times 12 \times 24} = 0.9$$

Wertebereich: $[0, (M-1) \times n] = [0, 36\,]$

• Zwischenschritt zum Kontingenzkoeffizienten •

> Als rein deskriptives Maß eher unbedeutend

> Verwendung beim **Unabhängigkeitstest** (Kap. 11, Folie 809 ff.)

Chi-Quadrat-Koeffizient

Gegeben sei eine $(k \times l)$-Kontingenztabelle der absoluten Häufigkeiten n_{ij} mit positiven Randhäufigkeiten, d.h.

$$n_{i\bullet} > 0 \text{ und } n_{\bullet j} > 0 \quad \text{für alle } i = 1, \ldots, k \text{ und } j = 1, \ldots, l.$$

Dann ist der **Chi-Quadrat-Koeffizient** definiert als

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{\left(n_{ij} - \frac{n_{i\bullet} n_{\bullet j}}{n}\right)^2}{\frac{n_{i\bullet} n_{\bullet j}}{n}}.$$

Speziell für (2×2)-Tabellen gilt:

$$\chi^2 = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_{\bullet 1} n_{\bullet 2} n_{1\bullet} n_{2\bullet}}.$$

Es gilt: $\chi^2 \in [0, (M-1) \cdot n]$, wobei $M = \min(k, l)$.

Mittlere quadratische Kontingenz

● Definition und Rechenbeispiel ●

> Definition: $\phi^2 = \frac{1}{n}\chi^2$

> Es gilt: $\phi^2 = \sum_{i=1}^{k}\sum_{j=1}^{l}\frac{\left(f_{ij} - f_{i\bullet}f_{\bullet j}\right)^2}{f_{i\bullet}f_{\bullet j}}$

Beweis:

$$\phi^2 = \frac{1}{n}\chi^2 = \frac{n}{n^2}\sum_{i=1}^{k}\sum_{j=1}^{l}\frac{\left(n_{ij} - \frac{n_{i\bullet}n_{\bullet j}}{n}\right)^2}{\frac{n_{i\bullet}n_{\bullet j}}{n}} = \sum_{i=1}^{k}\sum_{j=1}^{l}\frac{\left(\frac{n_{ij}}{n} - \frac{n_{i\bullet}n_{\bullet j}}{n\times n}\right)^2}{\frac{n_{i\bullet}n_{\bullet j}}{n\times n}}$$

> Beispiel zu Folien 167 (164):

$$\phi^2 = \frac{1}{n}\chi^2 = \frac{1}{100}\times 1.875 = 0.01875$$

5.1 Zusammenhänge zwischen kategorialen Merkmalen

● Wertebereich und Interpretation ●

> Allgemein: $[0, \underbrace{(M-1)}_{\text{Theoretischer Maximalwert}}]$, wobei $M = \min(k, l)$

> Beispiel zu Folie 164: $M = \min(2,3) = 2 \quad \Rightarrow \quad [0, (2-1)] = [0, 1]$

> Interpretation: Kein Unterschied zum Chiquadrat-Koeffizienten;
Ermittlung des Durchschnitts der Abweichungen

Mittlere quadratische Kontingenz

Die mittlere quadratische Kontingenz ist definiert als

$$\phi^2 = \frac{1}{n}\chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{l}\frac{(f_{ij} - f_{i\bullet}f_{\bullet j})^2}{f_{i\bullet}f_{\bullet j}} .$$

Dabei bezeichnet χ^2 den Chi-Quadrat-Koeffizienten.
Es gilt: $\phi^2 \in [0, (M-1)]$, wobei $M = \min(k, l)$.

Kontingenzkoeffizient nach Pearson

● Definition und Wertebereich ●

> Definition: $C = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{\phi^2}{\phi^2 + 1}}$

> Allgemein: $\left[0, \underbrace{\sqrt{(M-1)/M}}_{\text{Theoretischer Maximalwert}}\right]$, wobei $M = \min(k, l)$

> Beispiel zu Folie 164: $M = \min(2,3) = 2 \quad \Rightarrow \quad \left[0, \sqrt{1/2}\right] = [0, 0.7071]$

● Korrigierter Kontingenzkoeffizient ●

> Definition: $C_K = C \times \sqrt{M/(M-1)}$

> Wertebereich: $[0,1]$

● Beispiel 5.1.2 ●

Beispiel zu Folie 164 (167):

$$\chi^2 = 1.875 \quad \Rightarrow \quad C = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{1.875}{1.875 + 100}} \approx 0.1357$$

$$M = \min(2,3) = 2 \quad \Rightarrow \quad C_K = 0.1357 \times \sqrt{2/1} \approx 0.1919$$

● Interpretation ●

> $C_K = 0$, falls keine Abhängigkeit besteht, sonst > 0.

> $C_K = 1$, bei maximaler Abhängigkeit

> Anhand des Wertes von C oder C_K lässt sich nicht erkennen, in welcher Weise Abhängigkeiten bestehen (Vergleich bedingter Verteilungen!)

● Anmerkung zur Einordnung der Stärke von Zusammenhängen ●

Kontingenzkoeffizient nach Pearson

Gegeben sei eine $(k \times l)$-Kontingenztabelle der absoluten Häufigkeiten n_{ij} mit positiven Randhäufigkeiten, d.h.

$$n_{i\bullet} > 0 \text{ und } n_{\bullet j} > 0 \quad \text{für alle } i = 1, \ldots, k \text{ und } j = 1, \ldots, l.$$

Dann ist der Kontingenzkoeffizient nach Pearson definiert als

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{\phi^2}{\phi^2 + 1}}, \quad \text{wobei } \chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{\left(n_{ij} - \frac{n_{i\bullet} n_{\bullet j}}{n}\right)^2}{\frac{n_{i\bullet} n_{\bullet j}}{n}}$$

Es gilt $C \in [0, \sqrt{(M-1)/M}]$, wobei $M = \min(k, l)$.
Der korrigierte Kontingenzkoeffizient nach Pearson ist definiert als

$$C_K = C\sqrt{M/(M-1)}$$

Es gilt: $C_K \in [0, 1]$.

Transformationseigenschaften der Zusammenhangsmaße

- **Invarianz gegenüber Zeilen- und Spaltenvertauschungen**
- **Äquivarianz des Chiquadrat-Koeffizienten gegenüber Multiplikationen**

> Beispiel:

Tab. 5.1.12: Gleichstarker Zusammenhang bei unterschiedlichen Besetzungshäufigkeiten

	Raucher	Nichtraucher
weiblich	10	90
männlich	20	80

	Raucher	Nichtraucher
weiblich	1	9
männlich	2	8

> Nachweis:

$$\chi^2_{Trafo} = \sum_{i=1}^{k}\sum_{j=1}^{l} \frac{\left(cn_{ij} - \frac{cn_{i\bullet}cn_{\bullet j}}{cn}\right)^2}{\frac{cn_{i\bullet}cn_{\bullet j}}{cn}} = c\sum_{i=1}^{k}\sum_{j=1}^{l} \frac{\left(n_{ij} - \frac{n_{i\bullet}n_{\bullet j}}{n}\right)^2}{\frac{n_{i\bullet}n_{\bullet j}}{n}}$$

$$= c\chi^2$$

● Invarianz aller anderen Maße gegenüber Multiplikationen ●

> Mittlere quadratische Kontingenz:

$$\phi^2 = \frac{1}{n}\chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{l}\frac{\left(f_{ij} - f_{i\bullet}f_{\bullet j}\right)^2}{f_{i\bullet}f_{\bullet j}}$$

> (Korrigierter) Kontingenzkoeffizient

$$C = \sqrt{\frac{\phi^2}{\phi^2 + 1}} \quad \Rightarrow \quad C_K = C \times \sqrt{M/(M-1)}$$

> Beispiel (Folie 177): Links und rechts jeweils $C_K = 0.2863$

● Invarianz und statistisches Problem ●

> Invarianz gegenüber Multiplikationen erzeugt statistisches Problem

> Starker Zusammenhang gemäß *C* bei wenigen Beobachtungen wenig aussagekräftig

Invarianzeigenschaften der Zusammenhangsmaße

Gegeben sei eine $(k \times l)$-Kontingenztabelle der absoluten Häufigkeiten n_{ij} mit positiven Randhäufigkeiten, d.h.

$$n_{i\bullet} > 0 \text{ und } n_{\bullet j} > 0 \quad \text{für alle } i = 1, \ldots, k \text{ und } j = 1, \ldots, l.$$

(i) Chi-Quadrat-Koeffizient χ^2, mittlere quadratische Kontingenz ϕ^2 und Kontingenzkoeffizient (C und C_K) sind invariant gegenüber Zeilen- und Spaltenvertauschungen.

(ii) Gegenüber einer Multiplikation aller Häufigkeiten n_{ij} mit $c > 0$ erweist sich χ^2 als äquivariant. Die Maße ϕ^2, C und C_K sind dagegen invariant.

5.1 Zusammenhänge zwischen kategorialen Merkmalen

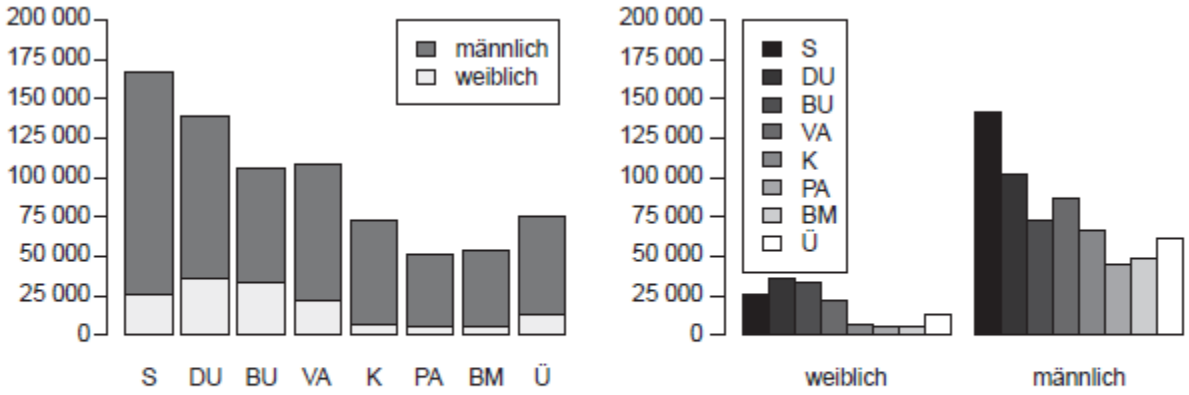
5.1.3 Grafische Analysemöglichkeiten

Gestapelte und gruppierte Säulendiagramme

- **Konzept und Beispiel**

S	Straftaten im Straßenverkehr
DU	Diebstahl und Unterschlagung
BU	Betrug und Urkundenfälschung
VA	Andere gegen das Vermögen (Raub, Erpressung, Begünstigung, Hehlerei, Sachbeschädigung, usw.)
K	Körperverletzung
PA	Andere gegen die Person (Mord, Totschlag, gegen die sexuelle Bestimmung, Beleidigung, usw.)
BM	Straftaten gegen das Betäubungsmittelgesetz
Ü	Übrige Straftaten

5.1 Zusammenhänge zwischen kategorialen Merkmalen



Daten: Statistisches Bundesamt [2014c]

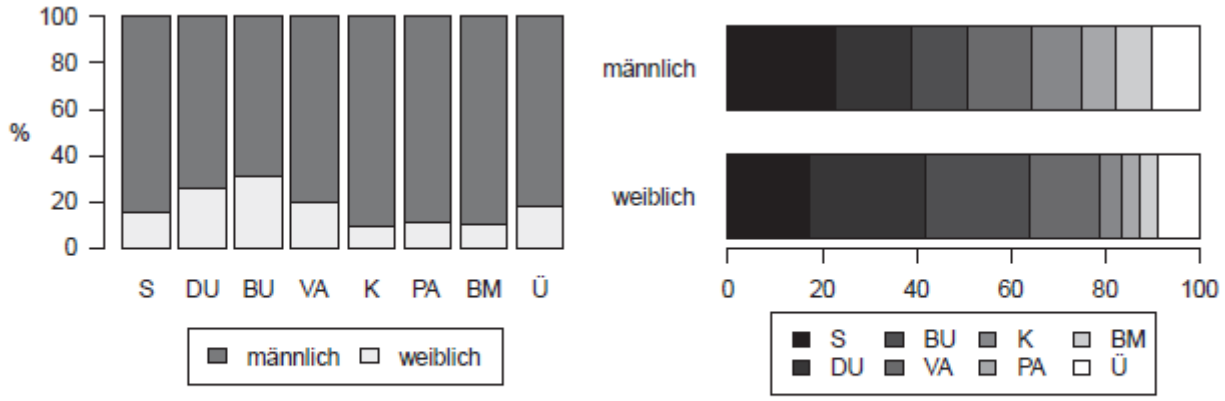
Abb. 5.1.4: Gestapeltes und gruppiertes Säulendiagramm für absolute Häufigkeitsverteilungen

● Stapelung vs. Gruppierung ●

> Stapelung: **Randverteilung** eines Merkmals leichter ersichtlich

> Gruppierung: (Absolute) **Bedingte Verteilungen** leichter ersichtlich

Segmentierte Säulen- und Balkendiagramme



Daten: Statistisches Bundesamt [2014c]

Abb. 5.1.5: Segmentiertes Säulen- und Balkendiagramm für bedingte Verteilungen

Darstellung der (relativen) **bedingten Verteilungen**

Assoziationsplots

● **Konzept** ●

> „Abweichungen von der Unabhängigkeit“: $n_{ij} - \frac{n_{i\bullet} n_{\bullet j}}{n}$

> **Pearson-Residuen**: $\frac{n_{ij} - \frac{n_{i\bullet} n_{\bullet j}}{n}}{\sqrt{\frac{n_{i\bullet} n_{\bullet j}}{n}}}$

> **Assoziationsplot**:
- Richtung der Säulen nach Vorzeichen der Residuen (nach oben oder unten)
- Höhe der Säulen proportional zu absoluten Residuen
- Breite der Säulen proportional zu Termen des Nenners
- Flächen der Säulen entsprechen den absoluten Abweichungen

> Unterschiede der bedingten Verteilungen werden hervorgehoben

5.1 Zusammenhänge zwischen kategorialen Merkmalen

> Rechenbeispiel von Folie 164 und 165

Tab. 5.1.9: Tatsächliche und zu erwartende Häufigkeiten mit jeweiligen Abweichungen

	Raucher	Gelegenheits-raucher	Nichtraucher	Summe
weiblich	4 (6.4)	8 (8)	28 (25.6)	40
	−2.4	**0**	**2.4**	
männlich	12 (9.6)	12 (12)	36 (38.4)	60
	2.4	**0**	**−2.4**	
Summe	16	20	64	100

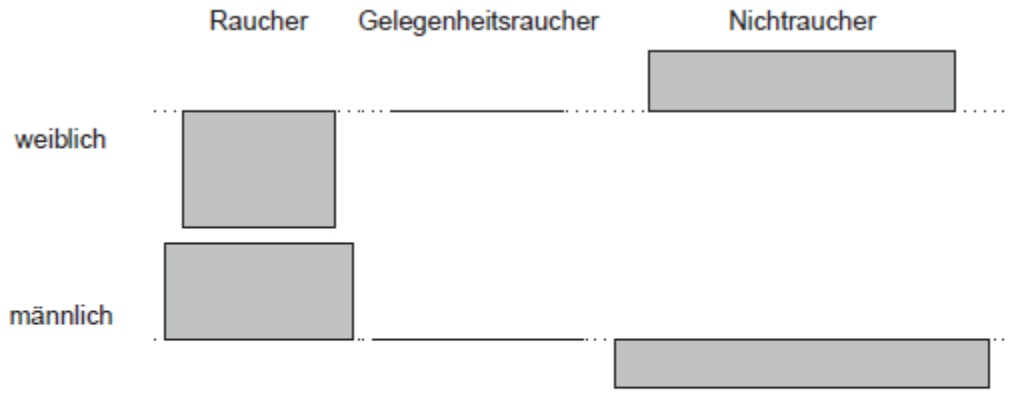
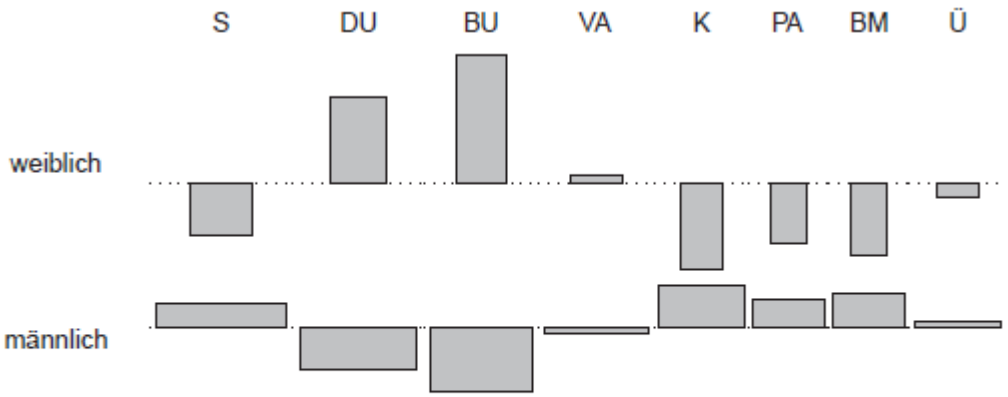


Abb. 5.1.3: Assoziationsplot der Abweichungen von der Unabhängigkeit

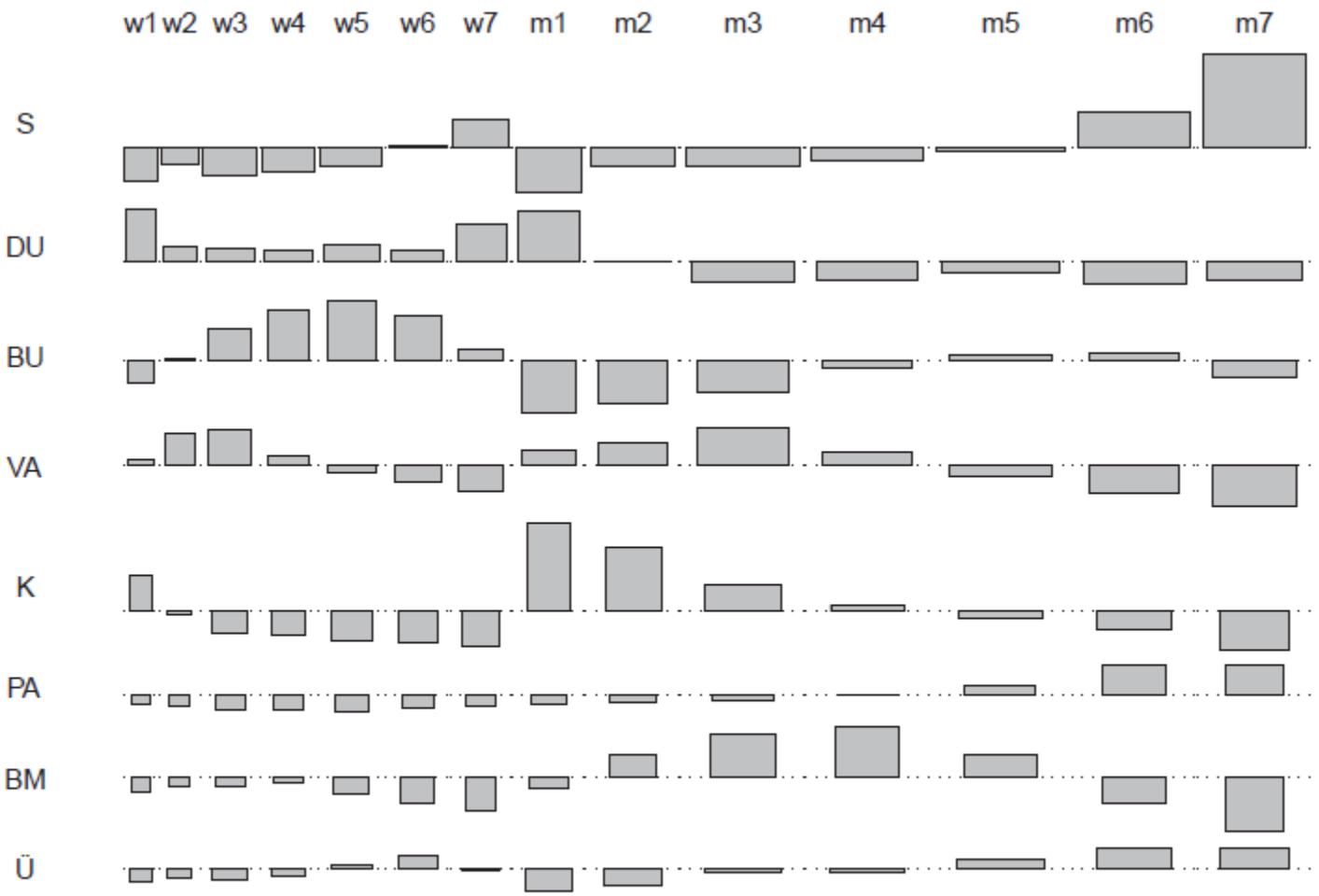
● Beispiel 5.1.3 ●



Daten: Statistisches Bundesamt [2014c]

Abb. 5.1.6: Assoziationsplot für Abweichungen von der Unabhängigkeit

5.1 Zusammenhänge zwischen kategorialen Merkmalen



Daten: Statistisches Bundesamt [2014c]

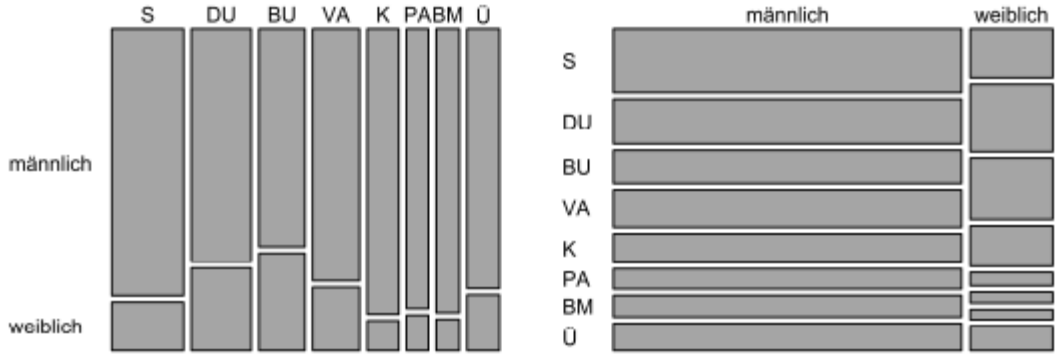
Abb. 5.3.11: Assoziationsplot: Arten von Straftaten und Altersgruppen nach Geschlecht

Mosaikplots

● Konzept ●

Breiten der Säulen (Balken) spiegeln Randverteilung eines Merkmals wider. Dadurch entsprechen Flächen der gemeinsamen relativen Verteilung (f_{ij})

● Beispiel 5.1.4 ●



Daten: Statistisches Bundesamt [2014c]

Abb. 5.1.7: Mosaikplots für gemeinsame Verteilungen

Höhe: $\frac{n_{11}}{n_{\bullet 1}} = f_{11}^{X|Y}$

Breite: $\frac{n_{\bullet 1}}{n} = f_{\bullet 1}$

Fläche: $\frac{n_{\bullet 1} n_{11}}{n \times n_{\bullet 1}} = f_{11}$

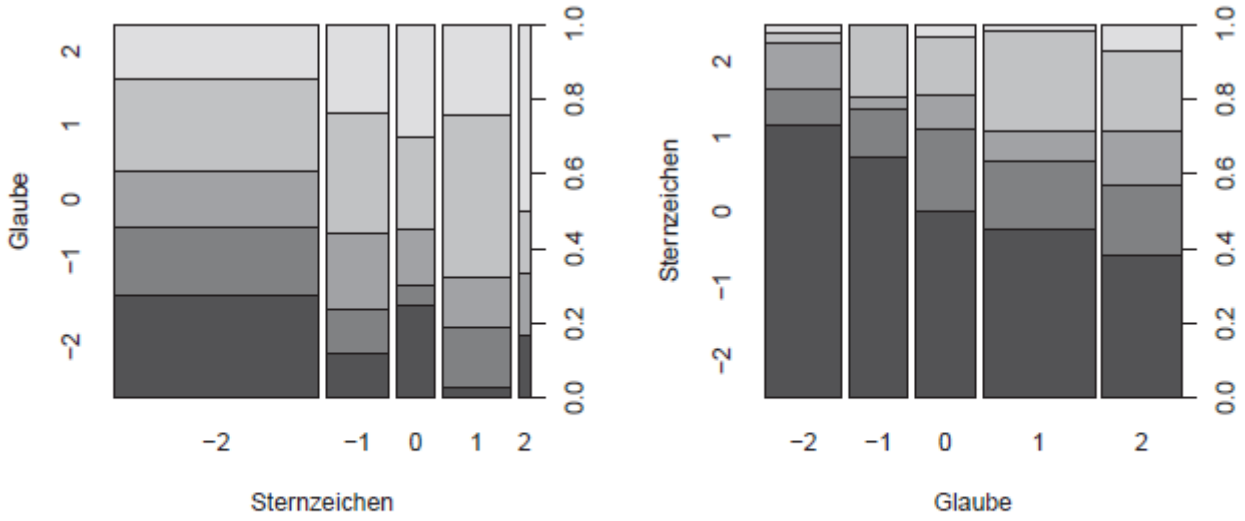
● Höherdimensionale Mosaikplots ●

Spineplots

● Konzept ●

Ähnlich wie Mosaikplots, jedoch Lücken zwischen den Säulen und Skalenachse

● Beispiel 5.1.5 ●



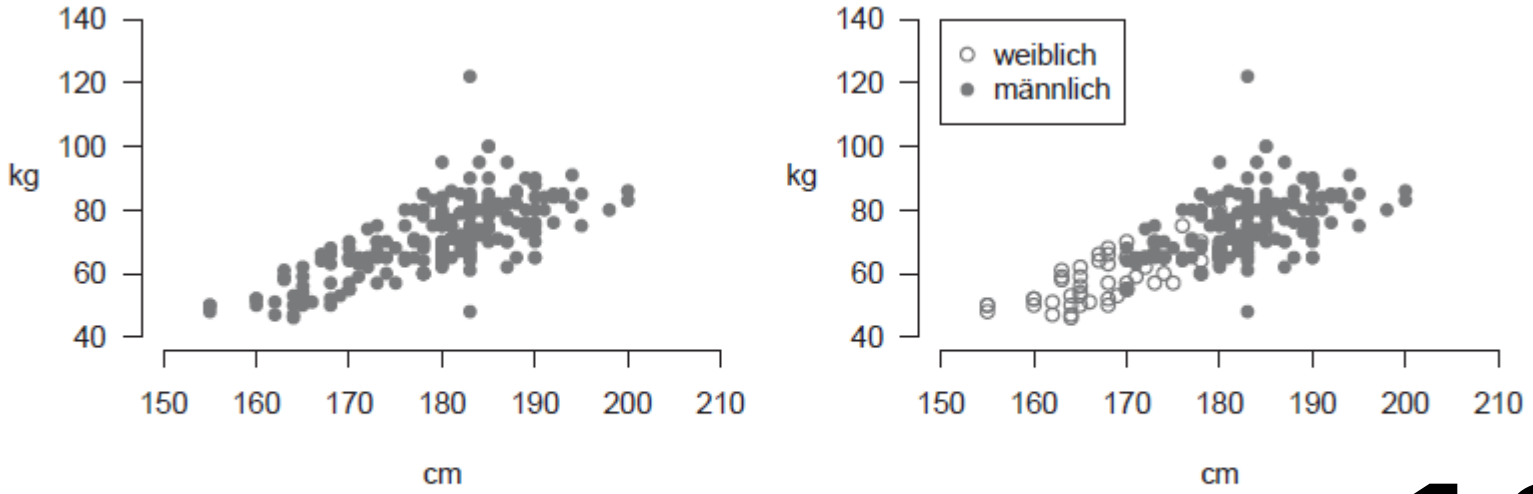
Daten: Eigene Erhebungen

Abb. 5.1.8: Spineplots: Glaube an Gott und Einstellung zu Sternzeichen

5.2.1 Grafische Analysemöglichkeiten

Streudiagramme

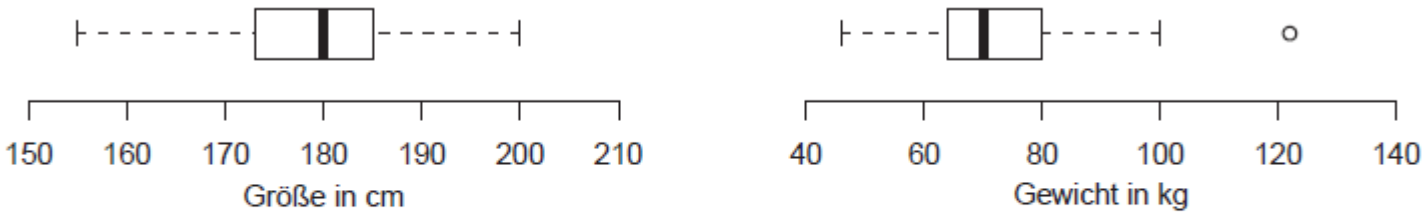
- Konzept
- Beispiel 5.2.1



Daten: Eigene Erhebungen

Abb. 5.2.1: Einfaches Streudiagramm und Streudiagramm mit kategorialer Einfärbung

5.2 Zusammenhänge zwischen metrischen Merkmalen



Daten: Eigene Erhebungen

Abb. 5.2.2: Darstellung der Randverteilungen

● Richtung von Zusammenhängen ●

> Positiv: Größere *x*-Werte gehen tendenziell mit größeren *y*-Werten einher

> Negativ: Größere *x*-Werte gehen tendenziell mit kleineren *y*-Werten einher

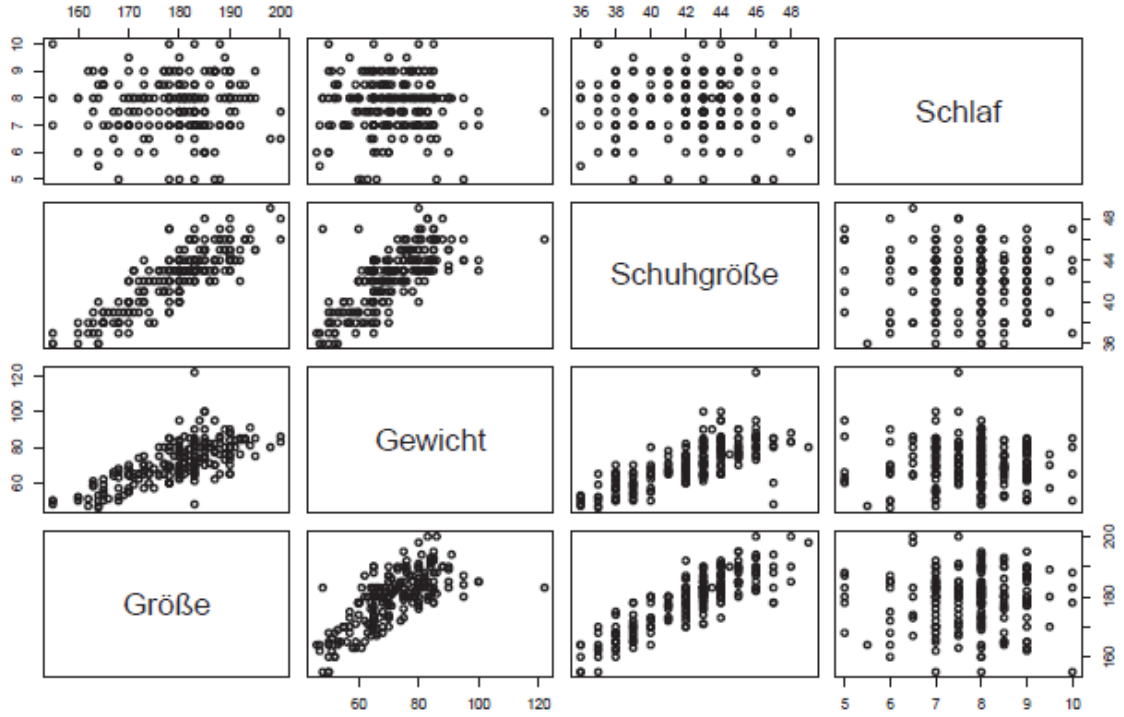
● Zweidimensionale Ausreißer ●

Höherdimensionale Ausreißer nicht zwingend in niedrigeren Dimensionen sichtbar (vgl. Abb. 5.2.1 und 5.2.2)

5.2 Zusammenhänge zwischen metrischen Merkmalen

Streudiagramm-Matrizen

- Konzept
- Beispiel 5.2.2



Daten: Eigene Erhebungen

Abb. 5.2.3: Streudiagramm-Matrix

- Hilfsmittel für „Datenscreening“ und Modellbildung

Hexagonalplots

- Konzept
- Beispiel 5.2.3

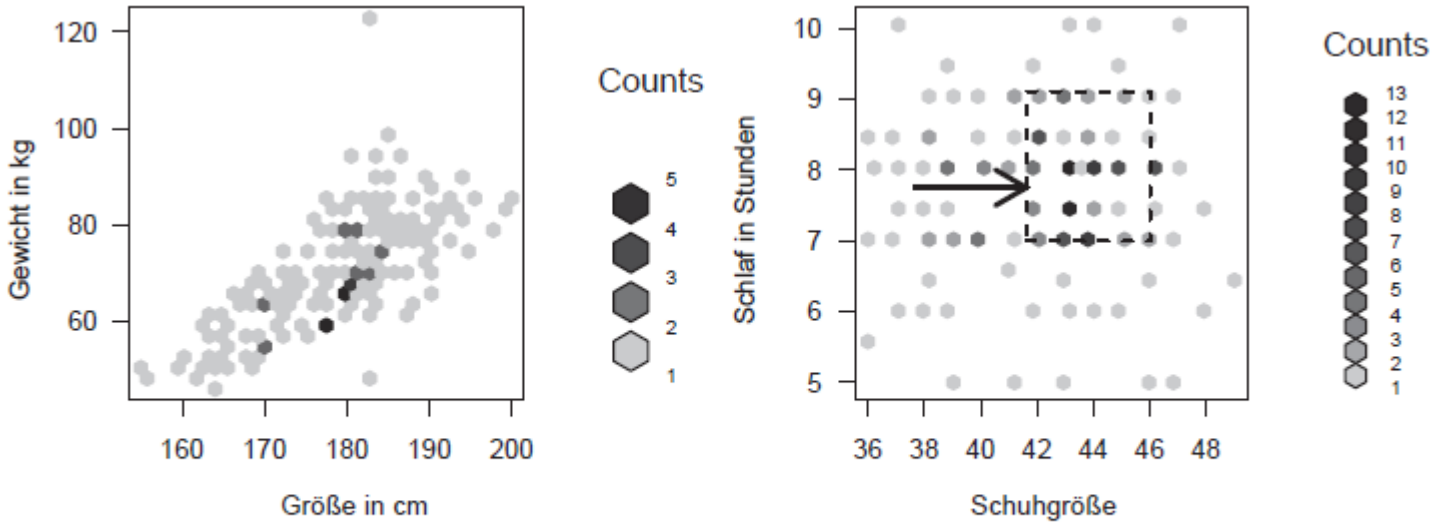


Abb. 5.2.4: Hexagonalplots zur Darstellung von Punktedichten

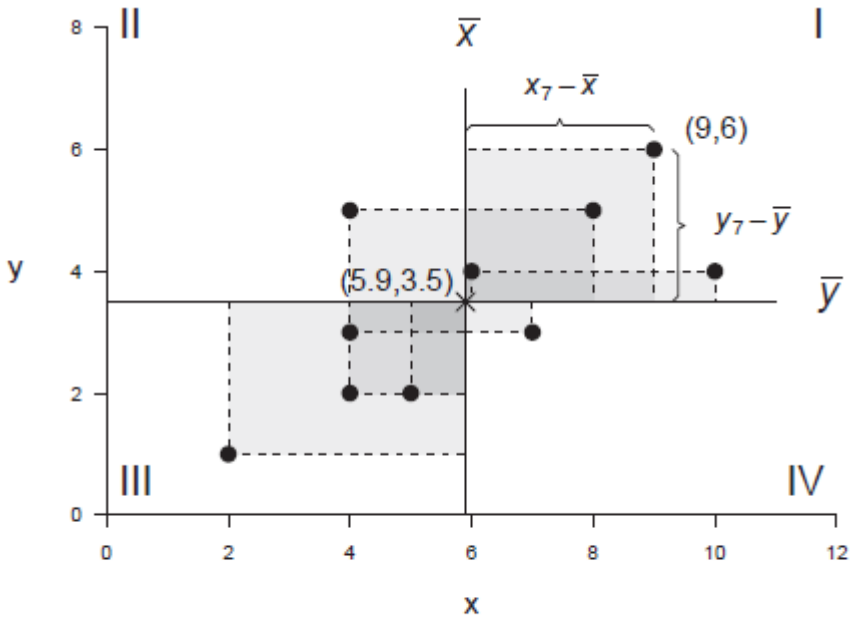
Mehr als die Hälfte aller etwa 200 Studierenden besitzt eine Schuhgröße zwischen 42 und 46 und schläft zwischen 7 und 9 Stunden (ohne Färbung nicht ersichtlich)

5.2.2 Zusammenhangsmaße für metrische Merkmale

Empirische Kovarianz

● **Definition und Interpretation** ●

> Gegeben: (2, 1), (4, 2), (4, 3), (5, 2), (6, 4), (8, 5), (9, 6), (10, 4), (4, 5), (7, 3)



$\bar{x} = 5.9 \qquad \bar{y} = 3.5$

Rechteckfläche z. B.:

$(x_7 - \bar{x}) \times (y_7 - \bar{y})$

$= (9 - 5.9) \times (6 - 3.5)$

$= 3.1 \times 2.5$

$= 7.75$

Abb. 5.2.5: Grafische Illustration zur Konzeption der Kovarianz

5.2 Zusammenhänge zwischen metrischen Merkmalen

> **Empirische Kovarianz**: $\tilde{s}_{XY} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$

> Als „durchschnittliche Rechteckfläche“ (< 0, = 0 oder > 0) interpretierbar

● **Beispiel 5.2.4** ●

Tab. 5.2.1: Ermittlung aller Rechteckflächen

i	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
1	2	1	−3.9	−2.5	9.75
2	4	2	−1.9	−1.5	2.85
3	4	3	−1.9	−0.5	0.95
4	5	2	−0.9	−1.5	1.35
5	6	4	0.1	0.5	0.05
6	8	5	2.1	1.5	3.15
7	9	6	3.1	2.5	7.75
8	10	4	4.1	0.5	2.05
9	4	5	−1.9	1.5	−2.85
10	7	3	1.1	−0.5	−0.55
Summe	59	35	0	0	24.50

Hier gilt:

$$\tilde{s}_{XY} = 24.5/10 = 2.45$$

● Verschiebungsformel für die empirische Kovarianz ●

> Es gilt:

$$\tilde{s}_{XY} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}$$

> Beweis:

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n}(x_i y_i - \bar{y}x_i - \bar{x}y_i + \bar{x}\bar{y})$$

$$= \sum_{i=1}^{n} x_i y_i - \bar{y}\sum_{i=1}^{n} x_i - \bar{x}\sum_{i=1}^{n} y_i + \sum_{i=1}^{n}\bar{x}\bar{y} = \sum_{i=1}^{n} x_i y_i - \bar{y}n\bar{x} - \bar{x}n\bar{y} + n\bar{x}\bar{y}$$

$$= \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$$

Division durch n

● Transformationseigenschaften ●

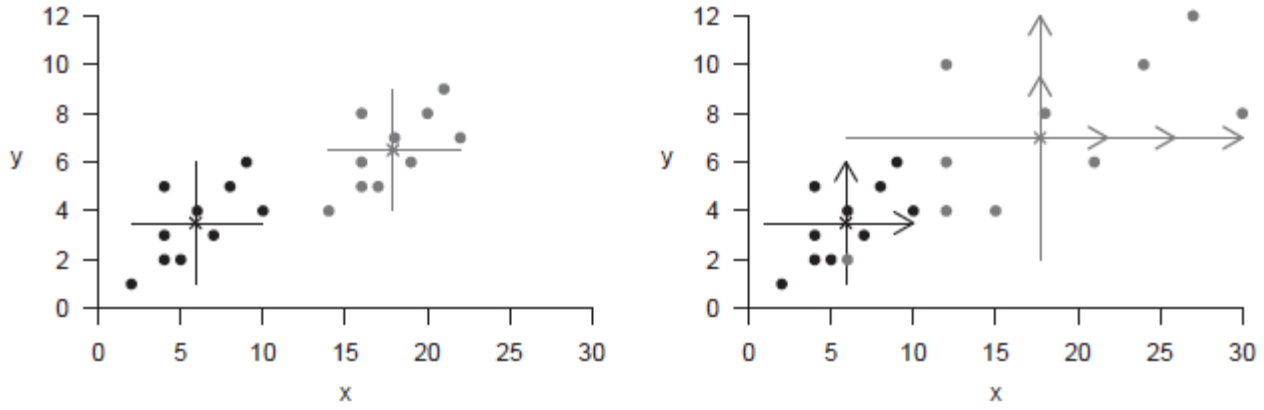


Abb. 5.2.6: Verschiebungen und Umskalierungen im zweidimensionalen Fall

● Verschiebungen ●

> Transformation: $(u_i, v_i) = (x_i + c_X, y_i + c_Y)$ für $i = 1, \ldots, n$

> Rechteckflächen bei Verschiebung sind invariant:

$$(u_i - \bar{u})(v_i - \bar{v}) = (x_i + c_X - \bar{x} - c_X)(y_i + c_Y - \bar{y} - c_Y)$$
$$= (x_i - \bar{x})(y_i - \bar{y})$$

⇒ Kovarianz ist verschiebungsinvariant

● Umskalierungen ●

> Transformation: $(u_i, v_i) = (c_X x_i, c_Y y_i)$ für $i = 1, \ldots, n$ und $c_X > 0, c_Y > 0$

> Rechteckflächen bei Umskalierung sind äquivariant im Sinne von ...

$$(u_i - \bar{u})(v_i - \bar{v}) = (c_X x_i - c_X \bar{x})(c_Y y_i - c_Y \bar{y}) = c_X c_Y (x_i - \bar{x})(y_i - \bar{y})$$

⇒ Kovarianz ist äquivariant im Sinne von...

$$\tilde{s}_{UV} = \frac{1}{n}\sum_{i=1}^{n}(u_i - \bar{u})(v_i - \bar{v}) = c_X c_Y \times \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = c_X c_Y \tilde{s}_{XY}$$

5.2 Zusammenhänge zwischen metrischen Merkmalen

Empirische Kovarianz

Für metrische Beobachtungswerte $(x_1, y_1), \ldots, (x_n, y_n)$ bezüglich zweier Merkmale X und Y berechnet sich die empirische Kovarianz als

$$\tilde{s}_{XY} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{xy}. \qquad (5.2.1)$$

Die Kovarianz ist verschiebungsinvariant und auch skalenäquivariant, sofern nur eines der beiden Merkmale umskaliert wird.

- **Zwischenschritt zum Korrelationskoeffizienten** -

> Problem: Kovarianz als Zusammenhangsmaß nicht geeignet, da skalenabhängig und Wertebereich unbeschränkt (nichtnormiertes Maß)

> Lösung: Berechne Kovarianz der standardisierten Werte

Empirischer Korrelationskoeffizient nach Pearson

● Kovarianz standardisierter Werte ●

> Unterziehe *x*- und *y*-Werte einer *Z*-Standardisierung

$$u_i = \frac{x_i - \bar{x}}{\tilde{s}_X} = \frac{1}{\tilde{s}_X} x_i - \frac{\bar{x}}{\tilde{s}_X}, \qquad v_i = \frac{y_i - \bar{y}}{\tilde{s}_Y} = \frac{1}{\tilde{s}_Y} y_i - \frac{\bar{y}}{\tilde{s}_Y}$$

> Dann gilt (vgl. Folie 115):

$$\bar{u} = 0, \qquad \bar{v} = 0, \qquad \tilde{s}_U^2 = 1, \qquad \tilde{s}_V^2 = 1$$

> Daraus folgt dann:

$$\tilde{s}_{UV} = \frac{1}{n}\sum_{i=1}^{n}(u_i - \bar{u})(v_i - \bar{v}) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\tilde{s}_X} - 0\right)\left(\frac{y_i - \bar{y}}{\tilde{s}_Y} - 0\right)$$

$$= \frac{1}{\tilde{s}_X \tilde{s}_Y} \times \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{\tilde{s}_{XY}}{\tilde{s}_X \tilde{s}_Y}$$

Korrelationskoeffizient

5.2 Zusammenhänge zwischen metrischen Merkmalen

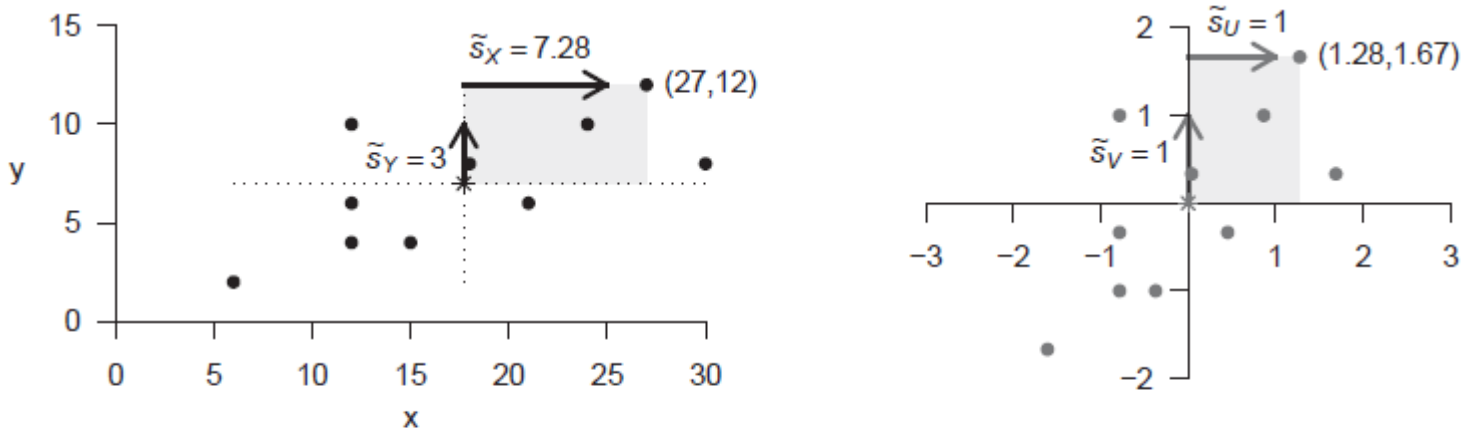


Abb. 5.2.7: Streudiagramm ursprünglicher und standardisierter Werte

● Definition – äquivalente Varianten ●

> Variante 1: $$r_{XY} = \frac{\tilde{s}_{XY}}{\tilde{s}_X \tilde{s}_Y} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Mittelwertbildungen

$$= \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i{}^2 - \bar{x}^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n} y^2 - \bar{y}^2}}$$

> Variante 2: $$r_{XY} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Summenbildungen

$$= \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sqrt{\sum_{i=1}^{n} x_i{}^2 - n\bar{x}^2}\sqrt{\sum_{i=1}^{n} y^2 - n\bar{y}^2}}$$

● Verschiebungs- und skaleninvariant ●

> Korrelationskoeffizient ist **verschiebungsinvariant** sind, da dies Kovarianz und Varianz sind.

> Der Korrelationskoeffizient ist auch **skaleninvariant**:

$$r_{UV} = \frac{\tilde{s}_{UV}}{\tilde{s}_U \tilde{s}_V} = \frac{c_X c_Y \tilde{s}_{XY}}{c_X \tilde{s}_X c_Y \tilde{s}_Y} = \frac{\tilde{s}_{XY}}{\tilde{s}_X \tilde{s}_Y}$$

● Wertebereich ●

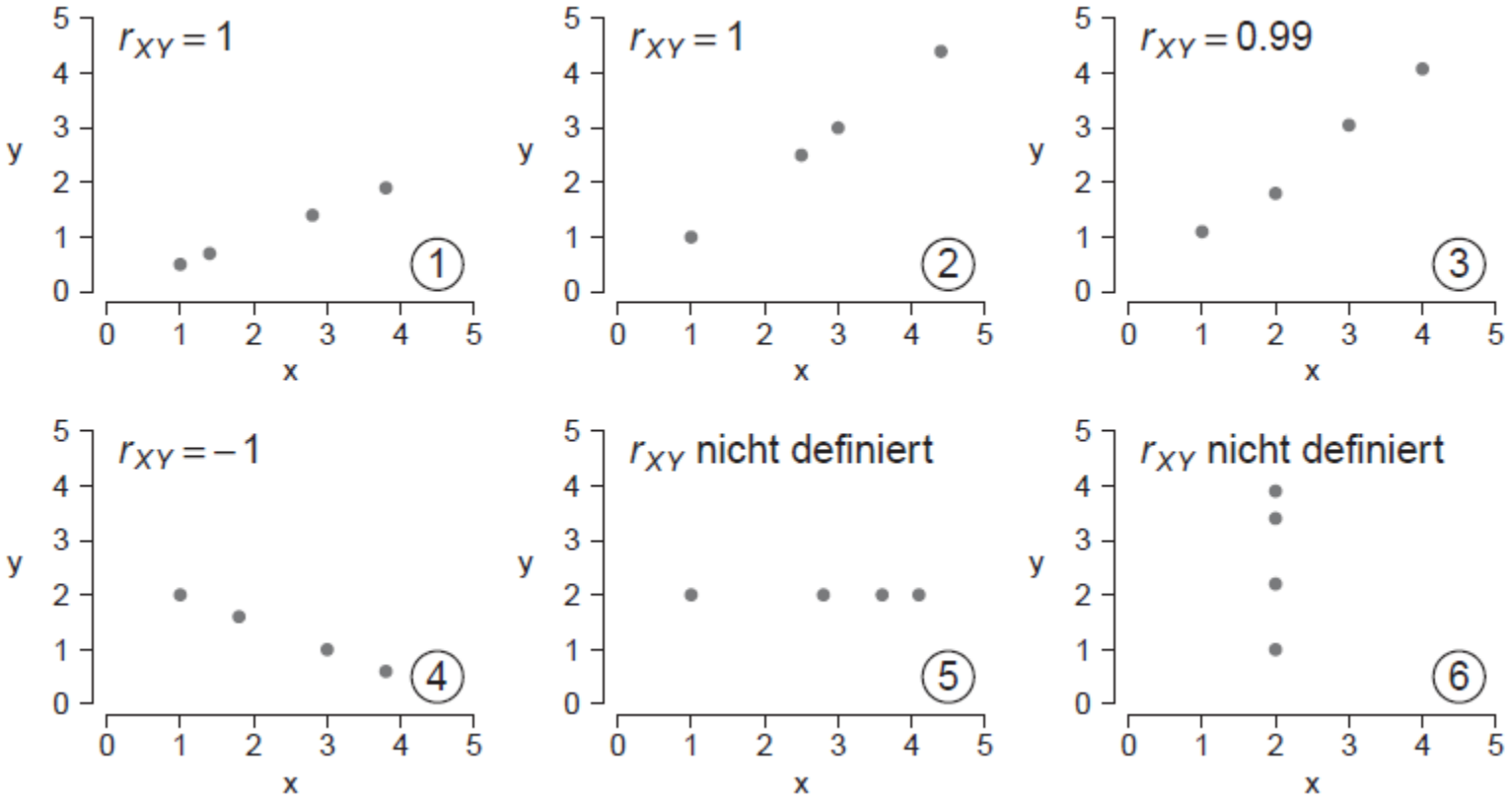
> Es lässt sich zeigen: $r_{XY} \in [-1,1]$

> Die Korrelation ist betragsmäßig genau dann gleich 1, falls die *y*-Werte perfekt linear von den *x*-Werten abhängen im Sinne von

$$(y_i - \bar{y}) = c(x_i - \bar{x}) \quad \text{für ein } c \neq 0$$
$$\Leftrightarrow \quad y_i = \bar{y} - c\bar{x} + cx_i$$
$$\Leftrightarrow \quad y_i = b_0 + b_1 x_i \quad \text{mit} \quad b_0 = \bar{y} - c\bar{x}, \quad b_1 = c$$

5.2 Zusammenhänge zwischen metrischen Merkmalen

- **Interpretation** -



5.2 Zusammenhänge zwischen metrischen Merkmalen

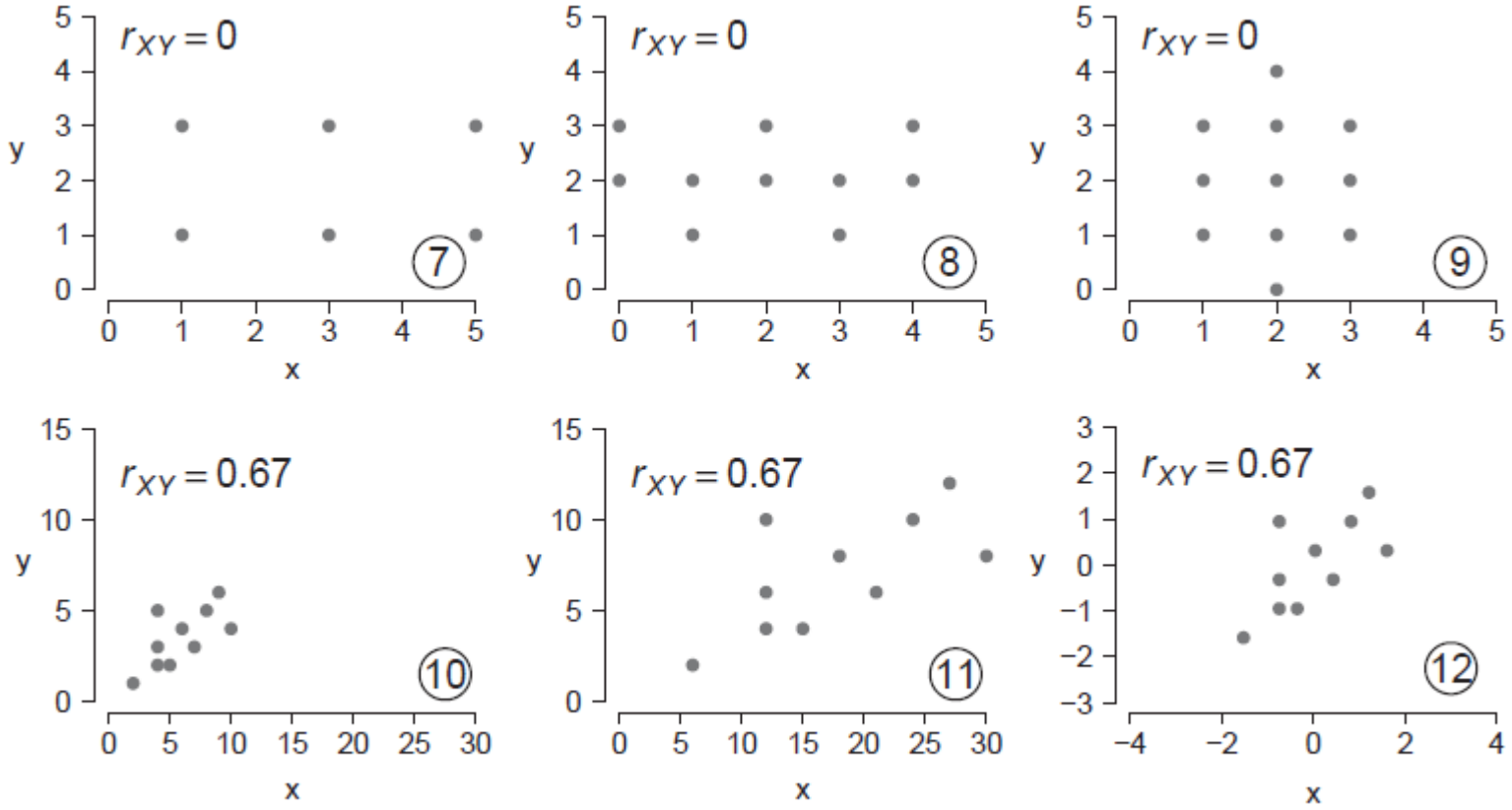


Abb. 5.2.8: Beispiele unterschiedlicher Korrelationen

Definition 5.2.1: Empirische Unkorreliertheit

x- und y-Werte heißen **(empirisch) unkorreliert,** wenn $\tilde{s}_{XY} = 0$ ist.

5.2 Zusammenhänge zwischen metrischen Merkmalen

- **Korrelation misst nur lineare Abhängigkeit**
- **Reine Optik kann täuschen**
- **Beispiel 5.2.5**

Gegeben: (2, 1), (4, 2), (4, 3), (5, 2), (6, 4), (8, 5), (9, 6), (10, 4), (4, 5), (7, 3)

Tab. 5.2.2: Arbeitstabelle zur Berechnung der Korrelation nach Pearson

i	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	2	1	4	1	2
2	4	2	16	4	8
3	4	3	16	9	12
4	5	2	25	4	10
5	6	4	36	16	24
6	8	5	64	25	40
7	9	6	81	36	54
8	10	4	100	16	40
9	4	5	16	25	20
10	7	3	49	9	21
Summe	59	35	407	145	231

$$\bar{x} = \frac{59}{10} = 5.9$$

$$\bar{y} = \frac{35}{10} = 3.5$$

$$r_{XY} = \frac{\tilde{s}_{XY}}{\tilde{s}_X \tilde{s}_Y} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} {x_i}^2 - \bar{x}^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n} y^2 - \bar{y}^2}}$$

$$= \frac{\frac{231}{10} - 5.9 \times 3.5}{\sqrt{\frac{407}{10} - 5.9^2}\sqrt{\frac{145}{10} - 3.5^2}}$$

$$\approx 0.6730$$

5.2 Zusammenhänge zwischen metrischen Merkmalen

Definition 5.2.2: Empirischer Korrelationskoeffizient nach Pearson

Für metrische Beobachtungswerte $(x_1, y_1), \ldots, (x_n, y_n)$ bezüglich zweier Merkmale X und Y berechnet sich der empirische Korrelationskoeffizient als

$$r_{XY} = \frac{\widetilde{s}_{XY}}{\widetilde{s}_X \widetilde{s}_Y} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{xy}}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i^2 - \overline{y}^2\right)}}$$

wobei $\widetilde{s}_X > 0$ und $\widetilde{s}_Y > 0$ vorausgesetzt wird. Es gilt: $r_{XY} \in [-1, 1]$. Der Korrelationskoeffizient ist verschiebungs- und skaleninvariant und misst die Stärke der linearen Abhängigkeit.

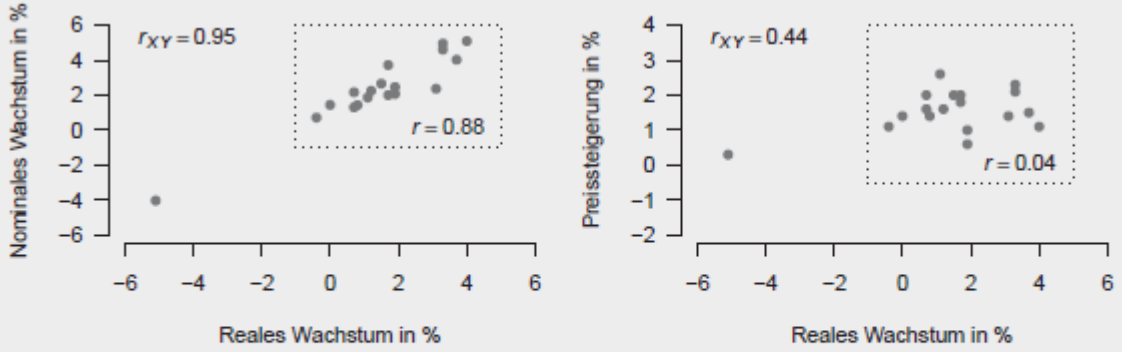
- **Verhältnis zur Kausalität**

> Keine Ursache-Wirkungs-Aussage

> Weitere Ausführungen in Abschnitt 5.3.3.

- **Zur Historie und Namensgebung des Korrelationskoeffizienten**

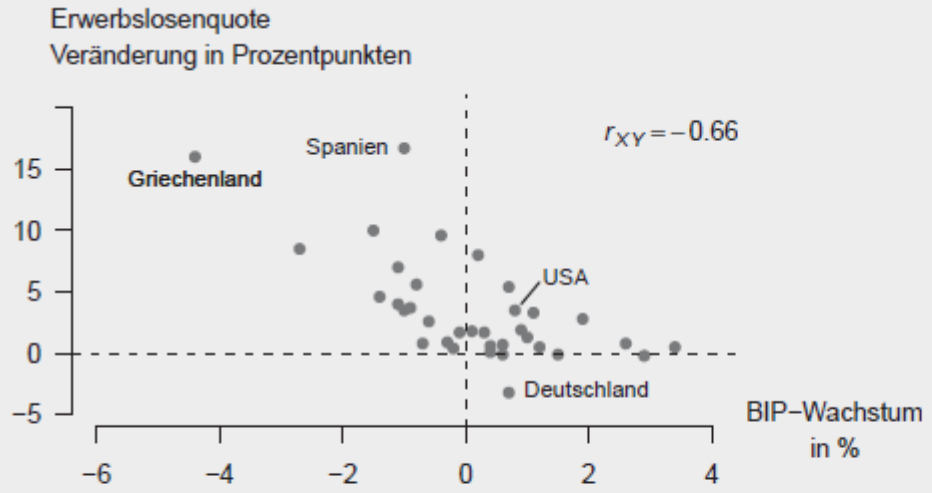
Beispiel 5.2.6: Korrelation ökonomischer Indikatoren



Daten: Statistisches Bundesamt [2014b und 2014d]

Abb. 5.2.9: Reales und nominales Wachstum, reales Wachstum und Verbraucherpreise

5.2 Zusammenhänge zwischen metrischen Merkmalen



Daten: Sachverständigenrat [2013, Schaubild 65]

Abb. 5.2.10: Wirtschafts- und Arbeitsmarktentwicklung von 2007 bis 2012 (nachgestellt)

Empirische Kovarianz- und Korrelationsmatrizen

- **Kovarianzmatrix**

	Größe	Gewicht	Schuhgröße	Schlaf
Größe	**81.1961**	77.5797	21.4765	0.1880
Gewicht	77.5797	**145.0541**	24.4409	−0.0700
Schuhgröße	21.4765	24.4409	**7.9675**	−0.0474
Schlaf	0.1880	−0.0700	−0.0474	**0.9996**

Beachte: $\tilde{s}_{XX} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x}) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \tilde{s}_X^2.$

„Kovarianz mit sich selbst“ = Varianz

- **Korrelationsmatrix**

	Größe	Gewicht	Schuhgröße	Schlaf
Größe	**1.0000**	0.7149	0.8444	0.0209
Gewicht	0.7149	**1.0000**	0.7189	−0.0058
Schuhgröße	0.8444	0.7189	**1.0000**	−0.0168
Schlaf	0.0209	−0.0058	−0.0168	**1.0000**

Beachte: $r_{XX} = \frac{\tilde{s}_{XX}}{\tilde{s}_X \tilde{s}_X} = \frac{\tilde{s}_X^2}{\tilde{s}_X^2} = 1.$ „Korrelation mit sich selbst“ = 1

Korrelationskoeffizient nach Spearman

● Hintergrund ●

> Betrachte folgende Daten: (1, 1), (8, 0.125), (4, 0.25), (2, 0.5)

> Feststellung: Perfekter monotoner Zusammenhang, jedoch Korrelation < 1

=> Korrelation misst nur Stärke der linearen Abhängigkeit

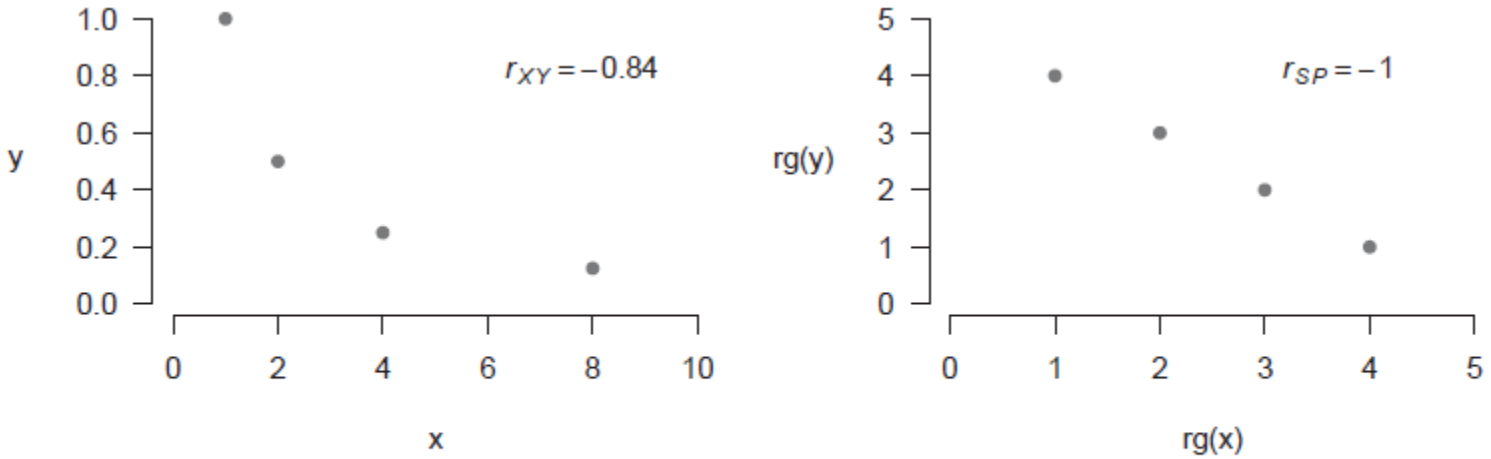


Abb. 5.2.11: Streudiagramm von Ursprungswerten und zugehörigen Rangwerten

5.2 Zusammenhänge zwischen metrischen Merkmalen

● Rangzahlen ●

> Originalwerte: $x_1 = 1, x_2 = 8, x_3 = 4, x_4 = 2$

> Geordnete Werte: $x_{(1)} = 1, x_{(2)} = 2, x_{(3)} = 4, x_{(4)} = 8$

> **Rangwerte**: $rg(x_1) = 1,\ rg(x_2) = 4,\ \ rg(x_3) = 3, rg(x_4) = 2$

> Bei **Bindungen** werden Durchschnittsränge gebildet (vgl. Tab. 5.2.3)

● Definition und Interpretation ●

> Der **Korrelationskoeffizient nach Spearman** (Rangkorrelationskoeffizient) ist der für die Rangwerte ermittelte Korrelationskoeffizient nach Pearson

$$r_S = \frac{\frac{1}{n}\sum_{i=1}^{n} rg(x_i)rg(y_i) - \overline{rg}_X\overline{rg}_Y}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} rg(x_i)^2 - \overline{rg}_X^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n} rg(y_i)^2 - \overline{rg}_Y^2}}$$

5.2 Zusammenhänge zwischen metrischen Merkmalen

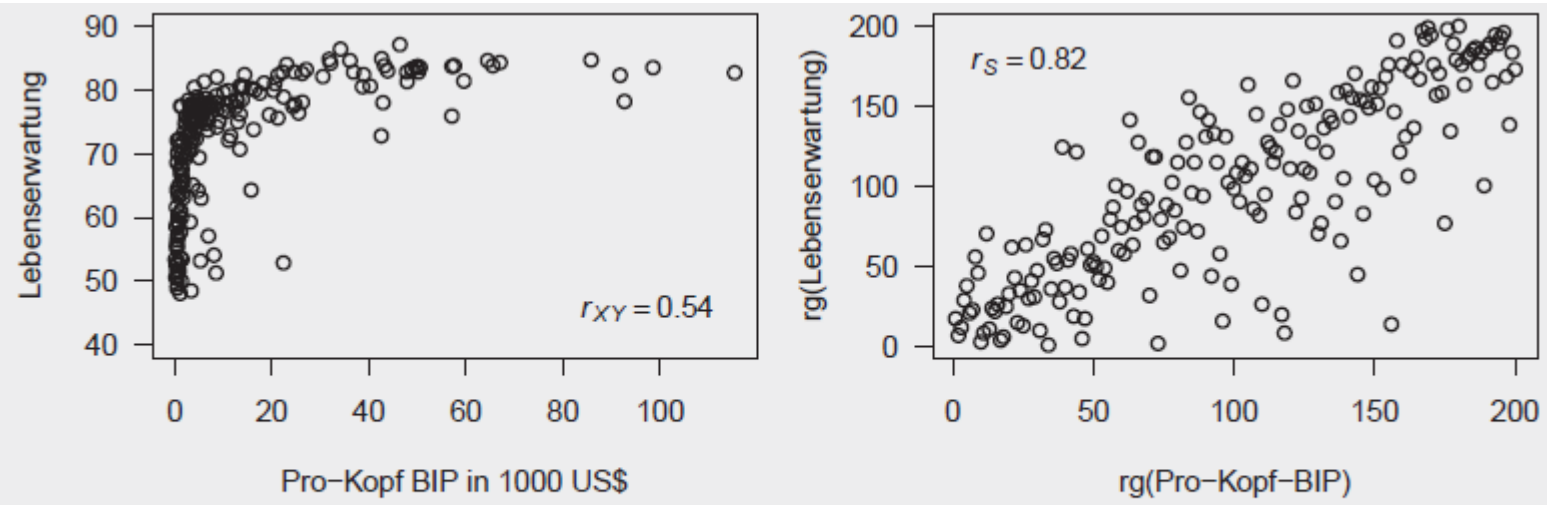
- Sinnvolle Ergänzung zur gewöhnlichen Korrelation
- Transformationseigenschaften
- Auch für Ordinalskala geeignet
- Beispiel 5.2.7

Tab. 5.2.3: Arbeitstabelle zur Berechnung des Rangkorrelationskoeffizienten

i	x_i	y_i	$rg(x_i)$	$rg(y_i)$	$rg(x_i)^2$	$rg(y_i)^2$	$rg(x_i)rg(y_i)$
1	2	1	1	1	1	1	1
2	4	2	3	2.5	9	6.25	7.5
3	4	3	3	4.5	9	20.25	13.5
4	5	2	5	2.5	25	6.25	12.5
5	6	4	6	6.5	36	42.25	39
6	8	5	8	8.5	64	72.25	68
7	9	6	9	10	81	100	90
8	10	4	10	6.5	100	42.25	65
9	4	5	3	8.5	9	72.25	25.5
10	7	3	7	4.5	49	20.25	31.5
Summe	59	35	55	55	383	383	353.5

$$r_S = \frac{353.5/10 - 55/10 \times 55/10}{\sqrt{383/10 - (55/10)^2}\sqrt{383/10 - (55/10)^2}} \approx 0.6335$$

Beispiel 5.2.8: Korrelation von Wirtschaftsleistung und Lebenserwartung



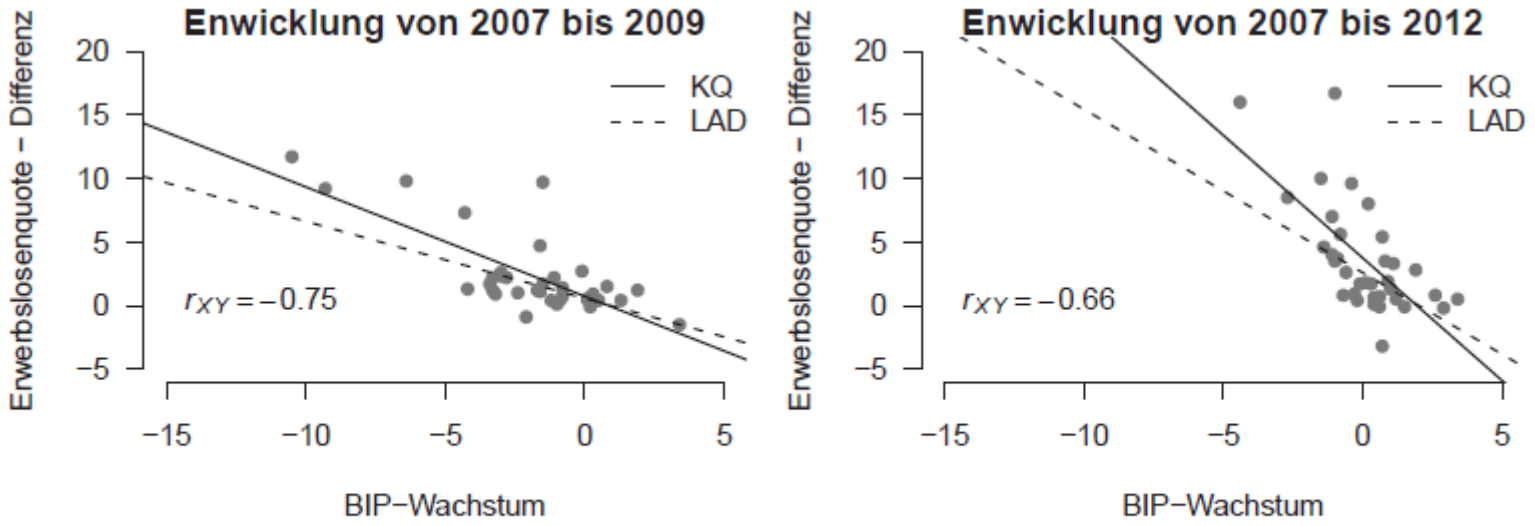
Daten: United Nations Statistics Division [2014]

Abb. 5.2.12: Zusammenhang von Wirtschaftsleistung und Lebenserwartung

5.2.3 Einfache lineare Regression

Zweck und allgemeine Vorgehensweise

● Hintergrund ●



Quelle: Sachverständigenrat [2013, Schaubild 65]

Abb. 5.2.13: Wirtschafts- und Arbeitsmarktentwicklung in unterschiedlichen Zeiträumen

5.2 Zusammenhänge zwischen metrischen Merkmalen

Vorteile einer funktionalen Beschreibung von Zusammenhängen:
Quantifizierung kausaler Effekte, Prognosen, ...

● Lineares Regressionsproblem ●

> Zusammenhang zwischen *X* und *Y* soll mittels einer Geraden

$$y = b_0 + b_1 x$$

beschrieben werden.

> Regressionsproblem: Welche Gerade passt am besten?

● Mathematisch formaler Rahmen ●

> Gegeben Beobachtungswerte: $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$

> Es gelte die Beziehung: $y_i = b_0 + b_1 x_i + u_i$ für $i = 1, \dots, n$

> Ausführlich notiert:

$$y_1 = b_0 + b_1 x_1 + u_1,$$
$$y_2 = b_0 + b_1 x_2 + u_2,$$
$$\ldots$$
$$y_n = b_0 + b_1 x_n + u_n.$$

Y X U Involvierte Variablen (Merkmale)

Y ... **Zielvariable**, **abhängige Variable**, **Regressand**, ...

X ... **erklärende Variable**, **unabhängige Variable**, **Regressor**, ...

U ... **Fehler**(variable), **Residuum**, ...

● **Interpretation von Koeffizienten und Prognosen** ●

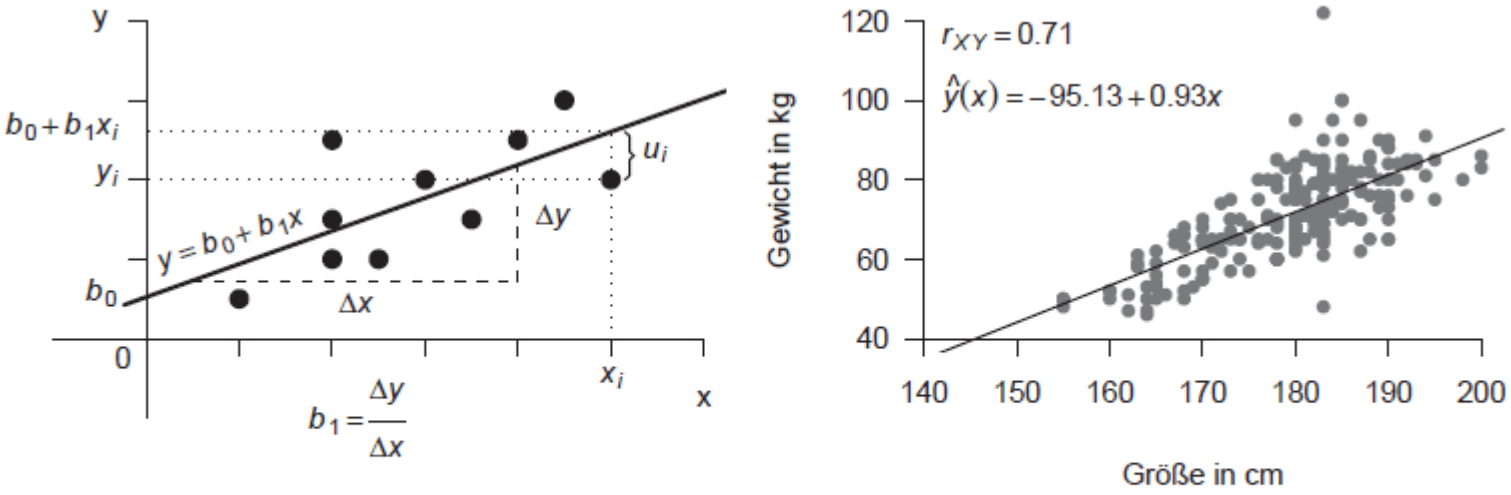
> b_1 ... Steigungskoeffizient:
Veränderung in X um Δx Einheiten geht „tendenziell“ mit einer Veränderung in Y um $b_1 \times \Delta x$ Einheiten einher

5.2 Zusammenhänge zwischen metrischen Merkmalen

> b_0 ... Achsenabschnitt:
der für Y „prognostizierte Wert für $X = 0$ (inhaltlich nicht immer sinnvoll)

> Beispiel einer Prognose (Prädiktion) des Körpergewichts einer 180 cm großen Person (Abb. 5.2.14):

$$-95.13 + 0.93 \times 180 = 72.27$$



Quelle: Eigenerhebung

Abb. 5.2.14: Einfache lineare Regression: Interpretation und Beispiel

5.2 Zusammenhänge zwischen metrischen Merkmalen

● Berechnungsmethoden ●

Minimiere $\sum_{i=1}^{n} |u_i|$ oder alternativ: $\sum_{i=1}^{n} u_i^2$

LAD-Gerade
(Least Absolute Deviation)

KQ-Gerade
(Kleinste Quadrate)

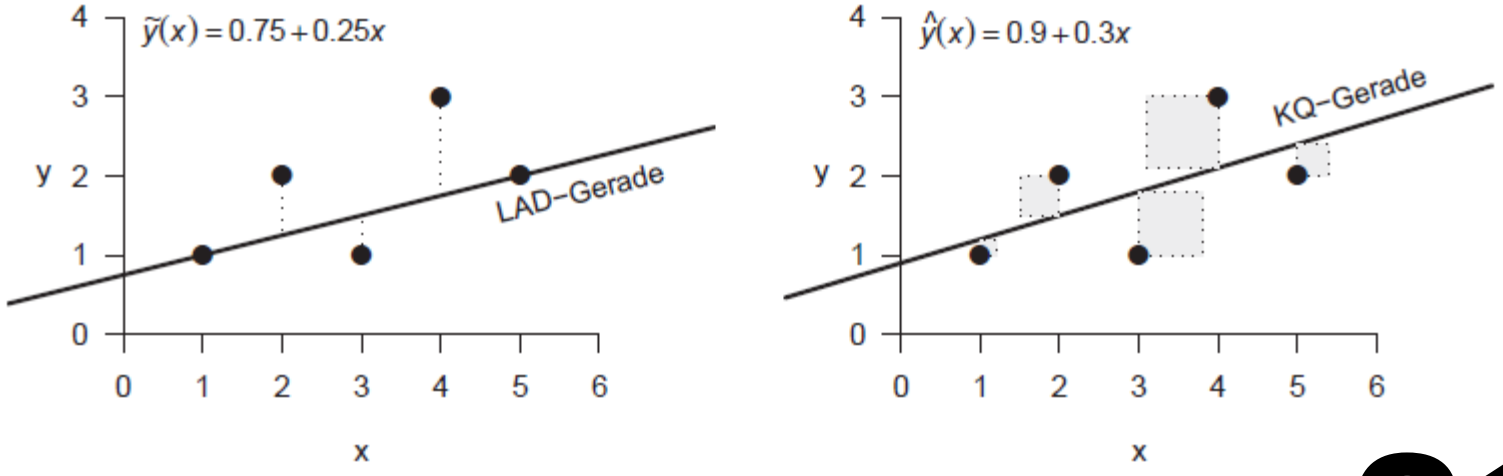


Abb. 5.2.15: Minimierung absoluter oder quadratischer Abweichungen

● Frage der Auswahl ●

> Kleinste-Quadrate-Gerade am populärsten aus praktischen und theoretische Gründen

> Allerdings: LAD-Gerade robuster gegenüber Ausreißern

> Beziehung von KQ-Gerade zu LAD-Gerade ähnlich wie arithmetisches Mittel zu Median

KQ−Methode (L_2−Regression)

● **Minimierungsproblem** ●

> Minimiere Summe der quadratischen Abweichungen:

$$\sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$$

> Definiere dazu analytische Funktion:

$$\hat{Q}(b_0, b_1) = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$$

> Analytisches Minimierungsproblem:

$$\min_{b_0, b_1} \hat{Q}(b_0, b_1)$$

> Beispiel:

Beobachtungswerte: (1, 1), (2, 2), (3, 1), (4, 3), (5, 2) (vgl. Folie 220)

$$\hat{Q}(b_0, b_1) = \sum_{i=1}^{5} (y_i - b_0 - b_1 x_i)^2$$

$$= (y_1 - b_0 - b_1 x_1)^2 + (y_2 - b_0 - b_1 x_2)^2 + (y_3 - b_0 - b_1 x_3)^2$$

$$+ (y_4 - b_0 - b_1 x_4)^2 + (y_5 - b_0 - b_1 x_5)^2$$

$$= (1 - b_0 - b_1)^2 + (2 - b_0 - 2b_1)^2 + (1 - b_0 - 3b_1)^2$$

$$+ (3 - b_0 - 4b_1)^2 + (2 - b_0 - 5b_1)^2$$

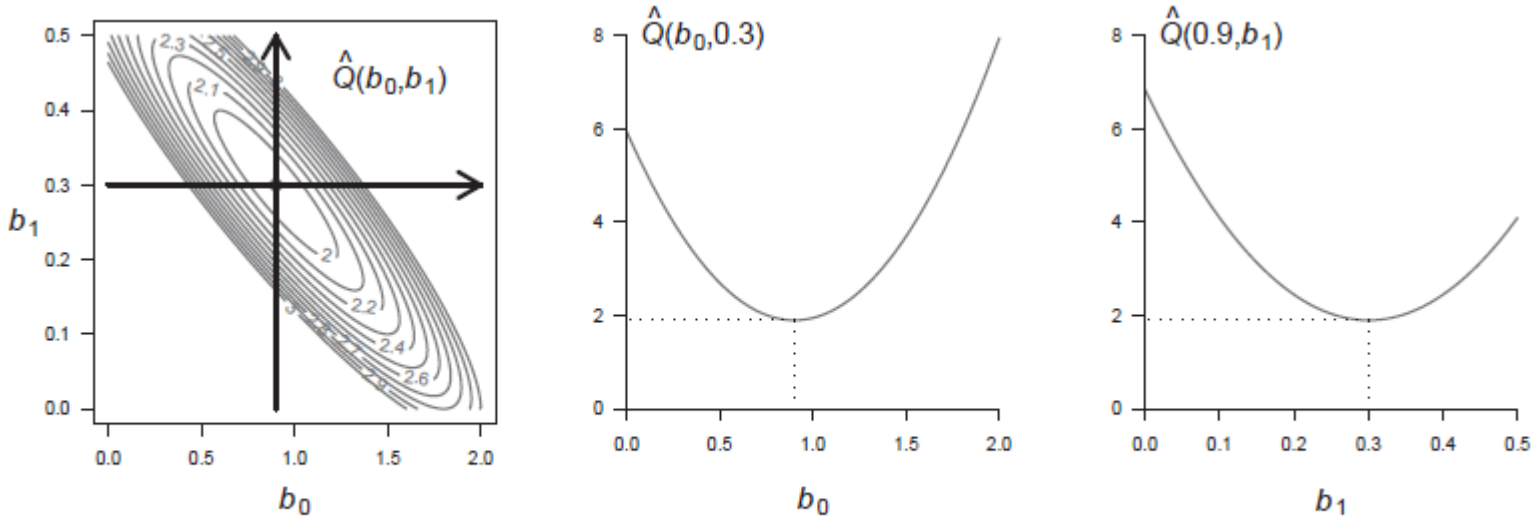


Abb. 5.2.16: L_2-Regression: Minimierung der Zielfunktion

> Im vorliegenden Beispiel lauten die Lösungen:

$$\hat{b}_0 = 0.9 \text{ und } \hat{b}_1 = 0.3 \text{ mit } \hat{Q}(0.9, 0.3) = 1.9$$

● Lösung und Berechnung ●

> Allgemeine Lösungsformeln (werden in Kap. 12 hergeleitet, Folie 851):

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x} \quad \text{und} \quad \hat{b}_1 = \frac{\tilde{s}_{XY}}{\tilde{s}_X^2}$$

> Für $\tilde{s}_X^2 = 0$ sind die Lösungen nicht eindeutig (Abb. 5.2.17)

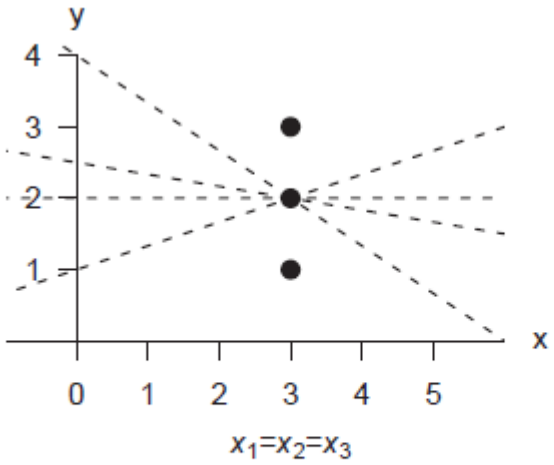


Abb. 5.2.17: L_2-Regression: Beispiel einer nicht eindeutigen Lösung

● Bezeichnungen ●

> **KQ-Gerade**: $\hat{y}(x) = \hat{b}_0 + \hat{b}_1 x$

> **Gefittete Werte**: $\hat{y}_i = \hat{y}(x_i) = \hat{b}_0 + \hat{b}_1 x_i$

> **KQ-Residuen**: $\hat{u}_i = y_i - \hat{y}_i$

● Eigenschaften der KQ-Gerade ●

> (1) KQ-Gerade verläuft stets durch den Schwerpunkt $(\bar{x}, \bar{y})$

(2) Die Summe der gefitteten Werte ist gleich der Summe der *y*-Werte:

$$\sum_{i=1}^{n} \hat{y}_i = \sum_{i=1}^{n} y_i$$

(3) Die Summe der KQ-Residuen ist gleich 0:

$$\sum_{i=1}^{n} \hat{u}_i = 0$$

> Nachweis erfolgt in Kap. 12, Folie 855 ff.

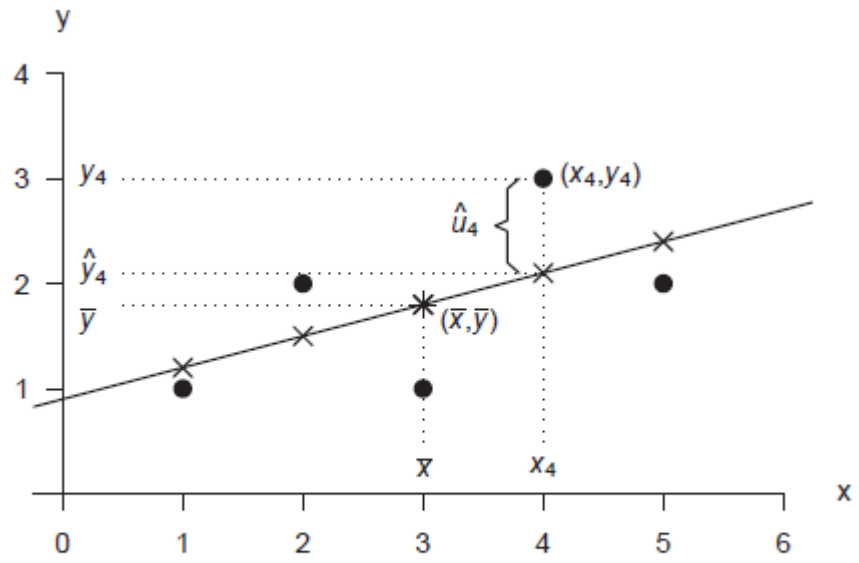


Abb. 5.2.18: L_2-Regression durch den Schwerpunkt

● Beispiel 5.2.9 ●

Beobachtungswerte: (1, 1), (2, 2), (3, 1), (4, 3), (5, 2)

Tab. 5.2.4: Arbeitstabelle zur Bestimmung einer KQ-Geraden

i	x_i	y_i	$x_i y_i$	x_i^2	$\hat{y}_i$	$\hat{u}_i$	$\lvert\hat{u}_i\rvert$	$\hat{u}_i^2$
1	1	1	1	1	1.2	−0.2	0.2	0.04
2	2	2	4	4	1.5	0.5	0.5	0.25
3	3	1	3	9	1.8	−0.8	0.8	0.64
4	4	3	12	16	2.1	0.9	0.9	0.81
5	5	2	10	25	2.4	−0.4	0.4	0.16
Summe	15	9	30	55	9	0	2.8	1.9

$$\bar{x} = \frac{15}{5} = 3$$

$$\bar{y} = \frac{9}{5} = 1.8$$

$$\hat{b}_1 = \frac{\tilde{s}_{XY}}{\tilde{s}_X^2} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2} = \frac{30/5 - 3 \times 1.8}{55/5 - 3^2} = 0.3$$

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x} = 1.8 - 0.3 \times 3 = 0.9$$

$$\Rightarrow \quad \hat{y}(x) = \hat{b}_0 + \hat{b}_1 x = 0.9 + 0.3x$$

● Interpretation der KQ-Gerade ●

> Deutung von Steigung und Achsenabschnitt wie auf Folien 217-218

> „Durchschnittsgerade“ (Schwerpunkteigenschaft)

Beispiel von Folie 218:

„*Mit jedem Zentimeter nimmt das Gewicht* ***durchschnittlich*** *um 0.93 Kilogramm zu*“

„*Eine 180 cm große Person sollte* ***im Durchschnitt*** *72.27 kg wiegen*“

● Bezug zum Korrelationskoeffizienten ●

Es gilt: $\hat{b}_1 = \frac{\tilde{s}_{XY}}{\tilde{s}_X^2}$ und $r_{XY} = \frac{\tilde{s}_{XY}}{\tilde{s}_X \tilde{s}_Y}$

Daraus folgt:

$$\hat{b}_1 < 0 \quad \Leftrightarrow \quad r_{XY} < 0$$

$$\hat{b}_1 = 0 \quad \Leftrightarrow \quad r_{XY} = 0$$

$$\hat{b}_1 > 0 \quad \Leftrightarrow \quad r_{XY} > 0$$

● Bezug zum arithmetischen Mittel ●

Betrachtet man Minimierungsproblem mit Restriktion $b_1 = 0$, erhält man

$$\min_{b_0} \hat{f}(b_0) \quad \text{mit} \quad \hat{f}(b_0) = \sum_{i=1}^{n} (y_i - b_0)^2$$

Die Lösung hierzu lautet bekanntlich: $\hat{b}_0 = \bar{y}$

● Vor- und Nachteile der KQ-Regression ●

> Einfache Berechenbarkeit, gute Interpretierbarkeit, Eindeutigkeit

> In gewisser Weise optimal (**Gauß-Markov-Theorem**)

> Jedoch: Empfindlich gegenüber Ausreißern (nicht robust)

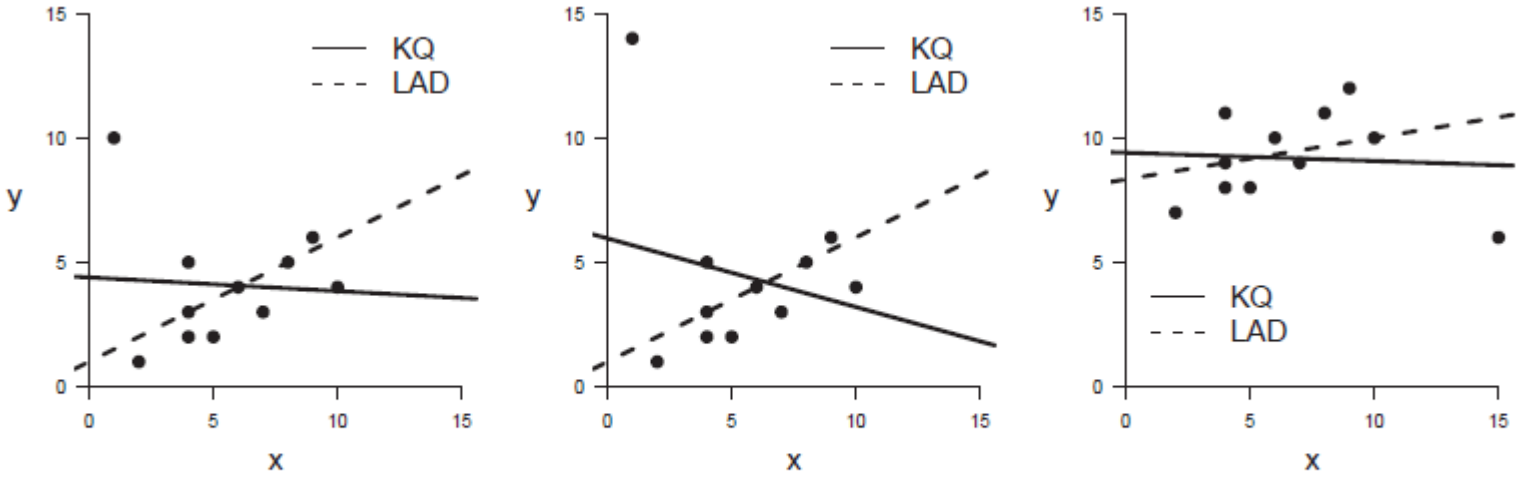


Abb. 5.2.19: KQ-Gerade und LAD-Gerade bei Ausreißern

● Anmerkungen zur Historie ●

LAD−Methode (L_1−Regression)

● **Minimierungsproblem** ●

> Minimiere Summe der absoluten Abweichungen:

$$\sum_{i=1}^{n} |u_i| = \sum_{i=1}^{n} |y_i - b_0 - b_1 x_i|$$

> Definiere dazu Funktion:

$$\tilde{Q}(b_0, b_1) = \sum_{i=1}^{n} |y_i - b_0 - b_1 x_i|$$

> Minimierungsproblem:

$$\min_{b_0, b_1} \tilde{Q}(b_0, b_1)$$

> Beispiel:

Beobachtungswerte: (1, 1), (2, 2), (3, 1), (4, 3), (5, 2) (vgl. Folie 219)

$$\tilde{Q}(b_0, b_1) = \sum_{i=1}^{5} |y_i - b_0 - b_1 x_i|$$

$$= |y_1 - b_0 - b_1 x_1| + |y_2 - b_0 - b_1 x_2| + |y_3 - b_0 - b_1 x_3|$$

$$+ |y_4 - b_0 - b_1 x_4| + |y_5 - b_0 - b_1 x_5|$$

$$= |1 - b_0 - b_1| + |2 - b_0 - 2b_1| + |1 - b_0 - 3b_1|$$

$$+ |3 - b_0 - 4b_1| + |2 - b_0 - 5b_1|$$

5.2 Zusammenhänge zwischen metrischen Merkmalen

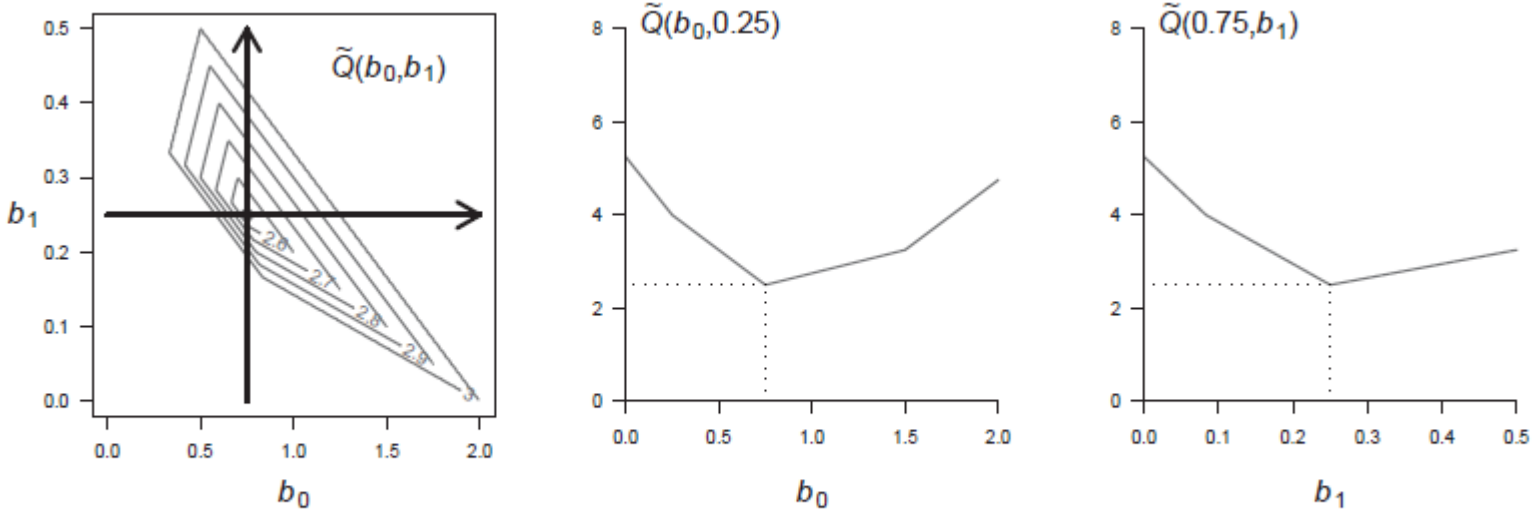


Abb. 5.2.20: L_1-Regression: Minimierung der Zielfunktion

> Im vorliegenden Beispiel lauten die Lösungen:

$$\tilde{b}_0 = 0.75 \text{ und } \tilde{b}_1 = 0.24 \text{ mit } \tilde{Q}(0.75, 0.25) = 2.5$$

- **Lösung und Berechnung**

Analytisch nicht handhabbar; stattdessen: numerische Bestimmung

- **Eindeutigkeitsprobleme**

(siehe nächste Folie)

5.2 Zusammenhänge zwischen metrischen Merkmalen

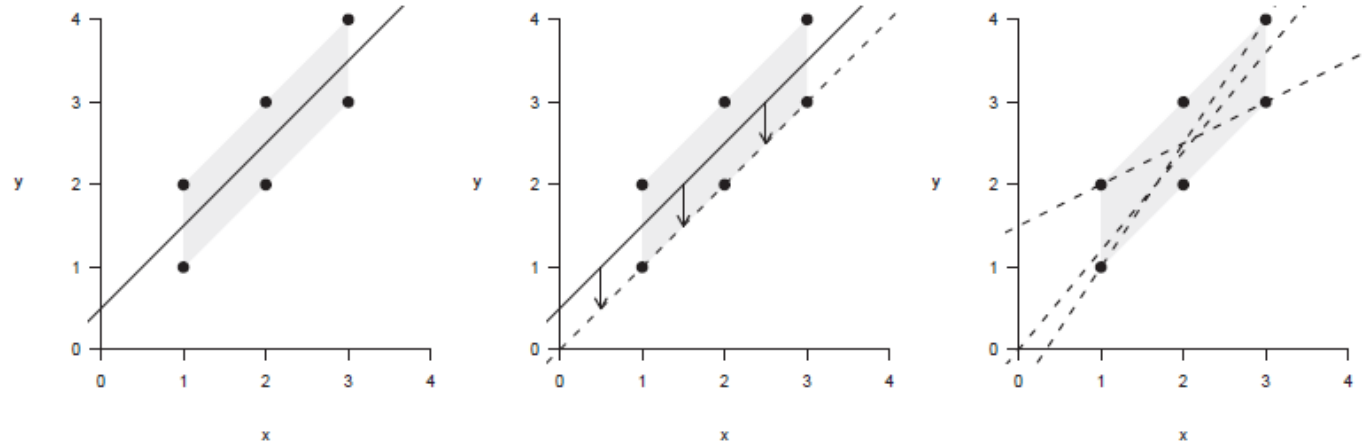


Abb. 5.2.21: L_1-Regression: Beispiel einer nicht eindeutigen Lösung

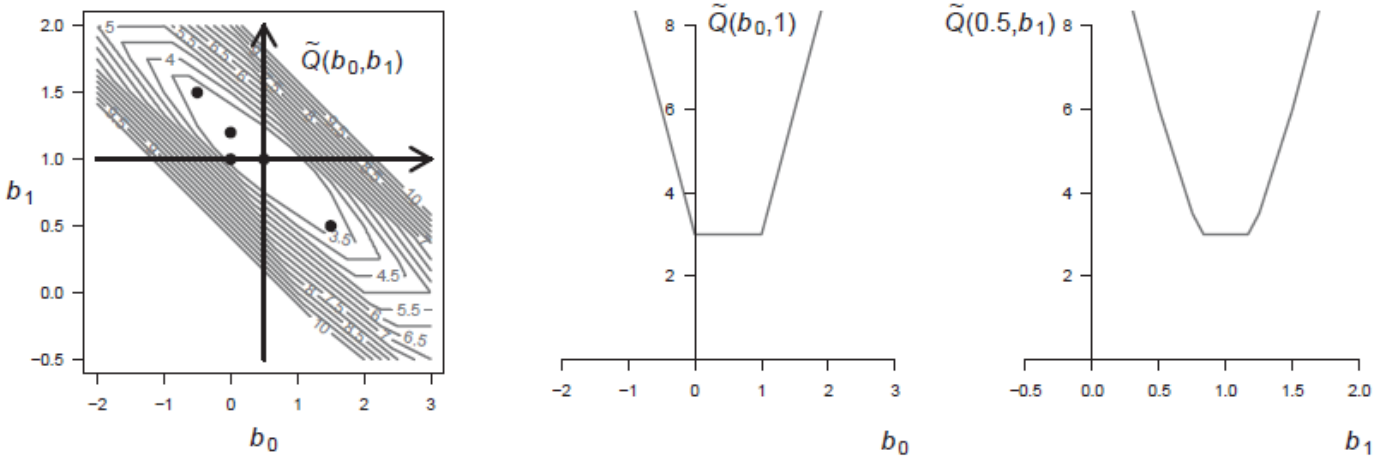


Abb. 5.2.22: Minimierung der Zielfunktion bei nicht eindeutiger Lösung

● Bezeichnungen ●

> **LAD-Gerade**: $\tilde{y}(x) = \tilde{b}_0 + \tilde{b}_1 x$

> **Gefittete Werte**: $\tilde{y}_i = \tilde{y}(x_i) = \tilde{b}_0 + \tilde{b}_1 x_i$

> **LAD-Residuen**: $\tilde{u}_i = y_i - \tilde{y}_i$

● Eigenschaften der LAD-Gerade ●

> 3 Eigenschaften der KQ-Gerade (Folie 225) gelten nicht mehr!

> „Mediangerade “ in folgendem Sinne:
Etwa 50% der Beobachtungen liegt oberhalb und 50% unterhalb der Geraden

> Verallgemeinerung im Rahmen der sog. **Quantilsregression**

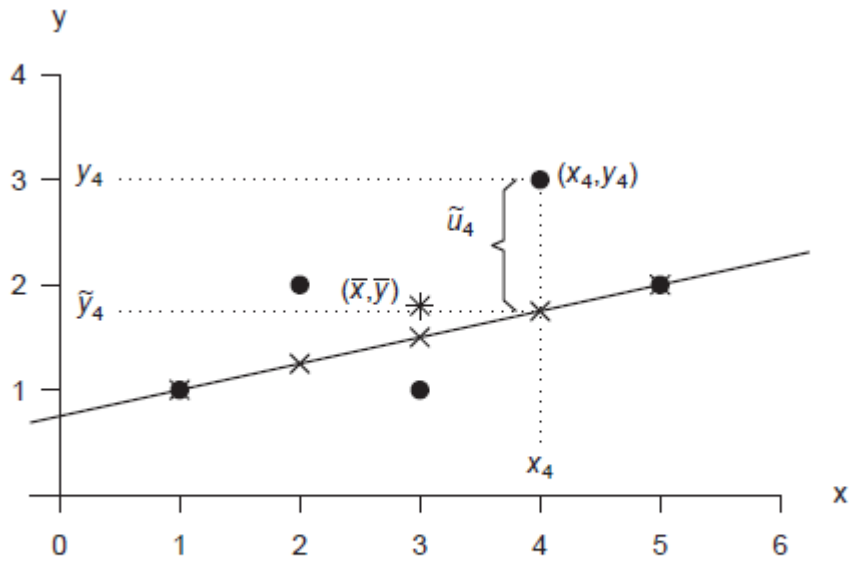


Abb. 5.2.23: LAD-Gerade – verläuft hier nicht durch Schwerpunkt

● Beispiel 5.2.10 ●

Beobachtungswerte: (1, 1), (2, 2), (3, 1), (4, 3), (5, 2)

Numerische Ermittlung der LAD-Geraden (z. B. mit Software R):

$$\tilde{y}(x) = \tilde{b}_0 + \tilde{b}_1 x = 0.75 + 0.25x$$

Tab. 5.2.5: Ergebnisse der L_1-Regression

i	x_i	y_i	$\tilde{y}_i$	$\tilde{u}_i$	$\lvert\tilde{u}_i\rvert$	$\tilde{u}_i^2$
1	1	1	1	0	0	0
2	2	2	1.25	0.75	0.75	0.5625
3	3	1	1.5	−0.5	0.5	0.25
4	4	3	1.75	1.25	1.25	1.5625
5	5	2	2	0	0	0
Summe	15	9	7.5	1.5	2.5	2.125

Wohlgemerkt:

$$\sum_{i=1}^{n} \tilde{y}_i \neq \sum_{i=1}^{n} y_i$$

$$\sum_{i=1}^{n} \tilde{u}_i \neq 0$$

Beispielsweise gilt hier:

$$\tilde{y}_1 = \tilde{y}(x_1) = \tilde{b}_0 + \tilde{b}_1 x_1 = 0.75 + 0.25 \times 1 = 1$$

$$\tilde{y}_2 = \tilde{y}(x_2) = \tilde{b}_0 + \tilde{b}_1 x_2 = 0.75 + 0.25 \times 2 = 1.25$$

$$\vdots$$

$$\tilde{u}_1 = y_1 - \tilde{y}_1 = 1 - 1 = 0$$

$$\tilde{u}_2 = y_2 - \tilde{y}_2 = 2 - 1.25 = 0.75$$

$$\vdots$$

Außerdem zu beobachten (kein Zufall!):

$$\sum_{i=1}^{n} \hat{u}_i^2 \leq \sum_{i=1}^{n} \tilde{u}_i^2 \quad \text{und} \quad \sum_{i=1}^{n} |\tilde{u}_i| \leq \sum_{i=1}^{n} |\hat{u}_i|$$

● Interpretation der LAD-Gerade ●

> Deutung von Steigung und Achsenabschnitt wie auf Folien 217-218

> „Mediangerade “ (Folie 235)

> Wesentlich sind die Robustheitseigenschaften

● Bezug zum Median ●

Betrachtet man Minimierungsproblem mit Restriktion $b_1 = 0$, erhält man

$$\min_{b_0} \tilde{f}(b_0) \quad \text{mit} \quad \tilde{f}(b_0) = \sum_{i=1}^{n} |y_i - b_0|$$

Die Lösung hierzu lautet bekanntlich: $\tilde{b}_0 = \tilde{y}_{0.5}$

- **Vor- und Nachteile der LAD-Regression**

> Unempfindlich(er) gegenüber Ausreißern (**robust**)

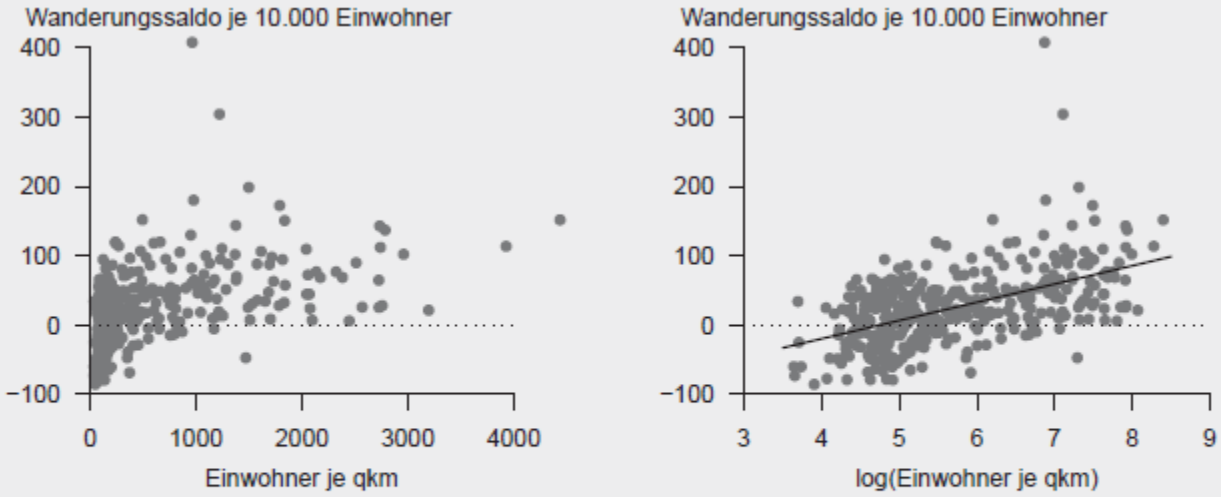
> Schwieriger zu berechnen (numerisches Problem), Eindeutigkeits-probleme können auftreten

- **Anmerkungen zur Historie**

Kritische Punkte und Alternativen

- **Frage der Richtung der Abhängigkeit**
- **Multiple lineare Regression**
- **Nichtlineare Regressionsansätze**

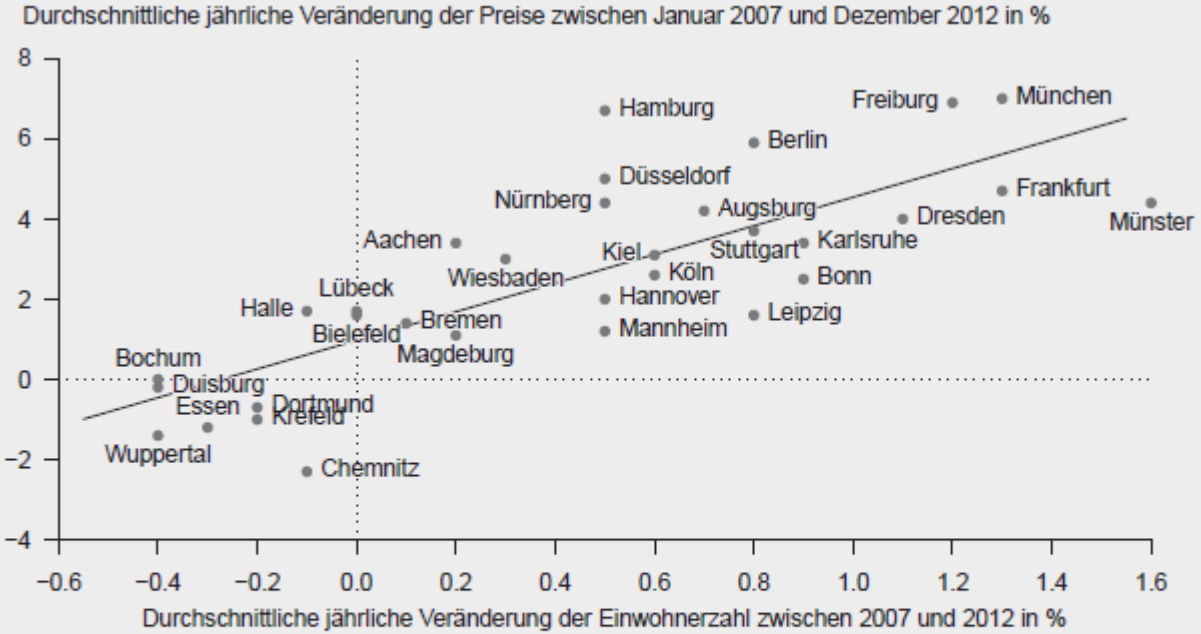
Beispiel 5.2.11: Entwicklung von Wohnungspreisen in Deutschland



Daten: Statistische Ämter des Bundes und der Länder [2014]

Abb. 5.2.24: Bevölkerungsdichte und Wanderungssaldo in Deutschland (2011)

5.2 Zusammenhänge zwischen metrischen Merkmalen



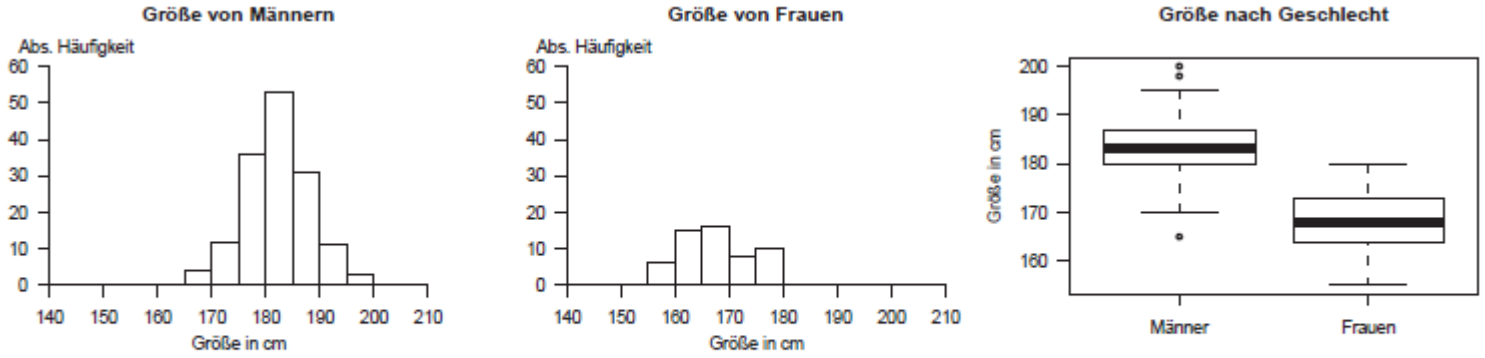
Daten: Sachverständigenrat [2013, Schaubild 115]

Abb. 5.2.25: Entwicklung von Einwohnerzahlen und Wohnungspreisen deutscher Großstädte (nachgestellt)

5.3.1 Zusammenhänge in anderen Skalierungsfällen

Grafische Analysemöglichkeiten

- **Metrisches vs. kategoriales Merkmal** -



Daten: Eigene Erhebungen

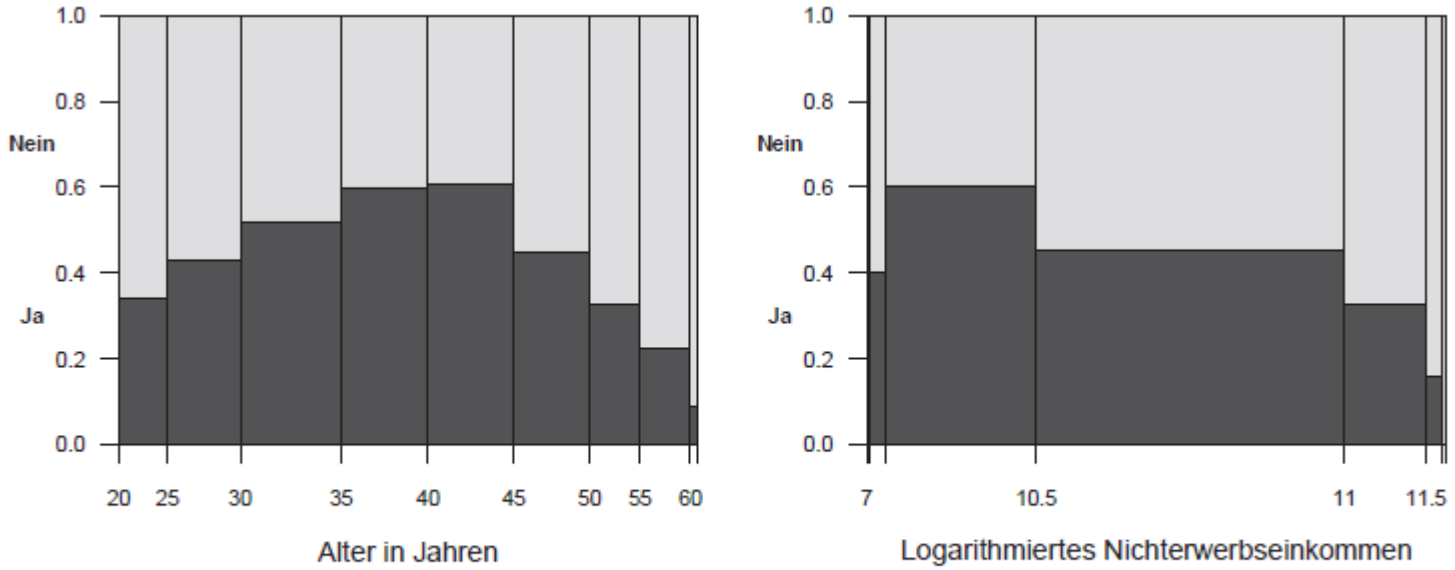
Abb. 5.3.1: Säulendiagramme und Boxplots: Größe nach Geschlecht

5.3 Ergänzende und vertiefende Themen

● Kategoriales vs. metrisches Merkmal ●

Tab. 5.3.1: Arbeitsmarktbeteiligung nach Alter

Alter	20–25	25–30	30–35	35–40	40–45	45–50	50–55	55–60	60–65
Ja	25	47	79	78	78	51	25	17	1
Nein	48	63	73	53	51	63	51	59	10

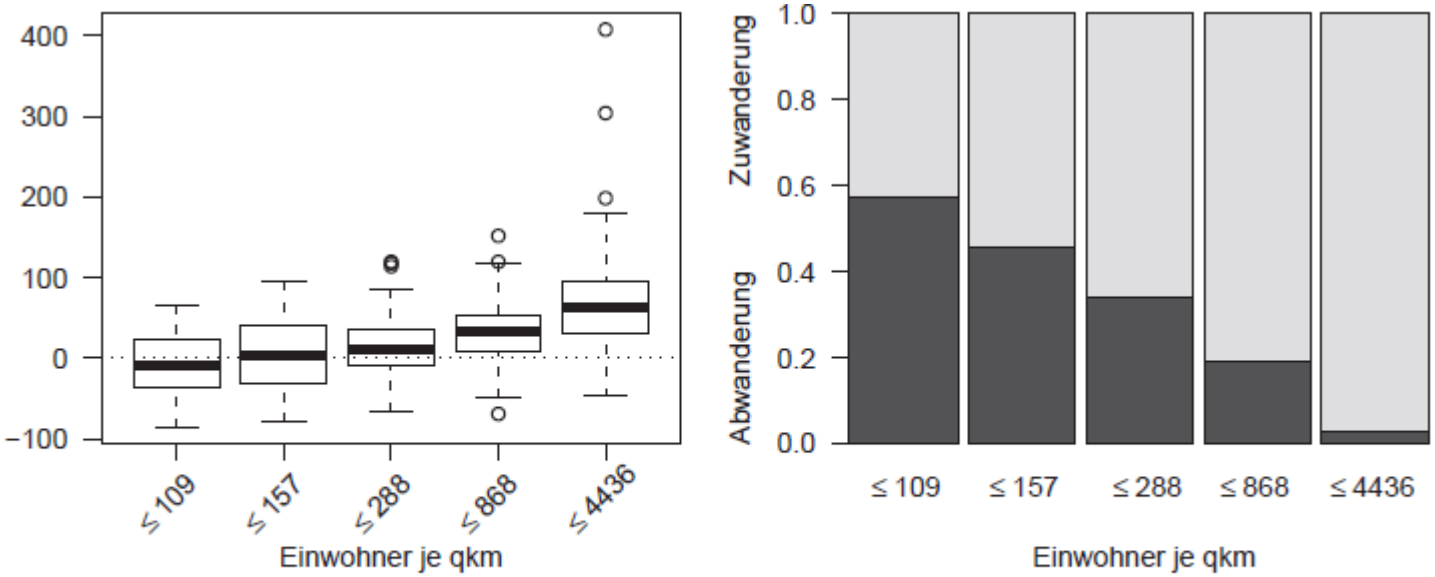


Daten: R-Paket „AER“ (Datensatz „SwissLabor“)

Abb. 5.3.2: Spinogramme: Arbeitsmarktbeteiligung nach Alter und Nichterwerbseinkommen

5.3 Ergänzende und vertiefende Themen

- **Kategorisierte und dichotomisierte Merkmale** -



Daten: Statistische Ämter des Bundes und der Länder [2014]

Abb. 5.3.3: Bevölkerungsdichte und Wanderungssaldo – alternative Darstellungen

Zusammenhangsmaße

- **Zusammenhangsmaße für ordinale Merkmale** -

Tab. 5.3.2: Glaube an Gott und Glaube an Sternzeichen

Sternzeichen / Glaube	-2	-1	0	1	2	Σ
−2	30	4	5	1	1	41
−1	20	4	1	6	0	31
0	16	7	3	5	1	32
1	27	11	5	16	1	60
2	16	8	6	9	3	42
Σ	109	34	20	37	6	206

Quelle: Eigenerhebung

> Hier gilt: $C_K = 0.3585,$ $r_{XY} = 0.2515,$ $r_S = 0.2369$

> Berechnung von r_{XY} hier eigentlich nicht ganz sinnvoll, da ordinal skaliertes Merkmal vorliegt

> Beispiel mit größerem Unterschied zwischen r_{XY} und r_S

Tab. 5.3.3: „Monotone nichtlineare“ Struktur innerhalb einer Kontingenztabelle

Sternzeichen / Glaube	-2	-1	0	1	2	Σ
−2	0	0	10	10	10	30
−1	0	10	0	0	0	10
0	10	0	0	0	0	10
1	10	0	0	0	0	10
2	10	0	0	0	0	10
Σ	30	10	10	10	10	70

> Hier gilt: $C_K = 0.9129,$ $r_{XY} = -0.8455,$ $r_S = -0.9231$

> Es existieren zahlreiche alternative Zusammenhangsmaße für ordinal skalierte Merkmale

Spezielle Regressionsmodelle

- **Hintergrund**
- **Anpassungs- und Prognosemaße**
- **Metrisches vs. kategoriales Merkmal: Varianzanalysemodelle**

> Modellierung der Streuungszerlegung im Rahmen eines (multiplen) linearen Regressionsmodells

> Zur Erinnerung nochmal ein altes Beispiel ...

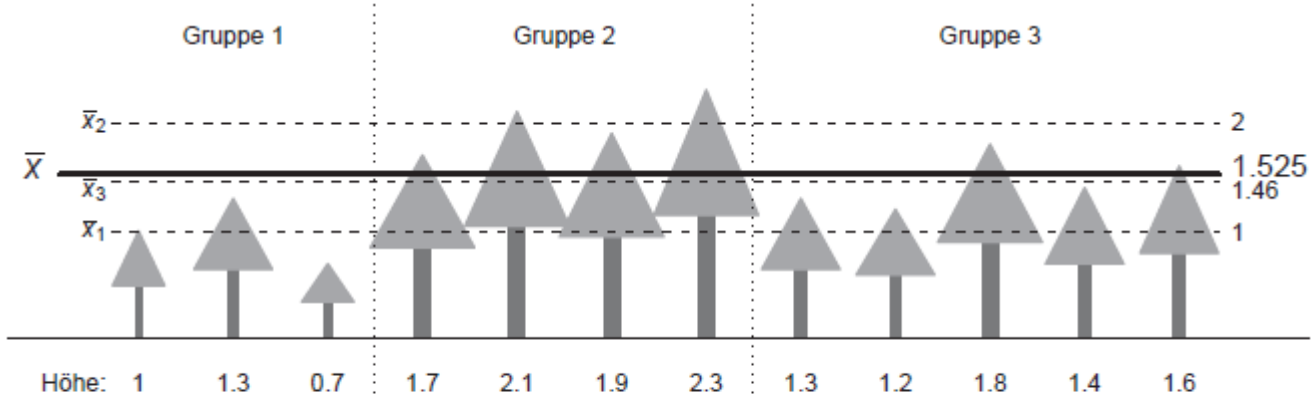
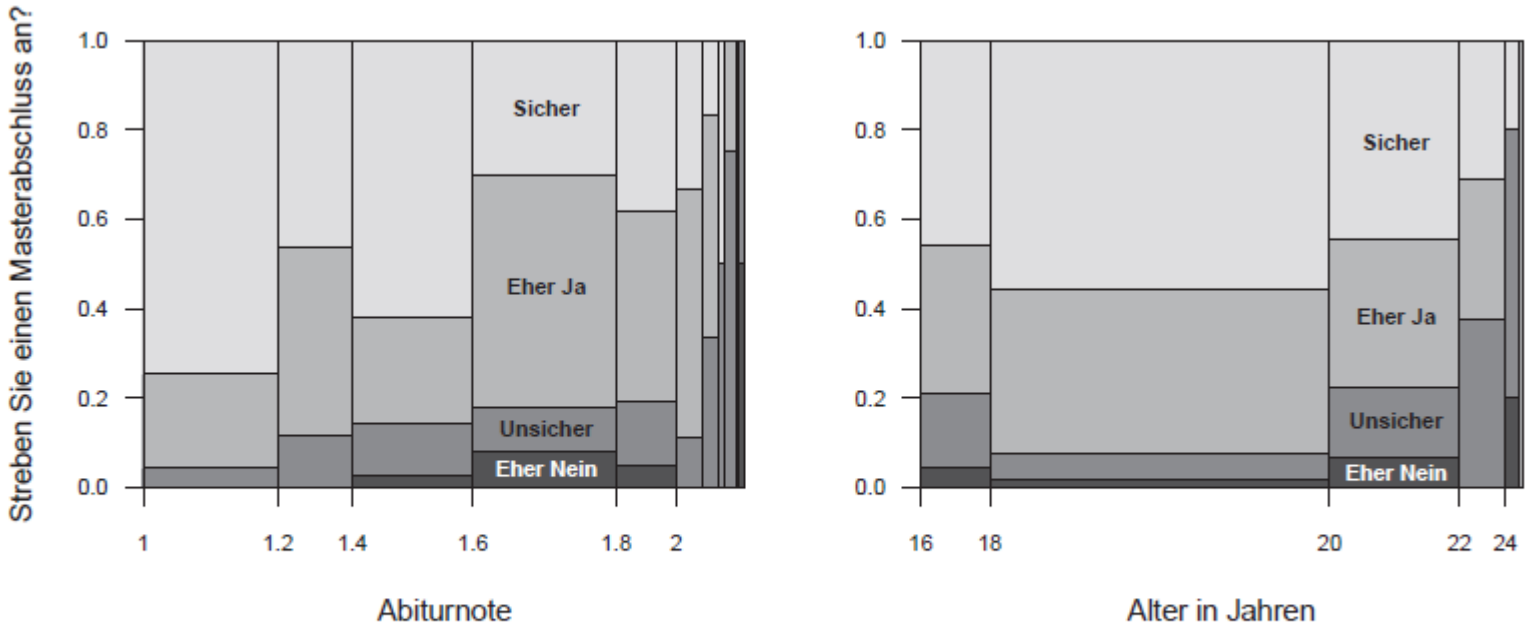


Abb. 4.5.1: Varianz bei gruppierten Daten – Streuungszerlegung

5.3 Ergänzende und vertiefende Themen

- **Kategoriales vs. metrisches Merkmal: Logistische Regressionsmodelle**



Daten: Eigene Erhebungen

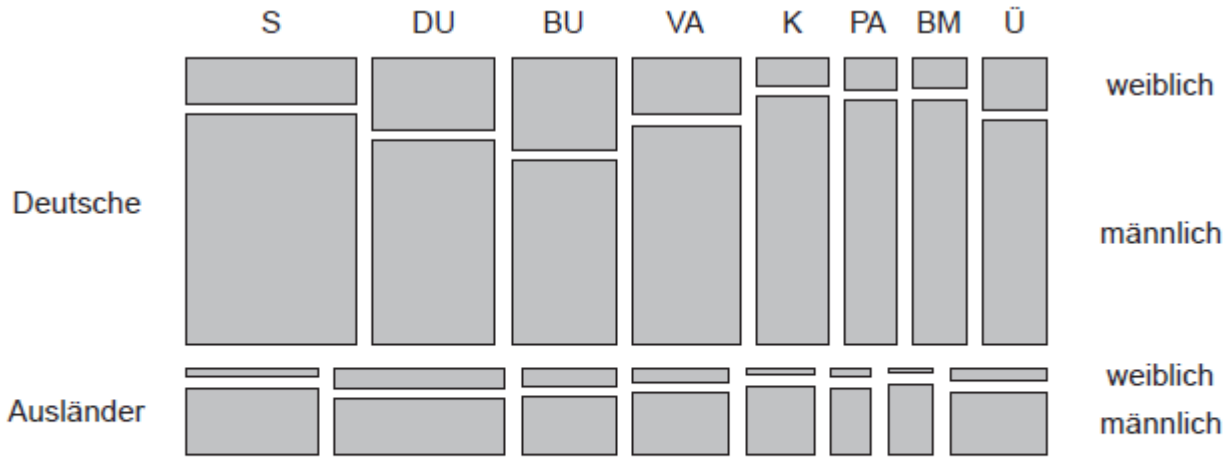
Abb. 5.3.4: Spinogramme: Masterambition, Abiturnote und Alter

- **Multiple Regression**
- **Generalisierte lineare Modelle**

5.3.2 Grafische Darstellung höherdimensionaler Daten

Höherdimensionale Mosaikplots

- Konzept
- Beispiel 5.3.1: 3-dimensionale Mosaikplots

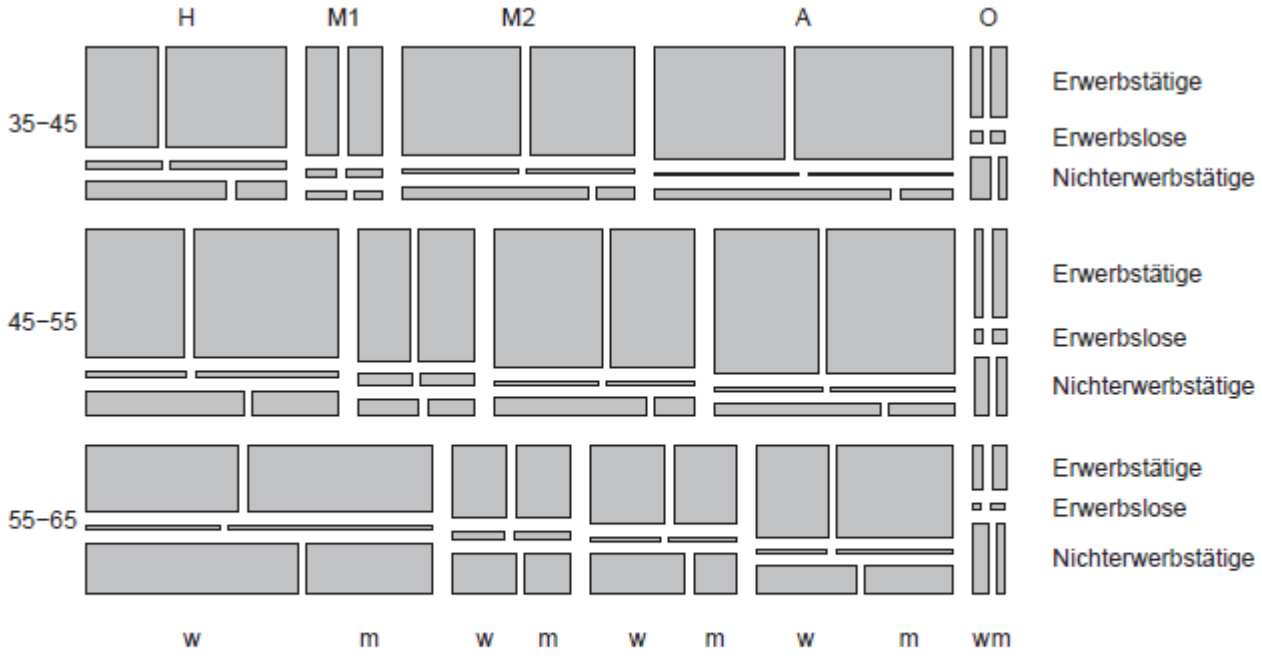


Daten: Statistisches Bundesamt [2014c]

Abb. 5.3.5: Mosaikplot: Staatsangehörigkeit, Art der Straftat und Geschlecht

5.3 Ergänzende und vertiefende Themen

- **Beispiel 5.3.2: 4-dimensionale Mosaikplots**



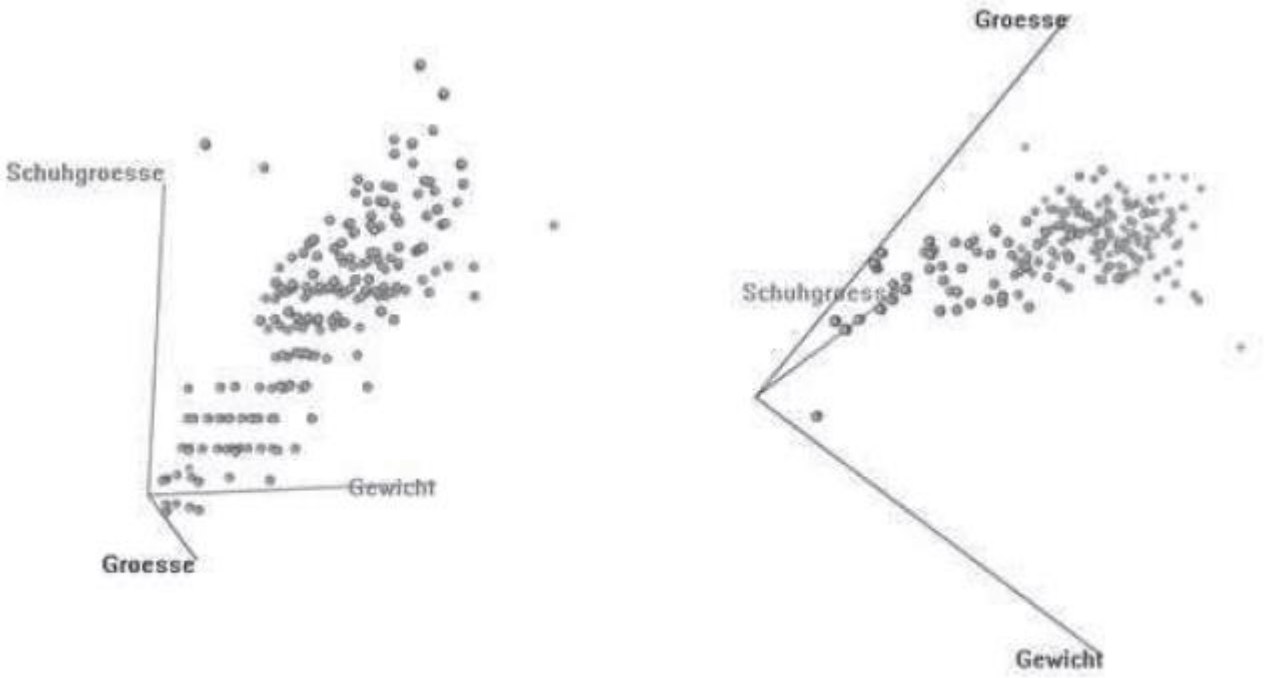
Daten: Statistisches Bundesamt [2013b, S. 79–80]

Abb. 5.3.6: Mosaikplot: Alter, Schulbildung, Erwerbsstatus und Geschlecht

- **Weitere Möglichkeiten**

3D-Punktwolken

- Konzept
- Beispiel 5.3.3

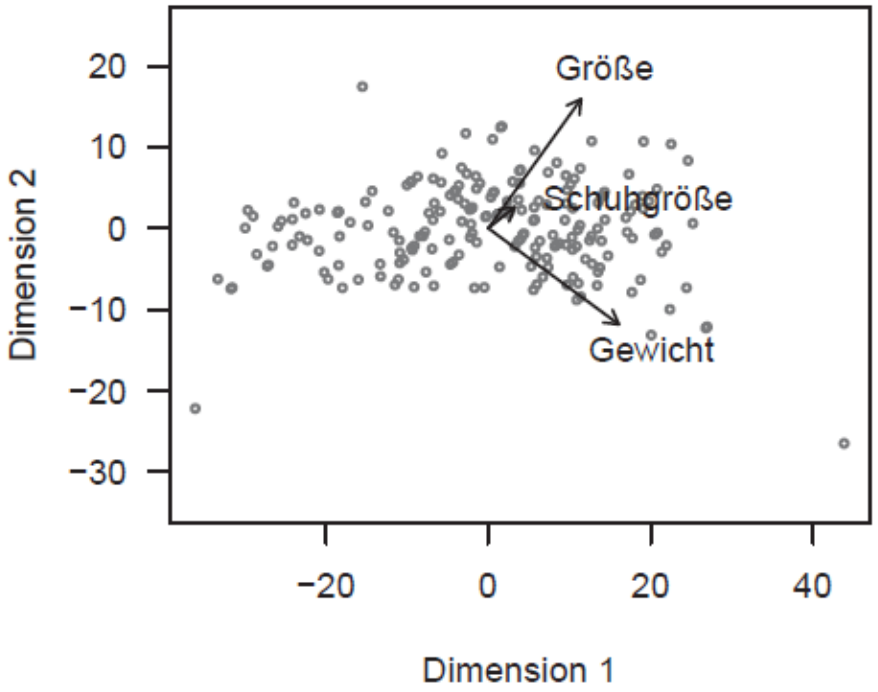


Daten: Eigene Erhebungen

Abb. 5.3.7: 3D-Punktwolken: Größe, Gewicht und Schuhgröße von Studierenden

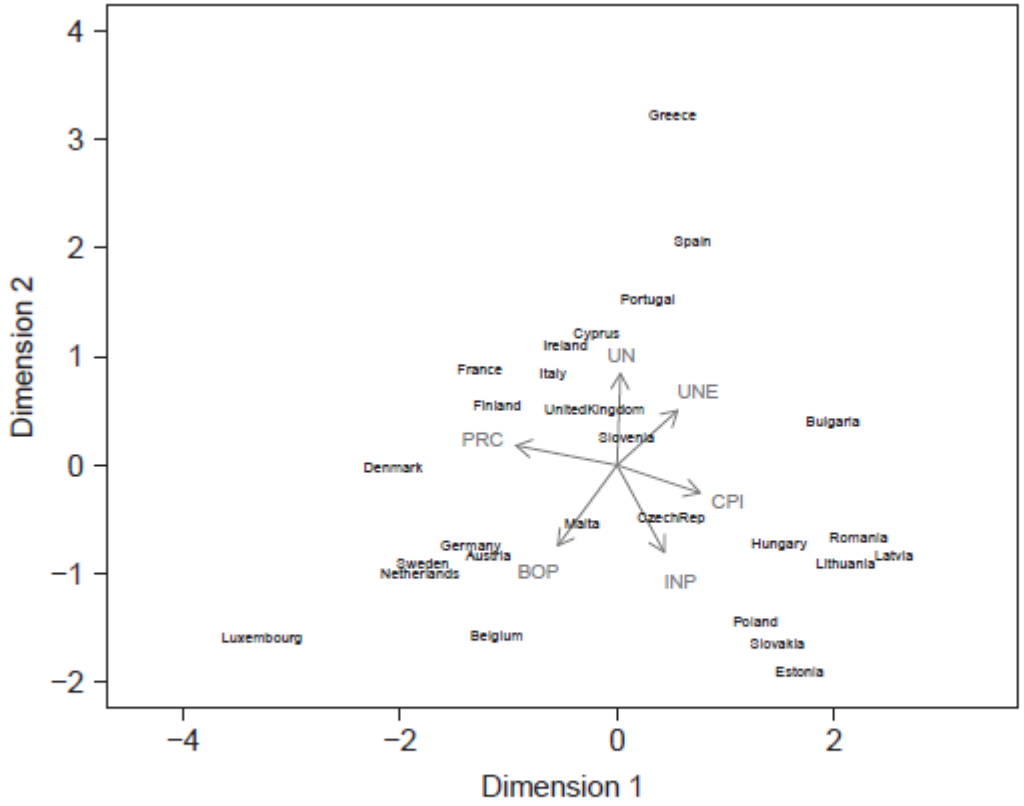
Biplots

- Konzept
- Hauptkomponentenanalyse-Biplots



Daten: Eigene Erhebungen

Abb. 5.3.8: PCA-Biplot: Größe, Gewicht und Schuhgröße von Studierenden

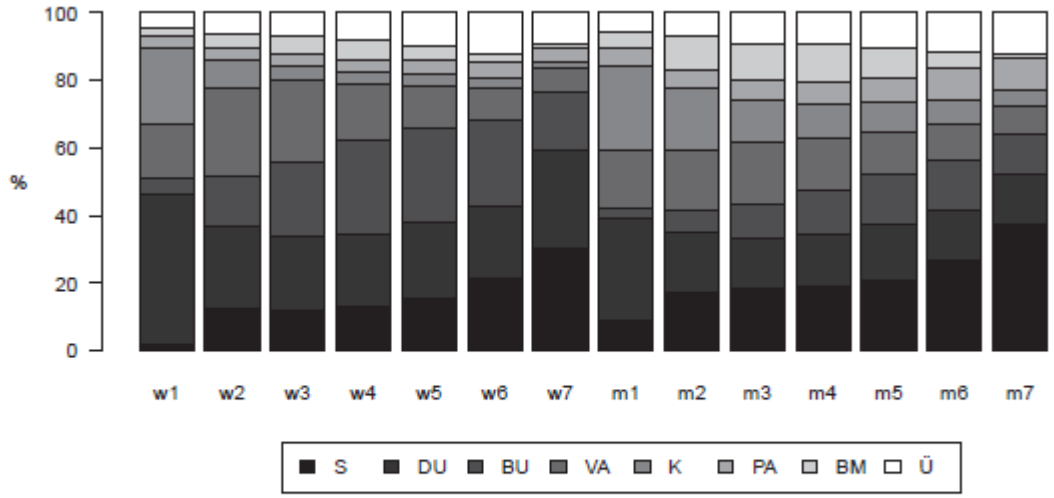


Daten: Greenacre [2012]

Abb. 5.3.9: PCA-Biplot: Ökonomische Indikatoren der EU-Staaten (2011) (nachgestellt)

5.3 Ergänzende und vertiefende Themen

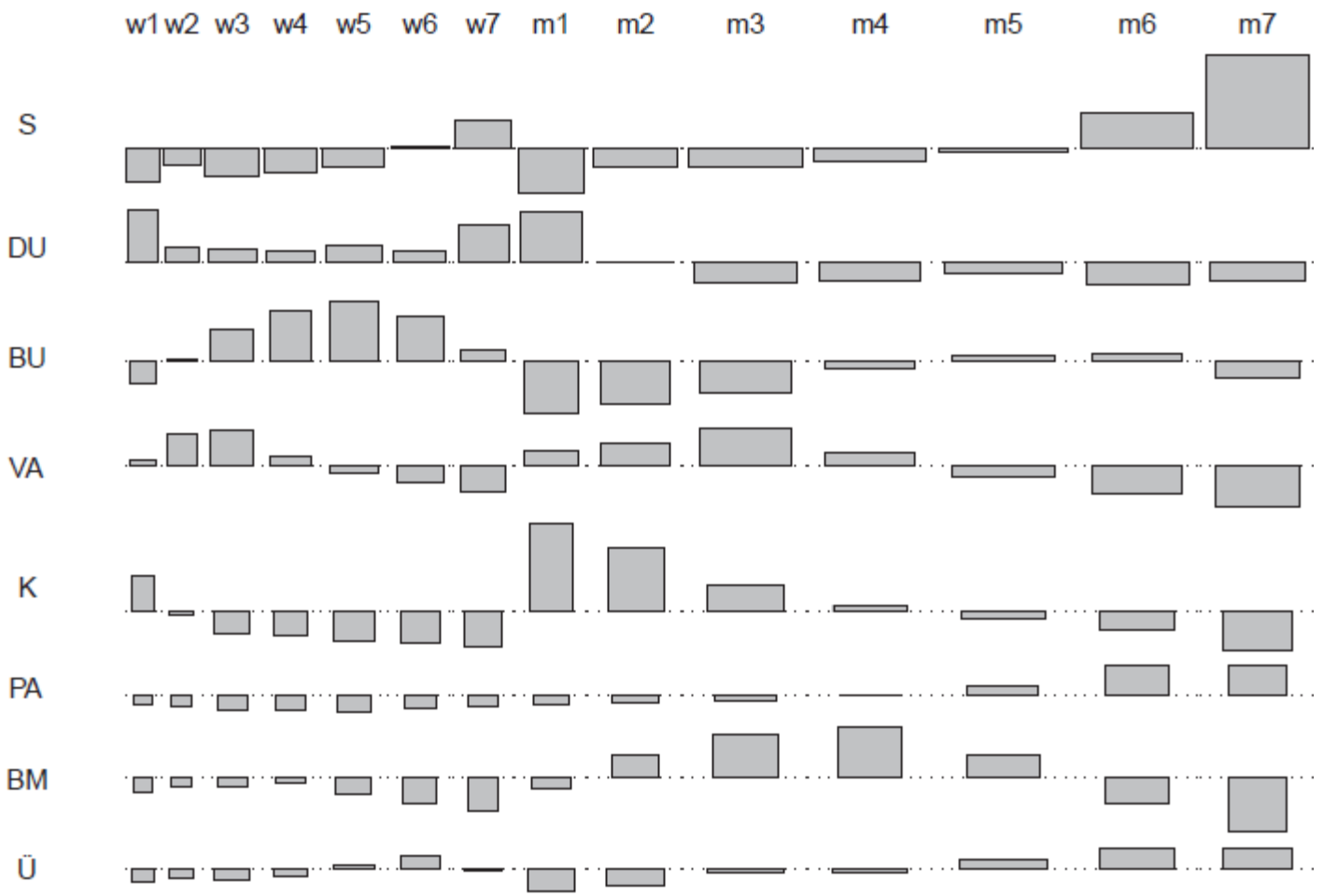
- **Korrespondenzanalyse-Biplots** -



Daten: Statistisches Bundesamt [2014c]

Abb. 5.3.10: Straftaten nach geschlechtsspezifischen Altersgruppen

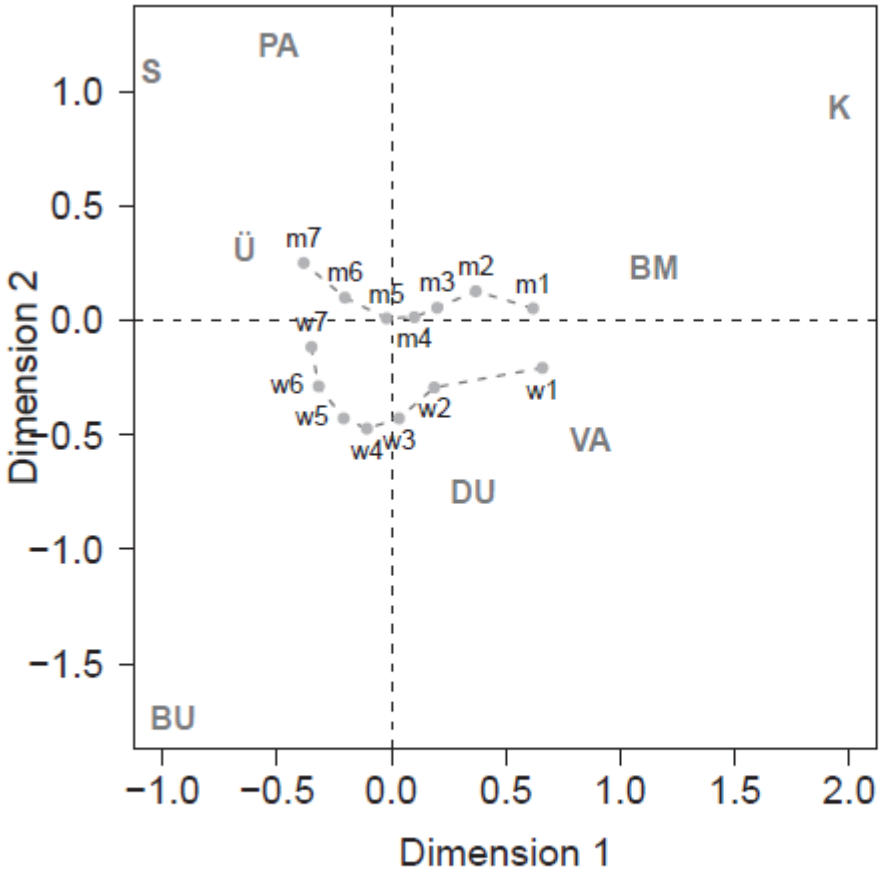
5.3 Ergänzende und vertiefende Themen



Daten: Statistisches Bundesamt [2014c]

Abb. 5.3.11: Assoziationsplot: Arten von Straftaten und Altersgruppen nach Geschlecht

5.3 Ergänzende und vertiefende Themen



Daten: Statistisches Bundesamt [2014c]

Abb. 5.3.12: CA-Biplot: Arten von Straftaten und geschlechtsspezifische Altersgruppen

5.3.3 Wichtige Aspekte bei der Analyse empirischer Zusammenhänge

Empirischer Zusammenhänge und Kausalität

- **Zum Kausalitätsbegriff**
- **Kausalität in der Statistik**

> „*Force as a cause of motion is exactly on the same footing as a tree-god as cause of growth*“ (Karl Pearson [1900b])

> Ein „*kognitives Konstrukt* “ (Kälble [1997])

> Zu *Causal Inference* und *Structural Causal Models*
vgl. etwa Cox und Wermuth [2004] oder Pearl [2009]

- **Korrelation impliziert keine Kausalität**

Empirische Zusammenhangsmaße können stets verzerrt sein durch

(i) systematische Verzerrungen

(ii) zufällige Schwankungen

5.3 Ergänzende und vertiefende Themen

- **Nachweis und Messung kausaler Effekte**
- **Experimentalstudien vs. Erhebungsstudien**
- **Einflussschema bei Kausalanalysen**

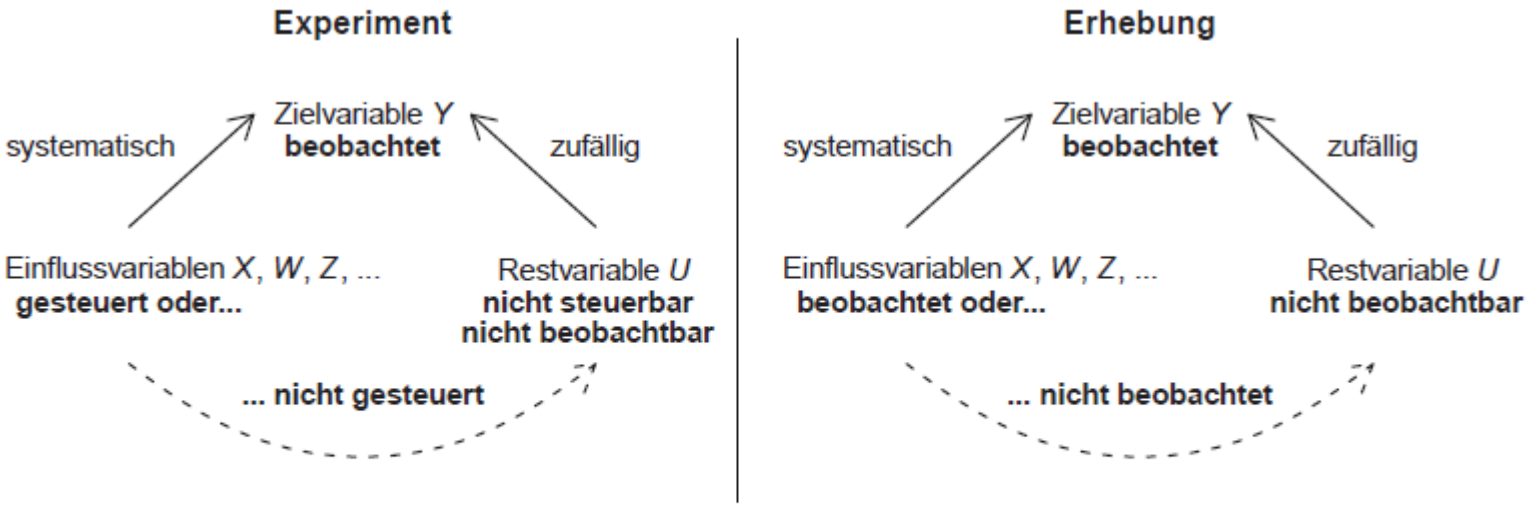


Abb. 5.3.13: Einflussschema bei Experimental- und Erhebungsstudien

Spezielle Formen von Kausalität

● Direkte und indirekte kausale Effekte ●

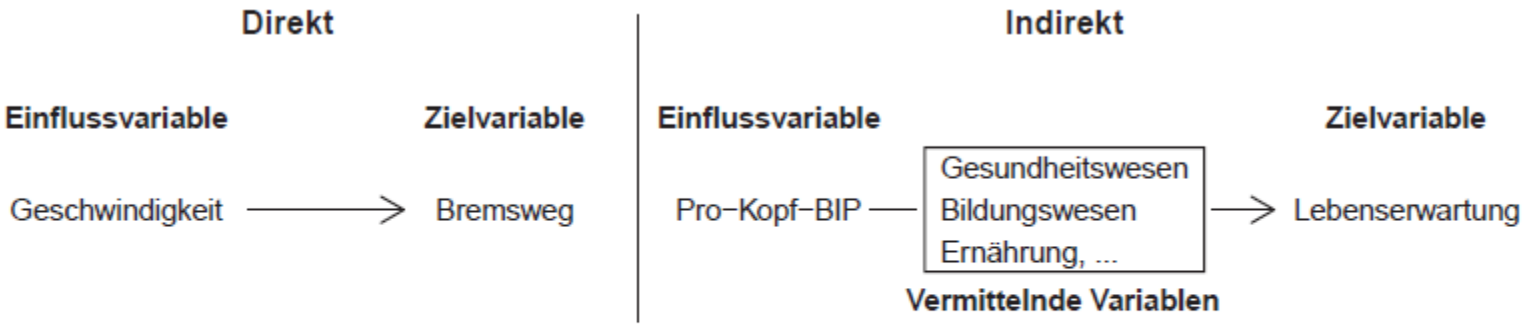


Abb. 5.3.14: Direkte und indirekte kausale Effekte

● Dynamische und simultane Abhängigkeit ●

> **Dynamische kausale Effekte:** Anpassungsvorgänge vollziehen sich zeitlich verzögert und schrittweise

> **Simultane Kausalität:** *X* beeinflusst *Y* und umgekehrt

Systematische Verzerrungen

- Hintergrund
- Beispiele systematischer Verzerrungen

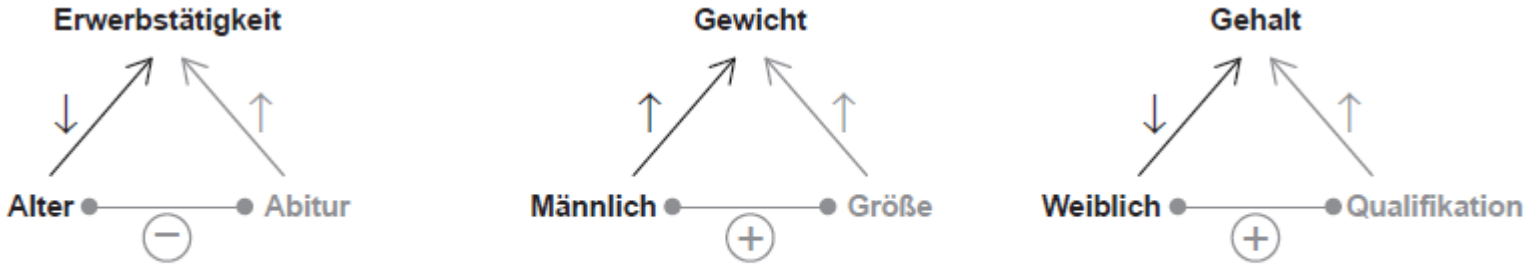


Abb. 5.3.15: Schematische Beispiele systematischer Verzerrungen

Kodierung binärer Variablen

Abitur: 0 = kein Abitur, 1 = Abitur vorhanden

Männlich: 0 = weiblich, 1 = männlich

Weiblich: 0 = männlich, 1 = weiblich

● Merkregeln für den 3-Variablen-Fall ●



Abb. 5.3.16: Richtung systematischer Verzerrungen – Merkregeln kompakt

> **Direkter Effekt** von *X* auf *Y* ist positiv oder negativ

> **Direkter Effekt** von *Z* auf *Y* ist positiv oder negativ

> **Indirekter Effekt** von *X* über *Z* auf *Y*

negativ ⇔ *Corr*(*X*, *Z*) und direkter *Z*-Effekt gegensätzlich

positiv ⇔ *Corr*(*X*, *Z*) und direkter *Z*-Effekt gleichgerichtet

0 ⇔ *Corr*(*X*, *Z*) = 0 und/oder direkter *Z*-Effekt = 0

> **Gesamteffekt** (**totaler Effekt**) von X auf Y spiegelt direkten X-Effekt systematisch **verzerrt** wider, sofern indirekter X-Effekt $\neq 0$

> Direkter X-Effekt wird

kompensiert $\Leftrightarrow$ direkter und indirekter X-Effekt gegensätzlich

verstärkt $\Leftrightarrow$ direkter und indirekter X-Effekt gleichgerichtet

● Weitere Bemerkungen ●

> Wird direkter Effekt von X auf Y verzerrt, so auch der direkte Effekt von Z auf Y

> Indirekter Effekt muss nicht kausaler Natur sein

> $Corr(X, Z)$ muss nicht kausal begründet sein

> Besonders schwerwiegende Verzerrung bei

Überkompensation $\Leftrightarrow$ Indirekter Effekt stärker als direkter Effekt

5.3 Ergänzende und vertiefende Themen

- **Scheinabhängigkeiten**

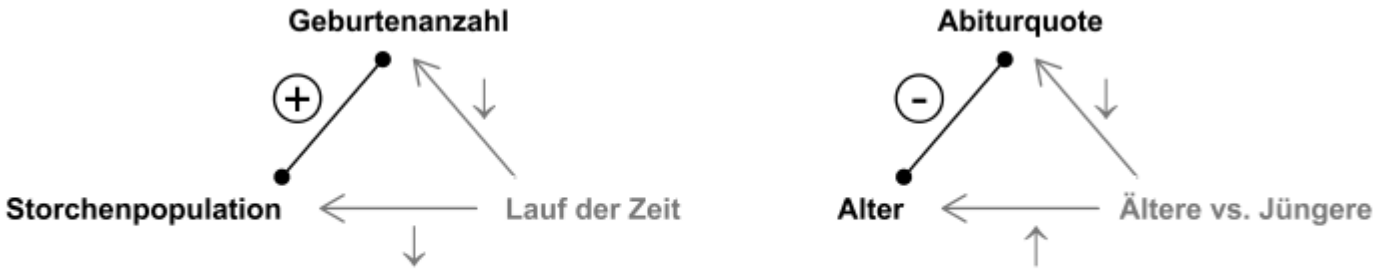


Abb. 5.3.17: Schematische Beispiele von Scheinabhängigkeiten

- **Methodischer Umgang mit systematischen Verzerrungen**
- **Versuchsplanung**
- **Randomisierung**
- **Untersuchung in homogeneren Untergruppen**
- **Multiple lineare Regression**
- **Multiple Regression in anderen Modellen**

● Beispiel 5.3.4 ●

> **Fall 1**: Erwerbstätigkeit (Zielvariable), Alter und Schulabschluss

Tab. 5.3.4: Erwerbstätigkeit, Alter und Schulabschluss erwerbsfähiger Personen

Erwerbstätigenquote nach Alter

35–45	84.9%
45–55	84.1%
55–65	61.3%

Erwerbstätigenquote nach Schulabschluss

ohne Abitur	73.7%
mit Abitur	86.3%

Anteil von Abiturienten nach Alter

35–45	35.4%
45–55	28.6%
55–65	23.5%

Erwerbstätigenquote nach Alter und Schulabschluss

	35–45	45–55	55–65
ohne Abitur	82.4%	81.6%	57.0%
mit Abitur	89.4%	90.4%	75.1%

Mögliches Konstrukt:
Verstärkung des direkten Effekts von Alter
durch indirekten Effekt von Alter über Abitur

5.3 Ergänzende und vertiefende Themen

> **Fall 2**: Gewicht (Zielvariable), Geschlecht und Größe

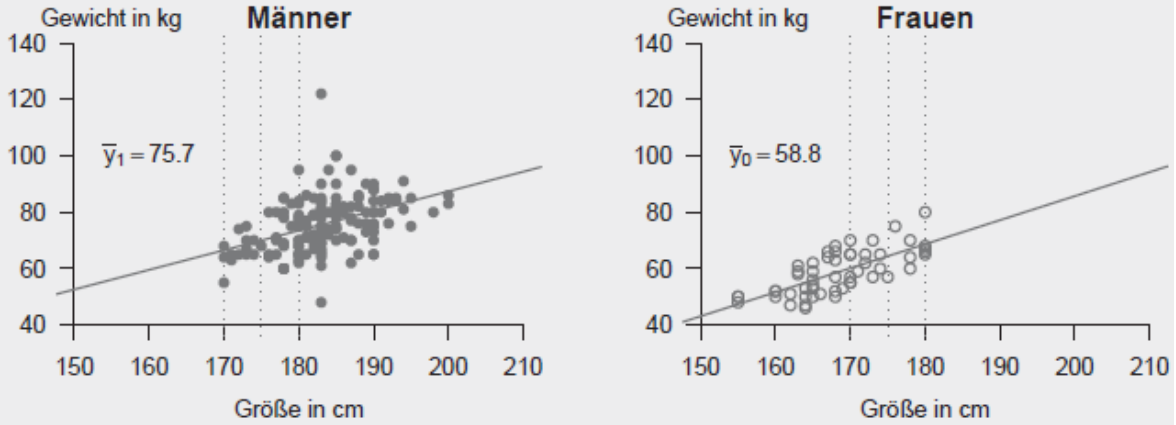


Abb. 5.3.18: Zusammenhang von Größe und Gewicht nach Geschlecht

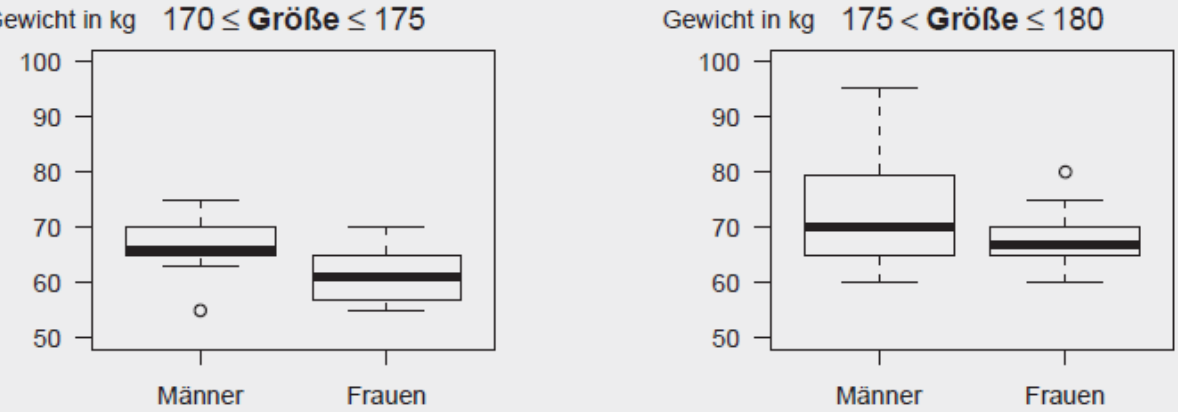


Abb. 5.3.19: Zusammenhang von Geschlecht und Gewicht nach Größe

Mögliches Konstrukt: Verstärkung des direkten Effekts von „männlich“ durch indirekten Effekt von „männlich“ über Größe

Zufällige Schwankungen

- **Hintergrund**
- **Beispiele rein zufälliger Abhängigkeiten**

> Beispiel 1: Man nehme einen weißen und einen schwarzen Würfel und werfe beide jeweils $n = 10$ Mal:

weiß: 1, 3, 5, 5, 4, 6, 4, 6, 3, 6
schwarz: 5, 4, 6, 1, 2, 3, 4, 1, 4, 5

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

$\Longrightarrow \quad r_{XY} = -0.34$

Jetzt 1000-fache Wiederholung dieses „Experiments“...

5.3 Ergänzende und vertiefende Themen

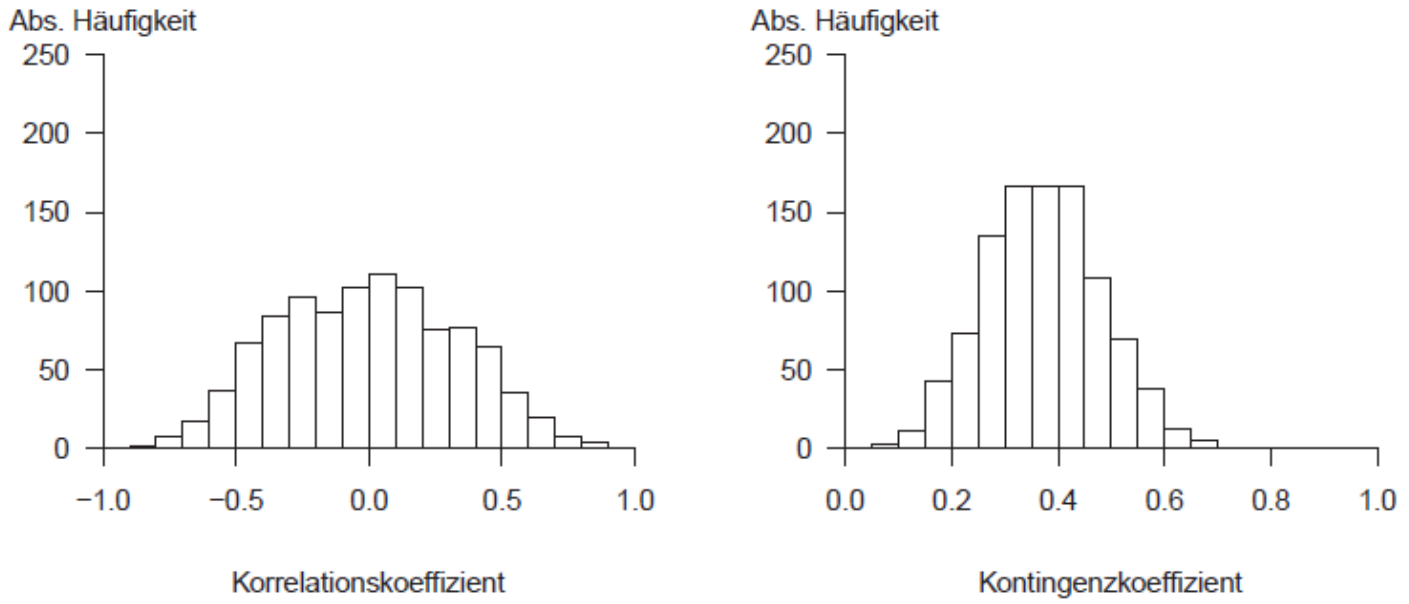


Abb. 5.3.20: Würfelexperimente: Empirische Zusammenhänge bei reiner Zufälligkeit

> Beispiel 2: Man nehme einen weißen und einen schwarzen Würfel und werfe beide jeweils 60 Mal

Tab. 5.3.5: Würfelexperiment: Weißer und schwarzer Würfel

Farbe \ Zahl	1	2	3	4	5	6
Weiß	9	6	14	10	8	13
Schwarz	9	16	10	6	7	12

$\Longrightarrow C_K = 0.32$

Jetzt 1000-fache Wiederholung dieses „Experiments“...

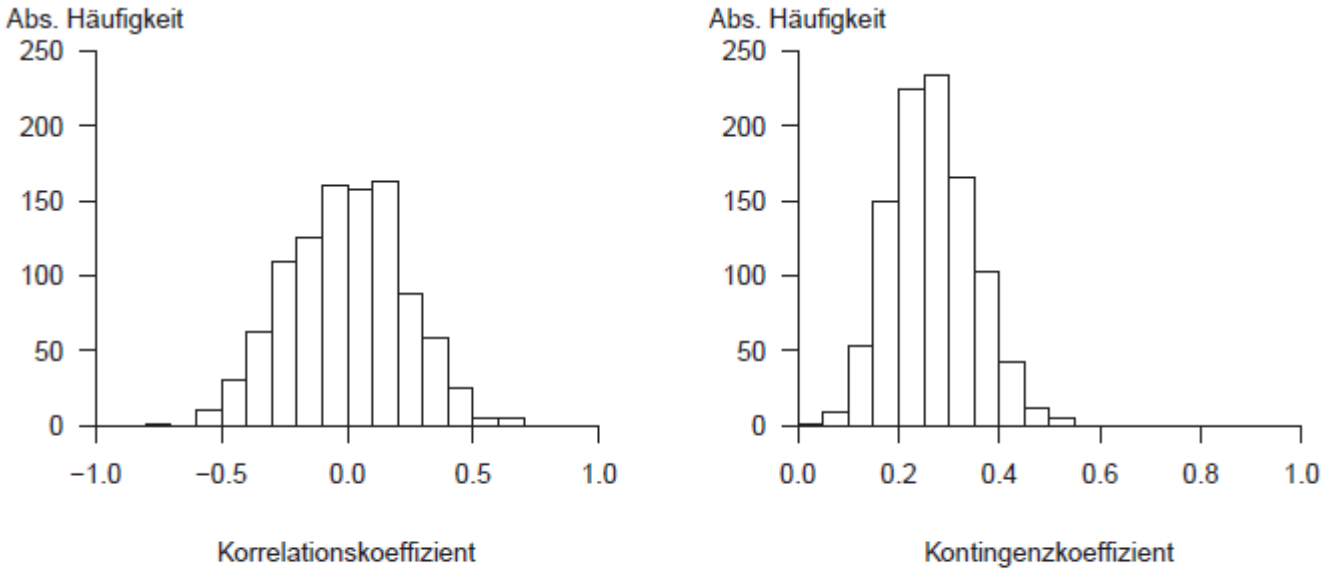


Abb. 5.3.21: Würfelexperimente bei höherer Anzahl von Würfen

5.3 Ergänzende und vertiefende Themen

- System oder Zufall?
- Methodischer Umgang mit zufälligen Schwankungen
- Höhere Aussagekraft durch mehr Beobachtungen
- Induktive Methoden

Teil 2:
Wahrscheinlichkeitsrechnung - Theoretischer Kalkül

6 Einführung in die Wahrscheinlichkeitsrechnung

6 Einführung in die Wahrscheinlichkeitsrechnung

6.1 Wichtige Grundbegriffe und Regeln	273
6.1.1 Interpretation von Zufall und Wahrscheinlichkeiten	273
6.1.2 Elementare Mengenlehre	276
6.1.3 Axiomatischer Wahrscheinlichkeitsbegriff und Regeln	286
6.2 Rechnen mit abhängigen und unabhängigen Ereignissen	299
6.2.1 Bedingte Wahrscheinlichkeiten und Rechenregeln	299
6.2.2 Stochastische Unabhängigkeit von Ereignissen	314
6.2.3 Kalkül nach der Formel von Bayes	324

6.1 Wichtige Grundbegriffe und Regeln

6.1.1 Interpretation von Zufall und Wahrscheinlichkeiten

● Zufallsvorgang und Zufall ●

Ein **Zufallsvorgang** ist ein Vorgang mit mindestens zwei möglichen verschiedenen Ergebnissen, bei dem im Voraus nicht eindeutig bestimmbar ist, welches Ergebnis eintreten wird.

● Zufall ist perspektivisch bedingt ●

Ein Zufallsvorgang ergibt sich aus der Perspektive eines (partiell) Unwissenden

● Zufall folgt Gesetzmäßigkeiten ●

> keinesfalls mit Willkür oder völliger Unberechenbarkeit gleichzusetzen

> Zufallsvorgang ist nicht (genau) vorhersagbar, besitzt aber (vorhersagbare) Gesetzmäßigkeiten

● Beispiele für die Verwendung von Wahrscheinlichkeiten ●

(1) *„Die Wahrscheinlichkeit eine Sechs zu würfeln beträgt 1/6“*

(2) *„Die Wahrscheinlichkeit für 6 Richtige im Lotto beträgt 0.00000715%“*

(3) *„Die Kreditausfallwahrscheinlichkeit für diese Kunden liegt bei 1%“*

(4) *„Die Wahrscheinlichkeit für eine Mädchengeburt beträgt laut Statistik 48.8%“*

(5) *„Mit einer Wahrscheinlichkeit von 95% liegt der Stimmenanteil dieser Partei nach Berechnungen zwischen 29.8% und 31.4%“*

(6) *„Nach Berechnungen wir die Bevölkerung im Jahr 2050 mit einer Wahrscheinlichkeit von 90% bei unter 50 Millionen liegen“*

(7) *„Dieser Patient wird mit einer Wahrscheinlichkeit von 70% sterben“*

(8) *„Ich bin mir mit 95%iger Wahrscheinlichkeit sicher, dass dieses Jahr ein Schaltjahr ist“*

6.1 Wichtige Grundbegriffe und Regeln

- **Klassische Wahrscheinlichkeit** (1), (2)
- **Statistische Wahrscheinlichkeit - empirisch** (3), (4), (7)?
- **Statistische Wahrscheinlichkeit - theoretisch** (5), (6), (7)?
- **Frequentistischer Deutungsansatz**
- **Objektive vs. subjektive Wahrscheinlichkeiten** (1) bis (6) vs. (7)? u. (8)
- **Klassische vs. statistische Wahrscheinlichkeiten**
- **Zielsetzung der Wahrscheinlichkeitsrechnung**

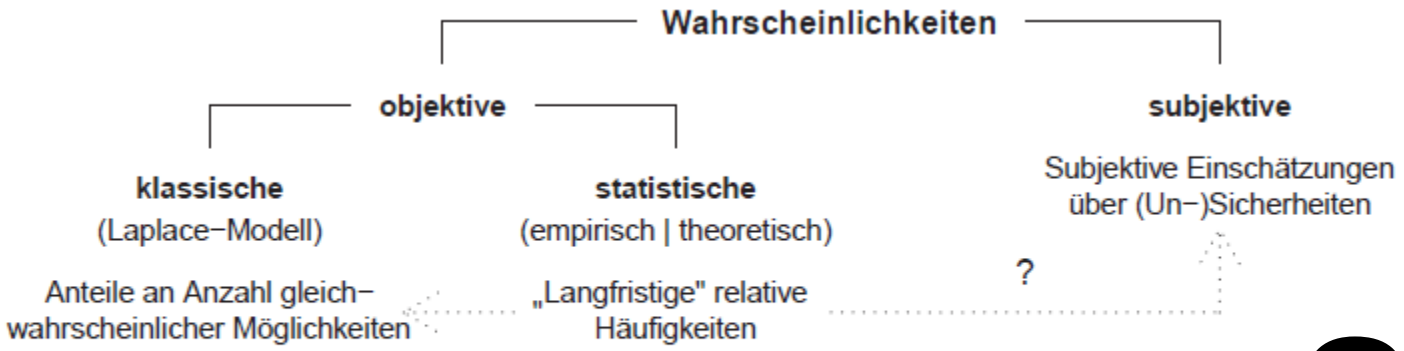


Abb. 6.1.1: Verwendung und Interpretation von Wahrscheinlichkeiten

6.1.2 Elementare Mengenlehre

● Mengen und Elemente ●

> Beispiel 1: Menge Z, welche die Zahlen 2, 4 und 6 enthält:

$$Z = \{2, 4, 6\}$$

> Beispiel 2: Menge W mit 4 verschiedenen Wetterzuständen:

$$W = \{sonnig, teils\ bedeckt, bedeckt, regnerisch\}$$

> Notation für Elemente, welche bestimmten Mengen angehören oder nicht:

$$1 \notin Z, \quad 2 \in Z, \quad 2 \notin W, \quad sonnig \in W$$

> Beschreibungen von Mengen anstelle von Einzelauflistung:

Statt $M_{10} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ z. B.

$$M_{10} = \{x: x \in \mathbb{N}\ mit\ 1 \leq x \leq 10\}$$

6.1 Wichtige Grundbegriffe und Regeln

● Wichtige Standardmengen ●

Wichtige Standardmengen	
$\mathbb{N} = \{1, 2, 3, \ldots\}$	Menge der natürlichen Zahlen.
$\mathbb{N}_0 = \{0, 1, 2, 3, \ldots\}$	Menge der natürlichen Zahlen inklusive 0.
$\mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, \ldots\}$	Menge der ganzen Zahlen.
$\mathbb{Q}$	Menge der rationalen Zahlen (siehe später).
$\mathbb{R}$	Menge der reellen Zahlen (siehe später).
$\emptyset$	Leere Menge.

● Teilmenge ●

> Gegeben seien die Mengen

$$A = \{2, 4, 6\}, \quad B = \{1, 2, 3, 4\} \text{ und } C = \{1, 2, 3, 4, 5, 6\}$$

Dann gilt z. B.: $A \subset C$, $A \not\subset B$, $B \subset C$

> „Definitorisch“ gilt:

Die leere Menge ist Teilmenge jeder Menge.

● Intervalle ●

> (a, b) **offen**: von ausschließlich a bis ausschließlich b

> $[a, b]$ **abgeschlossen**: von einschließlich a bis einschließlich b

> $[a, b)$ **halboffen rechts**: von einschließlich a bis ausschließlich b

> $(a, b]$ **halboffen links**: von ausschließlich a bis einschließlich b

● Schnittmenge ●

Beispiele:

$$A \cap B = \{2, 4\} \qquad A \cap C = \{2, 4, 6\}$$

$$B \cap C = \{1, 2, 3, 4\}$$

$$\bigcap_{i=1}^{4} M_i = M_1 \cap M_2 \cap M_3 \cap M_4$$

6.1 Wichtige Grundbegriffe und Regeln

• Vereinigungsmenge •

Beispiele: $A \cup B = \{1, 2, 3, 4, 6\}$ 2 und 4 jeweils nur einmal!

$$\bigcup_{i=1}^{4} M_i = M_1 \cup M_2 \cup M_3 \cup M_4$$

• Differenzmenge •

Beispiele: $A \setminus B = \{6\}$ $B \setminus A = \{1, 3\}$ $A \setminus C = \emptyset$

• Komplementärmenge •

Beispiele mit Teilmengen der **Grundmenge** C:

$\bar{A} = C \setminus A = \{1, 3, 5\}$ $\bar{B} = C \setminus B = \{5, 6\}$ $\bar{C} = C \setminus C = \emptyset$

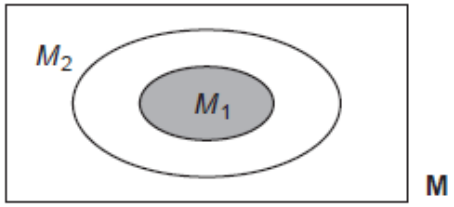
• Disjunkte Mengen •

M_1 und M_2 **disjunkt** $\Leftrightarrow$ $M_1 \cap M_2 = \emptyset$

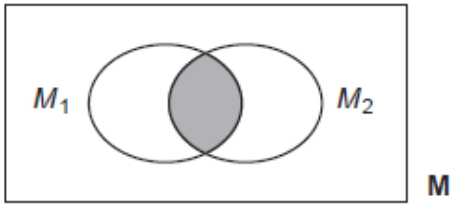
6.1 Wichtige Grundbegriffe und Regeln

• Venn-Diagramme •

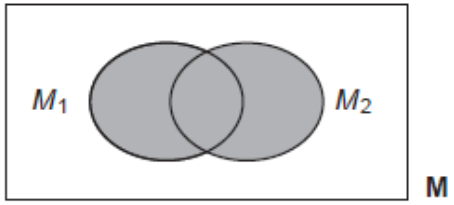
$M_1 \subset M_2$: Teilmenge



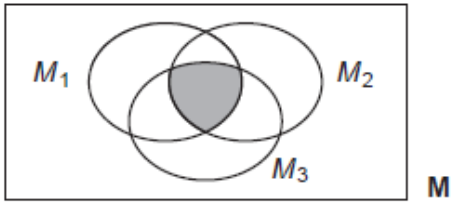
$M_1 \cap M_2$: Schnittmenge zweier Mengen



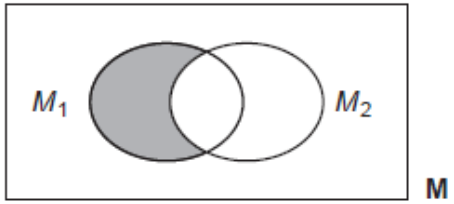
$M_1 \cup M_2$: Vereinigungsmenge



$M_1 \cap M_2 \cap M_3$: Schnittmenge dreier Mengen



$M_1 \setminus M_2$: Differenzmenge



$\overline{M} = \mathbf{M} \setminus M$: Komplementärmenge

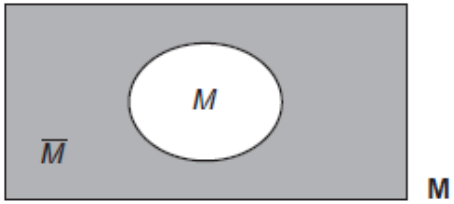


Abb. 6.1.2: Venn-Diagramme zur Illustration von Mengen und Mengenoperationen

• Elementare Regeln für Mengenoperationen •

Elementare Regeln für Mengenoperationen

Kommutativgesetze:

$$M_1 \cap M_2 = M_2 \cap M_1, \quad M_1 \cup M_2 = M_2 \cup M_1.$$

Assoziativgesetze:

$$(M_1 \cap M_2) \cap M_3 = M_1 \cap (M_2 \cap M_3),$$
$$(M_1 \cup M_2) \cup M_3 = M_1 \cup (M_2 \cup M_3).$$

Distributivgesetze:

$$(M_1 \cap M_2) \cup M_3 = (M_1 \cup M_3) \cap (M_2 \cup M_3),$$
$$(M_1 \cup M_2) \cap M_3 = (M_1 \cap M_3) \cup (M_2 \cap M_3).$$

● Potenzmenge ●

> **Potenzmenge** einer Menge ist die Menge aller möglichen Teilmenge dieser Menge

> Beispiel 1: $Z = \{2, 4, 6\}$

$$\mathfrak{P}(Z) = \{\{2\}, \{4\}, \{6\}, \{2,4\}, \{2,6\}, \{4,6\}, \{2,4,6\}, \emptyset\}$$

> Beispiel 2:

$$W = \{\boldsymbol{s}onnig, \boldsymbol{t}eils\ \boldsymbol{b}edeckt, \boldsymbol{b}edeckt, \boldsymbol{r}egnerisch\}$$

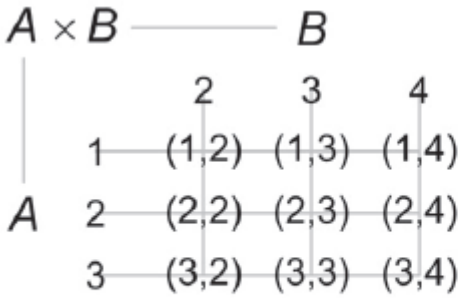
$$\mathfrak{P}(W) = \{\{\boldsymbol{s}\}, \{\boldsymbol{tb}\}, \{\boldsymbol{b}\}, \{\boldsymbol{r}\},$$
$$\{\boldsymbol{s}, \boldsymbol{tb}\}, \{\boldsymbol{s}, \boldsymbol{b}\}, \{\boldsymbol{s}, \boldsymbol{r}\}, \{\boldsymbol{tb}, \boldsymbol{b}\}, \{\boldsymbol{tb}, \boldsymbol{r}\}, \{\boldsymbol{b}, \boldsymbol{r}\},$$
$$\{\boldsymbol{s}, \boldsymbol{tb}, \boldsymbol{b}\}, \{\boldsymbol{tb}, \boldsymbol{b}, \boldsymbol{r}\}, \{\boldsymbol{s}, \boldsymbol{b}, \boldsymbol{r}\}, \{\boldsymbol{s}, \boldsymbol{tb}, \boldsymbol{r}\},$$
$$\{\boldsymbol{s}, \boldsymbol{tb}, \boldsymbol{b}, \boldsymbol{r}\}, \emptyset\}$$

● Produktmenge ●

Beispiel mit $M_1 = \{1, 2, 3\}$ und $M_2 = \{2, 3, 4\}$

$M_1 \times M_2 = \{(1,2), (1,3), (1,4), (2,2), (2,3), (2,4), (3,2), (3,3), (3,4)\}$

$M_2 \times M_1 = \{(2,1), (2,2), (2,3), (3,1), (3,2), (3,3), (4,1), (4,2), (4,3)\}$



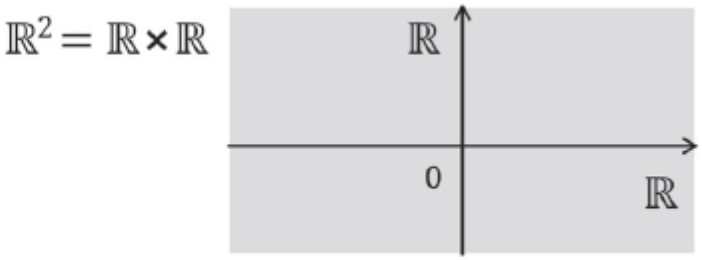


Abb. 6.1.3: Produktmengen

● Mächtigkeit einer Menge ●

> Beispiele: $Z = \{2, 4, 6\} \Rightarrow |Z| = 3$

$W = \{sonnig, teils\ bedeckt, bedeckt, regnerisch\}$
$\Rightarrow |W| = 4$

> Falls $|M| < \infty$ (endliche Menge), so gilt: $|\mathfrak{P}(M)| = 2^{|M|}$

Beispiele: $|\mathfrak{P}(Z)| = 2^{|Z|} = 2^3 = 8$ und $|\mathfrak{P}(W)| = 2^{|W|} = 2^4 = 16$

● Vergleich der Mächtigkeit von Mengen ●

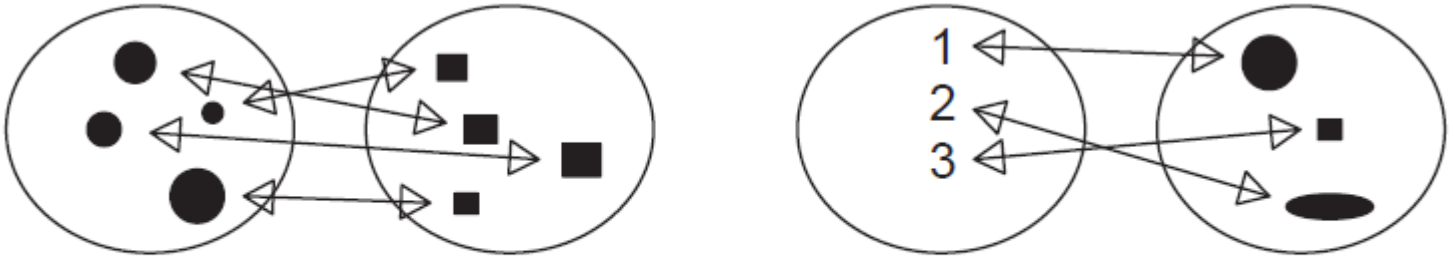


Abb. 6.1.4: Bijektion zwischen endlichen Mengen

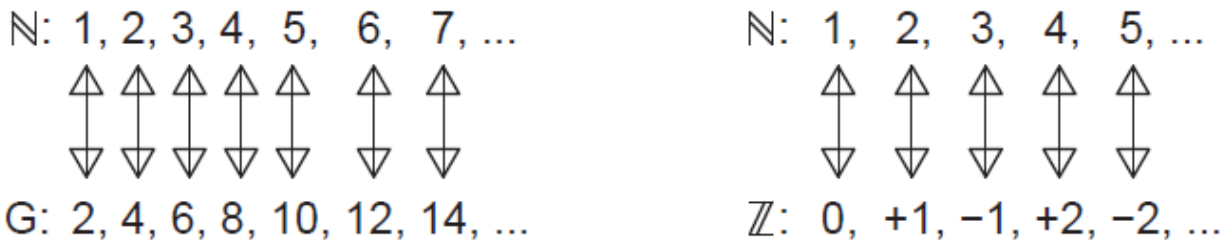


Abb. 6.1.5: Bijektion zwischen abzählbar unendlichen Mengen

Zwei Mengen sind gleichmächtig, wenn eine **Bijektion** zwischen beiden Mengen existiert

● Abzählbare und überabzählbare Mengen ●

Jede Menge, welche die gleiche Mächtigkeit wie ℕ aufweist, gilt als **abzählbar unendlich** oder kurz **abzählbar**

Tab. 6.1.1: Mächtigkeit von Mengen – Beispiele

Mächtigkeit	Beispiele
endlich	{1, 2, 3, 4}, {sonnig, bewölkt, regnerisch}, ...
abzählbar unendlich	ℕ, ℤ, ℚ, ...
überabzählbar	ℝ, ℝ \ ℚ, [0, 1), [2, 4.6), ...

6.1.3 Axiomatischer Wahrscheinlichkeitsbegriff und Regeln

● Ergebnisräume und Ereignisse ●

Der **Ergebnisraum** ist die Menge aller möglichen Ergebnisse eines Zufallsvorgangs und wird üblicherweise mit Ω (lies: *Omega*) bezeichnet. Besteht Ω aus endlich oder abzählbar unendlich vielen Elementen, so sprechen wir von einem **diskreten Ergebnisraum**, sonst von einem **überabzählbaren Ergebnisraum**.

Teilmengen von Ω, denen Wahrscheinlichkeiten in wohldefinierter Weise zugeordnet werden können, heißen **Zufallsereignisse** oder kurz **Ereignisse**. Die Elemente von Ω heißen **Elementarereignisse**. Der Ergebnisraum Ω selbst wird als **sicheres Ereignis** und das zugehörige Komplement $\overline{\Omega} = \emptyset$ als **unmögliches Ereignis** bezeichnet. Sind A und B sich gegenseitig ausschließende Ereignisse, d.h. gilt $A \cap B = \emptyset$, so spricht man von **disjunkten Ereignissen**.

● Beispiele ●

> Würfelwurf: $\Omega = \{1, 2, 3, 4, 5, 6\}$ **Ergebnisraum**

$\varpi_1 = 1, \varpi_2 = 2, \ldots, \varpi_6 = 6$ **Elementarereignisse**

$G = \{2, 4, 6\}$ G ... **Ereignis** „gerade Zahl“

6.1 Wichtige Grundbegriffe und Regeln

> Wetter am nächsten Tag:

$\Omega = \{sonnig, bedeckt, teils\ bedeckt, regnerisch\}$

$S = \{sonnig, teils\ bedeckt\}$ S ... Ereignis „zumindest teilweise sonnig“

> Zweimaliger Münzwurf:

$\Omega = \{(Zahl, Zahl), (Zahl, Wappen), (Wappen, Zahl), (Wappen, Wappen)\}$

$Z_1 = \{(Zahl, Zahl), (Zahl, Wappen), (Wappen, Zahl)\}$

Z_1 ... Ereignis „mindestens einmal Zahl“

> Würfeln bis zur ersten Sechs:

$\Omega = \{1, 2, 3, \dots\} = \mathbb{N}$

$W_{11} = \{11, 12, 13, \dots\}$ W_{11} ... Ereignis „mindestens 11 Versuche“

> Lebensdauer einer Computer-Festplatte:

$\Omega = [0, \infty)$

$L_1 = [10000, \infty)$

L_1 ... Ereignis „mindestens 10 000 Stunden“

$L_2 = [10\,000, 20\,000]$

L_2 ... Ereignis „zwischen 10 000 und 20 000 Stunden

• Wahrscheinlichkeiten und Ereignisse diskreter Ergebnisräume •

Definition 6.1.1: Diskretes Wahrscheinlichkeitsmaß

Ordnet eine Funktion P den Teilmengen eines diskreten Ergebnisraumes Ω Zahlen unter Einhaltung folgender Regeln zu:

(K1) $P(A) \geq 0$ für $A \subset \Omega$,

(K2) $P(\Omega) = 1$,

(K3) falls $A_1, A_2, A_3, \cdots \subset \Omega$ ***paarweise disjunkt*** sind, d.h. $A_i \cap A_j = \emptyset$ für $i \neq j$ und $i, j = 1, 2, 3, \ldots$, dann gilt

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i),$$

so nennt man P ein ***diskretes Wahrscheinlichkeitsmaß*** auf der Potenzmenge von Ω und die Funktionswerte von P heißen ***Wahrscheinlichkeiten***.

6.1 Wichtige Grundbegriffe und Regeln

• Allgemeine Anmerkungen zur Axiomatik •

> Vorliegende Definition gilt zunächst einmal nur für **diskrete Ergebnisräume**

> Überabzählbare Ergebnisräume erzeugen mathematisches Problem bei der Definition von Wahrscheinlichkeiten (siehe später)

> Sämtliche Rechenregeln letztlich auf diese 3 Axiome zurückführbar

> Axiome geben keine Anhaltspunkte darüber, wie konkrete Wahrscheinlichkeiten berechnet werden sollen

• Wahrscheinlichkeiten als Funktionswerte •

Mathematisch sind Wahrscheinlichkeiten Funktionswerte einer Funktion P, die den Teilmengen einer Grundmenge Ω unter Einhaltung der Regeln (K1)–(K3) Zahlen zwischen 0 und 1 zuordnet. Dabei wird P als **Wahrscheinlichkeitsmaß** und Ω als **Ergebnisraum** bezeichnet. Im Falle diskreter Ergebnisräume lässt sich dies auch formal ausdrücken als $P : \mathfrak{P}(\Omega) \to [0, 1] \subset \mathbb{R}$.

Beispiel:
In einer Urne liegen 3 Kugeln, die mit den Zahlen 2, 4 und 6 beschriftet sind. Es wird zufällig eine Kugel gezogen. Dann gilt:

Ergebnisraum: $\Omega = \{2, 4, 6\}$

Definitionsbereich von *P*:

$$\mathfrak{P}(Z) = \{\{2\}, \{4\}, \{6\}, \{2,4\}, \{2,6\}, \{4,6\}, \{2,4,6\}, \emptyset\}$$

Wertebereich von *P*:

[0, 1] oder genauer eine Teilmenge davon wie etwa $\{0, 1/3, 2/3, 1\}$

● Die Bedeutung der Axiome im Einzelnen ●

> (K1): Wahrscheinlichkeiten sind nicht negativ

> (K2): Normierungsaxiom (Beschränkung auf Wertebereich [0, 1])

> (K3): Einzige Rechenregel (Additionssatz für disjunkte Ereignisse)

6.1 Wichtige Grundbegriffe und Regeln

• Schlussfolgerungen aus den Axiomen •

Elementare Regeln für Wahrscheinlichkeiten

Seien Ω ein Ergebnisraum und $A, B, A_1, A_2, \ldots, A_n$ beliebige Ereignisse.
Dann gelten:

1. $0 \leq P(A) \leq 1$.
2. $P(\emptyset) = 0$.
3. Wenn $A \subset B$, dann ist $P(A) \leq P(B)$.
4. $P(\overline{A}) = 1 - P(A)$.
5. Wenn $A_1, A_2 \ldots, A_n$ ***paarweise disjunkt*** sind, d.h. $A_i \cap A_j = \emptyset$ für $i \neq j$ und $i, j = 1, 2, \ldots, n$, dann ist

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i).$$

6. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
7. $P(A \cup B) = P(A) + P(B)$, wenn $A \cap B = \emptyset$ ist.

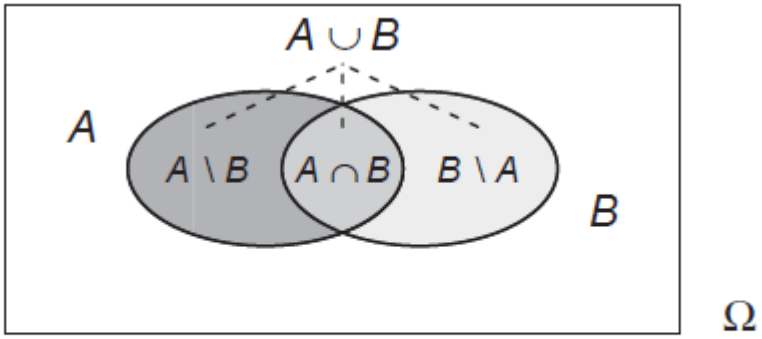


Abb. 6.1.6: Illustration zum axiomatischen Nachweis des Additionssatzes

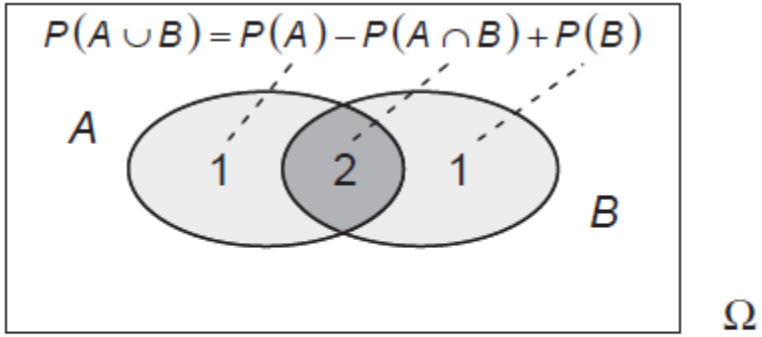


Abb. 6.1.7: „Geometrischer Nachweis" des Additionssatzes für zwei Ereignisse

● Beispiel 6.1.1: Würfelwurf ●

$\Omega = \{1, 2, 3, 4, 5, 6\}$

Betrachte Ereignisse $A = \{2, 4, 6\}$ und $D = \{6\}$

Gemäß Regel Nr. 3 (Folie 291) muss dann gelten: $P(D) \leq P(A)$

Gemäß Regel Nr. 4 muss gelten: $P(\bar{A}) = 1 - P(A)$

Man beachte, dass aus den bisherigen Rechenregeln nicht hervorgeht welche konkreten Wahrscheinlichkeiten hier gelten sollten.
Beispielsweise könnte im vorliegenden Fall gelten:

$$P(D) = \frac{1}{6} \leq P(A) = \frac{1}{2} = P(\bar{A})$$

sofern man jedem Elementarereignis die Wahrscheinlichkeit 1/6 zuordnet.

● Beispiel 6.1.2: Additionssatz für 3 Ereignisse ●

$$P(A \cup B \cup C) = P(A) + P(B) + P(C)$$
$$-P(A \cap B) - P(B \cap C) - P(A \cap C)$$
$$+P(A \cap B \cap C)$$

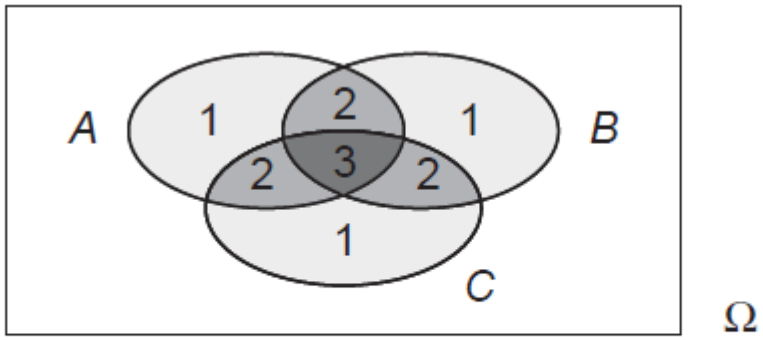


Abb. 6.1.8: „Geometrischer Nachweis“ des Additionssatzes für drei Ereignisse

6.1 Wichtige Grundbegriffe und Regeln

• Additionskalkül für diskrete Ergebnisräume •

Satz 6.1.1: Additionskalkül für diskrete Ergebnisräume

Sei Ω ein diskreter Ergebnisraum mit endlich oder abzählbar unendlich vielen Elementarereignissen $\omega_1, \omega_2, \ldots, \omega_k$ bzw. $\omega_1, \omega_2, \omega_3, \ldots$ Sei $A \subset \Omega$ ein beliebiges Ereignis. Weiter gelte:

(i) $P(\{\omega_i\}) \geq 0$ für alle i.
(ii) $\sum_i P(\{\omega_i\}) = 1$.
(iii) $P(A) = P(\{\omega_i : \omega_i \in A\}) = \sum_{\omega_i \in A} P(\{\omega_i\})$.

Dann ist P ein diskretes Wahrscheinlichkeitsmaß auf $\mathfrak{P}(\Omega)$.

> Unter Kenntnis der Wahrscheinlichkeiten der Elementarereignisse lässt sich im Falle eines diskreten Ergebnisraumes die Wahrscheinlichkeit für jedes beliebige Ereignis über Addition bestimmen.

> Beispiel: **Laplace Modell** (**klassische Wahrscheinlichkeiten**)

$$P(A) = \frac{|A|}{|\Omega|} \quad \text{für} \quad |\Omega| < \infty$$

● **Beispiel 6.1.1 fortgesetzt** ●

Einmaliger Würfelwurf mit $\Omega = \{1, 2, 3, 4, 5, 6\}$

Betrachte folgende Ereignisse:

$$A = \{2, 4, 6\}, \qquad B = \{2, 3, 6\}, \qquad C = \{1, 2, 3, 4\}, \qquad D = \{6\}$$

Dann gilt z. B. (gemäß Additionskalkül und Rechenregeln):

$$P(A) = \frac{|A|}{|\Omega|} = \frac{3}{6} = 1/2 \qquad P(B) = \frac{|B|}{|\Omega|} = \frac{3}{6} = 1/2$$

$$P(C) = \frac{|C|}{|\Omega|} = \frac{4}{6} = 2/3 \qquad P(D) = \frac{|D|}{|\Omega|} = 1/6$$

$$A \cup C = \{1, 2, 3, 4, 6\} \qquad A \cap C = \{2, 4\}$$

$$\Rightarrow \quad P(A \cup C) = 5/6 \qquad \Rightarrow \quad P(A \cap C) = 2/6 = 1/3$$

$$P(A \cup C) = P(A) + P(C) - P(A \cap C) = \frac{1}{2} + \frac{2}{3} - \frac{1}{3} = 5/6$$

6.1 Wichtige Grundbegriffe und Regeln

● Das Problem überabzählbarer Ergebnisräume ●

> Additionskalkül funktioniert nicht: Es ist mathematisch nicht möglich, jedem Elementarereignis eines überabzählbaren Ergebnisraumes eine positive Wahrscheinlichkeit zuzuordnen, ohne die Axiomatik zu verletzen.

> Eine (nichttriviale) Konsequenz davon ist, dass als Definitionsbereich eines Wahrscheinlichkeitsmaße eine „deutlich kleinere“ Menge als die Potenzmenge gewählt werden muss (sog. „Sigma-Algebren)

● Integrationskalkül für überabzählbare Ergebnisräume ●

> Additionskalkül wird durch **Integrationskalkül** ersetzt: Wahrscheinlichkeiten werden nicht über Addition sondern über Integration bestimmt

> Integrationskalkül erzeugt jedoch **Messbarkeitsproblem**: Nicht mehr jeder Teilmenge von Ω kann eine Wahrscheinlichkeit zugeordnet werden (Verkleinerung der Potenzmenge)

> Weiteres hierzu später bei **stetigen Zufallsvariablen**

6.1 Wichtige Grundbegriffe und Regeln

Additionskalkül vs. Integrationskalkül und Rechenregeln

Im Rahmen diskreter Ergebnisräume lassen sich Ereigniswahrscheinlichkeiten über einen **Additionskalkül** bestimmen, innerhalb dessen anteilige Summen von Einzelwahrscheinlichkeiten bestimmt werden. Für überabzählbare Ergebnisräume wird dagegen ein **Integrationskalkül** verwendet, innerhalb dessen anteilige Flächen von Wahrscheinlichkeitsdichten zur Bestimmung von Ereigniswahrscheinlichkeiten bestimmt werden.

Alle bisherigen und zukünftigen Rechenregeln für Wahrscheinlichkeiten gelten unabhängig von der Beschaffenheit des Ergebnisraumes, sofern keine Einschränkungen vorgenommen werden.

- **Maßtheoretische Verallgemeinerungen** -

> Axiome und Rechenregeln auf überabzählbare Ergebnisräume erweiterbar

> Im Grunde bleibt alles erhalten, lediglich für Definitionsmenge des W‘maßes (auch **Ereignisraum** genannt), wird i. d. R. nicht mehr die Potenzmenge, sondern eine sog. **Sigma-Algebra** gewählt („kleineres“ Teilmengensystem)

> Primär rein theoretisches Problem; keine praktische Relevanz

6.2.1 Bedingte Wahrscheinlichkeiten und Rechenregeln

• Definition •

Definition 6.2.1: Bedingte Wahrscheinlichkeit

Seien A und B Ereignisse eines Ergebnisraumes Ω mit $P(B) > 0$. Dann ist die ***bedingte Wahrscheinlichkeit von A unter B*** definiert als

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

Die bedingte Wahrscheinlichkeit ist für $P(B) = 0$ nicht definiert.

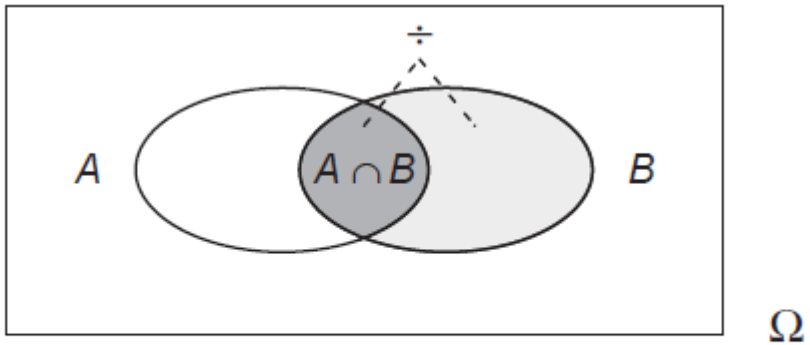


Abb. 6.2.1: Illustration zur bedingten Wahrscheinlichkeit von A unter B

● Beispiel 6.2.1 ●

Tab. 6.2.1: Gemeinsame relative Verteilung von Geschlecht und Raucherstatus

	Raucher	Gelegenheits-raucher	Nichtraucher	Summe
weiblich	0.04	0.08	0.28	0.40
männlich	0.12	0.12	0.36	0.60
Summe	0.16	0.20	0.64	1.00

> Empirischer Kalkül (bisher):

Raucheranteil insgesamt: $f_{\bullet 1} = 0.16$

Raucheranteil bei Frauen: $f_{11}^{Y|X} = \frac{f_{11}}{f_{1\bullet}} = \frac{0.04}{0.40} = 0.10$

> Theoretischer Kalkül (jetzt):

Zufallsvorgang: Eine Person werde zufällig gezogen

$\Omega = \{(w, R), (w, GR), (w, NR), (m, R), (m, GR), (m, NR)\}$

$w \ldots weiblich, m \ldots männlich$
$R \ \ldots Raucher, GR \ldots Gelegenheitsraucher, NR \ldots Nichtraucher$

Betrachte Ereignisse:

$R = \{(w, R), (m, R)\}$ ... ein Raucher wird gezogen

$F = \{(w, R), (w, GR), (w, NR)\}$ ... eine Frau wird gezogen

Dann gilt jetzt z. B.:

$$P(R) = 0.16$$

$$P(R|F) = \frac{P(R \cap F)}{P(F)} = \frac{0.04}{0.40} = 0.10$$

- **Arithmetik bedingter Wahrscheinlichkeiten** -

Für bedingte Wahrscheinlichkeiten gelten die gleichen Rechenregeln wie für „gewöhnliche“ (unbedingte)

● Beispiel 6.1.1 fortgesetzt ●

Einmaliger Würfelwurf mit $\Omega = \{1, 2, 3, 4, 5, 6\}$

Betrachte folgende Ereignisse:

$$A = \{2, 4, 6\}, \qquad B = \{2, 3, 6\}, \qquad C = \{1, 2, 3, 4\}$$

Dann gilt z. B. :

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{2, 6\})}{P(B)} = \frac{2/6}{3/6} = 2/3$$

$$P(C|B) = \frac{P(C \cap B)}{P(B)} = \frac{P(\{2, 3\})}{P(B)} = \frac{2/6}{3/6} = 2/3$$

$$P(A \cap C|B) = \frac{P\big((A \cap C) \cap B\big)}{P(B)} = \frac{P(\{2\})}{P(B)} = \frac{1/6}{3/6} = \frac{1}{3}$$

Einerseits gilt:

$$P(A \cup C|B) = \frac{P\big((A \cup C) \cap B\big)}{P(B)} = \frac{P(\{2,3,6\})}{P(B)} = \frac{3/6}{3/6} = 1$$

Andererseits gilt mit Rechenregel Nr. 6:

$$P(A \cup C|B) = P(A|B) + P(C|B) - P(A \cap C|B) = \frac{2}{3} + \frac{2}{3} - \frac{1}{3} = 1$$

Einerseits:

$$P(\bar{C}|B) = \frac{P(\bar{C} \cap B)}{P(B)} = \frac{P(\{6\})}{P(B)} = \frac{1/6}{3/6} = \frac{1}{3}$$

Andererseits gilt mit Rechenregel Nr. 4:

$$P(\bar{C}|B) = 1 - P(C|B) = 1 - \frac{2}{3} = \frac{1}{3}$$

Bedingtes Wahrscheinlichkeitsmaß und gewöhnliche Rechenregeln

Sei Ω ein Ergebnisraum und B ein Ereignis mit $P(B) > 0$. Dann wird durch $P(\bullet|B)$ ein Wahrscheinlichkeitsmaß definiert, das auch als ***bedingtes Wahrscheinlichkeitsmaß unter B*** bezeichnet wird. Somit gelten für bedingte Wahrscheinlichkeiten die gewöhnlichen Rechenregeln für Wahrscheinlichkeiten.

• Multiplikationsregel •

Satz 6.2.1: Multiplikationsregel

Seien $A_1, A_2, \ldots, A_n$ mit $n \geq 2$ Ereignisse eines Ergebnisraumes Ω, wobei $P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$ sei. Dann gilt:

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdot \ldots \cdot P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}).$$

2 Ereignisse: $$P(A_2|A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)}$$

$$\Leftrightarrow \quad P(A_1 \cap A_2) = P(A_2|A_1)P(A_1)$$

3 Ereignisse: $$P(A_3|A_1 \cap A_2) = \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)}$$

$$\Leftrightarrow \quad P(A_1 \cap A_2 \cap A_3) = P(A_3|A_1 \cap A_2)P(A_1 \cap A_2)$$

$$\Rightarrow \quad P(A_1 \cap A_2 \cap A_3) = P(A_3|A_1 \cap A_2)P(A_2|A_1)P(A_1)$$

... analog geht es immer weiter. Für 4 Ereignisse erhält man dann

$$P(A_1 \cap A_2 \cap A_3 \cap A_4)$$
$$= P(A_4|A_1 \cap A_2 \cap A_3)P(A_3|A_1 \cap A_2)P(A_2|A_1)P(A_1)$$

● Beispiel 6.2.2 ● aus Hartung und Heine [2004]

Die Entwicklungsabteilung eines Produzenten von Haushaltsgeräten ist in **90%** der Fälle für die Markteinführung der von ihr entwickelten Geräte. Ein positives Votum der Entwicklungsabteilung führt mit einer Wahrscheinlichkeit von **0.7** bei der Marketingabteilung ebenfalls zu einem positiven Votum. Sind beide Abteilungen für die Markteinführung des neuen Gerätes, so entscheidet die Geschäftsleitung dennoch mit einer Wahrscheinlichkeit von **0.2** dagegen. Ist die Marketingabteilung gegen die Markteinführung, die Entwicklungsabteilung aber dafür, so stimmt die Geschäftsleitung nur mit einer Wahrscheinlichkeit von **0.4** zu.

a) Wie groß ist die Wahrscheinlichkeit, dass die Markteinführung eines neuen Produktes sowohl von der Geschäftsleitung als auch von der Entwicklungs- und der Marketingabteilung getragen wird?

b) Mit welcher Wahrscheinlichkeit entscheiden sich Geschäftsleitung und Entwicklungsabteilung für die Markteinführung eines neuen Produktes?

Definiere Ereignisse

A_1 = Entwicklungsabteilung ist für die Markteinführung

A_2 = Marketingabteilung ist für die Markteinführung

A_3 = Geschäftsleitung ist für die Markteinführung

Aus Text geht hervor:

$P(A_1) = 0.9$ $\qquad$ $P(A_2|A_1) = 0.7$

$P(\overline{A_3}|A_1 \cap A_2) = 0.2$ $\qquad$ $P(A_3|A_1 \cap \overline{A_2}) = 0.4$

Die gefragten Wahrscheinlichkeiten sind:

a) $P(A_1 \cap A_2 \cap A_3)$ und b) $P(A_1 \cap A_3)$

Zu a):

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$$

$$= P(A_1)P(A_2|A_1)[1 - P(\overline{A_3}|A_1 \cap A_2)]$$

$$= 0.9 \times 0.7 \times [1 - 0.2] = 0.504$$

Zu b):

$$A_1 \cap A_3 = (A_1 \cap A_2 \cap A_3) \cup (A_1 \cap \overline{A_2} \cap A_3)$$

$$\Rightarrow \quad P(A_1 \cap A_3) = P(A_1 \cap A_2 \cap A_3) + P(A_1 \cap \overline{A_2} \cap A_3)$$

$$= 0.504 + P(A_1)P(\overline{A_2}|A_1)P(A_3|A_1 \cap \overline{A_2})$$

$$= 0.504 + 0.9 \times (1 - 0.7) \times 0.4 = 0.612$$

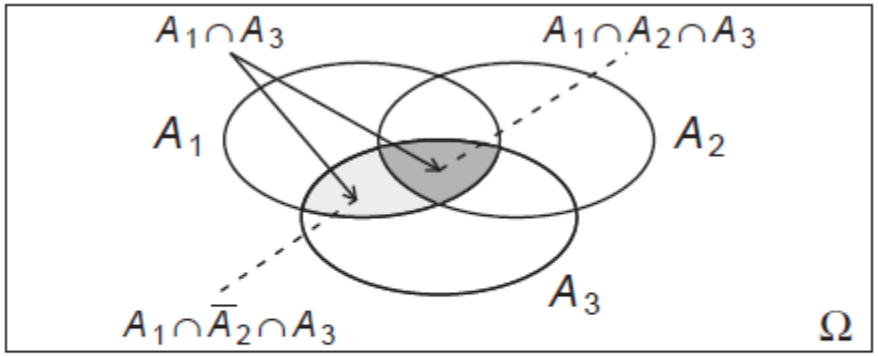


Abb. 6.2.2: Illustration zur Bestimmung von $P(A_1 \cap A_3)$

• Wahrscheinlichkeitsbäume •

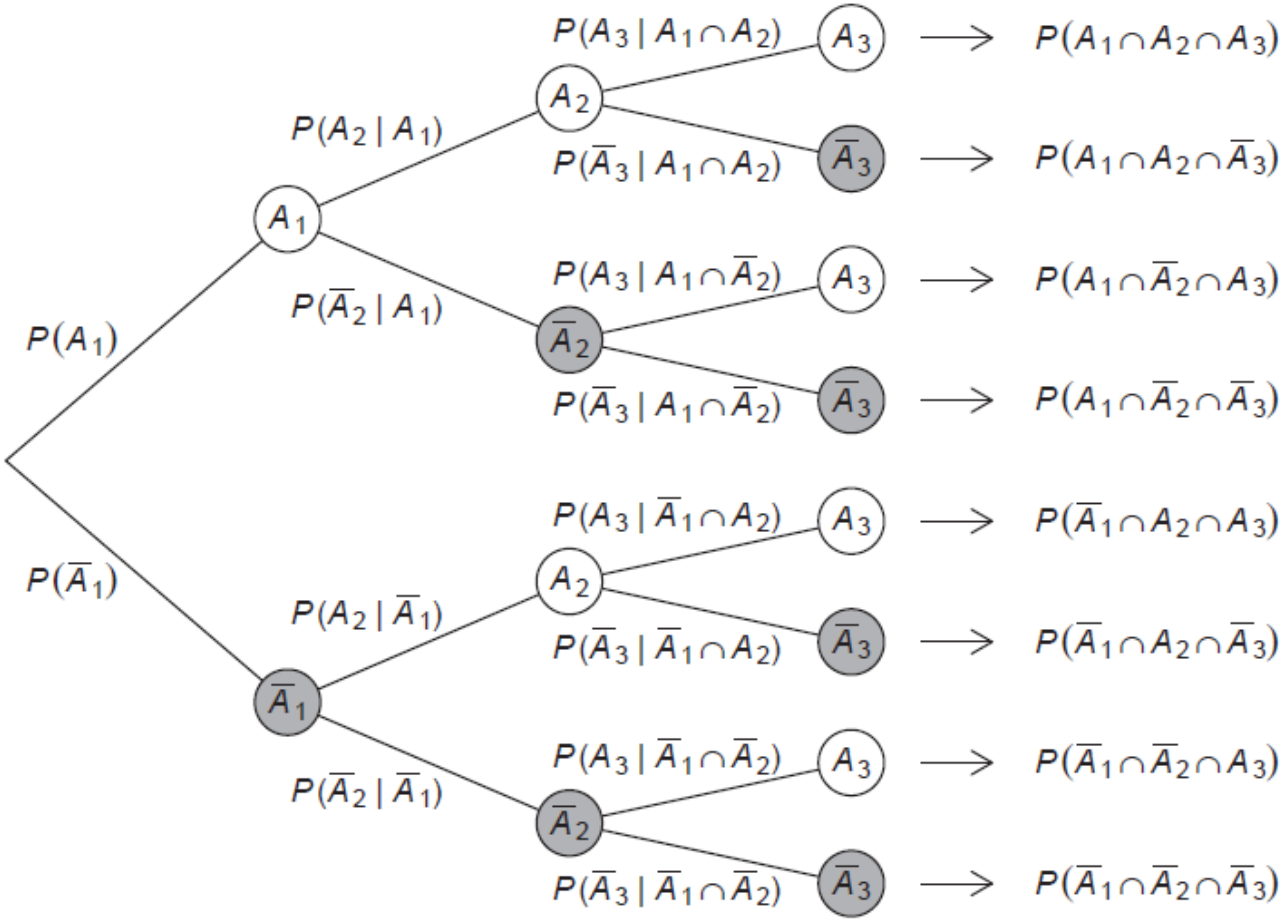


Abb. 6.2.3: Beispiel einer Pfadstruktur eines Wahrscheinlichkeitsbaumes

6.2 Rechnen mit abhängigen und unabhängigen Ereignissen

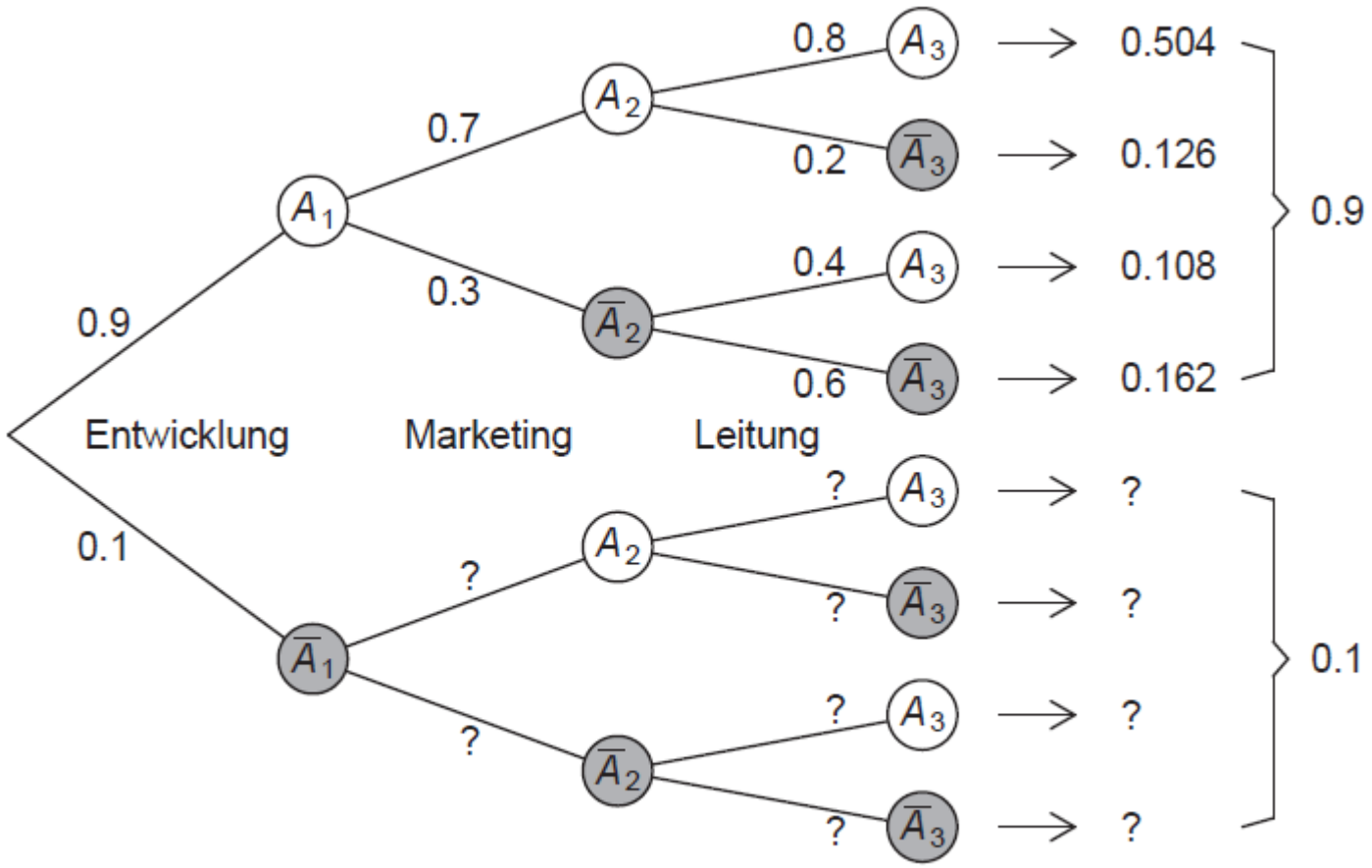


Abb. 6.2.4: Entscheidungsstruktur für die Markteinführung eines Produktes (Beispiel 6.2.2)

Entscheidend für a) und b): Pfade Nr. 1 und Nr. 3

● Beispiel 6.2.3 ●

Angenommen, in einem Raum befinden sich 30 Personen. Mit welcher Wahrscheinlichkeit haben dann wenigstens zwei von diesen Personen am gleichen Tag Geburtstag? Zur Beantwortung dieser Frage wird die vereinfachende Annahme getroffen, dass Geburtstage allgemein über die 365 Tage des Jahres gleichmäßig verteilt sind.

Definiere Ereignisse

B = Alle 30 Personen haben unterschiedlichen Tagen Geburtstag

A_2 = die 2. Person hat an einem anderen Tag Geburtstag als die 1. Person

A_3 = die 3. Person hat an einem anderen Tag Geburtstag als die ersten beiden

$\vdots$

A_n = die n-te Person hat an einem anderen Tag Geburtstag als die ersten n Personen

Die gefragte Wahrscheinlichkeit ist: $P(\bar{B})$

Es gilt dann:

$$B = A_2 \cap A_3 \cap \cdots \cap A_{30}$$

$$\Rightarrow \quad P(B) = P(A_2)P(A_3|A_2)P(A_4|A_2 \cap A_3) \ldots P(A_{30}|A_2 \cap A_3 \cap \cdots \cap A_{29})$$

Gemäß der Annahme gleichwahrscheinlicher Geburtstage ergibt dies:

$$P(B) = \frac{364}{365} \times \frac{363}{365} \times \frac{362}{365} \times \cdots \times \frac{336}{365} \approx 0.294$$

$$\Rightarrow \quad P(\bar{B}) = 1 - P(B) = 1 - 0.294 = 0.706$$

Die Wahrscheinlichkeit, dass unter 30 Personen mindestens 2 Personen am gleichen Tag Geburtstag haben, beträgt also 70.6%.

Bei n = 70 Personen beträgt die Wahrscheinlichkeit dafür bereits 99%.

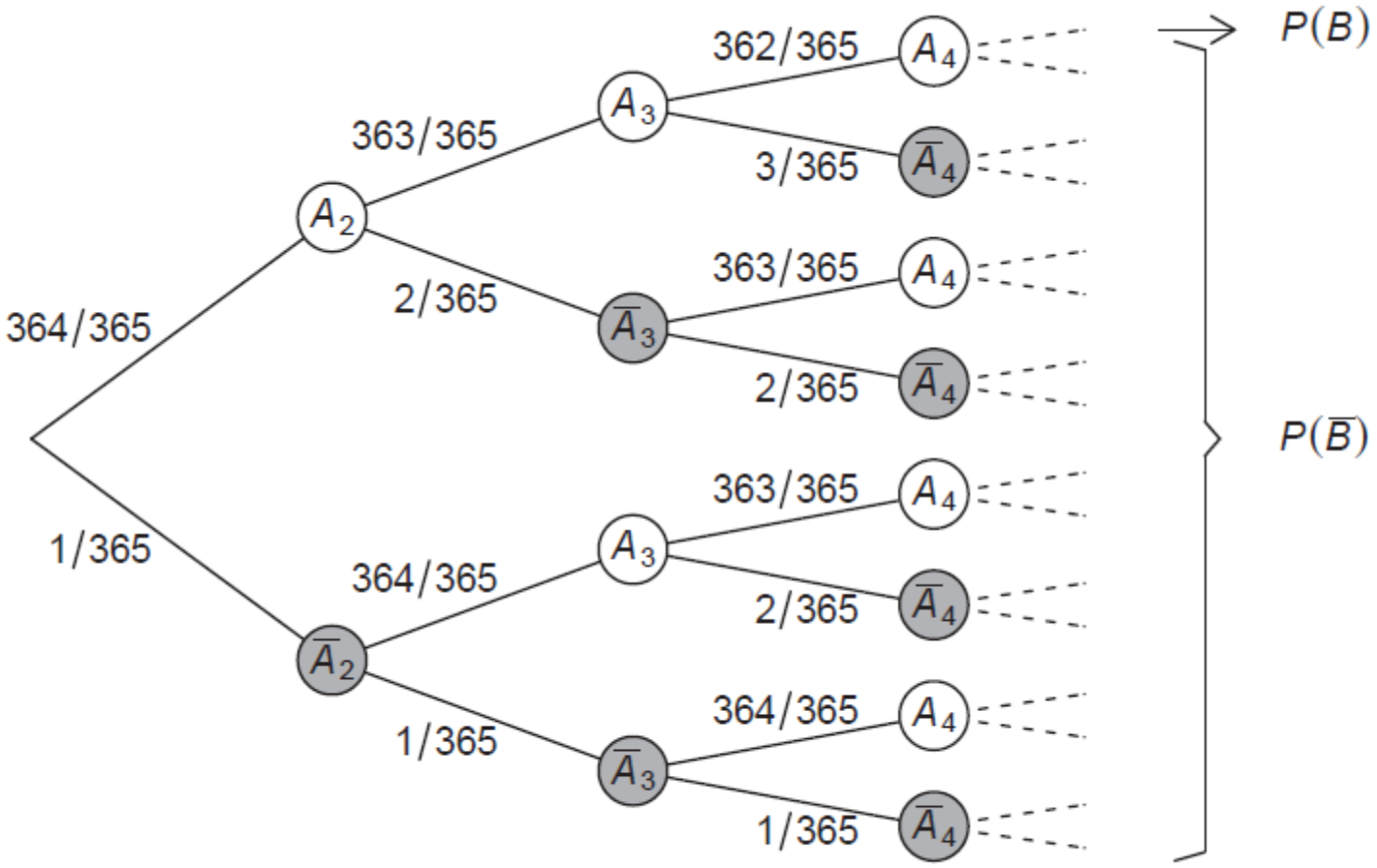


Abb. 6.2.5: Geburtstag am gleichen Tag von n Personen (Beispiel 6.2.3)

6.2.2 Stochastische Unabhängigkeit von Ereignissen

• Definition •

Definition 6.2.2: Abhängigkeit und Unabhängigkeit von zwei Ereignissen

Seien A und B Ereignisse eines Ergebnisraumes mit $P(A) > 0$ und $P(B) > 0$. Dann sind A und B ***(stochastisch) unabhängig*** (kurz ***st.u.***), falls eines der folgenden äquivalenten Kriterien erfüllt ist:

(i) $P(A|B) = P(A)$.

(ii) $P(B|A) = P(B)$.

(iii) $P(A \cap B) = P(A)P(B)$. (***Multiplikationskriterium***)

Für $P(A) = 0$ oder $P(B) = 0$ definieren wir A und B ebenfalls als (stochastisch) unabhängig. In allen anderen Fällen sind A und B ***(stochastisch) abhängig***.

Kriterium (iii) ist die stochastische Variante des (empirischen) Multiplikationskriteriums $f_{ij} = f_{i\bullet} f_{\bullet j}$

● Implizierte Unabhängigkeit von Gegenereignissen ●

Die Unabhängigkeit von A und B impliziert weitere paarweise Unabhängigkeiten in Bezug auf die Gegenereignisse. Beispielweise gilt:

$$P(A|B) = P(A)$$

$$\Rightarrow \quad P(\bar{A}|B) = 1 - P(A|B) = P(\bar{A}) \quad \Rightarrow \quad \bar{A} \text{ und } B \text{ unabhängig}$$

Analog folgt auch die Unabhängigkeit von A und $\bar{B}$ und $\bar{A}$ und $\bar{B}$

Tab. 6.2.2: Implizierte Kontingenztabelle für zwei unabhängige Ereignisse A und B

	B	$\bar{B}$	Summe
A	$P(A)P(B)$	$P(A)P(\bar{B})$	$P(A)$
$\bar{A}$	$P(\bar{A})P(B)$	$P(\bar{A})P(\bar{B})$	$P(\bar{A})$
Summe	$P(B)$	$P(\bar{B})$	1.00

• Interpretation •

Interpretation stochastischer Unabhängigkeit

Zwei Ereignisse A und B eines Ergebnisraumes Ω sind genau dann ***stochastisch unabhängig***, falls die auf A, $\overline{A}$, B und $\overline{B}$ beruhende Kontingenztabelle mit Wahrscheinlichkeiten das ***stochastische Multiplikationskriterium*** erfüllt (vgl. Tab.6.2.2). Für A und $\overline{A}$ impliziert dies beispielsweise:

$$P(A) = P(A|B) = P(A|\overline{B}) \quad \text{und} \quad P(\overline{A}) = P(\overline{A}|B) = P(\overline{A}|\overline{B}),$$

sofern $0 < P(B) < 1$. Sowohl die Wahrscheinlichkeit für das Eintreten von A als auch die Wahrscheinlichkeit für das Nichteintreten von A ist unabhängig davon, ob B eintritt oder nicht eintritt. Analoges gilt für B und $\overline{B}$.

● Beispiel 6.2.4 ●

Tab. 6.2.1: Gemeinsame relative Verteilung von Geschlecht und Raucherstatus

	Raucher	Gelegenheits-raucher	Nichtraucher	Summe
weiblich	0.04	0.08	0.28	0.40
männlich	0.12	0.12	0.36	0.60
Summe	0.16	0.20	0.64	1.00

Zufallsvorgang: Eine Person werde zufällig gezogen

Definiere Ereignisse

R = Ein Raucher wird gezogen
GR = Ein Gelegenheitsraucher wird gezogen
NR = Ein Nichtraucher wird gezogen
F = Eine Frau wird gezogen
M = Ein Mann wird gezogen

Dann sind R und F abhängig. Denn es gilt:

$$P(R|F) = \frac{P(R \cap F)}{P(F)} = \frac{0.04}{0.40} = 0.1 \qquad P(F|R) = \frac{P(R \cap F)}{P(R)} = \frac{0.04}{0.16} = 0.25$$

(i) $0.1 = P(R|F) \neq P(R) = 0.16$ (ii) $0.25 = P(F|R) \neq P(F) = 0.40$

(iii) $0.04 = P(R \cap F) \neq P(R)P(F) = 0.064$

⇒ Die Wahrscheinlichkeit, ob jemand Raucher ist oder nicht, hängt also vom Geschlecht ab.

Für GR und F erhalten wir hingegen Unabhängigkeit. Die Wahrscheinlichkeit ob jemand Gelegenheitsraucher ist oder nicht, ist von Geschlecht unabhängig.

Tab. 6.2.3: Implizierte Kontingenztabelle für GR und F

	F	M	Summe
GR	0.08	0.12	0.20
$R \cup NR$	0.32	0.48	0.80
Summe	0.40	0.60	1.00

Multiplikationskriterium (iii) erfüllt

● Disjunkte Ereignisse sind abhängig ●

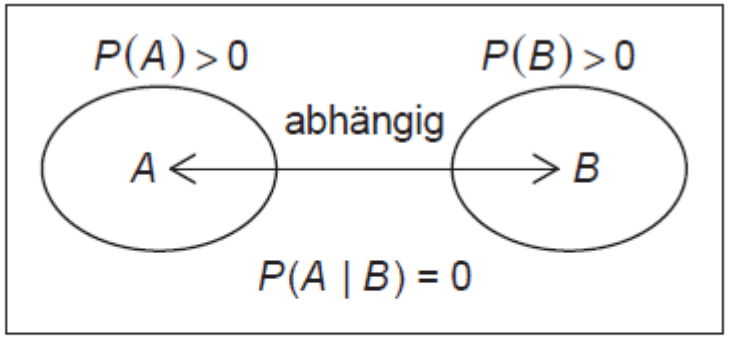


Abb. 6.2.6: Disjunkte Ereignisse sind abhängig

● Unabhängigkeit ist nicht transitiv ●

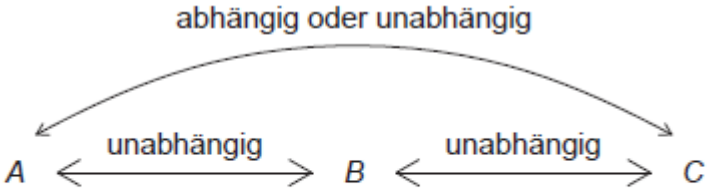


Abb. 6.2.7: Unabhängigkeit ist nicht transitiv

Betrachte als Beispiel

A, B und $C = \bar{A}$,

wobei A und B unabhängig sind

● Unabhängigkeit von mehreren Ereignissen ●

> Wären in Abb. 6.2.7 auch A und C unabhängig, so läge eine Situation **paarweise Unabhängigkeit** vor.

> Paarweise Unabhängigkeit impliziert jedoch keine „vollständige Unabhängigkeit.

> Bei paarweiser Unabhängigkeit von A, B und C gilt lediglich:

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

$$= \frac{P(A \cap B \cap C)}{P(A)P(B)} = \quad ?$$

> Bei **(vollständiger) Unabhängigkeit** von A, B und C gilt auch:

$$P(A \cap B \cap C) = P(A)P(B)P(C)$$

$$\Rightarrow \quad P(C|A \cap B) = P(C)$$

• Unabhängigkeit von mehreren Ereignissen •

Definition 6.2.3: Abhängigkeit und Unabhängigkeit von mehreren Ereignissen

Seien $A_1, A_2, \ldots, A_n$ $(n \geq 2)$ Ereignisse eines Ergebnisraumes Ω. Dann sind diese Ereignisse ***(stochastisch) unabhängig***, falls gilt:

$$P(A_i \cap A_j) = P(A_i)P(A_j) \quad \text{für } i \neq j,$$
$$P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k) \quad \text{für } i \neq j,\ j \neq k,\ i \neq k$$
$$\vdots \qquad\qquad \vdots$$
$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2)\ldots P(A_n).$$

Sind nicht alle dieser Kriterien gleichzeitig erfüllt, sind die Ereignisse ***(stochastisch) abhängig***. Ist das erste Kriterium erfüllt, sind die Ereignisse ***paarweise (stochastisch) unabhängig***.

Die Wahrscheinlichkeit des „Schnittereignisses“ muss für jede Teilmenge bestehend aus bis zu n Ereignissen dem Produkt der Wahrscheinlichkeiten der jeweiligen einzelnen Ereignisse entsprechen.

• Beispiel 6.2.5 •

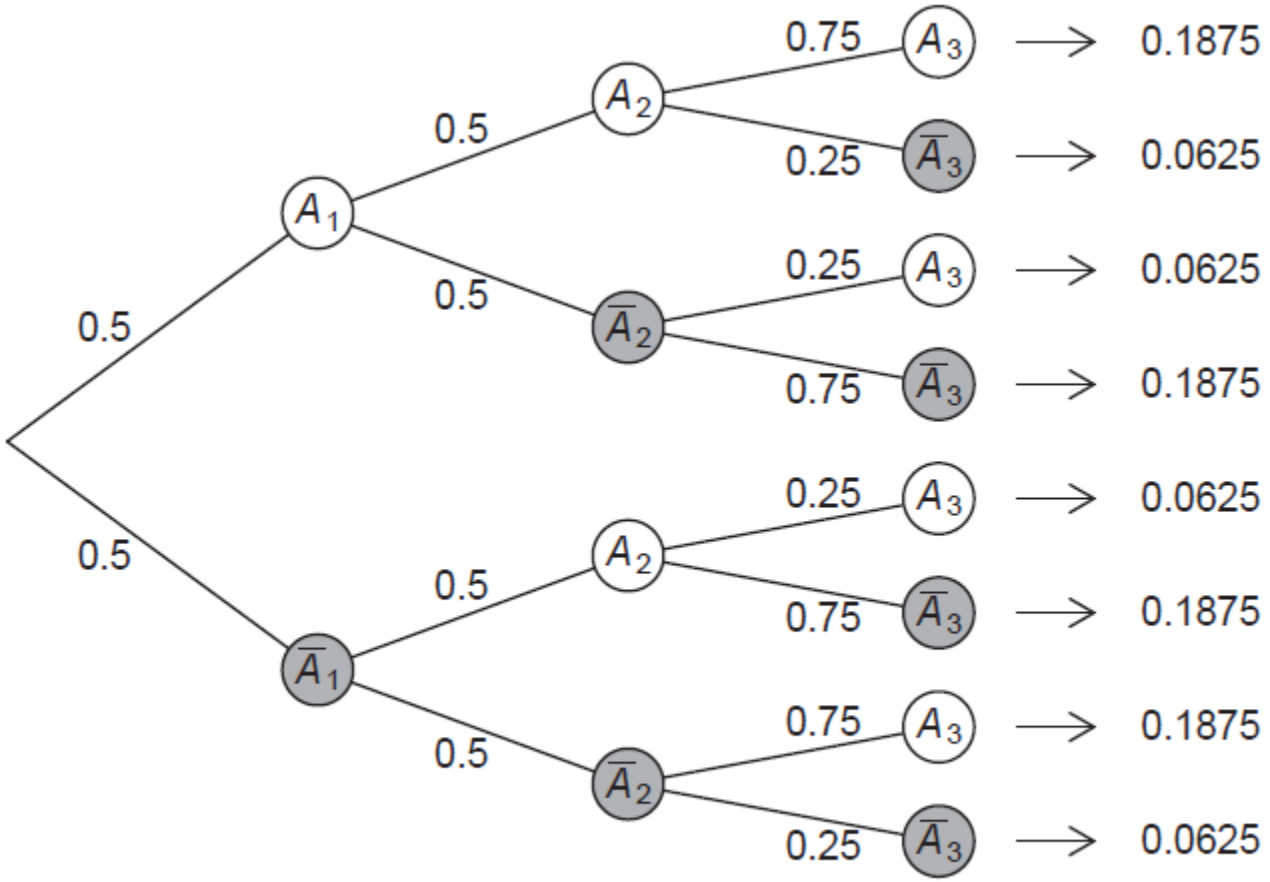


Abb. 6.2.8: Paarweise aber nicht vollständig unabhängige Ereignisse

• Implizierte Unabhängigkeit weiterer Ereignisse •

Implizierte Unabhängigkeit weiterer Ereignisse

Gegeben seien $n = n_1 + n_2 + \cdots + n_p$ unabhängige Ereignisse eines Ergebnisraumes Ω, die sich wie folgt in p Reihen anordnen lassen:

$$\begin{array}{c} A_{11}, A_{12}, \ldots, A_{1n_1}, \\ A_{21}, A_{22}, \ldots, A_{2n_2}, \\ \vdots \quad \vdots \quad \ldots \quad \vdots \\ A_{p1}, A_{p2}, \ldots, A_{pn_p}. \end{array}$$

Werden jeweils nur aus den Ereignissen der einzelnen Reihen neue Ereignisse $A_1, A_2, \ldots, A_p$ gebildet (durch Bildung von Komplementen, Schnitten oder Vereinigungen), so sind auch die so gebildeten Ereignisse unabhängig.

6.2.3 Kalkül nach der Formel von Bayes

• Satz von der totalen Wahrscheinlichkeit •

Satz von der totalen Wahrscheinlichkeit

Seien $A_1, A_2, \ldots, A_n$ ***paarweise disjunkte*** Ereignisse eines Ergebnisraumes Ω, d.h. für $i \neq j$ sei $A_i \cap A_j = \emptyset$. Weiter gelte $A_1 \cup A_2 \cup \cdots \cup A_n = \Omega$ und $P(A_i) > 0$ für $i = 1, \ldots, n$. Dann gilt für jedes Ereignis $B \subset \Omega$:

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i).$$

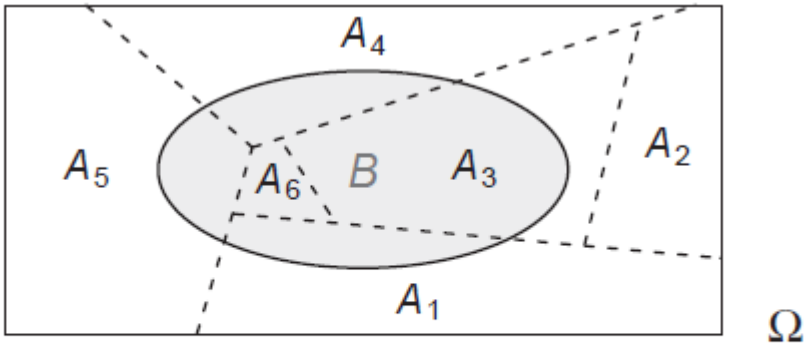


Abb. 6.2.9: Illustration des Satzes von der totalen Wahrscheinlichkeit

> Eine derart beschriebene Menge von Ereignissen heißt auch **disjunkte Zerlegung** oder **Partition** des Ergebnisraumes.

> Es gilt:

$$B = (A_1 \cap B) \cup (A_2 \cap B) \cup \cdots \cup (A_n \cap B) = \bigcup_{i=1}^{n} (A_i \cap B)$$

$$\Rightarrow \quad P(B) = P\left[\bigcup_{i=1}^{n} (A_i \cap B)\right] = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

> Man beachte, dass die Ereignisse $A_1 \cap B$, $A_2 \cap B$, ..., $A_n \cap B$, disjunkt sind

● Bayes-Formel ●

> Allgemein gilt nun für die j-te Zelle der Partition

$$\text{(i)} \quad P(A_j|B) = \frac{P(A_j \cap B)}{P(B)}$$

$$\text{(ii)} \quad P(B|A_j) = \frac{P(A_j \cap B)}{P(A_j)} \quad \Leftrightarrow \quad P(A_j \cap B) = P(B|A_j)P(A_j)$$

> Einsetzen von (ii) im Zähler von (i) und Einsetzen des **Satzes von der totalen Wahrscheinlichkeitscheinlichkeit** im Nenner von (i) ergibt...

Satz 6.2.2: Satz von Bayes

Angenommen, die Ereignisse $A_1, A_2, \ldots, A_n$ bilden eine disjunkte Zerlegung des Ergebnisraumes Ω mit $P(A_i) > 0$ für $i = 1, \ldots, n$. Dann gilt für jedes Ereignis $B \subset \Omega$ mit $P(B) > 0$:

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)} \quad \text{für } i = 1, \ldots, n.$$

● Beispiel 6.2.6 ●

In Australien wird ein Patient, der von einer unbekannten Giftschlange gebissen wurde, in die Notaufnahme einer Klinik gebracht. Es muss nun möglichst schnell das richtige Antiserum gespritzt werden. Angenommen, es kommen im vorliegenden Fall nur drei verschiedene Schlangenarten in Frage, die wir hier mit A1, A2 und A3 bezeichnen. Nach hauseigenen Statistiken des betroffenen Krankenhauses seien in der Vergangenheit **70%** aller Bisse auf Schlange A1, **20%** auf Schlange A2 und **10%** auf Schlange A3 zurückzuführen gewesen. Der Patient weist nun aber ein ganz bestimmtes Vergiftungssymptom auf, das einer landesweiten Studie zu Folge mit einer Wahrscheinlichkeit von **5%** bei A1, mit **20%** bei A2 und mit **30%** bei A3 auftritt. Wie sollte nun entschieden werden?

Definiere Ereignisse

A_i = Patient wurde von Schlange A_i gebissen (i = 1, 2, 3)

B = Das betreffende Symptom tritt beim Patienten auf

Aus Text geht hervor:

$P(A_1) = 0.7$ $P(B|A_1) = 0.05$

$P(A_2) = 0.2$ $P(B|A_2) = 0.20$

$P(A_3) = 0.1$ $P(B|A_3) = 0.30$

Mit Bayes-Formel folgt:

$P(A_1|B) = 0.33$

$P(A_2|B) = 0.38$

$P(A_3|B) = 0.29$

Beispielrechnung für $P(A_1|B)$:

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{\sum_{i=1}^{3} P(B|A_i)P(A_i)}$$

$$= \frac{0.05 \times 0.7}{0.05 \times 0.7 + 0.2 \times 0.2 + 0.3 \times 0.1} = 0.33$$

Entscheidung für ...

A_1 A_3 A_2

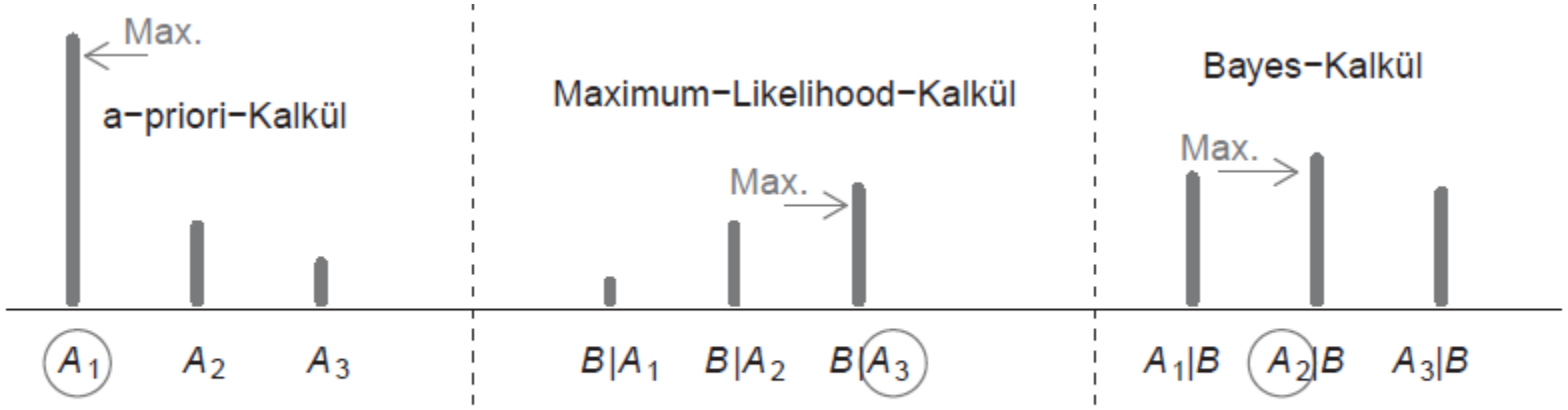


Abb. 6.2.10: Beispiel: Illustration der verschiedenen kalkulatorischen Ansätze

• Bayes-Kalkül vs. Maximum-Likelihood-Kalkül •

Bayes-Kalkül vs. Maximum-Likelihood-Kalkül

Grundfrage des Bayes-Kalküls:

Welcher Zustand ist unter den beobachteten Gegebenheiten am wahrscheinlichsten?

Grundfrage des Maximum-Likelihood-Kalküls:

Unter welchem Zustand sind die beobachteten Gegebenheiten am wahrscheinlichsten?

7 Theoretische Verteilungen und Abhängigkeiten

7 Theoretische Verteilungen und Abhängigkeiten

7.1 Zufallsvariablen und deren Wahrscheinlichkeitsverteilungen	335
7.1.1 Eindimensionale Zufallsvariablen	335
Konzept und Ausblick	335
Diskrete Zufallsvariablen und Additionskalkül	337
Stetige Zufallsvariablen und Integrationskalkül	341
Theoretische Verteilungsfunktion	349
Wahrscheinlichkeitsverteilung	356
7.1.2 Mehrdimensionale Zufallsvariablen	359
Konzept und Ausblick	359
Zweidimensionale diskrete Zufallsvariablen	360
Zweidimensionale stetige Zufallsvariablen	369
Stochastische Abhängigkeit und Unabhängigkeit	385
Höherdimensionale Wahrscheinlichkeitsverteilungen	389

7 Theoretische Verteilungen und Abhängigkeiten

7.1.3 Verteilung von Funktionen von Zufallsvariablen	403
Funktionen einer Zufallsvariablen	403
Funktionen aus mehreren Zufallsvariablen	413
7.2 Theoretische Kennwerte	421
7.2.1 Kennwerte in Bezug auf Lage und Streuung	421
Erwartungswert	421
Theoretische Varianz und Standardabweichung	445
Theoretische Quantile und theoretischer Median	454
7.2.2 Kennwerte in Bezug auf Abhängigkeiten	459
Bedingte Erwartungswerte und Varianzen	459
Theoretische Kovarianz und Korrelation	472
Theoretische Regressionskoeffizienten	482

7 Theoretische Verteilungen und Abhängigkeiten

7.2.3 Spezifische Eigenschaften theoretischer Kennwerte 485

Minimumeigenschaften von Lagekennwerten 485

Wichtige Transformationseigenschaften 487

Endliche und nicht endliche theoretische Momente 490

7.3 Spezielle eindimensionale Verteilungen 492

7.3.1 Spezielle diskrete Verteilungen 492

Elementare Kombinatorik 492

Einpunktverteilung 496

Bernoulli-Verteilung 497

Binomialverteilung 499

Poisson-Verteilung 505

7 Theoretische Verteilungen und Abhängigkeiten

7.3.2 Spezielle stetige Verteilungen	509
Stetige Gleichverteilung	509
Exponentialverteilung	513
Normalverteilung	519
7.4 Verteilung stochastischer Summen und Mittelwerte	530
7.4.1 Exakte Aussagen	530
Erwartungswerte und Varianzen	530
Verteilungen unter bestimmten Ausgangsverteilungen	535
7.4.2 Asymptotische und approximative Aussagen	543
Gesetz der großen Zahlen (GGZ)	543
Zentraler Grenzwertsatz (ZGWS)	548

7.1.1 Eindimensionale Zufallsvariablen

Konzept und Ausblick

● Zufallsvariablen und Realisationen ●

> Zufallsvariablen sind Funktionen, welche die Ergebnisse eines Zufalls-vorgangdurch Zahlen ausdrücken.

> Beispiel: Raucherstatus einer zufällig ausgewählten Person

Ursprünglicher Ergebnisraum:

$$\Omega_{Rauchen} = \{Nichtraucher, Gelegenheitsraucher, Raucher\}$$

Definierte nun Zufallsvariable X: $\Omega_{Rauchen} \longrightarrow \{0, 1, 2\}$ mit

$$X(Nichtraucher) = 0,$$
$$X(Gelegenheitsraucher) = 1,$$
$$X(Raucher) = 2$$

Neuer Ergebnisraum: $\Omega_{012} = \{0, 1, 2\}$

> Elementarereignisse des neuen Ergebnisraumes, welcher eine Teilmenge von $\mathbb{R}$ ist, heißen **Realisationen** (einschränkend: **Realisationsmöglichkeiten**

● Vorteil kurzer und operabler Schreibweisen ●

> $X = 0$ für Ereignis „Nichtraucher“ oder $X = 2$ für Ereignis „Raucher“

> $X > 0$ für Ereignis „kein Nichtraucher“

> Wahrscheinlichkeiten für Ereignisse werden dann kurz notiert
$P(X = 0), P(X = 2)$ oder $P(X > 0)$

> Man beachte: Bei mehreren Zufallsvariablen notiert man gewöhnlich
$P(X = 0, Y = 1)$ statt $P(X = 0 \cap Y = 1)$

● Bemerkung zur mathematischen Definition ●

● Wahrscheinlichkeitsverteilung einer Zufallsvariable ●

Art und Weise wie einzelne Realisationen wahrscheinlichkeitsmäßig (theoretisch) verteilt sind

Diskrete Zufallsvariablen und Additionskalkül

• Wahrscheinlichkeitsfunktion für diskrete Zufallsvariablen •

Definition 7.1.1: Diskrete Zufallsvariable und Wahrscheinlichkeitsfunktion

Eine Zufallsvariable X heißt diskret, falls es für endlich oder abzählbar unendlich viele Werte $a_1, a_2, \ldots, a_k, \ldots$ eine Funktion f_X gibt, für die gilt:

$$f_X(x) = \begin{cases} P(X = a_j) = p_j, & \text{für } x = a_j,\ j = 1, \ldots, k, \ldots \\ 0, & \text{sonst,} \end{cases}$$

wobei

(i) $p_j \geq 0$ für alle j und

(ii) $\sum_{j=1}^{\infty} p_j = 1$.

Die Funktion f_X heißt dann **Wahrscheinlichkeitsfunktion** von X. Die Menge $T_X = \{a_j : p_j > 0\}$ wird als Trägermenge von X bezeichnet. Ihre Elemente heißen **Realisationsmöglichkeiten**, **Träger-** oder **Massenpunkte** von X.

• Additionskalkül für diskrete Zufallsvariablen •

Satz 7.1.1: Additionskalkül für diskrete Zufallsvariablen

Sei X eine diskrete Zufallsvariable mit Realisationsmöglichkeiten $a_1, a_2, \ldots, a_k, \ldots$ und zugehörigen Wahrscheinlichkeiten $p_1, p_2, \ldots, p_k, \ldots$ Dann gilt für jedes Ereignis $X \in A$ mit $A \subset \mathbb{R}$:

$$P(X \in A) = \sum_{j:a_j \in A} p_j.$$

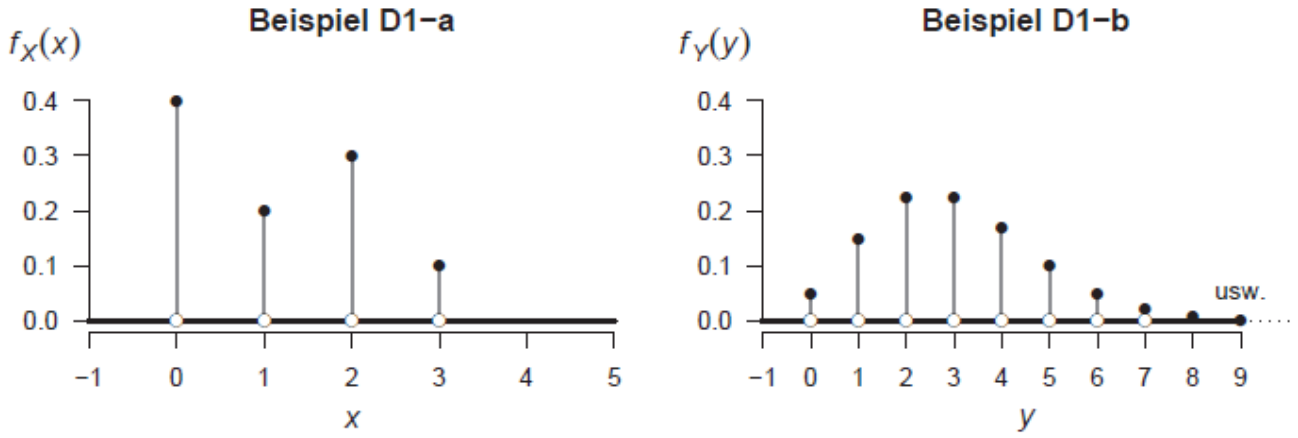


Abb. 7.1.1: Wahrscheinlichkeitsfunktionen diskreter Zufallsvariablen

● Beispiel D1-a ●

Sei X die Anzahl mitreisender Kinder bei gebuchten Pauschalreisen eines Reiseveranstalters. Beispielsweise könnte dann gelten:

$$P(X=0)=0.4,\ \ P(X=1)=0.2,\ \ P(X=2)=0.3,\ \ \ P(X=3)=0.1.$$

Damit lautet die Trägermenge $T_X=\{0,1,2,3\}$.

Gemäß Notation gilt dann weiter

$$p_1=0.4,\quad p_2=0.2,\quad p_3=0.3,\quad p_4=0.1$$

bzw. $f_X(0)=0.4,\ f_X(1)=0.2,\ f_X(2)=0.3,\ f_X(3)=0.1.$

Das Ereignis „Es reisen höchstens 2 Kinder mit“ wird dann berechnet als

$$\begin{aligned} P(X\le 2) &= P(X=0)+P(X=1)+P(X=2)\\ &= \quad 0.4 \quad + \quad 0.2 \quad + \quad 0.3 \quad = 0.9 \end{aligned}$$

oder alternativ über

$$\begin{aligned} P(X\le 2) &= 1-P(X>2) = 1-P(X=3)\\ &= 1-0.1=0.9 \end{aligned}$$

● Beispiel D1-b ●

Sei Y die Anzahl eingehender Notrufe an einem Rettungswagenstützpunkt während einer Stunde. Beispielsweise könnte dann gelten:

$$f_Y(y) = \begin{cases} \frac{3^y}{y!} e^{-3}, & y = 0, 1, 2, \dots, \\ 0, & sonst. \end{cases}$$

Bemerkung:
W‘funktion einer *Po*(3)-Verteilung
(nähere Hintergründe später)

Beispielsweise gilt dann:

$$P(Y = 0) = \frac{3^0}{0!} e^{-3} = e^{-3} \approx 0.05 \quad \text{oder} \quad P(Y = 2) = \frac{3^2}{2!} e^{-3} \approx 0.22$$

Damit lautet die Trägermenge $T_Y = \mathbb{N}_0$.

Das Ereignis „Es gehen mehr als 8 Notrufe ein“ wird dann berechnet als

$$P(Y > 8) = 1 - P(Y \leq 8) = 1 - \sum_{j=0}^{8} \frac{3^j}{j!} e^{-3} \approx 0.004$$

Stetige Zufallsvariablen und Integrationskalkül

● Wahrscheinlichkeitsdichtefunktion für stetige Zufallsvariablen ●

> Stetige Zufallsvariablen besitzen überabzählbar viele Realisationsmöglichkeiten.

> Typischerweise bilden sie die Ausprägungen stetig skalierter Merkmale ab.

> Wie bereits ausgeführt ist es nicht möglich, jedem einzelnen Elementarereignis eines überzahlbären Ergebnisraumes eine positive Wahrscheinlichkeit zuzuordnen ohne die Kolmogoroff‘sche Axiomatik zu verletzen

> Lösung: Es werden in erster Linie nur noch Intervallereignissen wie

$$\{X \leq b\}, \{X > a\}, \{a \leq X \leq b\} \text{ oder } \{a < X < b\}$$

positive Wahrscheinlichkeiten zugeordnet

> Die konkreten Wahrscheinlichkeiten für diese Ereignisse ergeben sich durch ein **Integrationskalkül** (anstelle eines Additionskalküls). Hierfür verwendet man **Wahrscheinlichkeitsdichtefunktionen**.

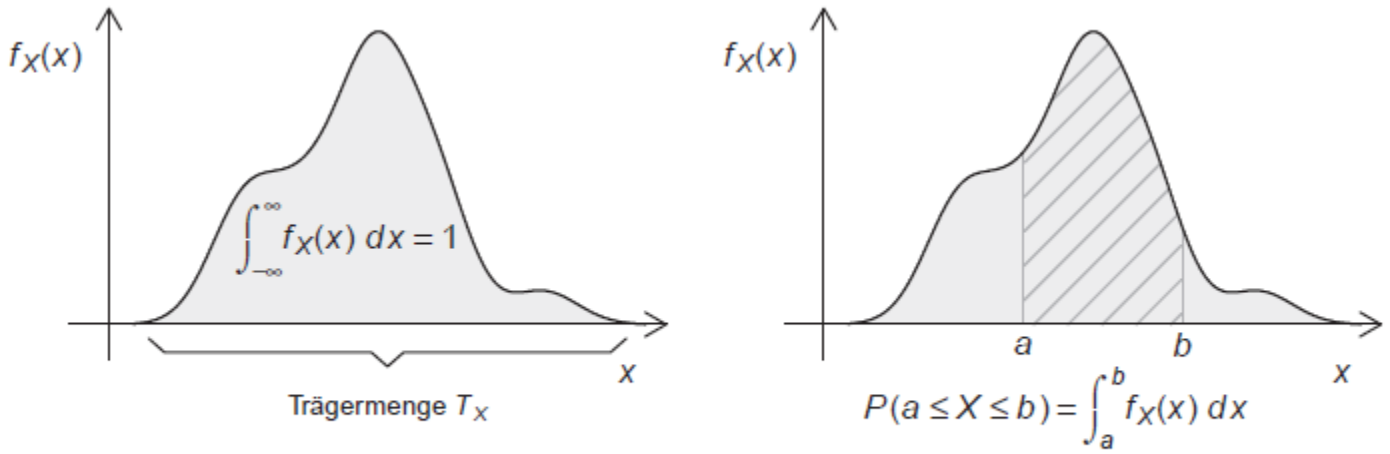


Abb. 7.1.2: Wahrscheinlichkeitsdichtefunktion und Integrationskalkül

Definition 7.1.2: Stetige Zufallsvariable und Wahrscheinlichkeitsdichtefunktion

Eine Zufallsvariable X heißt **stetig**, falls es eine Funktion $f_X(x)$ gibt, für die gilt:

$$P(a \le X \le b) = \int_a^b f_X(x)dx \quad \text{für alle } a \le b, \text{ wobei}$$

(i) $f_X(x) \ge 0$ für alle reellen x gilt und

(ii) $\int_{-\infty}^{\infty} f_X(x)dx = 1$ ist.

Die Funktion f_X heißt dann **Wahrscheinlichkeitsdichtefunktion, Dichtefunktion** oder **Dichte** von X. Die Menge $T_X = \{x : f_X(x) > 0\}$ wird als Trägermenge von X bezeichnet. Ihre Elemente heißen **Realisationsmöglichkeiten** von X.

Dichtewerte sind keine Wahrscheinlichkeiten!

● Integrationskalkül für stetige Zufallsvariablen ●

> Die Wahrscheinlichkeit für jede reelle Zahl ist gleich 0. Dies ergibt sich aus

$$P(X = 1) = P(X \in [a, a]) = \int_a^a f_X(x)dx = 0$$

> Es gilt außerdem stets: $P(X \in T_x) = 1$ und $P(X \in \overline{T_x}) = 0$

> Weiter sind Gleich- und Ungleichzeichen unerheblich, da z. B.

$$P(a < X \leq b) = P(a \leq X \leq b) - P(X = a)$$

$$= P(a \leq X \leq b)$$

Satz 7.1.2: Integrationskalkül für stetige Zufallsvariablen

Sei X eine stetige Zufallsvariable mit Dichtefunktion f_X und $A \subset \mathbb{R}$ eine aus abzählbar vielen Intervallen durch Vereinigung-, Schnitt- und Komplementbildung konstruierte oder konstruierbare Menge. Dann gilt

$$P(X \in A) = \int_A f_X(x)dx.$$

Speziell folgt daraus:

(i) $P(X = a) = 0$ für jedes $a \in \mathbb{R}$ und

(ii) $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$.

• Messbarkeitsproblem •

> Über den Integrationskalkül kann nicht jeder Teilmenge (jedem Ereignis) $A \subset \mathbb{R}$ eine Wahrscheinlichkeit zugeordnet werden

> Streng genommen werden dann auch nur die messbaren Mengen als Ereignisse bezeichnet.

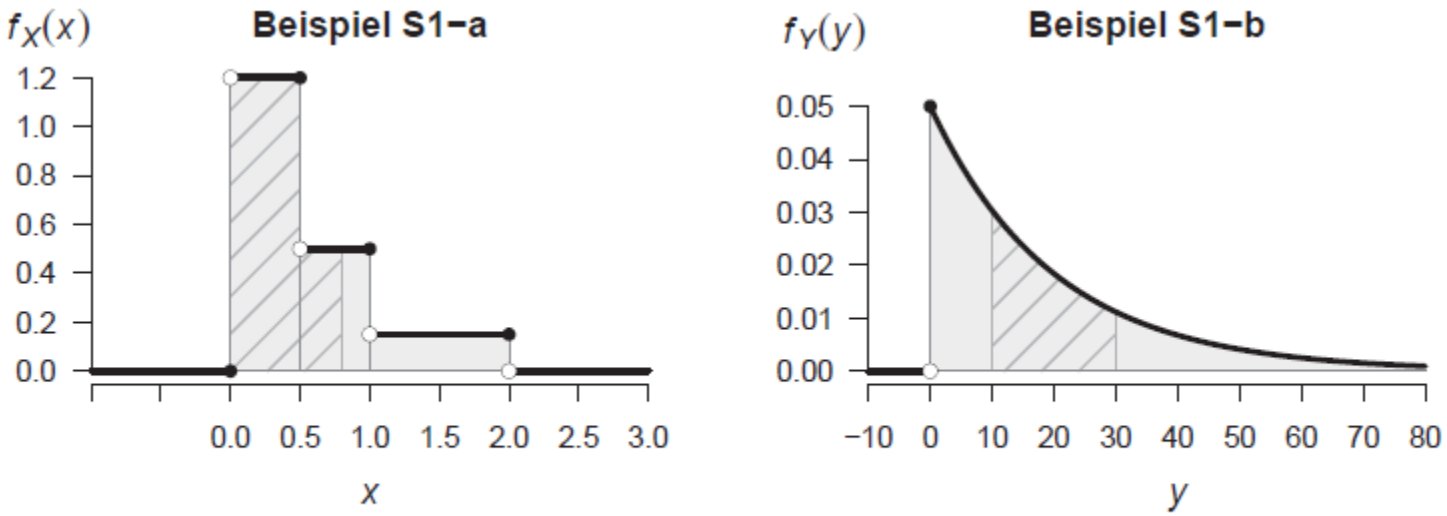


Abb. 7.1.3: Dichtefunktionen stetiger Zufallsvariablen

● Beispiel S1-a ●

Sei X die Höhe des Trinkgeldes, welches Servicekräfte eines Restaurant pro Abrechnung erhalten. Beispielsweise könnte dann gelten:

$$f_X(x) = \begin{cases} 1.2 & x \in (0, 0.5], \\ 0.5, & x \in (0.5, 1.0], \\ 0.15 & x \in (1, 2.0], \\ 0, & sonst. \end{cases}$$

> Konzeptionell kaum Unterschied zur **Häufigkeitsdichtefunktion** (**Histogramm**). Jetzt: Dichtefunktion für Wahrscheinlichkeiten (Wahrscheinlichkeitsdichte statt Häufigkeitsdichte)

> Beispielsweise beträgt der Dichtewert in (0, 0.5] konstant 1.2. Dies bedeutet im vorliegenden Fall, dass 120% Wahrscheinlichkeitsmasse auf einer Einheit (1 Euro) verteilt sind oder 60% Wahrscheinlichkeitsmasse auf einer halben Einheit (50 Cent)

> Der **Intergrationskalkül** kann hier leicht „im Kopf“ berechnet werden. Beispielsweise gilt dann:

$$P(X \leq 1) = \int_0^1 f_X(x)dx = 0.5 \times 1.2 + 0.5 \times 0.5 = 0.85$$

$$P(X \leq 0.8) = \int_0^{0.8} f_X(x)dx = 0.5 \times 1.2 + 0.3 \times 0.5 = 0.75$$

● Beispiel S1-b ●

Sei Y die Wartezeit (in Minuten) bis zum nächsten eingehenden Notruf an einem Rettungswagenstützpunkt. Beispielsweise könnte dann gelten:

$$f_Y(y) = \begin{cases} 0.05e^{-0.05y}, & y \geq 0, \\ 0, & y < 0. \end{cases}$$

Bemerkung:
Dichte einer *Exp*(0.05)-Verteilung
(nähere Hintergründe später)

Beispielsweise gilt dann:

$$P(10 \leq Y \leq 30) = \int_{10}^{30} f_Y(y)dy = \int_{10}^{30} 0.05e^{-0.05y}dy$$

$$= [-e^{-0.05y}]_{10}^{30} = -e^{-0.05\times 30} + e^{-0.05\times 10} \approx 0.38$$

oder

$$P(Y > 80) = 1 - P(Y \leq 80) = 1 - \int_{0}^{80} 0.05e^{-0.05y}dy$$

$$= 1 - [-e^{-0.05y}]_0^{80} = 1 - [-e^{-0.05\times 80} + 1] \approx 0.02$$

● Verwendung von Indikatorfunktionen ●

> Für eine vorgegebene Menge A nimmt die **Indikatorfunktion** $I_A(x)$ für $x \in A$ den Wert 1 an, und sonst den Wert 0, d. h.

$$I_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

> In der Praxis ist A häufig ein Intervall. Damit lassen sich abschnittsweise definierte Funktionen in geschlossener Form notieren.

> Beispiel 1: Dichtefunktion von Beispiel S-1a

$$f_X(x) = 1.2I_{(0,0.5]}(x) + 0.5I_{(0.5,1]}(x) + 0.15I_{(1,2]}(x)$$

Man beachte, dass dann z. B. für $x = 0.8$ gilt (wie es auch sein soll):

$$\begin{aligned} f_X(0.8) &= 1.2I_{(0,0.5]}(0.8) + 0.5I_{(0.5,1]}(0.8) + 0.15I_{(1,2]}(0.8) \\ &= 1.2 \times 0 + 0.5 \times 1 + 0.15 \times 0 = 0.5 \end{aligned}$$

> Beispiel 2: Dichtefunktion von Beispiel S-1b

$$f_Y(y) = 0.05e^{-0.05y}I_{[0,\infty)}(y)$$

Theoretische Verteilungsfunktion

- **Verteilungsfunktion diskreter und stetiger Zufallsvariablen** -

Definition 7.1.3: Theoretische Verteilungsfunktion

Sei X eine diskrete oder stetige Zufallsvariable. Dann ist die (theoretische) Verteilungsfunktion von X definiert als $F_X(x) = P(X \leq x)$. Daraus folgt:

(i) Falls X diskret ist, gilt: $F_X(x) = \sum_{j:a_j \leq x} p_j$.

(ii) Falls X stetig ist, gilt: $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$.

Theoretisches Pendant zur **empirischen Verteilungsfunktion** $F_n(x)$

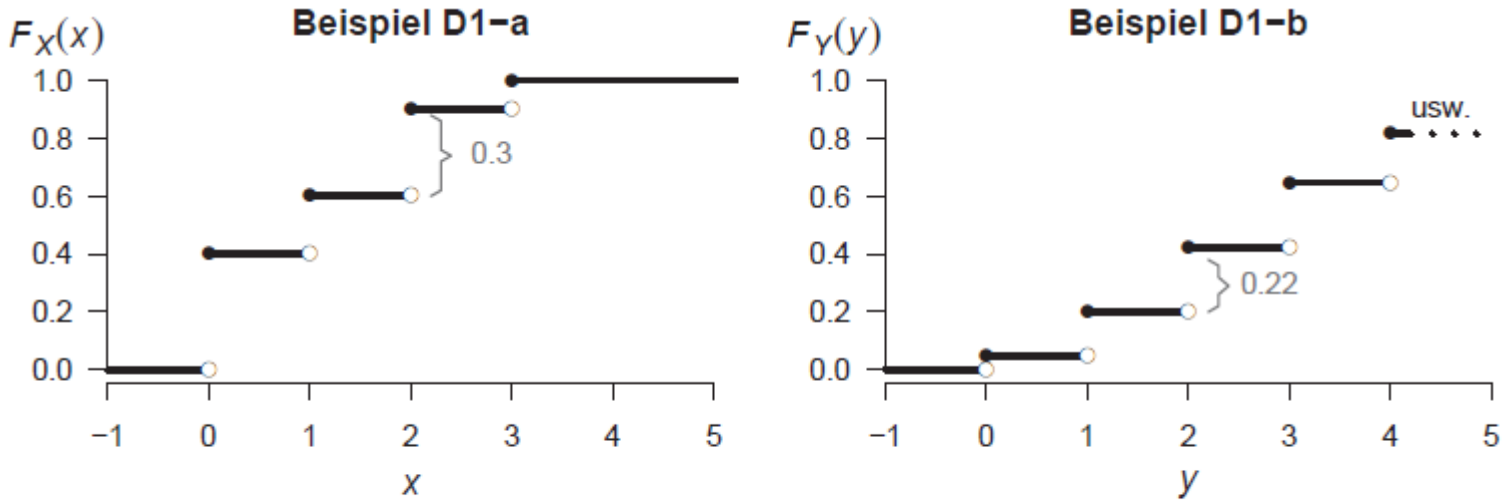


Abb. 7.1.4: Verteilungsfunktionen diskreter Zufallsvariablen

> Bei diskreten Zufallsvariablen ergibt sich die *Vf* durch Kumulieren der Einzelwahrscheinlichkeiten der Wahrscheinlichkeitsfunktion bis zu einem bestimmten Wert *x*.

> Daraus ergibt sich eine monoton steigende (rechtsstetige) Treppenfunktion

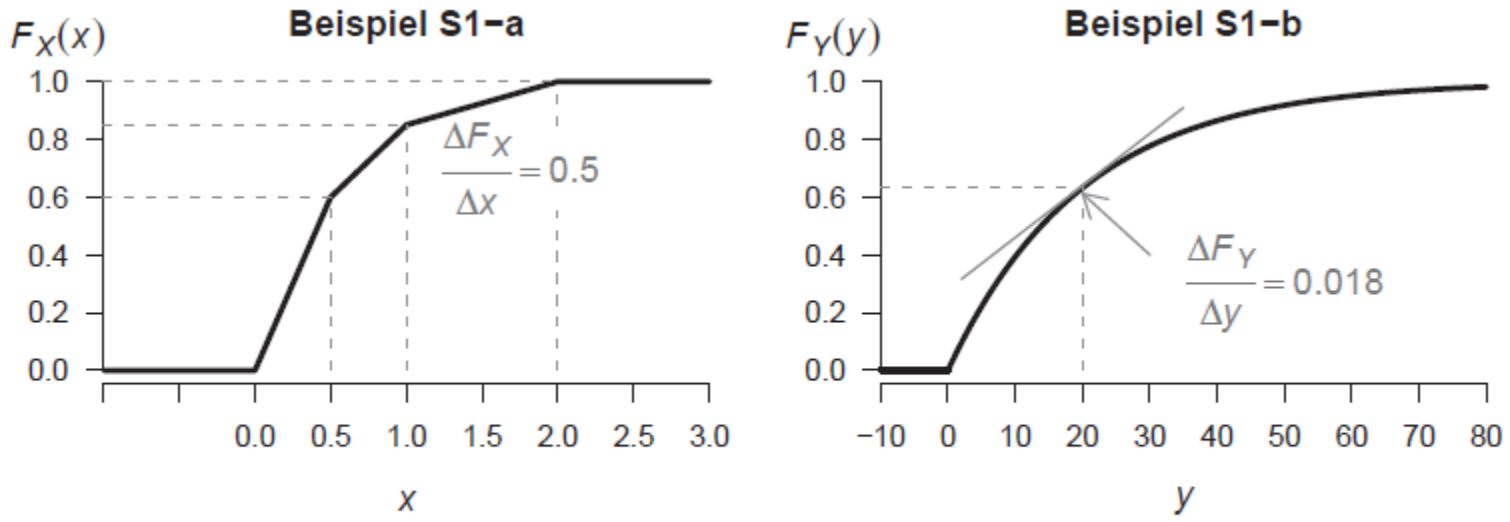


Abb. 7.1.5: Verteilungsfunktionen stetiger Zufallsvariablen

> Bei stetigen Zufallsvariablen ergibt sich die *Vf* durch Integration der Wahrscheinlichkeitsdichtefunktion bis zu einem bestimmten Wert *x*.

> Daraus ergibt sich eine monoton steigende stetige Funktion

> Ist die Dichtefunktion $f(x)$ an der Stelle *x* stetig, so gilt:

$$F'(x) = \frac{dF(x)}{dx} = f(x)$$

Eigenschaften von Verteilungsfunktionen

1. Für die Verteilungsfunktion F_X einer Zufallsvariablen X gilt stets:
   (a) $F_X(x) \in [0, 1]$ für jedes $x \in \mathbb{R}$.
   (b) F_X ist monoton steigend, d.h. $F_X(x_1) \leq F_X(x_2)$ für $x_1 < x_2$.
   (c) F_X ist rechtsstetig, d.h. $F_X(x + \frac{1}{n}) \to F_X(x)$ für $n \to \infty$.
   (d) $F_X(x) \to 0$ für $x \to -\infty$.
   (e) $F_X(x) \to 1$ für $x \to +\infty$.
2. Die Verteilungsfunktion einer diskreten Zufallsvariable ist eine Treppenfunktion mit Sprungstellen an den Realisationsmöglichkeiten und Sprunghöhen, die den jeweiligen Wahrscheinlichkeiten entsprechen.
3. Die Verteilungsfunktion F_X einer stetigen Zufallsvariable X mit Dichtefunktion f_X ist stetig. Sofern F_X an der Stelle x differenzierbar ist, gilt:

$$F'_X(x) = f_X(x).$$

● **Beispiel S1-a fortgesetzt** ●

Aus $f_X(x) = 1.2I_{(0,0.5]}(x) + 0.5I_{(0.5,1]}(x) + 0.15I_{(1,2]}(x)$

erhalten wir

$$F_X(x) = 1.2xI_{(0,0.5]}(x) + (0.5x + 0.35)I_{(0.5,1]}(x)$$
$$+(0.15x + 0.7)I_{(1,2]}(x) + I_{(2,\infty)}(x)$$

Man beachte, dass jeweils die kumulierte Wahrscheinlichkeitsmasse des vorhergehenden Abschnitts „mitgenommen“ werden muss. Deshalb wählt man z. B. als Stammfunktion des 2. Abschnitts (0.5, 1] nicht einfach $0.5x$, sondern $0.5x + 0.35$

VF ist außerhalb der „Knickstellen 0.5, 1 und 2 überall differenzierbar. Beispielsweise gilt für $x \in (0.5, 1]$:

$$F_X'(x) = f_X(x) = 0.5$$

● Beispiel S1-b fortgesetzt ●

Aus $f_Y(y) = 0.05e^{-0.05y} I_{[0,\infty)}(y)$

erhalten wir für

$y < 0$: $F_Y(y) = P(Y \leq y) = 0$

$y \geq 0$: $F_Y(y) = \int_0^y f_Y(t)dt = \int_0^y 0.05e^{-0.05t}dt$

$= [-e^{-0.05t}]_0^y = -e^{-0.05y} - (-e^{-0.05\times 0})$

$= 1 - e^{-0.05y}$

Insgesamt ergibt dies dann:

$$F_Y(y) = (1 - e^{-0.05y}) I_{[0,\infty)}(y)$$

Beispielsweise gilt:

$$F_Y'(20) = f_Y(20) = 0.05e^{-0.05\times 20} \approx 0.018$$

● Äquivalenz und Eindeutigkeit ●

> Diskrete Zufallsvariablen: Informationsgehalt von Wahrscheinlichkeitsfunktion und Verteilungsfunktion äquivalent, d. h. von dem einen kann jeweils auf das andere eindeutig geschlossen werden.

> Stetige Zufallsvariablen: Dichtefunktion und *Vf* „fast äquivalent“: Von Dichte kann eindeutig auf *Vf* geschlossen werden. Die Umkehrung gilt „nicht ganz“. Dichtefunktionen können sich an abzählbar vielen Stellen unterscheiden.

> Beispiel: Zwei „fast gleiche“ Dichtefunktionen für Beispiel S-1a, welche genau die gleiche *Vf* implizieren:

$$f_X(x) = 1.2I_{(0,0.5]}(x) + 0.5I_{(0.5,1]}(x) + 0.15I_{(1,2]}(x)$$

$$f_X^*(x) = 1.2I_{[0,0.5)}(x) + 0.5I_{[0.5,1)}(x) + 0.15I_{[1,2)}(x)$$

Hier gilt z. B. : $f_X(1) = 0.5$ aber $f_X^*(1) = 0.15$

> Dichtefunktion einer stetigen Zufallsvariable muss nicht zwingend stetig sein (vgl. Beispiel S-1a), die zugehörige *Vf* jedoch schon.

Wahrscheinlichkeitsverteilung

• Impliziertes Wahrscheinlichkeitsmaß und Wahrscheinlichkeitsverteilung •

Definition 7.1.4: Wahrscheinlichkeitsverteilung einer Zufallsvariable

Unter der Wahrscheinlichkeitsverteilung einer Zufallsvariable X versteht man das von X implizierte Wahrscheinlichkeitsmaß P_X auf der Menge der reellen Zahlen. Für messbare Mengen $A \subset \mathbb{R}$ gilt stets: $P(X \in A) = P_X(A)$.

• Diskretes und stetiges Wahrscheinlichkeitsmaß auf $\mathbb{R}$ •

> Unabhängig von der Existenz einer Zufallsvariable definiert jede Funktion $f\colon \mathbb{R} \to [0,1]$ eindeutig ein **diskretes W'maß** auf $\mathbb{R}$, falls für abzählbare Werte $a_1, a_2, \ldots, a_N, \ldots$ gilt:

(i) $f(a_i) \geq 0$ für alle i,

(ii) $\sum_{i=1}^{\infty} f(a_i) = 1$

> Unabhängig von der Existenz einer Zufallsvariable definiert jede Funktion $f\colon \mathbb{R} \to [0, \infty)$ eindeutig ein **stetiges W'maß** auf $\mathbb{R}$, falls gilt:

(i) $f(x) \geq 0$,

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

● Was man unter „identisch verteilt" versteht ●

> Zwei diskrete Zufallsvariablen sind genau dann **identisch verteilt**, falls

$$f_X = f_Y \quad \Leftrightarrow \quad F_X = F_Y$$

die Wahrscheinlichkeitsfunktionen oder äquivalent die Verteilungsfunktionen also übereinstimmen.

> Zwei stetige Zufallsvariablen sind genau dann **identisch verteilt**, falls

$$F_X = F_Y$$

Die Übereinstimmung der Dichten ist dazu **nicht** äquivalent (vgl. hierzu Folie 356). Eine „Fast-Übereinstimmung" der Dichten genügt bereits.

> Zusammengefasst (diskret und stetig) erhalten wir die Übereinstimmung der Verteilungsfunktionen als allgemein gültiges Kriterium:

$$X \text{ und } Y \text{ sind identisch verteilt} \quad \Leftrightarrow \quad F_X = F_Y$$

> „Identisch verteilt“ heißt **nicht**, dass X und Y identisch sind (stets identische Zufallsergebnisse liefern). Es bedeutet lediglich, dass X und Y identische Wahrscheinlichkeitsverteilungen besitzen.

> Beispiel: Einmaliger Münzwurf
Definiere (binäre) Zufallsvariablen X und Y mit

$$X = \begin{cases} 1, & falls\ Zahl \\ 0, & falls\ Wappen \end{cases} \qquad Y = \begin{cases} 0, & falls\ Zahl \\ 1, & falls\ Wappen \end{cases}$$

Dann gilt: $Y = 1 - X$. Damit können X und Y (im gleichen Wurf) nicht identische Ergebnisse liefern. Es gilt: $X \neq Y$. Gleichwohl sind X und Y identisch verteilt, da

$$P(X = 0) = P(Y = 0)(= 0.5) \text{ bzw. } P(X = 1) = P(Y = 1)(= 0.5).$$

7.1.2 Mehrdimensionale Zufallsvariablen

Konzept und Ausblick

● Hintergrund ●

Anstelle der Analyse **empirischer Abhängigkeiten** (Unabhängigkeiten)
jetzt Analyse **stochastischer Abhängigkeiten** (Unabhängigkeite)

● Zufallsvektoren ●

Betrachtung von Zufallsvektoren $\mathbf{X}^T = (X_1, X_2, \dots, X_n)$, (X, Y), $(X, Y)^T$ etc.

● Mehrdimensionale Wahrscheinlichkeitsverteilungen ●

Zu berechnende Ereigniswahrscheinlichkeiten lauten jetzt beispielsweise

$P(X_1 \in A_1, X_2 \in A_2)$ für $A_1, A_2 \subset \mathbb{R}$, $P(X_1 = x_1, X_2 = x_2, X_3 = x_3)$

oder auch

$P(X_1 \in A_1 | X_2 \in A_2)$ und $P(X_1 = x_1 | X_2 = x_2, X_3 = x_3)$

Zweidimensionale diskrete Zufallsvariablen

• Gemeinsame diskrete Verteilung und Randverteilungen •

Definition 7.1.5: Gemeinsame diskrete Verteilung

Zwei Zufallsvariablen X und Y sind gemeinsam diskret verteilt, falls es für endlich oder abzählbar unendlich viele 2-Tupel (a_i, b_j) mit $a_i \in \{a_1, a_2, \ldots, a_k, \ldots\}$ und $b_j \in \{b_1, b_2, \ldots, b_l, \ldots\}$ eine Funktion f_{XY} gibt, für die gilt:

$$f_{XY}(x, y) = \begin{cases} P(X = a_i, Y = b_j) = p_{ij}, & (x, y) = (a_i, b_j), \\ 0, & \text{sonst}, \end{cases}$$

wobei

(i) $p_{ij} \geq 0$ für alle i und j und

(ii) $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p_{ij} = 1.$

Die Funktion f_{XY} heißt dann **gemeinsame Wahrscheinlichkeitsfunktion** von X und Y. Sie legt die gemeinsame Verteilung von X und Y fest. Alternativ spricht man auch vom **diskreten Zufallsvektor** (X, Y) mit der (**zweidimensionalen**) **Wahr-**

scheinlichkeitsfunktion f_{XY}. Die Menge $T_{XY} = \{(a_i, b_j) : p_{ij} > 0\}$ wird als Trägermenge von (X, Y) bezeichnet. Ihre Elemente heißen **Realisationsmöglichkeiten, Träger-** oder **Massenpunkte** von (X, Y).

Tab. 7.1.1: Allgemeine Gestalt einer Kontingenztabelle mit Wahrscheinlichkeiten

X \ Y	b_1	b_2	$\dots$	b_j	$\dots$	b_l	$P(X = x)$
a_1	p_{11}	p_{12}	$\dots$	p_{1j}	$\dots$	p_{1l}	$p_{1\bullet}$
a_2	p_{21}	p_{22}	$\dots$	p_{2j}	$\dots$	p_{2l}	$p_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$			$\vdots$	$\vdots$
a_i	p_{i1}	p_{i2}	$\dots$	p_{ij}	$\dots$	p_{il}	$p_{i\bullet}$
$\vdots$	$\vdots$	$\vdots$			$\ddots$	$\vdots$	$\vdots$
a_k	p_{k1}	p_{k2}	$\dots$	p_{kj}	$\dots$	p_{kl}	$p_{k\bullet}$
$P(Y = y)$	$p_{\bullet 1}$	$p_{\bullet 2}$	$\dots$	$p_{\bullet j}$	$\dots$	$p_{\bullet l}$	1

Definition 7.1.6: Diskrete Randverteilungen

Sei (X, Y) ein diskreter Zufallsvektor mit Wahrscheinlichkeitsfunktion $f_{XY}(x, y)$. Dann sind die (eindimensionalen) Wahrscheinlichkeitsfunktionen von X und Y gegeben durch

$$f_X(x) = P(X = x) = \sum_j f_{XY}(x, b_j) \quad \text{bzw.}$$

$$f_Y(y) = P(Y = y) = \sum_i f_{XY}(a_i, y).$$

Die durch f_{XY} mittels f_X und f_Y implizierten Verteilungen von X bzw. Y werden auch als **Randverteilungen** von X bzw. Y bezeichnet.

• Bedingte diskrete Verteilungen •

Definition 7.1.7: Bedingte diskrete Verteilungen

Sei (X, Y) ein diskreter Zufallsvektor mit Wahrscheinlichkeitsfunktion $f_{XY}(x, y)$. Dann ist die **bedingte Wahrscheinlichkeitsfunktion von X gegeben $Y = y$** definiert als

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)} \quad \text{für } f_Y(y) > 0$$

und die **bedingte Wahrscheinlichkeitsfunktion von Y gegeben $X = x$** als

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)} \quad \text{für } f_X(x) > 0.$$

Die durch die bedingten Wahrscheinlichkeitsfunktionen implizierten Verteilungen von X und Y werden als **bedingte Verteilung von X gegeben $Y = y$** bzw. **bedingte Verteilung von Y gegeben $X = x$** bezeichnet.

Für $f_Y(y) = 0$ bzw. $f_X(x) = 0$ sind die bedingten Wahrscheinlichkeitsfunktionen bzw. Verteilungen jeweils nicht definiert.

Bei Kontingenztabelle auch **Zeilen**- bzw. **Spaltenverteilungen**

• Beispiel D2-a •

Tab. 7.1.2: Kontingenztabelle mit Wahrscheinlichkeiten für Beispiel D2-a

X \ Y	0	1	2	Summe
0	0.04	0.08	0.28	0.40
1	0.12	0.12	0.36	0.60
Summe	0.16	0.20	0.64	1.00

• Beispiel D2-b •

Tab. 7.1.5: Kontingenztabelle mit Wahrscheinlichkeiten für Beispiel D2-b

X \ Y	1	2	3	4	5	6	Summe
1	1/36	1/36	1/36	1/36	1/36	1/36	1/6
2	1/36	1/36	1/36	1/36	1/36	1/36	1/6
3	1/36	1/36	1/36	1/36	1/36	1/36	1/6
4	1/36	1/36	1/36	1/36	1/36	1/36	1/6
5	1/36	1/36	1/36	1/36	1/36	1/36	1/6
6	1/36	1/36	1/36	1/36	1/36	1/36	1/6
Summe	1/6	1/6	1/6	1/6	1/6	1/6	1.0

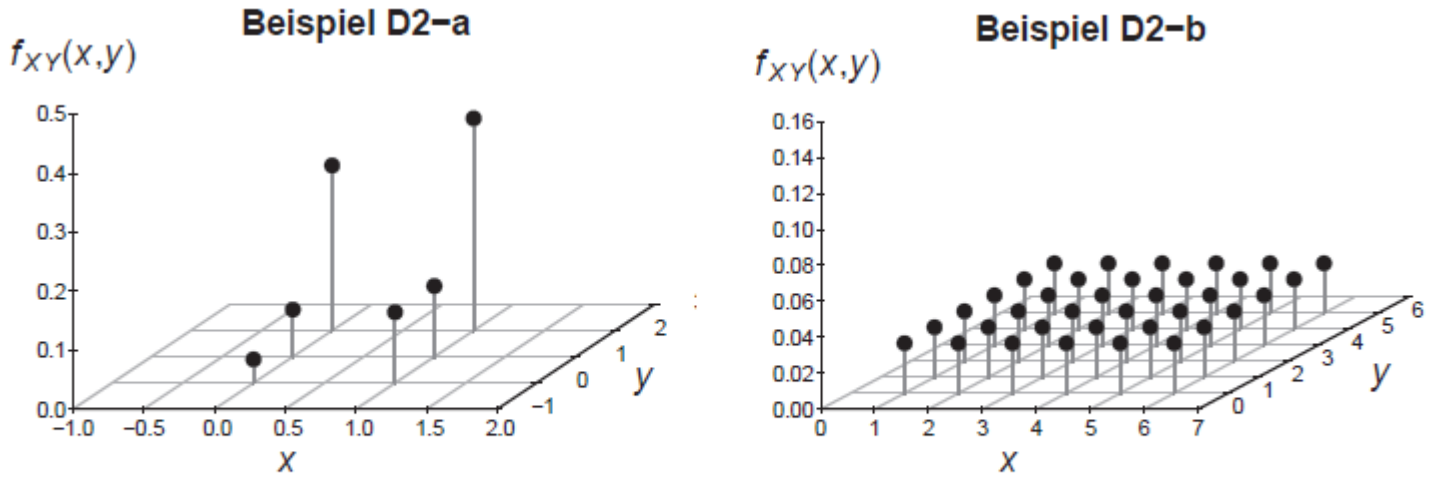


Abb. 7.1.6: Gemeinsame diskrete Verteilungen

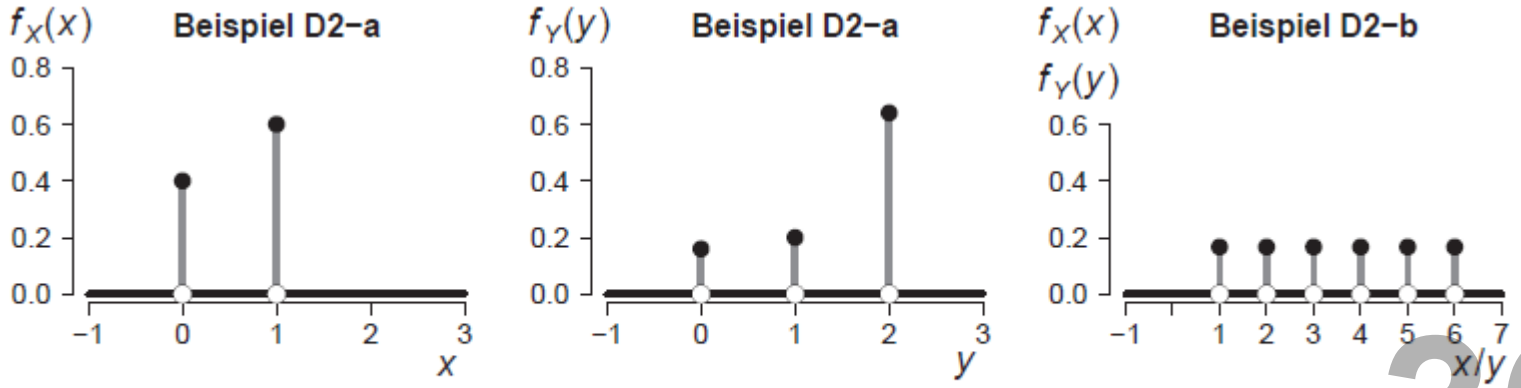


Abb. 7.1.7: Diskrete Randverteilungen

Tab. 7.1.3: Bedingte Verteilung von X unter $Y = y$ für Beispiel D2-a

X \ $Y = y$	$y = 0$	$y = 1$	$y = 2$
0	0.25	0.40	0.4375
1	0.75	0.60	0.5625
Summe	1.00	1.00	1.00

Beispielsweise gilt hier:

$$P(X = 0|Y = 0) = \frac{P(X = 0, Y = 0)}{P(Y = 0)} = \frac{0.04}{0.16} = 0.25$$

bzw. $$f_{X|Y}(0|0) = \frac{f_{XY}(0,0)}{f_Y(0)} = \frac{0.04}{0.16} = 0.25$$

Tab. 7.1.4: Bedingte Verteilung von Y unter $X = x$ für Beispiel D2-a

$X = x$ \ Y	0	1	2	Summe
0	0.1	0.2	0.7	1.0
1	0.2	0.2	0.6	1.0

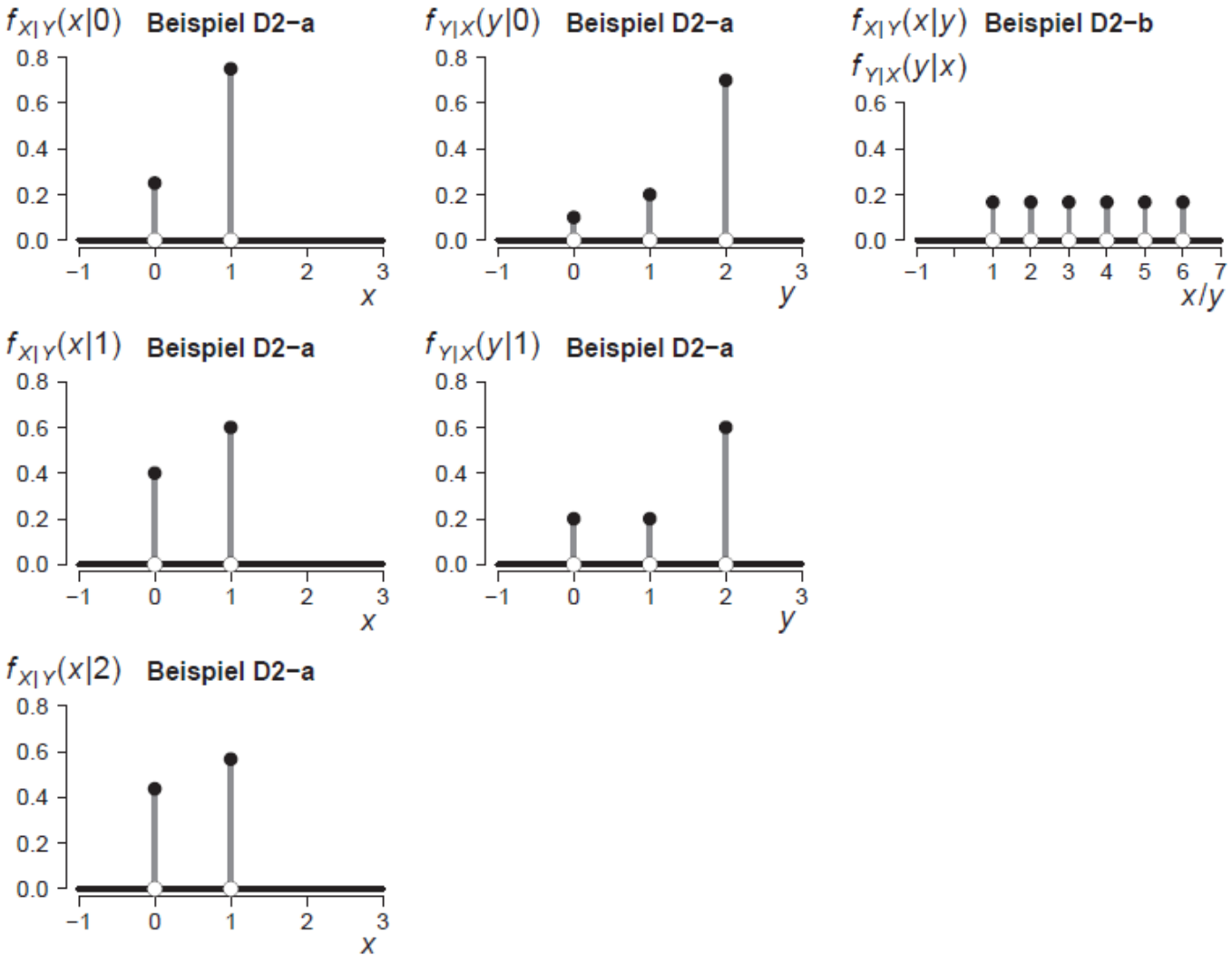


Abb. 7.1.8: Bedingte diskrete Verteilungen

• Zweidimensionale diskrete Verteilungsfunktion •

Zweidimensionale diskrete Verteilungsfunktion

Sei (X, Y) ein diskreter Zufallsvektor mit Wahrscheinlichkeitsfunktion $f_{XY}(x, y)$. Dann ist die Verteilungsfunktion von (X, Y) oder auch gemeinsame Verteilungsfunktion von X und Y gegeben durch

$$F_{XY}(x, y) = P(X \leq x, Y \leq y) = \sum_{i:a_i \leq x} \sum_{j:b_j \leq y} f_{XY}(a_i, b_j),$$

wobei (a_i, b_j) Realisationsmöglichkeiten von (X, Y) sind.

• Beispiel D2-a fortgesetzt •

Beispielsweise gilt dann:

$$F_{XY}(0,0) = 0.04, \quad F_{XY}(0,1) = 0.12, \quad F_{XY}(0,2) = 0.4,$$
$$F_{XY}(1,0) = 0.16, \quad F_{XY}(1,1) = 0.36, \quad F_{XY}(1,2) = 1,$$

$$F_{XY}(0,-0.1) = 0, \quad F_{XY}(0,0.1) = 0.04, \quad F_{XY}(1.2,2.4) = 1,$$
$$F_{XY}(0.8,1.5) = 0.12, \quad F_{XY}(5,0) = 0.16, \quad \text{usw.}$$

Zweidimensionale stetige Zufallsvariablen

• Gemeinsame stetige Verteilung und Randverteilungen •

Definition 7.1.8: Gemeinsame stetige Verteilung

Zwei Zufallsvariablen X und Y sind **gemeinsam stetig verteilt**, falls es eine Funktion f_{XY} gibt, sodass für alle reellen $a \leq b$ und $c \leq d$ gilt:

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f_{XY}(x, y) dy dx,$$

wobei

(i) $f_{XY}(x, y) \geq 0$ und

(ii) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dy dx = 1.$

Die Funktion f_{XY} heißt dann **gemeinsame Wahrscheinlichkeitsdichtefunktion**, **gemeinsame Dichtefunktion** oder **gemeinsame Dichte** von X und Y. Sie legt die **gemeinsame Verteilung** von X und Y fest. Alternativ spricht man auch vom **stetigen Zufallsvektor** (X, Y) mit der **(zweidimensionalen) Dichtefunktion** f_{XY}. Die Menge $T_{XY} = \{(x, y) : f_{XY}(x, y) > 0\}$ wird als **Trägermenge** von (X, Y) bezeichnet. Ihre Elemente bilden **Realisationsmöglichkeiten** von (X, Y).

Definition 7.1.9: Stetige Randverteilungen

Sei (X, Y) ein stetiger Zufallsvektor mit Dichtefunktion $f_{XY}(x, y)$. Dann sind die (eindimensionalen) Dichtefunktionen von X und Y gegeben durch

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)dy \quad \text{bzw.}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)dx.$$

Die durch f_{XY} mittels der Randdichten f_X und f_Y implizierten Verteilungen von X bzw. Y werden auch als Randverteilungen von X bzw. Y bezeichnet.

• Bedingte stetige Verteilungen •

Definition 7.1.10: Bedingte stetige Verteilungen

Sei (X, Y) ein stetiger Zufallsvektor mit Dichtefunktion $f_{XY}(x, y)$. Dann ist die **bedingte Dichte von X gegeben $Y = y$** definiert als

$$f_{X|Y} = \frac{f_{XY}(x, y)}{f_Y(y)} \quad \text{für } f_Y(y) > 0$$

und die **bedingte Dichte von Y gegeben $X = x$** als

$$f_{Y|X} = \frac{f_{XY}(x, y)}{f_X(x)} \quad \text{für } f_X(x) > 0.$$

Die durch die bedingten Dichten implizierten Verteilungen von X und Y werden als **bedingte Verteilung von X gegeben $Y = y$** bzw. **bedingte Verteilung von Y gegeben $X = x$** bezeichnet.

Für $f_Y(y) = 0$ bzw. $f_X(x) = 0$ sind die bedingten Dichten bzw. Verteilungen jeweils nicht definiert.

● Beispiel S2-a ●

> Seien X und Y gemeinsam stetig verteilt gemäß der Dichtefunktion

$$f_{XY}(x,y) = (0.5x + 0.25y)I_{[0,1]}(x)I_{[0,2]}(y)$$

> Beispielrechnung für eine Ereigniswahrscheinlichkeit:

$$P(0.1 \leq X \leq 0.5, 0.4 \leq Y \leq 0.9) = \int_{0.1}^{0.5}\int_{0.4}^{0.9} f_{XY}(x,y)\,dydx$$

$$= \int_{0.1}^{0.5}\int_{0.4}^{0.9} (0.5x + 0.25y)I_{[0,1]}(x)I_{[0,2]}(y)\,dydx$$

$$= \int_{0.1}^{0.5} I_{[0,1]}(x)\left(\int_{0.4}^{0.9} (0.5x + 0.25y)I_{[0,2]}(y)\,dy\right)dx$$

$$= \int_{0.1}^{0.5} I_{[0,1]}(x)[0.5xy + 0.125y^2]_{y=0.4}^{y=0.9} dx$$

$$= \int_{0.1}^{0.5} (0.45x + 0.125 \times 0.9^2 - 0.2x - 0.125 \times 0.4^2) dx$$

$$= \int_{0.1}^{0.5} (0.25x + 0.08125) dx = [0.125x^2 + 0.08125x]_{0.1}^{0.5}$$

$$= (0.03125 + 0.040625 - 0.00125 - 0.008125)$$

$$= 0.0625$$

> Die Randdichte von X errechnet sich als

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy$$

$$= \int_{-\infty}^{\infty} (0.5x + 0.25y) I_{[0,1]}(x) I_{[0,2]}(y)\, dy$$

$$= I_{[0,1]}(x) \int_{0}^{2} (0.5x + 0.25y)\, dy$$

$$= I_{[0,1]}(x) [0.5xy + 0.125y^2]_{y=0}^{y=2}$$

$$= I_{[0,1]}(x)(x + 0.5 - 0)$$

$$= (x + 0.5) I_{[0,1]}(x)$$

> Für die Randdichte von Y erhält man analog

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)dx = \cdots = (0.25y + 0.25)I_{[0,2]}(y)$$

> Die bedingte Dichte von X gegeben $Y = y$ lautet

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

$$= \frac{(0.5x + 0.25y)I_{[0,1]}(x)I_{[0,2]}(y)}{(0.25y + 0.25)I_{[0,2]}(y)}$$

$$= \frac{2x + y}{y + 1} I_{[0,1]}(x) \quad \text{für} \quad y \in [0,2]$$

Speziell ergibt sich daraus dann z. B. als

- bedingte Dichte von X gegeben $Y = 0$

$$f_{X|Y}(x|0) = \frac{2x+0}{0+1} I_{[0,1]}(x) = 2x I_{[0,1]}(x)$$

- bedingte Dichte von X gegeben $Y = 1$

$$f_{X|Y}(x|1) = \frac{2x+1}{1+1} I_{[0,1]}(x) = (x+0.5) I_{[0,1]}(x)$$

- bedingte Dichte von X gegeben $Y = 2$

$$f_{X|Y}(x|2) = \frac{2x+2}{2+1} I_{[0,1]}(x) = \left(\frac{2}{3}x + \frac{2}{3}\right) I_{[0,1]}(x)$$

> Für die bedingte Dichte von Y gegeben $X = x$ erhält man analog

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

$$= \frac{(0.5x + 0.25y)I_{[0,1]}(x)I_{[0,2]}(y)}{(x + 0.5)I_{[0,1]}(x)}$$

$$= \frac{0.5x + 0.25y}{x + 0.5} I_{[0,2]}(y) \quad \text{für } x \in [0,1]$$

Speziell ergibt sich daraus dann z. B. als

$$f_{Y|X}(y|0) = \frac{0 + 0.25y}{0 + 0.5} I_{[0,2]}(y) = 0.5y I_{[0,2]}(y)$$

$$f_{Y|X}(y|0.5) = \frac{0.25 + 0.25y}{0.5 + 0.5} I_{[0,2]}(y) = (0.25y + 0.25) I_{[0,2]}(y)$$

$$f_{Y|X}(y|1) = \frac{0.5 + 0.25y}{1 + 0.5} I_{[0,2]}(y) = \left(\frac{1}{6}y + \frac{1}{3}\right) I_{[0,2]}(y)$$

● Beispiel S2-b ●

> Seien X und Y gemeinsam stetig verteilt gemäß der Dichtefunktion

$$f_{XY}(x,y) = I_{[0,1]}(x) I_{[0,1]}(y)$$

> Beispielrechnung für eine Ereigniswahrscheinlichkeit:

$$P(0.1 \leq X \leq 0.5, 0.4 \leq Y \leq 0.9) = (0.5 - 0.1)(0.9 - 0.4) = 0.2$$
$$= (0.5 - 0.1)(0.9 - 0.4) = 0.2$$

Berechenbar „im Kopf“! Warum?

Formale Rechnung:

$$P(0.1 \leq X \leq 0.5, 0.4 \leq Y \leq 0.9) = \int_{0.1}^{0.5} \int_{0.4}^{0.9} f_{XY}(x,y)\, dy dx$$

$$= \int_{0.1}^{0.5} \int_{0.4}^{0.9} I_{[0,1]}(x) I_{[0,1]}(y)\, dy dx$$

$$= \int_{0.1}^{0.5} I_{[0,1]}(x) \left(\int_{0.4}^{0.9} 1\, dy \right) dx \quad = \int_{0.1}^{0.5} I_{[0,1]}(x) [y]_{y=0.4}^{y=0.9} dx$$

$$= \int_{0.1}^{0.5} (0.9 - 0.4) dx \quad = (0.9 - 0.4)[x]_{0.1}^{0.5}$$

$$= (0.5 - 0.1)(0.9 - 0.4)$$

$$= 0.2$$

> Die Randdichte von X errechnet sich als

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy = \int_{-\infty}^{\infty} I_{[0,1]}(x)I_{[0,1]}(y)\,dy$$

$$= I_{[0,1]}(x)\int_{0}^{1} 1\,dy = I_{[0,1]}(x)[y]_0^1$$

$$= I_{[0,1]}(x)(1-0) = I_{[0,1]}(x)$$

> Für die Randdichte von Y erhält man analog

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)dx = \cdots = I_{[0,1]}(y)$$

> Die bedingte Dichte von X gegeben $Y = y$ lautet

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{I_{[0,1]}(x) I_{[0,1]}(y)}{I_{[0,1]}(y)}$$

$$= I_{[0,1]}(x) \quad \text{für} \quad y \in [0,1]$$

> Für die bedingte Dichte von Y gegeben $X = x$ erhält man analog

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{I_{[0,1]}(x) I_{[0,1]}(y)}{I_{[0,1]}(x)}$$

$$= I_{[0,1]}(y) \quad \text{für} \quad x \in [0,1]$$

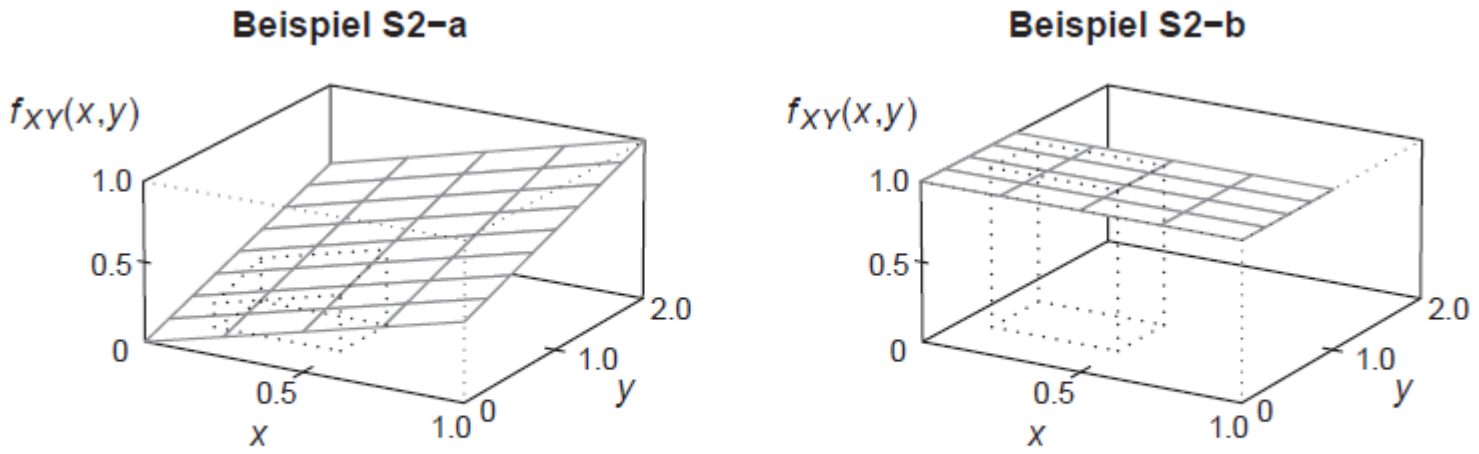


Abb. 7.1.9: Gemeinsame stetige Verteilungen

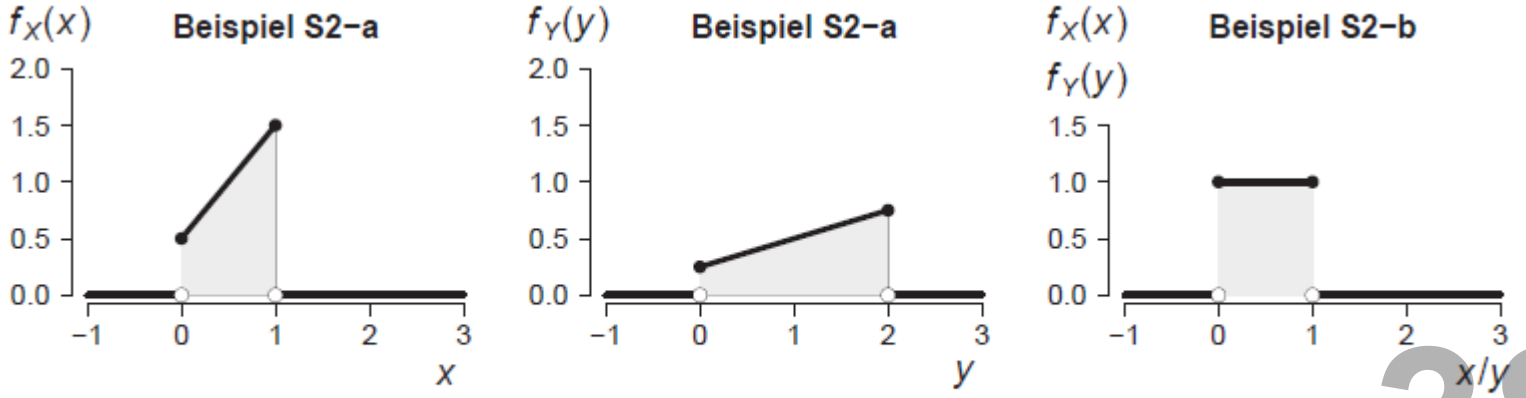


Abb. 7.1.10: Stetige Randverteilungen

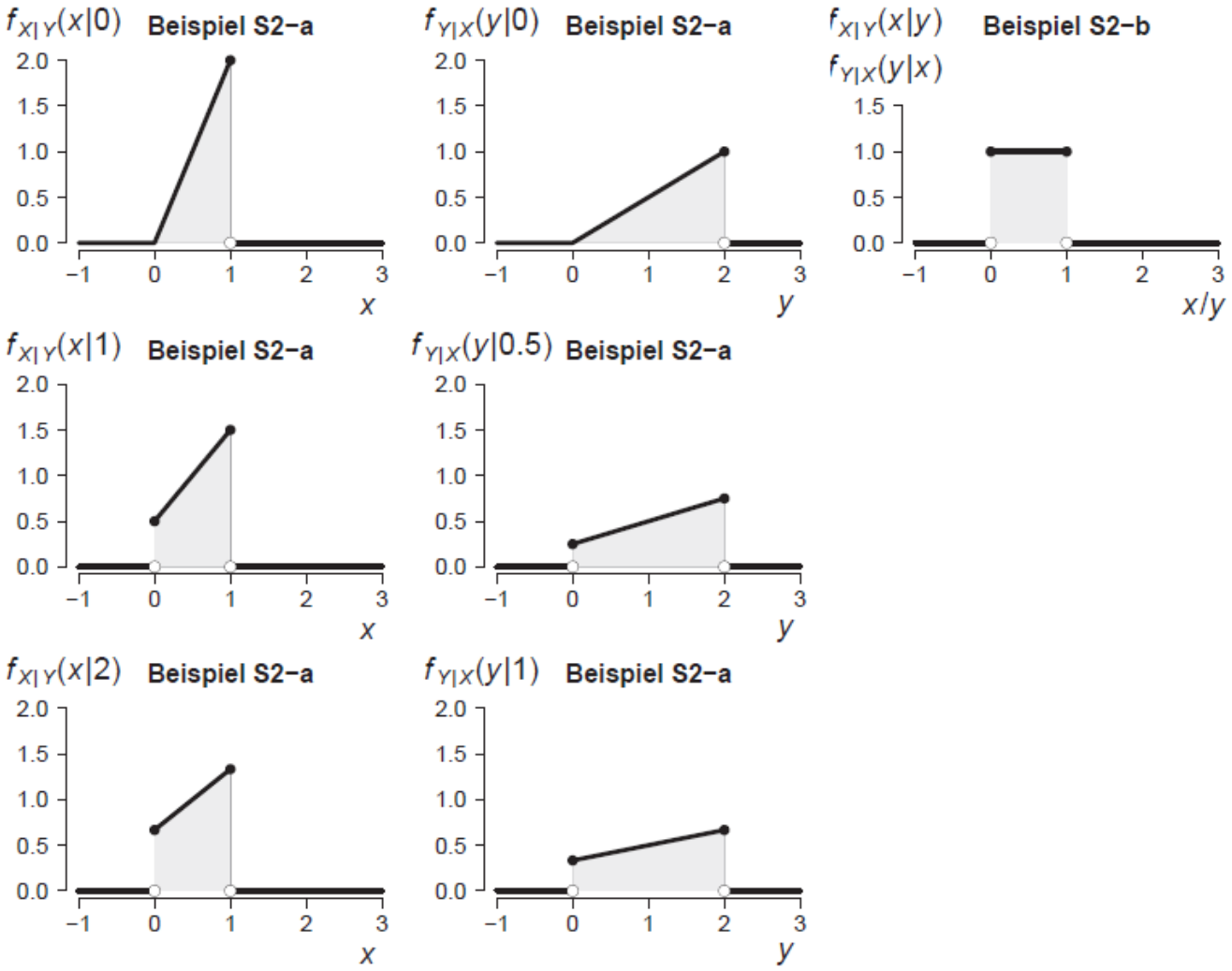


Abb. 7.1.11: Bedingte stetige Verteilungen (Beispiele)

● Zweidimensionale stetige Verteilungsfunktion ●

Definition 7.1.11: Zweidimensionale stetige Verteilungsfunktion

Sei (X, Y) ein stetiger Zufallsvektor mit Dichtefunktion $f_{XY}(x, y)$. Dann ist die Verteilungsfunktion von (X, Y) oder auch gemeinsame Verteilungsfunktion von X und Y gegeben durch

$$F_{XY}(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(u, v) dv du.$$

Ist $F_{XY}(x, y)$ in (x, y) differenzierbar, so gilt:

$$\frac{\partial^2 F_{XY}}{\partial x \partial y}(x, y) = f_{XY}(x, y).$$

● Beispiel S2-a fortgesetzt ●

Siehe Lehrbuch S. 304

Stochastische Abhängigkeit und Unabhängigkeit

• Formale Unabhängigkeitskriterien •

Definition 7.1.12: Stochastische Unabhängigkeit und Abhängigkeit

Gegeben seien zwei Zufallsvariablen X und Y (diskret oder stetig). Dann sind folgende Aussagen äquivalent.

(i) X und Y sind **stochastisch unabhängig**.

(ii) $f_{X|Y}(x|y) = f_X(x)$ für alle $x, y \in \mathbb{R}$ mit $f_Y(y) > 0$.

(iii) $f_{Y|X}(y|x) = f_Y(y)$ für alle $x, y \in \mathbb{R}$ mit $f_X(x) > 0$.

(iv) $f_{XY}(x, y) = f_X(x) f_Y(y)$ für alle $x, y \in \mathbb{R}$.

Kriterium (iv) ist das sog. **Multiplikationskriterium** gemäß dem sich die gemeinsame Verteilung aus dem Produkt der Randverteilungen ergibt. Trifft eine der vier Aussagen nicht zu, sind X und Y **stochastisch abhängig**.

- **Beispiel D2-a fortgesetzt** ⇒ abhängig
- **Beispiel D2-b fortgesetzt** ⇒ unabhängig
- **Beispiel S2-a fortgesetzt** ⇒ abhängig
- **Beispiel S2-b fortgesetzt** ⇒ unabhängig

- **Die Beziehung zwischen „unabhängig“ und „identisch verteilt“**

Die Beziehung zwischen „unabhängig“ und „identisch verteilt“

Zwei Zufallsvariablen können stochastisch unabhängig sein und dabei sowohl identisch als auch nicht identisch verteilt sein. Genauso können zwei Zufallsvariablen stochastisch abhängig sein und dabei sowohl identisch als auch nicht identisch verteilt sein.

vgl. hierzu das Beispiel von Folie 358:
X und Y sind („stark“) abhängig, jedoch identisch verteilt

• Implizierte Unabhängigkeit von Ereignissen •

Satz 7.1.3: Unabhängige Zufallsvariablen implizieren unabhängige Ereignisse

Sind X und Y unabhängige Zufallsvariablen und A und B beliebige (messbare) Teilmengen von $\mathbb{R}$, dann gilt

(i) $P(X \in A | Y \in B) = P(X \in A)$, falls $P(Y \in B) > 0$.

(ii) $P(Y \in B | X \in A) = P(Y \in B)$, falls $P(X \in A) > 0$.

(iii) $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$.

Beispiel:
Betrachte zwei Ereignisse $A = (-\infty. 2]$ und $B = [1.2, 2.8]$ und zwei Zufallsvariablen X und Y, die gemeinsam stetig verteilt sind. Dann gilt:

$$P(X \in A, Y \in B) = \int_{-\infty}^{2}\int_{1.2}^{2.8} f_{XY}(x,y)\,dydx = \int_{-\infty}^{2}\int_{1.2}^{2.8} f_X(x)f_Y(y)\,dydx$$

$$= \int_{-\infty}^{2} f_X(x)\left(\int_{1.2}^{2.8} f_Y(y)\,dy\right)dx$$

$$= \int_{-\infty}^{2} f_X(x)P(Y \in B)dx = \left(\int_{-\infty}^{2} f_X(x)dx\right)P(Y \in B)$$

$$= P(X \in A)P(Y \in B)$$

Höherdimensionale Wahrscheinlichkeitsverteilungen

● Allgemeines ●

Alle bisherigen Konzepte „in naheliegender Weise“ auf den *n*-dimensionalen Fall erweiterbar. Nur wenige neue Aspekte kommen hinzu.

● *n*-dimensionale gemeinsame Verteilungen ●

n-dimensionale Wahrscheinlichkeits- oder Dichtefunktion

$$f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) \qquad \text{(einheitlich notiert)}$$

Diskreter Fall:

$$f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

Stetiger Fall:

$$P(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2, \dots, a_n \leq X_n \leq b_n)$$

$$= \int_{a_1}^{b_1} \int_{a_2}^{b_2} \dots \int_{a_n}^{b_n} f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n)\, dx_n \dots dx_2 dx_1$$

• Randverteilungen •

Jetzt nicht mehr nur eindimensionale Randverteilungen wie

$$f_{X_1}(x_1), f_{X_2}(x_2), \dots, f_{X_n}(x_n)$$

sondern auch mehrdimensionale Randverteilungen wie z. B.

$$f_{X_1 X_2}(x_1, x_2), f_{X_1 X_3}(x_1, x_3) \text{ und } f_{X_2 X_3}(x_2, x_3)$$

als zweidimensionale Randverteilungen von $f_{X_1 X_2 X_3}(x_1, x_2, x_3)$

● Bedingte Verteilungen ●

> Allgemein: Division einer gemeinsamen Verteilung durch eine ein- oder mehrdimensionale (bedingende) Randverteilung

> Bedingte Verteilung von X_1 gegeben $X_2 = x_2, X_3 = x_3, \dots, X_n = x_n$:

$$f_{X_1|X_2X_3\dots X_n}(x_1|x_2, x_3, \dots, x_n) = \frac{f_{X_1X_2\dots X_n}(x_1, x_2, \dots, x_n)}{f_{X_2X_3\dots X_n}(x_2, x_3, \dots, x_n)}$$

> Bedingte Verteilung von $(X_1, X_2)^T$ gegeben $X_3 = x_3$:

$$f_{X_1X_2|X_3}(x_1, x_2|x_3) = \frac{f_{X_1X_2X_3}(x_1, x_2, x_3)}{f_{X_3}(x_3)}$$

> Bedingte Verteilung von $(X_1, X_2)^T$ gegeben $X_3 = x_3$ und $X_4 = x_4$

$$f_{X_1X_2|X_3X_4}(x_1, x_2|x_3, x_4) = \frac{f_{X_1X_2X_3X_4}(x_1, x_2, x_3, x_4)}{f_{X_3X_4}(x_3, x_4)}$$

● Stochastische Unabhängigkeit mehrerer Zufallsvariablen ●

Definition 7.1.13: Stochastische Unabhängigkeit mehrerer Zufallsvariablen

Die Zufallsvariablen $X_1, \ldots, X_n$ (diskret oder stetig) sind (**vollständig**) **stochastisch unabhängig**, falls deren gemeinsame Verteilung dem Produkt der eindimensionalen Randverteilungen entspricht, d.h. falls für alle $x_1, x_2, \ldots, x_n \in \mathbb{R}$ gilt:

$$f_{X_1 X_2 \ldots X_n}(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \ldots f_{X_n}(x_n).$$

Trifft dieses Multiplikationskriterium nicht zu, sind sie **stochastisch abhängig**.

> Lediglich paarweise Unabhängigkeitsbetrachtungen nicht ausreichend

> Man betrachte hierzu Beispiel 6.2.5 (Folie 322)

Formulierung nun unter Verwendung von 3 Zufallsvariablen:

X_1, X_1 und X_3 sind paarweise unabhängig, jedoch nicht vollständig unabhängig, da beispielsweise gilt (vgl. Folie 393):

$$P(X_3 = 1 | X_1 = 1, X_2 = 1) = 0.75 \neq P(X_3 = 1) = 0.5$$

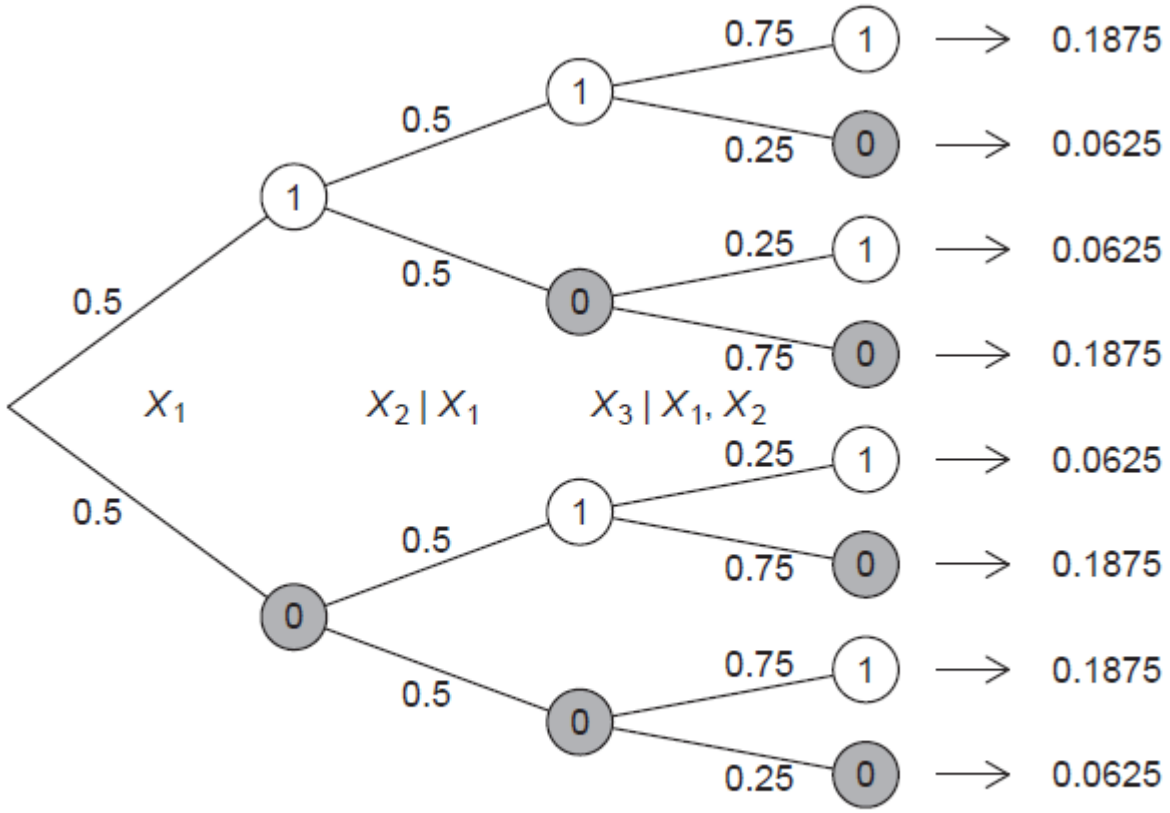


Abb. 7.1.12: Paarweise, aber nicht vollständig unabhängige Zufallsvariablen

• Implizierte Unabhängigkeiten •

Satz 7.1.4: Implizierte Unabhängigkeiten

Sind die Zufallsvariablen $X_1, \ldots, X_n$ (vollständig) unabhängig, so auch jede kleinere Teilauswahl aus diesen. Insbesondere folgt daraus die **paarweise Unabhängigkeit.**

Unabhängige Zufallsvariablen implizieren unabhängige Ereignisse. Für beliebige (messbare) Teilmengen $A_1, A_2, \ldots, A_n$ von $\mathbb{R}$ gilt deshalb stets:

$$P(X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n) = P(X_1 \in A_1)P(X_2 \in A_2)\ldots P(X_n \in A_n).$$

Im stetigen Fall ergibt sich dies aus

$$f_{X_1X_2}(x_1, x_2) = \int_{-\infty}^{\infty} f_{X_1X_2X_3}(x_1, x_2, x_3)dx_3 = \int_{-\infty}^{\infty} f_{X_1}(x_1)f_{X_2}(x_2)f_{X_3}(x_3)dx_3$$

$$= f_{X_1}(x_1)f_{X_2}(x_2)\int_{-\infty}^{\infty} f_{X_3}(x_3)dx_3 = f_{X_1}(x_1)f_{X_2}(x_2)$$

• Beispiel D3-a •

$X = 0$ für weiblich
$X = 1$ für männlich

$Y = 0$ für Raucher
$Y = 1$ für Gelegenheitsraucher
$Y = 2$ für Nichtraucher

$Z = 0$ für nichtvegetarisch
$Z = 1$ für vegetarisch

Tab. 7.1.6: Gemeinsame Verteilung von Beispiel D3-a

$Z = 0$

X \ Y	0	1	2	Summe
0	0.04	0.07	0.26	0.37
1	0.11	0.10	0.32	0.53
Summe	0.15	0.17	0.58	**0.90**

$Z = 1$

X \ Y	0	1	2	Summe
0	0.00	0.01	0.02	0.03
1	0.01	0.02	0.04	0.07
Summe	0.01	0.03	0.06	**0.10**

Tab. 7.1.7: Tupel-Darstellung für die gemeinsame Verteilung von Beispiel D3-a

Tupel	Wahrsch.	Tupel	Wahrsch.
(0, 0, 0)	0.04	(0, 0, 1)	0.00
(0, 1, 0)	0.07	(0, 1, 1)	0.01
(0, 2, 0)	0.26	(0, 2, 1)	0.02
(1, 0, 0)	0.11	(1, 0, 1)	0.01
(1, 1, 0)	0.10	(1, 1, 1)	0.02
(1, 2, 0)	0.32	(1, 2, 1)	0.04

Tab. 7.1.8: Zweidimensionale Randverteilungen von Beispiel D3-a

X \ Z	0	1	$P(X = x)$
0	0.37	0.03	0.40
1	0.53	0.07	0.60
$P(Z = z)$	0.90	0.10	1.00

Z \ Y	0	1	2	$P(Z = z)$
0	0.15	0.17	0.58	0.90
1	0.01	0.03	0.06	0.10
$P(Y = y)$	0.16	0.20	0.64	1.00

Tab. 7.1.9: Bedingte Verteilung von Z gegeben $X = x$ und $Y = y$

Z \ (x, y)	(0,0)	(0,1)	(0,2)	(1,0)	(1,1)	(1,2)
0	1.0	0.875	0.93	0.92	0.83	0.89
1	0.0	0.125	0.07	0.08	0.17	0.11
Summe	1.0	1.00	1.00	1.00	1.00	1.00

$$f_{Z|XY}(z\,|x,y) = \frac{f_{XYZ}(x,y,z\,)}{f_{XY}(\,x,y)}$$

Tab. 7.1.10: Bedingte Verteilung von $(X, Y)^T$ gegeben $Z = z$ von Beispiel D3-a

z=0 Y / X	0	1	2	Σ
0	0.04	0.08	0.29	0.41
1	0.12	0.11	0.36	0.59
Σ	0.16	0.19	0.65	1.00

z=1 Y / X	0	1	2	Σ
0	0.00	0.10	0.20	0.30
1	0.10	0.20	0.40	0.70
Σ	0.10	0.30	0.60	1.00

$$f_{XY|Z}(x,y\,|z) = \frac{f_{XYZ}(x,y,z\,)}{f_{Z}(\,z)}$$

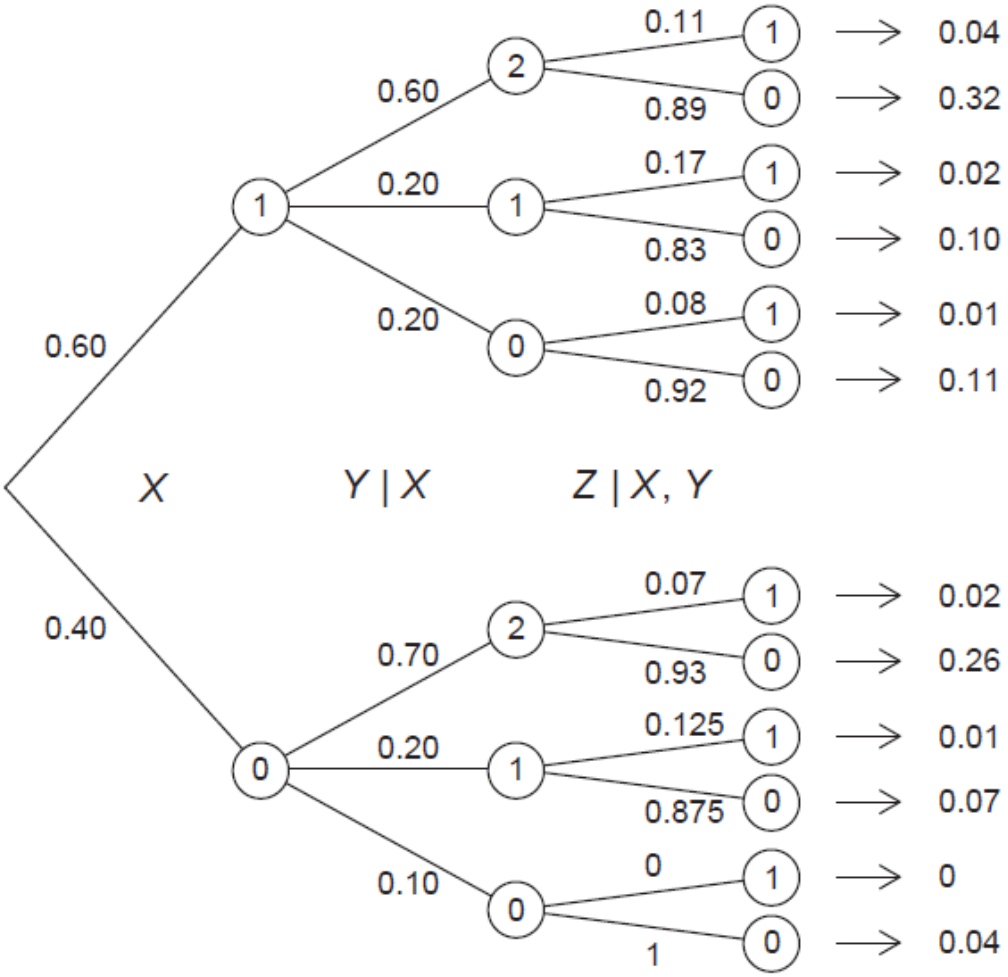


Abb. 7.1.13: Wahrscheinlichkeitsbaum für Beispiel D3-a

● Beispiel S3-a ●

> Seien X, Y und Z gemeinsam stetig verteilt gemäß Dichte

$$f_{XYZ}(x,y,z) = (0.5x + 0.5yz) I_{[0,1]}(x) I_{[0,2]}(y) I_{[0,1]}(z)$$

> Beispielrechnung für eine Ereigniswahrscheinlichkeit:

$$P(0 \le X \le 0.5, 0 \le Y \le 1, 0 \le Z \le 0.5)$$

$$= \int_0^{0.5} \int_0^{1} \int_0^{0.5} f_{XYZ}(x,y,z)\, dz dy dx = \cdots = 0.046875$$

> Die Randdichte von Z errechnet sich als

$$f_Z(z) = \int_0^{1} \int_0^{2} f_{XYZ}(x,y,z)\, dy dx = \cdots = (z + 0.5) I_{[0,1]}(z)$$

> Die zweidimensionale Dichte von X und Y errechnet sich als

$$f_{XY}(x,y) = \int_0^1 f_{XYZ}(x,y,z)\,dz = \cdots = (0.5x + 0.25y) I_{[0,1]}(x) I_{[0,2]}(y)$$

> Die bedingte Dichte von Z gegeben $X = x$ und $Y = y$ errechnet sich als

$$f_{Z|XY}(z\,|x,y) = \frac{f_{XYZ}(x,y,z)}{f_{XY}(x,y)} = \cdots = \frac{x + yz}{x + 0.5y} I_{[0,1]}(z)$$

für $x \in [0,1]$ und $y \in [0,2]$

Beispielsweise gilt dann:

$$f_{Z|XY}(z\,|0,2) = 2z I_{[0,1]}(z)$$

oder $$f_{Z|XY}(z\,|1,1) = \left(\frac{2}{3}z + \frac{2}{3}\right) I_{[0,1]}(z)$$

> Die bedingte Dichte von $(X, Y)^T$ gegeben $Z = z$ errechnet sich als

$$f_{XY|Z}(x, y\,|z) = \frac{f_{XYZ}(x, y, z\,)}{f_Z(z)} = \cdots = \frac{x + yz}{2z + 1} I_{[0,1]}(x) I_{[0,2]}(y)$$

für $z \in [0, 1]$

Beispielsweise gilt dann:

$$f_{XY|Z}(x, y\,|0) = x I_{[0,1]}(z) I_{[0,2]}(y)$$

oder $$f_{XY|Z}(x, y\,|1) = \frac{x + y}{3} I_{[0,1]}(x) I_{[0,2]}(y)$$

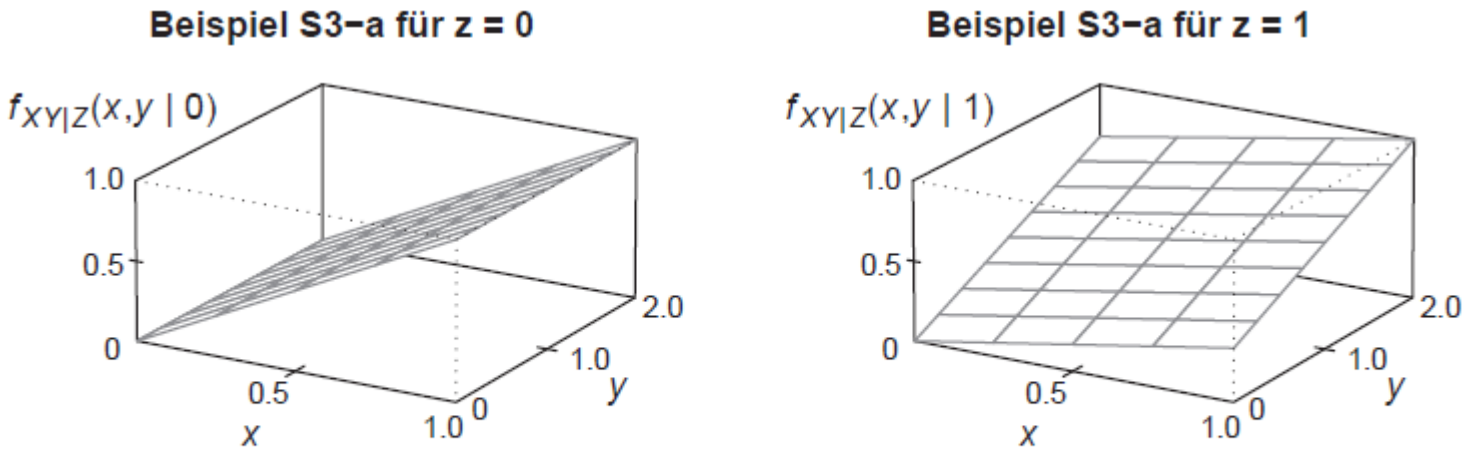


Abb. 7.1.14: Bedingte Verteilung von $(X, Y)^T$ gegeben $Z = z$ von Beispiel S3-a

Zwischen X, Y und Z bestehen keinerlei Unabhängigkeiten:

$$f_{XY} \neq f_X\, f_Y,\ f_{XZ} \neq f_X\, f_Z,\quad f_{YZ} \neq f_Y\, f_Z \quad \text{und} \quad f_{XYZ} \neq f_X\, f_Y\, f_Z$$

7.1.3 Verteilung von Funktionen von Zufallsvariablen

Funktionen einer Zufallsvariable

● Hintergrund ●

Funktionen von Zufallsvariablen sind i.A. wiederum Zufallsvariablen.

● Beispiele: Funktionen einer Zufallsvariable ●

> ***0-1-Variable und $U = X^2$:***

$$P(X = 0) = 0.9 \quad \text{und} \quad P(X = 1) = 0.1$$

Betrachte Funktion $U = X^2$. Dann gilt:

$$P(U = 0) = 0.9 \quad \text{und} \quad P(U = 1) = 0.1$$

$\Rightarrow$ X und U identisch verteilt

> ***0-2-Variable und $U = X^2$:***

$$P(X = 0) = 0.9 \quad \text{und} \quad P(X = 2) = 0.1$$

Betrachte Funktion $U = X^2$. Dann gilt:

$$P(U = 0) = 0.9 \quad \text{und} \quad P(U = 4) = 0.1$$

$\Rightarrow$ X und U nicht identisch verteilt, da unterschiedliche Trägerpunkte

> ***Diskrete Gleichverteilung mit negativen Trägerpunkten und $U = X^2$:***

$$P(X = -2) = P(X = -1) = P(X = 0) = P(X = 1) = P(X = 2) = 0.2$$

Betrachte Funktion $U = X^2$. Dann gilt:

$$P(U = 0) = 0.2, P(U = 1) = 0.4 \text{ und } P(U = 4) = 0.4$$

> ***Diskrete Gleichverteilung des Würfelwurfs und $U = X^2$:***

$$P(X = 1) = P(X = 2) = \cdots = P(X = 6) = 1/6$$

Betrachte Funktion $U = X^2$. Dann gilt:

$$P(U = 1) = P(U = 4) = \cdots = P(U = 36) = 1/6$$

> ***Stetige Gleichverteilung auf [0, 1] und $U = X^2$:***

$$f_X(x) = I_{[0,1]}(x)$$

Betrachte Funktion $U = X^2$. Dann gilt **nicht!** etwa

$$f_U(u) = f_X(u)^2 = I_{[0,1]}(u)$$

Stattdessen ist hier eine Herleitung über die Verteilungsfunktion möglich.

Dabei gilt zunächst einmal:

$$F_X(x) = xI_{[0,1]}(x) + I_{(1,\infty)}(x)$$

U besitzt offensichtlich die gleiche Trägermenge wie *X*. Dabei gilt:

Dabei gilt für $u \in [0, 1]$:

$$F_U(u) = P(U \leq u) = P(X^2 \leq u)$$

$$= P(X \leq \sqrt{u}) = F_X(\sqrt{u})$$

$$= \sqrt{u}\, I_{[0,1]}(\sqrt{u}) = \sqrt{u} I_{[0,1]}(u)$$

Mit $F_U(u) = 0$ für $u < 0$ und $F_U(u) = 1$ für $u > 1$ gilt insgesamt:

$$F_U(u) = \sqrt{u}\, I_{[0,1]}(u) + I_{(1,\infty)}(u)$$

Mit dem Zusammenhang zwischen Verteilungsfunktion und Dichte folgt für $u \in (0,1)$:

$$F_U' = \frac{1}{2\sqrt{u}}$$

Eine zulässige Dichte wäre dann

$$f_U(u) = \frac{1}{2\sqrt{u}}\, I_{(0,1)}(u)$$

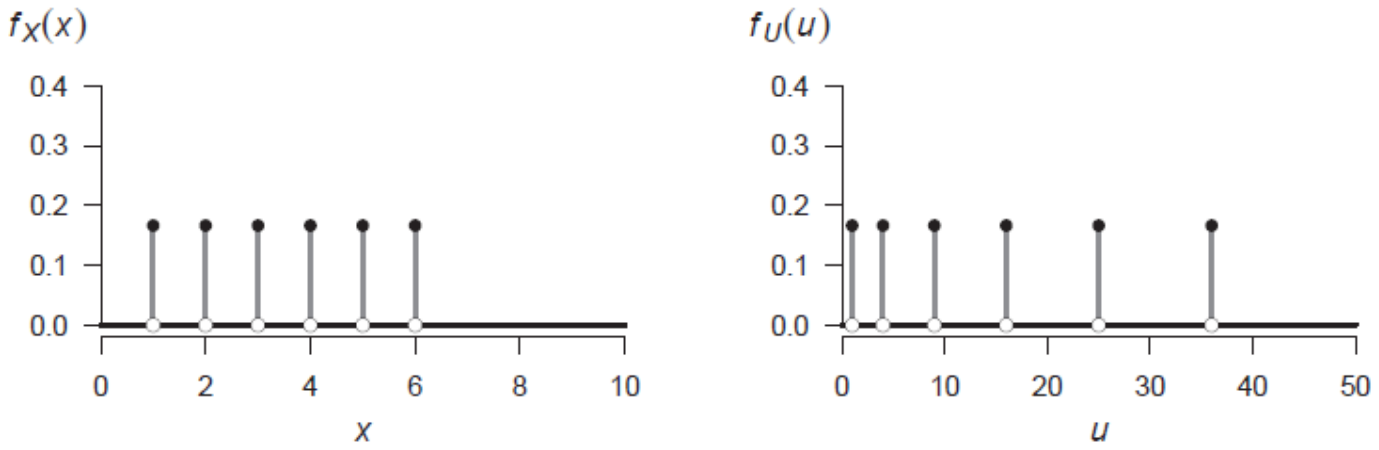


Abb. 7.1.15: Verteilung von X und $U = X^2$ bei einer diskreten Gleichverteilung

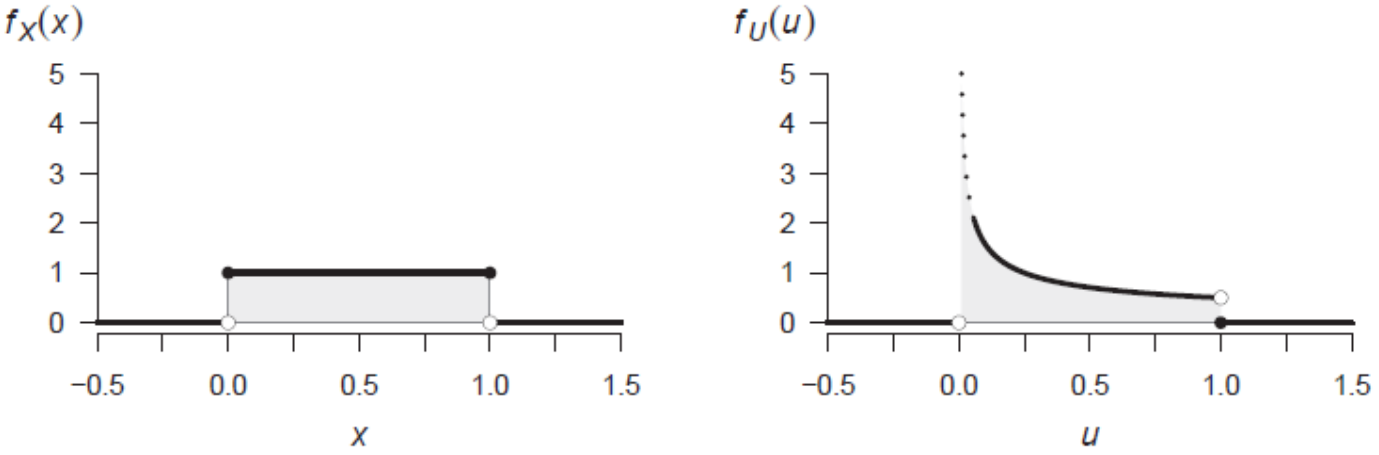


Abb. 7.1.16: Verteilung von X und $U = X^2$ bei einer stetigen Gleichverteilung

> ***Stetige Gleichverteilung auf [0, 1] und $Y = 2X$:***

$$f_X(x) = I_{[0,1]}(x) \quad \text{mit} \quad F_X(x) = xI_{[0,1]}(x) + I_{(1,\infty)}(x)$$

Betrachte Funktion $Y = 2X$.

Dann gilt für $y \in [0, 2]$:

$$F_Y(y) = P(Y \leq y) = P(2X \leq y)$$

$$= P(X \leq 0.5y) = F_X(0.5y)$$

$$= 0.5y\, I_{[0,1]}(0.5y) = 0.5yI_{[0,2]}(y)$$

Insgesamt erhält man daraus dann

$$F_Y(y) = 0.5y\, I_{[0,2]}(y) + I_{(2,\infty)}(y)$$

und über die 1. Ableitung die Dichte

$$f_Y(y) = 0.5\, I_{(0,2)}(y)$$

> ***Stetige Gleichverteilung auf [0, 2] und $U = X^2$:***

$$f_X(x) = 0.5\, I_{[0,2]}(x) \text{ mit } F_X(x) = 0.5x\, I_{[0,2]}(x) + I_{(2,\infty)}(x)$$

Betrachte Funktion $U = X^2$.

Dann gilt für $u \in [0,4]$:

$$F_U(u) = P(U \leq u) = P(X^2 \leq u)$$

$$= P(X \leq \sqrt{u}) = F_X(\sqrt{u})$$

$$= 0.5\sqrt{u} I_{[0,2]}(\sqrt{u}) = 0.5\sqrt{u} I_{[0,4]}(u)$$

Insgesamt erhält man daraus dann

$$F_U(u) = 0.5\sqrt{u}\, I_{[0,4]}(u) + I_{(4,\infty)}(u)$$

und über die 1. Ableitung die Dichte

$$f_U(u) = \frac{1}{4\sqrt{u}}\, I_{(0,4)}(y)$$

> ***Exponentialverteilung und Lineartransformation:***

$$f_X(x) = 0.05e^{-0.05x}\, I_{[0,\infty)}(x) \quad \text{mit}$$

$$F_X(x) = (1 - e^{-0.05x})\, I_{[0,\infty)}(x)$$

Betrachte die lineare transformierte (standardisierte) Größe

$$Z = \frac{X - 20}{20} = 0.05X - 1$$

Dann gilt:

$$F_Z(z) = P(Z \leq z) = P(0.05X - 1 \leq z)$$

$$= P(X \leq 20z + 20) = F_X(20z + 20)$$

$$= \left(1 - e^{-0.05(20z+20)}\right) I_{[0,\infty)}(20z + 20)$$

$$= (1 - e^{-z-1})\, I_{[0,\infty)}(20z + 20)$$

$$= (1 - e^{-z-1})\, I_{[-1,\infty)}(z)$$

Über die 1. Ableitung erhält man daraus dann die Dichte

$$f_Z(z) = e^{-z-1}\, I_{[-1,\infty)}(z).$$

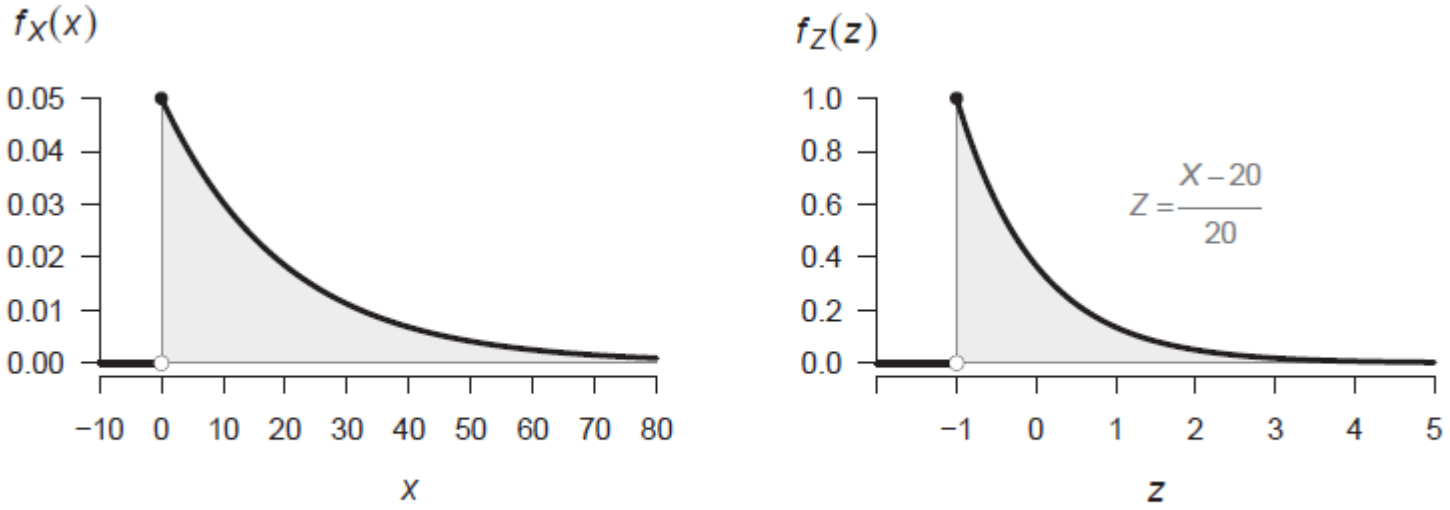


Abb. 7.1.17: Verteilung einer linear transformierten exponentialverteilten Zufallsvariable X

• Funktionen unabhängiger und identisch verteilter Zufallsvariablen •

Satz 7.1.5: Funktionen unabhängiger und identisch verteilter Zufallsvariablen

Für gegebene Zufallsvariablen $X_1, X_2, \ldots, X_n$ und beliebige (messbare) Funktionen $g_1, g_2, \ldots, g_n$ ($\mathbb{R} \to \mathbb{R}$) gilt: Sind $X_1, X_2, \ldots, X_n$

(i) stochastisch unabhängig, so auch $g_1(X_1), g_2(X_2), \ldots, g_n(X_n)$.
(ii) identisch verteilt, so auch $g_1(X_1), g_1(X_2), \ldots, g_1(X_n)$.
(iii) stochastisch unabhängig und identisch verteilt (u.i.v.), so auch $g_1(X_1), g_1(X_2), \ldots, g_1(X_n)$.

Funktionen aus mehreren Zufallsvariablen

● Hintergrund ●

Funktionen aus mehreren Zufallsvariablen sind i.A. wiederum Zufallsvariablen.

● Beispiele: Funktionen aus mehreren Zufallsvariablen ●

> ***Summen unabhängiger 0-1-Variablen:***

Die Betrachtung des stochastischen Verhaltens einer Summe kann unterschiedlich motiviert sein. Angenommen die Erfolgsquote bei Telefonbefragungen eines Meinungsforschungsinstituts liege bei 40%. d.h. bei 60% aller Anrufversuche legt eine angerufene Person entweder wieder auf oder ist im Weiteren nicht zu einer Befragung bereit. Hieraus könnte sich nun die Frage ergeben, wie viele Anrufe theoretisch getätigt werden müssen, um mit einer bestimmten Wahrscheinlichkeit z.B. eine bestimmte Anzahl erfolgreiche Befragungen zu erhalten.

Sofern nun n Anrufe getätigt werden, lassen sich diese mittels n identisch verteilter Zufallsvariablen $X_1, X_2, \ldots, X_n$ beschreiben. Dabei gilt dann:

$$P(X_i = 0) = 0.6 \text{ und } P(X_i = 1) = 0.4 \text{ für } i = 1, \ldots, n$$

Betrachte zunächst den Fall $n = 2$:

2-Tupel	Summe	Wahrschein.
(0, 0)	0	0.36
(1, 0)	1	0.24
(0, 1)	1	0.24
(1, 1)	2	0.16

Dies impliziert für die Verteilung der Summe $S_2 = X_1 + X_2$

s	0	1	2
$P(S_2 = s)$	0.36	0.48	0.16

Analog gilt weiter für $n = 3$:

3-Tupel	Summe	Wahrsch.	3-Tupel	Summe	Wahrsch.
(0, 0, 0)	0	0.216	(0, 1, 1)	2	0.096
(1, 0, 0)	1	0.144	(1, 0, 1)	2	0.096
(0, 1, 0)	1	0.144	(1, 1, 0)	2	0.096
(0, 0, 1)	1	0.144	(1, 1, 1)	3	0.064

s	0	1	2	3
$P(S_3 = s)$	0.216	0.432	0.288	0.064

und $n = 10$:

s	0	1	2	3	4	5
$P(S_{10} = s)$	0.01	0.04	0.12	0.21	0.25	0.20
s	6	7	8	9	10	
$P(S_{10} = s)$	0.11	0.04	0.01	0.00	0.00	

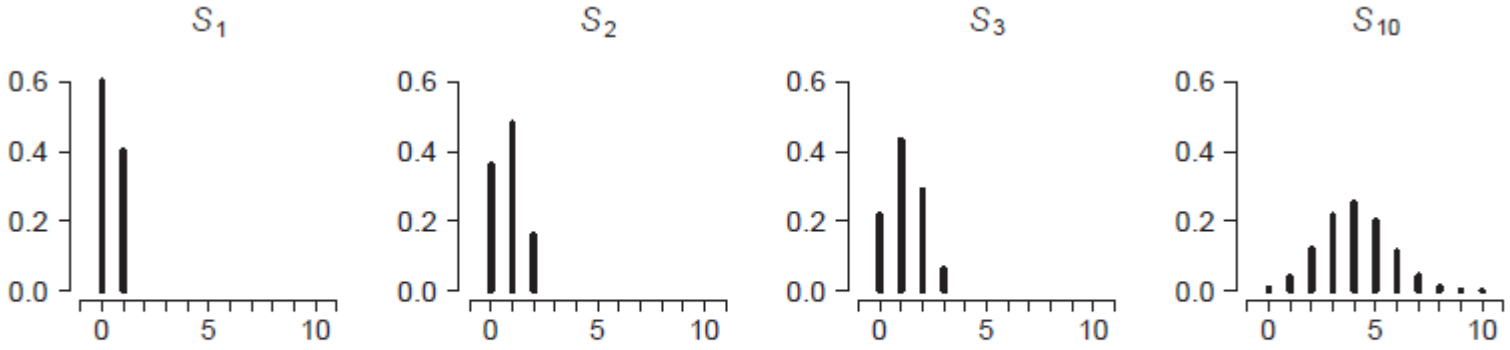


Abb. 7.1.18: Verteilung von Summen aus 0-1-Variablen

> *Summe und Produkt bei diskreter Gleichverteilung*

Tab. 7.1.13: Zweimaliges Würfeln – Summe und Produkt der Augenzahlen

X_1 \ X_2 s|u	1	2	3	4	5	6
1	**2|1**	**3|2**	**4|3**	**5|4**	**6|5**	**7|6**
	1/36	1/36	1/36	1/36	1/36	1/36
2	**3|2**	**4|4**	**5|6**	**6|8**	**7|10**	**8|12**
	1/36	1/36	1/36	1/36	1/36	1/36
3	**4|3**	**5|6**	**6|9**	**7|12**	**8|15**	**9|18**
	1/36	1/36	1/36	1/36	1/36	1/36
4	**5|4**	**6|8**	**7|12**	**8|16**	**9|20**	**10|24**
	1/36	1/36	1/36	1/36	1/36	1/36
5	**6|5**	**7|10**	**8|15**	**9|20**	**10|25**	**11|30**
	1/36	1/36	1/36	1/36	1/36	1/36
6	**7|6**	**8|12**	**9|18**	**10|24**	**11|30**	**12|36**
	1/36	1/36	1/36	1/36	1/36	1/36

Tab. 7.1.11: Zweimaliges Würfeln – Verteilung der Summe Augenzahlen

s	2	3	4	5	6	7
$P(S_2 = s)$	1/36	2/36	3/36	4/36	5/36	6/36
s	8	9	10	11	12	
$P(S_2 = s)$	5/36	4/36	3/36	2/36	1/36	

Tab. 7.1.12: Zweimaliges Würfeln – Verteilung des Produkts der Augenzahlen

u	1	2	3	4	5	6	8	9	10
$P(U = u)$	1/36	2/36	2/36	3/36	2/36	4/36	2/36	1/36	2/36
u	12	15	16	18	20	24	25	30	36
$P(U = u)$	4/36	2/36	1/36	2/36	2/36	2/36	1/36	2/36	1/36

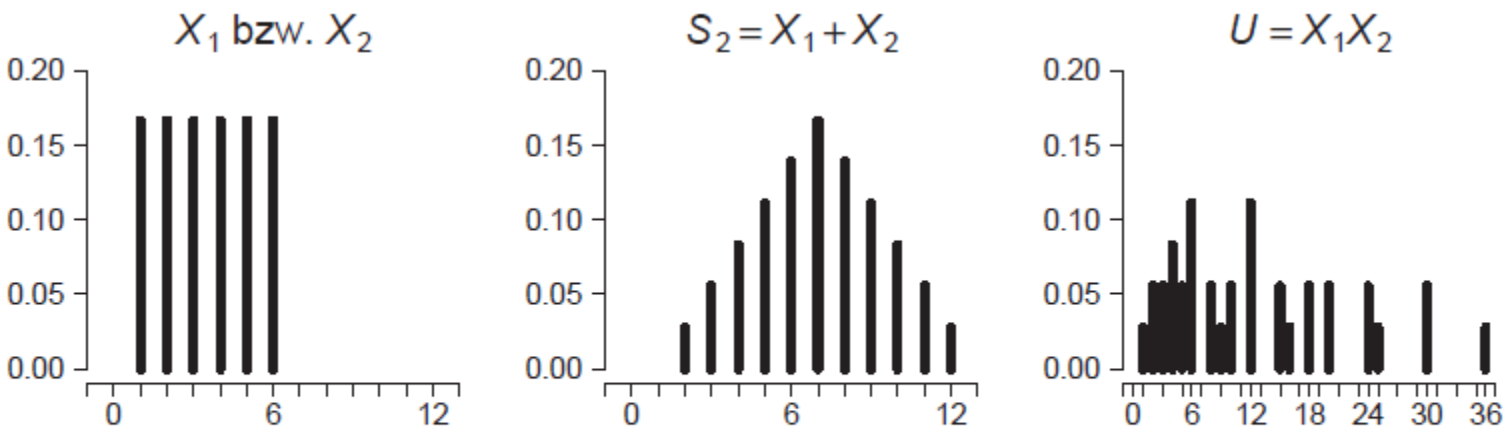


Abb. 7.1.19: Verteilung von Summe und Produkt zweier Würfelergebnisse

> *Summe und Produkt bei stetiger Gleichverteilung*

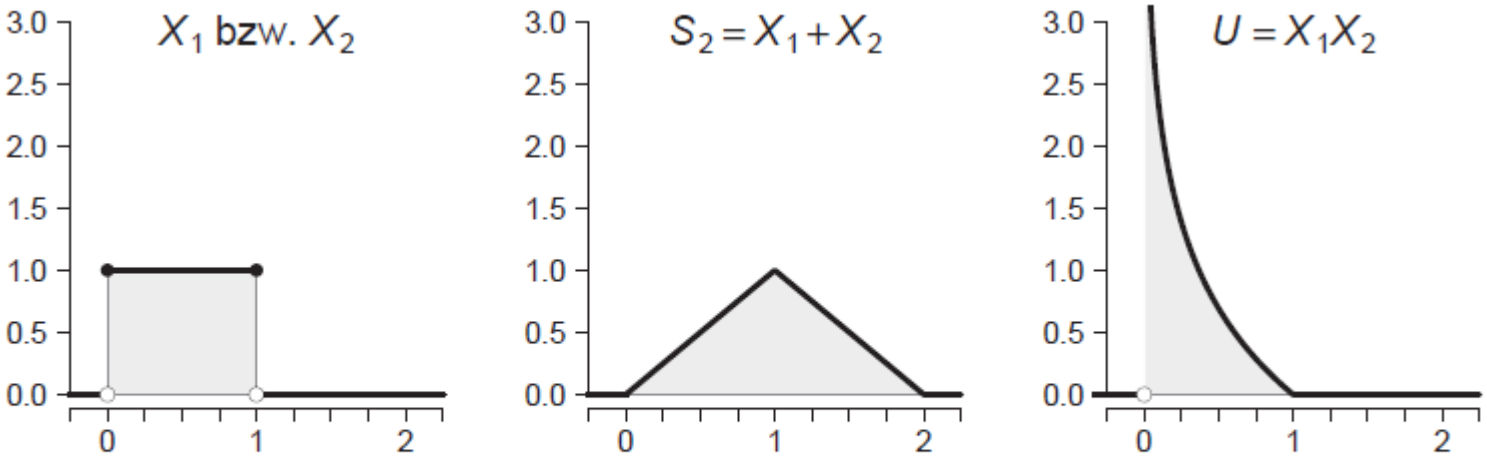


Abb. 7.1.20: Verteilung von Summe und Produkt zweier stetig gleichverteilter Zufallsvariablen

Herleitung „anspruchsvoll“ über sog. **Transformations-** und **Faltungssätze**

$$f_{S_2}(s) = s\, I_{[0,1]}(s) + (2-s) I_{(1,2]}(s) \qquad f_U(u) = -ln(u) I_{(0,1]}(u)$$

7.2.1 Kennwerte in Bezug auf Lage und Streuung

Erwartungswert

- **Definition und Notation** -

Definition 7.2.1: Erwartungswert

Sei X eine diskrete oder stetige Zufallsvariable mit Wahrscheinlichkeits- bzw. Dichtefunktion f_X. Dann ist der Erwartungswert von X definiert als

(i) $\mu_X = E(X) = \sum_j a_j P(X = a_j) = \sum_j a_j f_X(a_j)$

für diskretes X mit Realisationsmöglichkeiten $a_1, a_2, \ldots, a_k, \ldots$ und

(ii) $\mu_X = E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$

für stetiges X.

Empirisches Pendant: $\bar{x}$ Arithmetisches Mittel

• Interpretation des Erwartungswertes •

Interpretation des Erwartungswertes

Der Erwartungswert ist ein „theoretisch erwarteter Mittelwert“, der sich „langfristig“ einstellen sollte, sofern immer wieder erneut metrische Beobachtungswerte aus der gleichen unterstellten Wahrscheinlichkeitsverteilung gewonnen werden. Informal ausgedrückt gilt also:

$$\bar{x} \to \mu_X \text{ für wachsendes } n \text{ oder } \bar{x} \approx \mu_X \text{ für großes } n.$$

Formal präzise wird dies durch das ***Gesetz der großen Zahlen*** beschrieben.

Beispiel: X sei das Ergebnis beim Würfelwurf. Dann gilt:

$$E(X) = \sum_{j=1}^{6} j\, P(X = j)$$

$$= 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = 3.5$$

7.2 Theoretische Kennwerte

● Begründung des Rechenkalküls für stetige Verteilungen ●

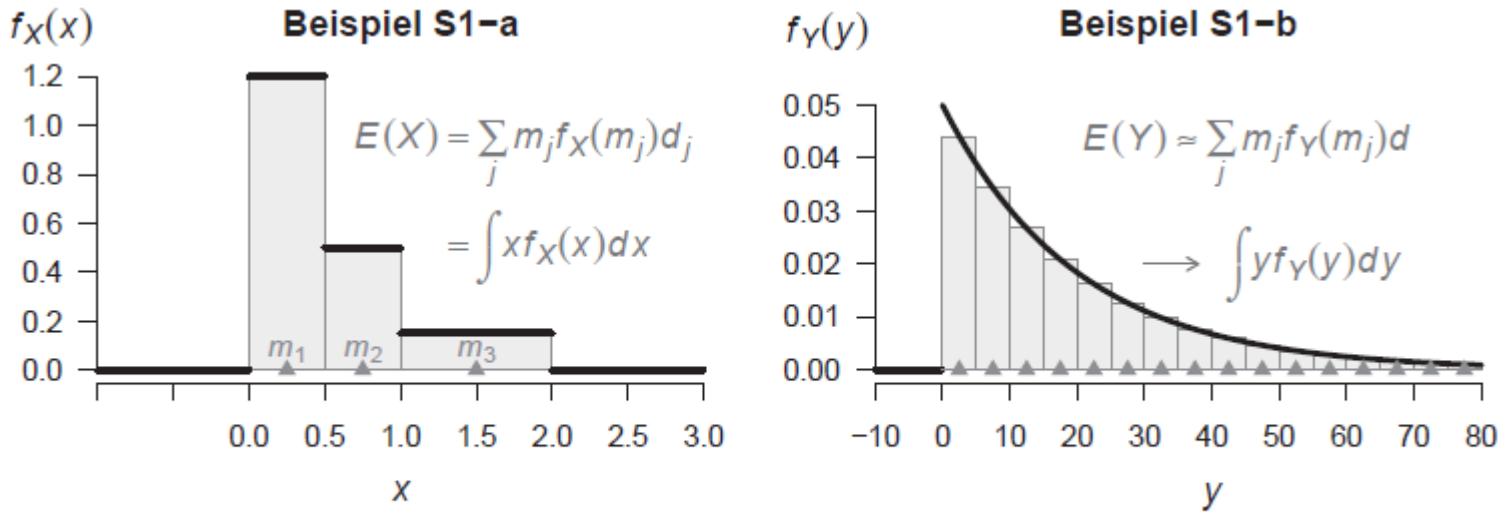


Abb. 7.2.1: Begründung des Rechenkalküls für stetige Verteilungen

Skizzenhafte Begründungen für die beiden Fälle

(i) Dichte besitzt Form eines Wahrscheinlichkeitshistogramms

(ii) Dichte besitzt anderweitige Form

nachlesbar im Lehrbuch (S. 333-335)

• Beispiel D1-a fortgesetzt •

X ... Anzahl mitreisender Kinder bei Pauschalreisen

$P(X = 0) = 0.4, P(X = 1) = 0.2, P(X = 2) = 0.3$ und $P(X = 3) = 0.1$

$$E(X) = 0 \times 0.4 + 1 \times 0.2 + 2 \times 0.3 + 3 \times 0.1 = 1.1$$

• Beispiel D1-b fortgesetzt •

Y ... Stündliche Anzahl von Notrufen an einem Rettungswagen-Stützpunkt

$$f_Y(y) = \begin{cases} \frac{3^y}{y!} e^{-3}, & y = 0, 1, 2, \dots, \\ 0, & sonst. \end{cases}$$

Bemerkung:
W‘funktion einer *Po*(3)-Verteilung
(nähere Hintergründe später)

$$E(Y) = \sum_{j=0}^{\infty} j \times \frac{3^j}{j!} e^{-3} = 0 + \frac{3}{1} e^{-3} + 2 \times \frac{3^2}{2!} e^{-3} + 3 \times \frac{3^3}{3!} e^{-3} + \dots = 3$$

kann mit einigem Aufwand gezeigt werden

● Beispiel S1-a fortgesetzt ●

X ... Höhe eines Trinkgelds

$$f_X(x) = 1.2I_{(0,0.5]}(x) + 0.5I_{(0.5,1]}(x) + 0.15I_{(1,2]}(x)$$

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

$$= \int_{0}^{0.5} 1.2x\, dx + \int_{0.5}^{1} 0.5x\, dx + \int_{1}^{2} 0.15x\, dx$$

$$= [0.6x^2]_0^{0.5} + [0.25x^2]_{0.5}^1 + [0.075x^2]_1^2$$

$$= 0.15 + 0.1875 + 0.225$$

$$= 0.5625$$

● **Beispiel S1-b fortgesetzt** ●

Y ... Wartezeit bis zum nächsten Notruf an einem Rettungswagen-Stützpunkt

$$f_Y(y) = 0.05e^{-0.05y} I_{[0,\infty)}(y)$$

Bemerkung: Dichte einer *Exp*(0.05)-Verteilung (nähere Hintergründe später)

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_{-\infty}^{\infty} 0.05 y e^{-0.05y} I_{[0,\infty)}(y)\, dy$$

$$= 0.05 \left[\frac{e^{-0.05y}}{0.05^2} (-0.05y - 1) \right]_0^{\infty}$$

kann mit einigem Aufwand gezeigt werden

$$= 0 - 0.05 \left(\frac{1}{0.05^2} (0 - 1) \right)$$

$$= 20$$

• Erwartungswert einer Funktion einer Zufallsvariable •

Satz 7.2.1: Erwartungswert einer Funktion einer Zufallsvariable

Sei X eine diskrete oder stetige Zufallsvariable mit Wahrscheinlichkeits- bzw. Dichtefunktion f_X und $g(x)$ eine (messbare) reellwertige Funktion, d.h. $g : \mathbb{R} \to \mathbb{R}$. Dann gilt für $Y = g(X)$:

(i) $E(Y) = E(g(X)) = \sum_j g(a_j)P(X = a_j) = \sum_j g(a_j)f_X(a_j)$

für diskretes X mit Realisationsmöglichkeiten $a_1, a_2, \ldots, a_k, \ldots$ und

(ii) $E(Y) = E(g(X)) = \int_{-\infty}^{\infty} g(x)f_X(x)dx$ für stetiges X.

Sei f_Y die Wahrscheinlichkeits- bzw. Dichtefunktion von Y. Alternativ zu (i) und (ii) können wir auch rechnen:

(i*) $E(Y) = \sum_j b_j P(Y = b_j) = \sum_j b_j f_Y(b_j)$

wobei $b_1, b_2, \ldots, b_l, \ldots$ die Realisationsmöglichkeiten von Y sind bzw.

(ii*) $E(Y) = \int_{-\infty}^{\infty} y f_Y(y)dy$.

Beispiel: Nachweis der Verschiebungs- und Skalenäquivarianz des Erwartungswertes

Diskreter Fall:

$$E(a+bX) = \sum_j (a+ba_j) P(X=a_j)$$

$$= \sum_j a\, P(X=a_j) + b \sum_j a_j\, P(X=a_j)$$

$$= a \underbrace{\sum_j P(X=a_j)}_{=1} + bE(X)$$

$$= a + bE(X)$$

Stetiger Fall:

$$E(a+bX) = \int_{-\infty}^{\infty} (a+bx) f_X(x)\, dx$$

$$= a \int_{-\infty}^{\infty} f_X(x)\, dx + b \int_{-\infty}^{\infty} x f_X(x)\, dx \quad = a \underbrace{\int_{-\infty}^{\infty} f_X(x)\, dx}_{=1} + bE(X)$$

$$= a + bE(X)$$

Empirisches Pendant:

Metrische Werte $x_1, \ldots, x_n$ werden linear transformiert gemäß $z_i = a + bx_i$

Dann gilt:

$$\bar{z} = a + b\bar{x}$$

● Beispiel D1-c ●

Sei X Würfelergebnis. Betrachte $U = X^2$ und $Z = 1 + 2X$.

x	1	2	3	4	5	6
$P(X = x)$	1/6	1/6	1/6	1/6	1/6	1/6
$u = x^2$	1	4	9	16	25	36
$P(U = u)$	1/6	1/6	1/6	1/6	1/6	1/6
$z = 1 + 2x$	3	5	7	9	11	13
$P(Z = z)$	1/6	1/6	1/6	1/6	1/6	1/6

Berechnungsvariante (i) von Satz 7.2.1 über die Verteilung von X:

$$E(U) = E(X^2) = \sum_{j=1}^{6} a_j^2 P(X = a_j) = 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + \cdots + 6^2 \times \frac{1}{6}$$

$$= \frac{91}{6} \approx 15.17$$

$$E(Z) = E(1 + 2X) = \sum_{j=1}^{6} (1 + 2a_j) P(X = a_j)$$

$$= (1 + 2) \times \frac{1}{6} + (1 + 4) \times \frac{1}{6} + \cdots + (1 + 12) \times \frac{1}{6} = 8$$

Berechnungsvariante (i)* von Satz 7.2.1 über die Verteilung von *U* bzw. *Z*:

$$E(U) = \sum_{j=1}^{6} b_j P(U = b_j) = 1 \times \frac{1}{6} + 4 \times \frac{1}{6} + \cdots + 36 \times \frac{1}{6} = \frac{91}{6}$$

$$E(Z) = \sum_{j=1}^{6} c_j P(Z = c_j) = 3 \times \frac{1}{6} + 5 \times \frac{1}{6} + \cdots + 13 \times \frac{1}{6} = 8$$

Außerdem ist der E-Wert linear (vgl. Folie 428), d. h.

$$E(1 + 2X) = 1 + 2\,E(X)$$

Mit dem Erwartungswert beim Würfelwurf

$$E(X) = \sum_{j=1}^{6} a_j P\left(X = a_j\right) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = 3.5$$

folgt daraus unmittelbar schon

$$E(Z) = 1 + 2\,E(X) = 8$$

Man beachte:

$$\frac{91}{6} = E(X^2) \neq E(X)^2 = 3.5^2 = \frac{49}{4}$$

● **Beispiel S1-c** ●

Sei X stetig gleichverteilt auf [0, 1]. Betrachte $U = X^2$ und $Z = 1 + 2X$.

$$f_X(x) = I_{[0,1]}(x)$$

Gemäß Vorarbeiten in Abschnitt 7.1.3 können folgende Dichten von U bzw. Z hergeleitet werden:

$$f_U(u) = \frac{1}{2\sqrt{u}} I_{(0,1)}(u) \quad \text{bzw.} \quad f_Z(z) = \frac{1}{2} I_{[1,3]}(z)$$

Berechnungsvariante (ii) von Satz 7.2.1 über die Verteilung von X:

$$E(U) = E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx$$

$$= \int_0^1 x^2\, dx = \left[\frac{1}{3}x^3\right]_0^1 = \frac{1}{3}$$

$$E(Z) = E(1+2X) = \int_{-\infty}^{\infty} (1+2x) f_X(x)\, dx = \int_{0}^{1} (1+2x)\, dx$$

$$= [x + x^2]_0^1 = 2$$

Berechnungsvariante (ii)* von Satz 7.2.1 über die Verteilung von *U* bzw. *Z*:

$$E(U) = \int_{-\infty}^{\infty} u f_U(u)\, du = \int_{0}^{1} \frac{u}{2\sqrt{u}}\, du = \left[\frac{1}{3} u^{3/2}\right]_0^1 = \frac{1}{3}$$

$$E(Z) = \int_{-\infty}^{\infty} z f_Z(z)\, dz = 0.5 \int_{1}^{3} z\, dz = 0.5 \left[\frac{1}{2} z^2\right]_1^3 = 2$$

Außerdem ist der E-Wert linear (vgl. Folie 428), d. h.

$$E(1 + 2X) = 1 + 2\,E(X)$$

Mit dem Erwartungswert von X

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_{0}^{1} x\,dx = \left[\frac{1}{2}x^2\right]_0^1 = 0.5$$

folgt daraus unmittelbar

$$E(Z) = 1 + 2\,E(X) = 2$$

• Erwartungswert einer Funktion aus mehreren Zufallsvariablen •

Satz 7.2.2: Erwartungswert einer Funktion aus mehreren Zufallsvariablen

Seien X und Y zwei diskrete oder zwei stetige Zufallsvariablen mit gemeinsamer Wahrscheinlichkeits- bzw. Dichtefunktion f_{XY} und $g(x, y)$ eine (messbare) reellwertige Funktion, d.h. $g : \mathbb{R}^2 \to \mathbb{R}$. Dann gilt für $Z = g(X, Y)$:

(i) $E(Z) = E(g(X, Y)) = \sum_i \sum_j g(a_i, b_j) P(X = a_i, Y = b_j)$
für diskretes $(X, Y)^T$ mit Realisationsmöglichkeiten $(a_1, b_1), \ldots, (a_k, b_l), \ldots$ und

(ii) $E(Z) = E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y) dy dx$
für stetiges $(X, Y)^T$.

Sei f_Z die Wahrscheinlichkeits- bzw. Dichtefunktion von Z. Alternativ zu (i) und (ii) können wir auch rechnen:

(i*) $E(Z) = \sum_j c_j P(Z = c_j) = \sum_j c_j f_Z(c_j)$,
wobei $c_1, c_2, \ldots, c_m, \ldots$ die Realisationsmöglichkeiten von Z sind, bzw.

(ii*) $E(Z) = \int_{-\infty}^{\infty} z f_Z(z) dz$.

• Spezialfall: Erwartungswert von Summen •

Satz 7.2.3: Erwartungswerte von Summen

Für n Zufallsvariablen $X_1, \ldots, X_n$ (diskret oder stetig) und Konstanten $c_0, c_1, \ldots, c_n$ gilt:

$$E(c_0 + c_1X_1 + c_2X_2 + \cdots + c_nX_n) = c_0 + c_1E(X_1) + c_2E(X_2) + \cdots + c_nE(X_n).$$

Speziell folgt daraus für zwei Zufallsvariablen X und Y:

$$E(X + Y) = E(X) + E(Y). \tag{7.2.13}$$

Man beachte hierzu den stetigen Fall mit $n = 2$:

$$E(X+Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f_{XY}(x,y)\,dydx$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf_{XY}(x,y)\,dydx + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}yf_{XY}(x,y)\,dxdy$$

$$= \int_{-\infty}^{\infty} x \underbrace{\left(\int_{-\infty}^{\infty} f_{XY}(x,y)\, dy \right)}_{f_X(x)} dx + \int_{-\infty}^{\infty} y \underbrace{\left(\int_{-\infty}^{\infty} f_{XY}(x,y)\, dx \right)}_{f_Y(y)} dy$$

$$= \int_{-\infty}^{\infty} x f_X(x)\, dx + \int_{-\infty}^{\infty} y f_Y(y)\, dy \quad = E(X) + E(Y)$$

Empirisches Pendant:

Metrische Werte $x_1, \ldots, x_n$ und $y_1, \ldots, y_n$. Dann gilt für $z_i = x_i + y_i$

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n}(x_i + y_i) = \frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{x} + \bar{y}$$

• Spezialfall: Erwartungswert von Produkten •

Satz 7.2.4: Multiplikationsregel bei Unabhängigkeit (und Unkorreliertheit)

Für n (vollständig) unabhängige Zufallsvariablen $X_1, \ldots, X_n$ (diskret oder stetig) gilt die **Multiplikationsregel** gemäß

$$E(X_1 \cdot X_2 \cdot \ldots \cdot X_n) = E(X_1) \cdot E(X_2) \cdot \ldots E(X_n).$$

Speziell folgt daraus für zwei unabhängige Zufallsvariablen X und Y:

$$E(X \cdot Y) = E(X) \cdot E(Y). \qquad (7.2.14)$$

Die Multiplikationsregel gilt auch bei **Unkorreliertheit** (Abschnitt 7.2.2).

Man beachte hierzu den stetigen Fall mit $n = 2$ unabhängigen Zufallsvariablen:

$$f_{XY}(x, y) = f_X(x) f_Y(y)$$

Dann gilt:

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy\, f_{XY}(x, y)\, dydx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_X(x) f_Y(y) dydx$$

$$= \int_{-\infty}^{\infty} x f_X(x) \underbrace{\left(\int_{-\infty}^{\infty} y f_Y(y)\, dy \right)}_{E(Y)} dx \quad = E(Y) \underbrace{\int_{-\infty}^{\infty} x f_X(x) dx}_{E(X)}$$

$$= E(X)E(Y)$$

● Beispiel D2-b fortgesetzt ●

Ein Würfel wird zwei Mal unabhängig voneinander geworfen...

Tab. 7.1.5: Kontingenztabelle mit Wahrscheinlichkeiten für Beispiel D2-b

X \ Y	1	2	3	4	5	6	Summe
1	1/36	1/36	1/36	1/36	1/36	1/36	1/6
2	1/36	1/36	1/36	1/36	1/36	1/36	1/6
3	1/36	1/36	1/36	1/36	1/36	1/36	1/6
4	1/36	1/36	1/36	1/36	1/36	1/36	1/6
5	1/36	1/36	1/36	1/36	1/36	1/36	1/6
6	1/36	1/36	1/36	1/36	1/36	1/36	1/6
Summe	1/6	1/6	1/6	1/6	1/6	1/6	1.0

Definiere jetzt: $X_1 = X$ und $X_2 = Y$

Einerseits gilt: $E(X_1 + X_2) = E(X_1) + E(X_2) = 3.5 + 3.5 = 7$

bzw. $E(X_1 X_2) = E(X_1)E(X_2) = 3.5 \times 3.5 = 12.25$

7.2 Theoretische Kennwerte

Tab. 7.1.11: Zweimaliges Würfeln – Verteilung der Summe Augenzahlen

s	2	3	4	5	6	7
$P(S_2 = s)$	1/36	2/36	3/36	4/36	5/36	6/36
s	8	9	10	11	12	
$P(S_2 = s)$	5/36	4/36	3/36	2/36	1/36	

Tab. 7.1.12: Zweimaliges Würfeln – Verteilung des Produkts der Augenzahlen

u	1	2	3	4	5	6	8	9	10
$P(U = u)$	1/36	2/36	2/36	3/36	2/36	4/36	2/36	1/36	2/36
u	12	15	16	18	20	24	25	30	36
$P(U = u)$	4/36	2/36	1/36	2/36	2/36	2/36	1/36	2/36	1/36

Andererseits folgt aus obigen Resultaten:

$$E(S_2) = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + \cdots + 12 \times \frac{1}{36} = 7$$

bzw. $$E(U) = 1 \times \frac{1}{36} + 2 \times \frac{2}{36} + 3 \times \frac{2}{36} + \cdots + 36 \times \frac{1}{36} = 12.25$$

● **Beispiel S2-b fortgesetzt** ●

Seien X und Y gemeinsam stetig verteilt gemäß der Dichtefunktion

$$f_{XY}(x,y) = I_{[0,1]}(x)I_{[0,1]}(y)$$

Definiere jetzt: $X_1 = X$ und $X_2 = Y$

Da der Erwartungswert von X_1 bzw. X_2 gleich 0.5 ist, gilt einerseits:

$$E(X_1 + X_2) = E(X_1) + E(X_2) = 0.5 + 0.5 = 1$$

bzw. $E(X_1X_2) = E(X_1)E(X_2) = 0.5 \times 0.5 = 0.25.$

Mit den Dichten für $S_2 = X_1 + X_2$ und $U = X_1X_2$ (vgl. Folie 420)

$$f_{S_2}(s) = s\, I_{[0,1]}(s) + (2-s)I_{(1,2]}(s)$$

bzw. $f_U(u) = -ln(u)I_{(0,1]}(u)$

folgt andererseits:

$$E(S_2) = \int_{-\infty}^{\infty} s f_{S_2}(s)\, ds = \int_0^1 s^2\, ds + \int_1^2 (2s - s^2)\, ds$$

$$= \left[\frac{1}{3}s^3\right]_0^1 + \left[s^2 - \frac{1}{3}s^3\right]_1^2 = 1$$

bzw. $$E(U) = \int_{-\infty}^{\infty} u f_U(u)\, du = \int_0^1 -u\, ln(u)\, du \underbrace{= \cdots = 0.25}$$

Partielles Integrieren

● Weitere Beispiele ●

Die Rechenregeln (i) und (ii) von Satz 7.2.1 gelten nicht nur für Summen und Produkte, sondern für nahezu beliebige Funktionen. Beispielhafte Ausführungen hierzu im LB S. 346-347.

Theoretische Varianz und Standardabweichung

• Definition und Notation •

Definition 7.2.2: Theoretische Varianz und Standardabweichung

Sei X eine diskrete oder stetige Zufallsvariable mit Erwartungswert $E(X) = \mu_X$. Dann ist die **(theoretische) Varianz** von X definiert als

$$\sigma_X^2 = Var(X) = E[(X - \mu_X)^2] = E(X^2) - \mu_X^2 \qquad (7.2.15)$$

und die **(theoretische) Standardabweichung** als

$$\sigma_X = \sqrt{\sigma_X^2}.$$

Empirisches Pendant: $\tilde{s}_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

• Verschiebungsformel für die theoretische Varianz •

Es gilt: $$Var(X) = E\left[\left(X - \underbrace{E(X)}_{\mu_X}\right)^2\right] = E(X^2) - \underbrace{E(X)^2}_{\mu_X^2}$$

Beweis:
$$E\left[\left(X - E(X)\right)^2\right] = E(X^2 - 2\mu_X X + \mu_X^2)$$
$$= E(X^2) - 2\mu_X E(X) + E(\mu_X^2)$$
$$= E(X^2) - 2\mu_X \mu_X + \mu_X^2$$
$$= E(X^2) - \mu_X^2$$

Empirisches Pendant: $$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$$

• Interpretation der theoretischen Varianz •

Interpretation der theoretischen Varianz und Standardabweichung

Die theoretische Varianz ist die „langfristig“ sich einstellende empirische Varianz, falls immer wieder erneut metrische Beobachtungswerte aus der gleichen unterstellten Wahrscheinlichkeitsverteilung gewonnen werden. Informal ausgedrückt gilt also:

$$\tilde{s}_X^2 \to \sigma_X^2 \text{ für wachsendes } n \text{ oder } \tilde{s}_X^2 \approx \sigma_X^2 \text{ für großes } n.$$

Für die theoretische Standardabweichung gilt analog:

$$\tilde{s}_X \to \sigma_X \text{ für wachsendes } n \text{ oder } \tilde{s}_X \approx \sigma_X \text{ für großes } n.$$

• Berechnung •

> Die Berechnung der Varianz einer Zufallsvariable X entspricht der Berechnung des Erwartungswertes der Funktion $g(X)$ mit $g(x) = (x - \mu_X)^2$.
> Im stetigen Fall etwa lautet die Rechnung über die originäre Formel dann:

$$Var(X) = E[(X - \mu_x)^2] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)\, dx$$

Dazu muss jedoch bereits der Erwartungswert von X, sprich μ_X vorliegen.

> Die Berechnung der Varianz via **Verschiebungsformel** lautet:

$$Var(X) = E(X^2) - \mu_X^2 \quad = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx - \mu_X^2$$

> In analoger Weise ist im diskreten Fall ein Additionskalkül zu verwenden.

• Spezialfall: Varianzen von Summen •

Satz 7.2.5: Varianzen von Summen bei Unabhängigkeit

Für n (vollständig) unabhängige Zufallsvariablen $X_1, \ldots, X_n$ (diskret oder stetig) und Konstanten $c_0, c_1, \ldots, c_n$ gilt:

$$Var(c_0 + c_1X_1 + \cdots + c_nX_n) = c_1^2 Var(X_1) + \cdots + c_n^2 Var(X_n).$$

Speziell folgt daraus für zwei unabhängige Zufallsvariablen X und Y:

$$Var(X + Y) = Var(X) + Var(Y).$$

Diese Resultate gelten auch bei **Unkorreliertheit** (Abschnitt 7.2.2).

Man beachte hierzu den Fall mit 2 unabhängigen Zufallsvariablen X und Y.

Definiere $Z = X + Y$. Dann gilt:

$$Var(X+Y) = Var(Z) = E\left[\left(Z - E(Z)\right)^2\right]$$

$$= E\left[\left(X + Y - E(X+Y)\right)^2\right]$$

$$= E[(X + Y - \mu_X - \mu_Y)^2] \quad = E[(X - \mu_X + Y - \mu_Y)^2]$$

$$= E[(X-\mu_X)^2 + (Y-\mu_Y)^2 + 2(X-\mu_X)(Y-\mu_Y)]$$

$$= \underbrace{E[(X-\mu_X)^2]}_{Var(X)} + \underbrace{E[(Y-\mu_Y)^2]}_{Var(Y)} + \underbrace{2E[(X-\mu_X)(Y-\mu_Y)]}_{=0}$$

Aufgrund der Unabhängigkeit gilt nämlich:

$$E[(X-\mu_X)(Y-\mu_Y)] = E[(X-\mu_X)]E[(Y-\mu_Y)] = 0 \times 0$$

● Beispiel D1-a fortgesetzt ●

X ... Anzahl mitreisender Kinder bei Pauschalreisen

$P(X = 0) = 0.4, P(X = 1) = 0.2, P(X = 2) = 0.3$ und $P(X = 3) = 0.1$

Berechnung von Varianz und Standardabweichung via Verschiebungsformel:

$$E(X) = 0 \times 0.4 + 1 \times 0.2 + 2 \times 0.3 + 3 \times 0.1 = 1.1$$

$$E(X^2) = 0^2 \times 0.4 + 1^2 \times 0.2 + 2^2 \times 0.3 + 3^2 \times 0.1 = 2.3$$

$$\Rightarrow \quad \sigma^2 = Var(X) = E(X^2) - E(X)^2 = 2.3 - 1.1^2 = 1.09$$

$$\Rightarrow \quad \sigma = \sqrt{\sigma^2} = \sqrt{1.09} \approx 1.04$$

● **Beispiel S1-a fortgesetzt** ●

X ... Höhe eines Trinkgelds

$$f_X(x) = 1.2I_{(0,0.5]}(x) + 0.5I_{(0.5,1]}(x) + 0.15I_{(1,2]}(x)$$

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)dx = \cdots = 0.5625 \quad \text{(vgl. Folie 425)}$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)dx = 1.2\int_0^{0.5} x^2dx + 0.5\int_{0.5}^{1} x^2dx + 0.15\int_1^2 x^2dx$$

$$= 1.2\left[\frac{1}{3}x^3\right]_0^{0.5} + 0.5\left[\frac{1}{3}x^3\right]_{0.5}^{1} + 0.15\left[\frac{1}{3}x^3\right]_1^2$$

$$\approx 0.5458$$

$$\Rightarrow \quad \sigma^2 = Var(X) = E(X^2) - E(X)^2 = 0.5458 - 0.5625^2 \approx 0.2294$$

$$\Rightarrow \quad \sigma = \sqrt{\sigma^2} = \sqrt{0.2294} \approx 0.48$$

● Beispiel D2-b fortgesetzt ●

Ein Würfel wird zwei Mal geworfen $\Rightarrow \quad X_1, X_2$

Betrachte die Augensumme $S_2 = X_1 + X_2$

Für den einmaligen Würfelwurf gilt (vgl. Beispiel D1-c, Folien 430-432):

$$E(X) = 3.5 \text{ und } E(X^2) \approx 15.17$$

$$\Rightarrow \quad \sigma^2 = Var(X) = E(X^2) - E(X)^2 \approx 2.92$$

Daraus folgt für die Augensumme:

$$Var(S_2) = Var(X_1 + X_2) = Var(X_1) + Var(X_2)$$
$$\approx 2 \times 2.92 = 5.84$$

● **Beispiel S2-b fortgesetzt** ●

Seien X und Y gemeinsam stetig verteilt gemäß der Dichtefunktion

$$f_{XY}(x,y) = I_{[0,1]}(x) I_{[0,1]}(y)$$

Betrachte die Summe $S_2 = X_1 + X_2$

Laut Beispiel S1-c (Folien 433-435) gilt:

$$E(X) = 0.5 \text{ und } E(X^2) = 1/3$$

$$\Rightarrow \quad \sigma^2 = Var(X) = E(X^2) - E(X)^2 = 1/12 \approx 0.08$$

Daraus folgt für die Summe:

$$Var(S_2) = Var(X_1 + X_2) = Var(X_1) + Var(X_2)$$
$$= 2 \times 1/12 = 1/6$$

Theoretische Quantile und theoretischer Median

• Definition •

Definition 7.2.3: Theoretische Quantile und theoretischer Median

Sei X eine stetige Zufallsvariable mit streng monotoner Verteilungsfunktion auf der Trägermenge. Dann ist das (theoretische) α-**Quantil** q_α bestimmt durch

$$P(X \leq q_\alpha) = \alpha \quad \text{für } \alpha \in (0, 1). \tag{7.2.19}$$

Insbesondere heißt $q_{0.5}$ **(theoretischer) Median.** Bei Zufallsvariablen mit nicht streng monotoner Verteilungsfunktion entstehen Eindeutigkeitsprobleme, die per Konvention unterschiedlich geregelt werden können.

Empirisches Pendant: $\tilde{x}_\alpha$ ... empirisches α-Quantil
speziell: $\tilde{x}_{0.5}$ ... empirischer Median

7.2 Theoretische Kennwerte

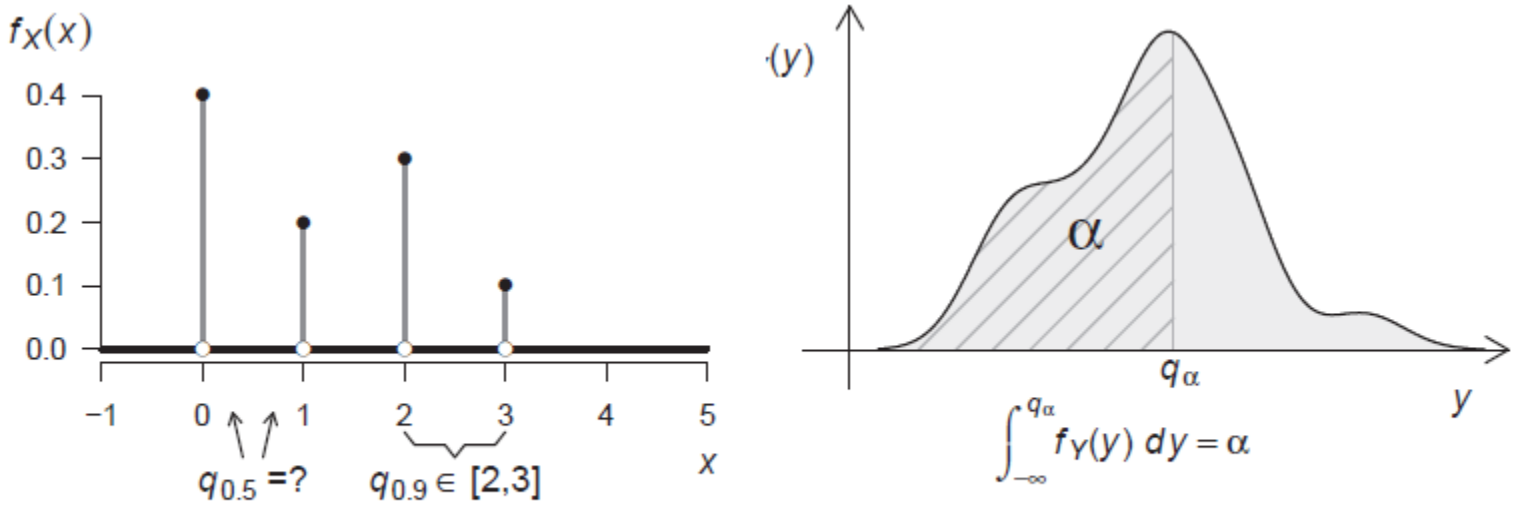


Abb. 7.2.2: Theoretische Quantile diskreter und stetiger Verteilungen

Eindeutigkeitsprobleme möglich (wie auch im empirischen Fall) bei diskreten Verteilungen oder bei stetigen Verteilungen mit „Intervalllücken“

● Interpretation der theoretischen Quantile ●

Bei unabhängiger identischer Wiederholung des gleichen Zufallsvorgangs streben die empirischen Quantile gegen die theoretischen Quantile

7.2 Theoretische Kennwerte

• Berechnung •

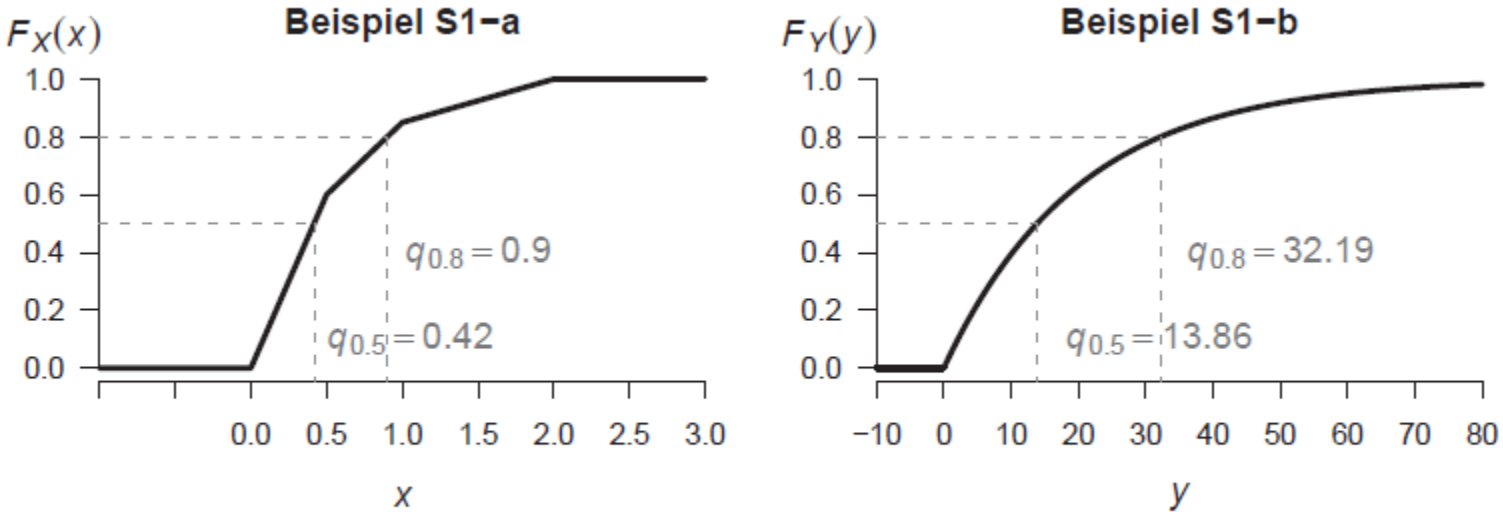


Abb. 7.2.3: Bestimmung theoretischer Quantile bei stetigen Verteilungen

Allgemein gilt: $F_X(q_\alpha) = \alpha$

Folglich gilt bei streng monotoner Verteilungsfunktion:

$$q_\alpha = F_X^{-1}(\alpha)$$

● **Beispiel S1-a fortgesetzt** ●

X ... Höhe eines Trinkgelds

$$f_X(x) = 1.2I_{(0,0.5]}(x) + 0.5I_{(0.5,1]}(x) + 0.15I_{(1,2]}(x)$$

$$F_X(x) = 1.2xI_{(0,0.5]}(x) + (0.5x + 0.35)I_{(0.5,1]}(x) + (0.15x + 0.7)I_{(1,2]}(x) + I_{(2,\infty)}(x)$$

(vgl. Folie 353)

Beispiel: Berechnung von Median und 0.8-Quantil

$$1.2q_{0.5} = 0.5 \quad \Leftrightarrow \quad q_{0.5} \approx 0.42 \qquad 0.5q_{0.8} + 0.35 = 0.8 \quad \Leftrightarrow \quad q_{0.8} = 0.9$$

● **Beispiel S1-b fortgesetzt** ●

Y ... Wartezeit bis zum nächsten Notruf an einem Rettungswagen-Stützpunkt

$$f_Y(y) = 0.05e^{-0.05y} I_{[0,\infty)}(y)$$

$$F_Y(y) = (1 - e^{-0.05y}) I_{[0,\infty)}(y)$$ (vgl. Folie 354)

Allgemeine Berechnungsformel für ein Quantil:

$$1 - e^{-0.05q_\alpha} = \alpha$$

$$\Leftrightarrow \quad e^{-0.05q_\alpha} = 1 - \alpha$$

$$\Leftrightarrow \quad -0.05q_\alpha = ln(1 - \alpha)$$

$$\Leftrightarrow \quad q_\alpha = -20\ ln(1 - \alpha) \quad \Rightarrow \quad F_Y^{-1}(\alpha) = -20\ ln(1 - \alpha)$$

Beispiel: Berechnung von Median und 0.8-Quantil:

$$q_{0.5} = -20\ ln(0.5) \approx 13.86$$

$$q_{0.8} = -20\ ln(0.2) \approx 32.19$$

7.2.2 Kennwerte in Bezug auf Abhängigkeiten

Bedingte Erwartungswerte und Varianzen

● Definition ●

Definition 7.2.4: Bedingte Erwartungswerte und Varianzen

Gegeben seien zwei Zufallsvariablen X und Y (diskret oder stetig). Dann werden Erwartungswert und Varianz der bedingten Verteilung von Y gegeben $X = x$ als **bedingter Erwartungswert von Y gegeben $X = x$** bzw. als **bedingte Varianz von Y gegeben $X = x$** bezeichnet. Dafür notiert man entsprechend

$$E(Y|X = x) \quad \text{bzw.} \quad Var(Y|X = x).$$

● Interpretation bedingter Erwartungswerte und Varianzen ●

Empirisches Pendant:

Gruppenmittelwerte und Gruppenvarianzen für gruppierte Daten

● Berechnung und Rechenregeln ●

Gleiche Rechenregeln wie für „gewöhnliche“ (unbedingte) Erwartungswerte und Varianzen; nur etwas mehr „Schreibaufwand“...

Beispiele:

$$E(Y|X=x) = \sum_j b_j f_{Y|X}(y|x)\,, \qquad E(Y|X=x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)\,dy\,,$$

$$Var(Y|X=x) = E\left[\left(Y - E(Y|X=x)\right)^2 | X=x\right]$$

$$= E[Y^2|X=x] - \left(E(Y|X=x)\right)^2$$

$$E(X+Y|Z=z) = E(X|Z=z) + E(Y|Z=z)$$

Satz 7.2.6: Rechenregeln für den bedingten Erwartungswert

Für die Zufallsvariablen X, Y, Z, die reellen Konstanten a, b, c und die (messbaren) Funktionen $g, h : \mathbb{R} \to \mathbb{R}$ gelten:

$$E(a + bX + cY|Z = z) = a + bE(X|Z = z) + cE(Y|Z = z), \tag{7.2.26}$$

$$E(g(X)h(Y)|X = x) = g(x)E(h(Y)|X = x), \tag{7.2.27}$$

sofern die angegebenen bedingten Erwartungswerte existieren.

● Bedingte Kennwerte bei Unabhängigkeit ●

Bei Unabhängigkeit stimmen bedingte Kennwerte mit unbedingten Kennwerten werten überein. Sind beispielsweise X und Y unabhängig, so gilt:

$$E(Y|X = x) = E(Y), \qquad E(X|Y = y) = E(X),$$

$$Var(Y|X = x) = Var(Y), \qquad Var(X|Y = y) = Var(X)$$

• Beispiel D2-a fortgesetzt •

Tab. 7.1.2: Kontingenztabelle mit Wahrscheinlichkeiten für Beispiel D2-a

X \ Y	0	1	2	Summe
0	0.04	0.08	0.28	0.40
1	0.12	0.12	0.36	0.60
Summe	0.16	0.20	0.64	1.00

Tab. 7.1.4: Bedingte Verteilung von Y unter $X = x$ für Beispiel D2-a

$X = x$ \ Y	0	1	2	Summe
0	0.1	0.2	0.7	1.0
1	0.2	0.2	0.6	1.0

Es gilt:

$$E(Y|X=0) = 0 \times 0.1 + 1 \times 0.2 + 2 \times 0.7 = 1.6,$$

$$E(Y^2|X=0) = 0^2 \times 0.1 + 1^2 \times 0.2 + 2^2 \times 0.7 = 3,$$

$$Var(Y|X=0) = E(Y^2|X=0) - E(Y|X=0)^2 = 3 - 1.6^2 = 0.44$$

Analog erhält man

$$E(Y|X=1) = 1.4 \quad \text{und} \quad Var(Y|X=1) = 0.64$$

● Beispiel D2-b fortgesetzt ●

Ein Würfel wird zwei Mal unabhängig voneinander geworfen...

Aufgrund der Unabhängigkeit gilt hier:

$$f_{X|Y} = f_X \quad \text{und} \quad f_{Y|X} = f_Y$$

Wie bereits ermittelt (Folien 432, 452) gilt:

$$E(X) = E(Y) = 3.5 \quad \text{und} \quad Var(X) = Var(Y) = 2.92$$

Mit der Unabhängigkeit folgt dann

$$E(Y|X=1) = E(Y|X=2) = \cdots = E(Y|X=6) = 3.5$$

$$Var(Y|X=1) = Var(Y|X=2) = \cdots = Var(Y|X=6) = 2.92$$

● **Beispiel S2-a fortgesetzt** ●

Seien X und Y gemeinsam stetig verteilt gemäß der Dichtefunktion

$$f_{XY}(x,y) = (0.5x + 0.25y)I_{[0,1]}(x)I_{[0,2]}(y)$$

Dann lautet die bedingte Dichte von Y unter $X = x$ (vgl. Folie 377):

$$f_{Y|X}(y|x) = \frac{0.5x + 0.25y}{x + 0.5} I_{[0,2]}(y) \quad \text{für} \quad x \in [0,1]$$

Daraus folgt allgemein

$$E(Y|X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)\, dy = \frac{1}{x + 0.5} \int_0^2 0.5xy + 0.25y^2 dy$$

$$= \frac{1}{x + 0.5} \left[\frac{1}{4} xy^2 + \frac{1}{12} y^3 \right]_{y=0}^{y=2}$$

$$= \frac{x + 2/3}{x + 0.5}$$

Speziell folgt daraus z. B.

$$E(Y|X=0) = 4/3, \quad E(Y|X=0.5) = 7/6, \quad E(Y|X=1) = 10/9$$

Weiter folgen mit

$$E(Y^2|X=x) = \int_{-\infty}^{\infty} y^2 f_{Y|X}(y|x)\, dy = \frac{1}{x+0.5}\int_0^2 0.5xy^2 + 0.25y^3 dy$$

$$= \frac{1}{x+0.5}\left[\frac{1}{6}xy^3 + \frac{1}{16}y^4\right]_{y=0}^{y=2}$$

$$= \frac{4/3x+1}{x+0.5}$$

mittels Verschiebungsformel z. B. die Resultate

$$Var(Y|X=0) \approx 0.22, \quad Var(Y|X=0.5) \approx 0.31,$$

$$Var(Y|X=1) \approx 0.32$$

● **Beispiel S2-b fortgesetzt** ●

Seien X und Y gemeinsam stetig verteilt gemäß der Dichtefunktion

$$f_{XY}(x,y) = I_{[0,1]}(x)I_{[0,1]}(y)$$

Aufgrund der Unabhängigkeit gilt hier:

$$f_{Y|X}(y|x) = f_Y(y) = I_{[0,1]}(y) \text{ für } x \in [0,1]$$

Wie bereits ermittelt (Folien 435, 453) gilt:

$$E(X) = E(Y) = 0.5 \text{ und } Var(X) = Var(Y) = 1/12$$

Mit der Unabhängigkeit folgt dann

$$E(Y|X = x) = 0.5 \text{ und } Var(Y|X = x) = 1/12 \text{ für } x \in [0,1]$$

● Bedingter Erwartungswert als Zufallsvariable ●

Normalerweise sind bedingte Erwartungswerte wie $E(Y|X = x)$ theoretische nichtstochastische Größen für vorgegebene feste x-Werte. Werden keine konkreten Werte gesetzt; schreibt man nur etwa $E(Y|X)$ oder $Var(Y|X)$. In diesem Fall sind diese Größen als stochastische Größen, d. h. als Zufallsvariablen, zu verstehen. Man beachte nachfolgende Beispiele.

● Beispiel D2-a fortgesetzt ●

Es gilt (vgl. Folie 462):

$$E(Y|X = 0) = 1.6 \text{ und } E(Y|X = 1) = 1.4$$

$$P(X = 0) = 0.4 \text{ und } P(X = 1) = 0.6$$

Schreibt man nur $E(Y|X)$, fasst man den Ausdruck als Funktion der ZV X auf. Damit ist $E(Y|X)$ selbst eine ZV in folgendem Sinne:

$$P(E(Y|X) = 1.4) = 0.6 = P(X = 1) \text{ und}$$

$$P(E(Y|X) = 1.6) = 0.4 = P(X = 0).$$

● Beispiel D2-b fortgesetzt ●

Wegen $E(Y|X = x) = 3.5$ für $x = 1, \ldots, 6$ gilt hier:

$$P(E(Y|X) = 3.5) = 1$$

● Beispiel S2-a fortgesetzt ●

Es gilt (vgl. Folie 464): $E(Y|X = x) = \frac{x + 2/3}{x + 0.5}$

Stochastisch aufgefasst : $E(Y|X) = \frac{X + 2/3}{X + 0.5}$ Funktion von X!

Setzen wir $U = E(Y|X)$, erhält man „mit etwas Aufwand“ als Dichte von U

$$f_U(u) = \frac{1}{36(u-1)^3} I_{[10/9,4/3]}(u)$$

● Beispiel S2-b fortgesetzt ●

Wegen $E(Y|X = x) = 0.5$ für jedes $x \in [0,1]$ gilt hier:

$$P(E(Y|X) = 0.5) = 1$$

7.2 Theoretische Kennwerte

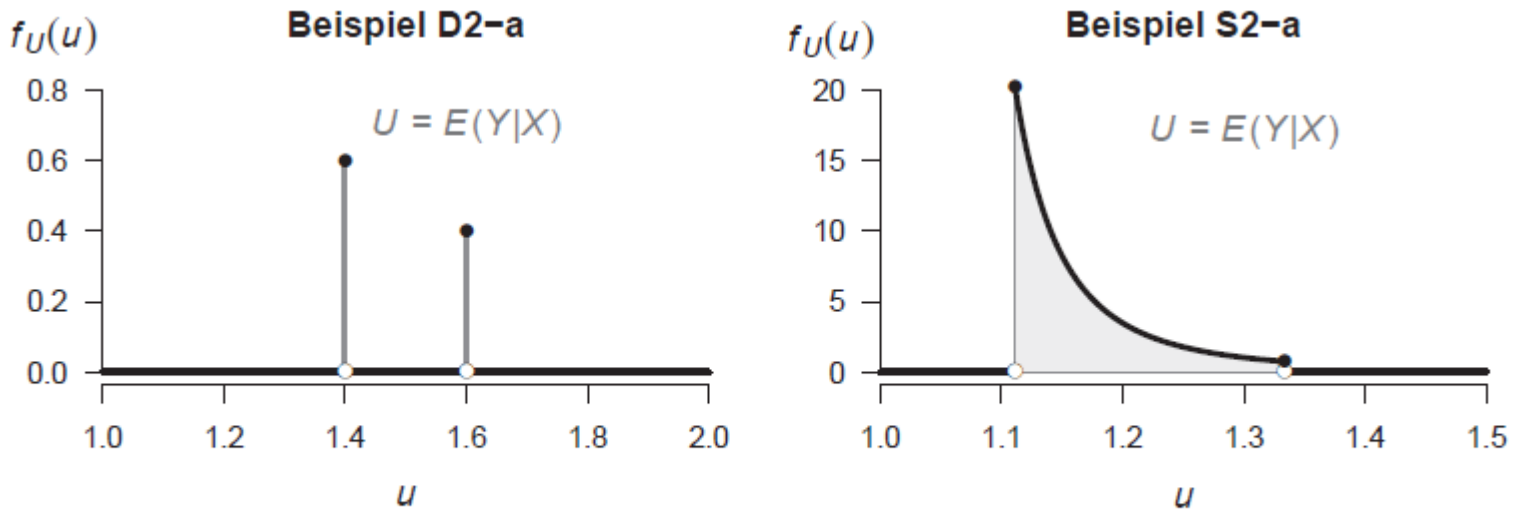


Abb. 7.2.4: Bedingter Erwartungswert als Zufallsvariable

• Iterierte Erwartungswertbildung •

Satz 7.2.7: Regeln bei iterierter Erwartungswertbildung

Gegeben seien zwei Zufallsvariablen X und Y (diskret oder stetig). Dann gilt:

(i) $E(Y) = E[E(Y|X)]$.

(ii) $Var(Y) = E[Var(Y|X)] + Var[E(Y|X)]$.

• Interpretation •

Empirisches Pendant:

Berechnung des arithmetischen Mittels bei gruppierten Daten bzw.
Berechnung der (empirischen Varianz) bei gruppierten Daten
(Streuungszerlegungsformel)

• Beispiel D2-a fortgesetzt •

Mit $P(Y = 0) = 0.16,\ P(Y = 1) = 20$ und $P(Y = 2) = 0.64$

folgt $E(Y) = 1.48$ und $Var(Y) = 0.5696$

Gemäß der zuvor ermittelten Verteilung von $U = E(Y|X)$ gilt

$$P(U = 1.4) = 0.6 \text{ und } P(U = 1.6) = 0.4.$$

Damit erhalten wir

$$E[E(Y|X)] = E(U) = 1.4 \times 0.6 + 1.6 \times 0.4 = \overbrace{1.48}^{E(Y)}$$

$$E[E(Y|X)^2] = E(U^2) = 1.4^2 \times 0.6 + 1.6^2 \times 0.4 = 2.2$$

$$\Rightarrow \quad Var[E(Y|X)] = Var(U) = E(U^2) - E(U)^2$$

$$= 2.2 - 1.48^2 = 0.0096$$

Wegen

$$Var(Y|X = 0) = 0.44 \quad \text{und} \quad Var(Y|X = 1) = 0.64$$

folgt dann weiter

$$E[Var(Y|X)] = Var(Y|X = 0)P(X = 0) + Var(Y|X = 1)P(X = 1)$$

$$= 0.44 \times 0.4 + 0.64 \times 0.6 = 0.56$$

Damit erhalten wir

$$E[Var(Y|X)] + Var[E(Y|X)] = 0.56 + 0.0096 = \underbrace{0.5696}_{Var(Y)}$$

Theoretische Kovarianz und Korrelation

• Definition und Notation •

Definition 7.2.5: Theoretische Kovarianz und Korrelation

Gegeben seien zwei Zufallsvariablen X und Y (diskret oder stetig) mit

$$E(X) = \mu_X,\ Var(X) = \sigma_X^2,\ E(Y) = \mu_Y,\ Var(Y) = \sigma_Y^2.$$

Dann ist die **(theoretische) Kovarianz** zwischen X und Y definiert als

$$\sigma_{XY} = Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X\mu_Y \tag{7.2.28}$$

und der **(theoretische) Korrelationskoeffizient** als

$$\varrho_{XY} = Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}.$$

Dabei gilt: $\varrho_{XY} \in [-1, 1]$.

• Verschiebungsformel für die theoretische Kovarianz •

Beweis: $Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$

$$= E(XY) - \mu_Y E(X) - \mu_X E(Y) + E(\mu_X \mu_Y)$$

$$= E(XY) - \mu_X\mu_Y - \mu_X\mu_Y + \mu_X\mu_Y$$

$$= E(XY) - \mu_X\mu_Y$$

• Interpretation von theoretischer Kovarianz und Korrelation •

Interpretation der theoretischen Kovarianz und Korrelation

Theoretische Kovarianz und Korrelation sind die sich „langfristig“ einstellende empirische Kovarianz bzw. Korrelation, falls immer wieder erneut metrische Beobachtungswerte aus der gleichen unterstellten gemeinsamen Wahrscheinlichkeitsverteilung zweier Zufallsvariablen X und Y gewonnen werden. Informal ausgedrückt gilt also:

$\tilde{s}_{XY} \to \sigma_{XY}$	und	$r_{XY} \to \varrho_{XY}$	für wachsendes n oder
$\tilde{s}_{XY} \approx \sigma_{XY}$	und	$r_{XY} \approx \varrho_{XY}$	für großes n.

● Abhängigkeit und Korrelation ●

> Unabhängigkeit impliziert Unkorreliertheit,
> Unkorreliertheit jedoch nicht Unabhängigkeit.

Man beachte, dass bei Unabhängigkeit gilt (vgl. Folie 440):

$$E(XY) = E(X)E(Y) \quad \Rightarrow \quad Cov(X,Y) = E(XY) - E(X)\,E(Y) = 0$$

● Beispiel D2-c ●

Tab. 7.2.1: Kontingenztabelle mit Wahrscheinlichkeiten für Beispiel D2-c

X \ Y	0	1	2	Summe
0	0.00	0.10	0.05	0.15
1	0.12	0.25	0.03	0.40
2	0.03	0.15	0.27	0.45
Summe	0.15	0.50	0.35	1.00

7.2 Theoretische Kennwerte

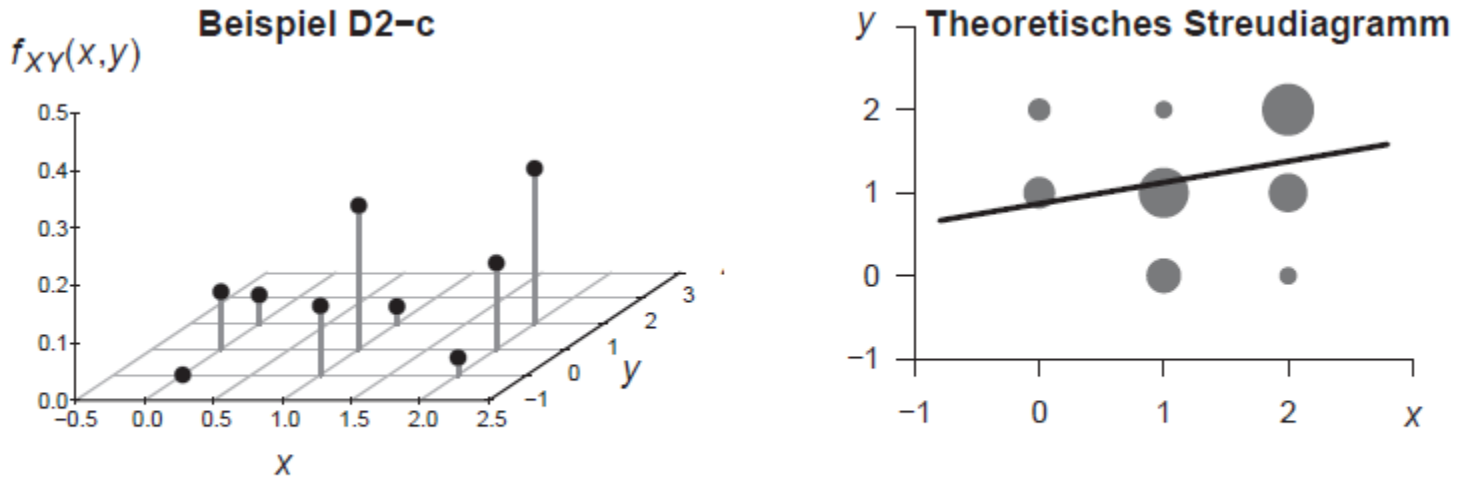


Abb. 7.2.5: Theoretischer Zusammenhang bei diskreter gemeinsamer Verteilung

Rechnerisch erhält man folgende Resultate:

$$E(X) = 0 \times 0.15 + 1 \times 0.40 + 2 \times 0.45 = 1.3$$

$$E(X^2) = 0^2 \times 0.15 + 1^2 \times 0.40 + 2^2 \times 0.45 = 2.2$$

$$E(Y) = 0 \times 0.15 + 1 \times 0.50 + 2 \times 0.35 = 1.2$$

$$E(Y^2) = 0^2 \times 0.15 + 1^2 \times 0.50 + 2^2 \times 0.35 = 1.9$$

Damit folgt:

$$Var(X) = E(X^2) - \big(E(X)\big)^2 = 2.2 - 1.3^2 = 0.51,$$

$$Var(Y) = E(Y^2) - \big(E(Y)\big)^2 = 1.9 - 1.2^2 = 0.46.$$

Weiter berechnet man gemäß den Rechenregeln für Erwartungswerte

$$E(XY) = \sum_{i=0}^{2}\sum_{j=0}^{2} i \times j \times P(X = i, Y = j)$$

$$= 0 \times 0 \times 0 + 0 \times 1 \times 0.1 + 0 \times 2 \times 0.05$$

$$+1 \times 0 \times 0.12 + 1 \times 1 \times 0.25 + 1 \times 2 \times 0.03$$

$$+2 \times 0 \times 0.03 + 2 \times 1 \times 0.15 + 2 \times 2 \times 0.27$$

$$= 1.69$$

Daraus folgt dann:

$$Cov(X,Y) = E(XY) - E(X)\,E(Y) = 1.69 - 1.3 \times 1.2 = 0.13$$

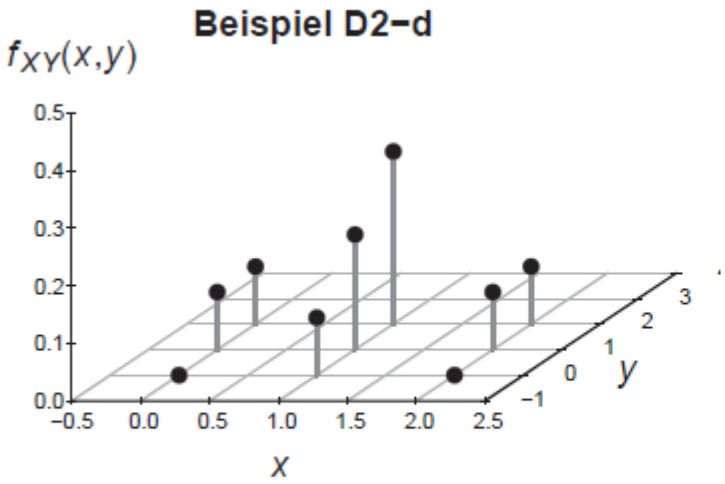
und

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{0.13}{\sqrt{0.51 \times 0.46}} \approx 0.2684$$

• Beispiel D2-d •

Tab. 7.2.2: Kontingenztabelle mit Wahrscheinlichkeiten für Beispiel D2-d

X \ Y	0	1	2	Summe
0	0.0	0.1	0.1	0.2
1	0.1	0.2	0.3	0.6
2	0.0	0.1	0.1	0.2
Summe	0.1	0.4	0.5	1.0



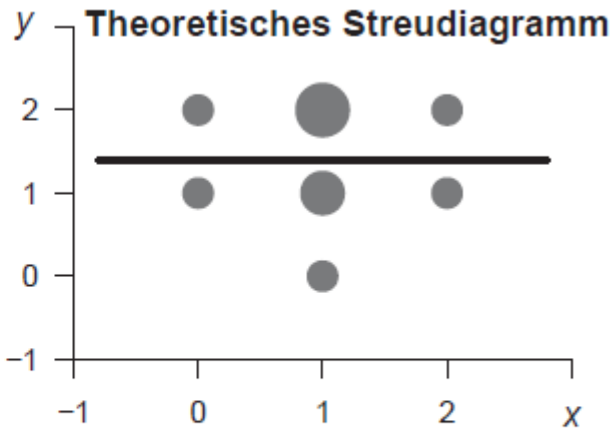


Abb. 7.2.6: Unkorreliert, aber nicht unabhängig

- **Beispiel S2-a fortgesetzt** -

$$f_{XY}(x,y) = (0.5x + 0.25y)I_{[0,1]}(x)I_{[0,2]}(y)$$

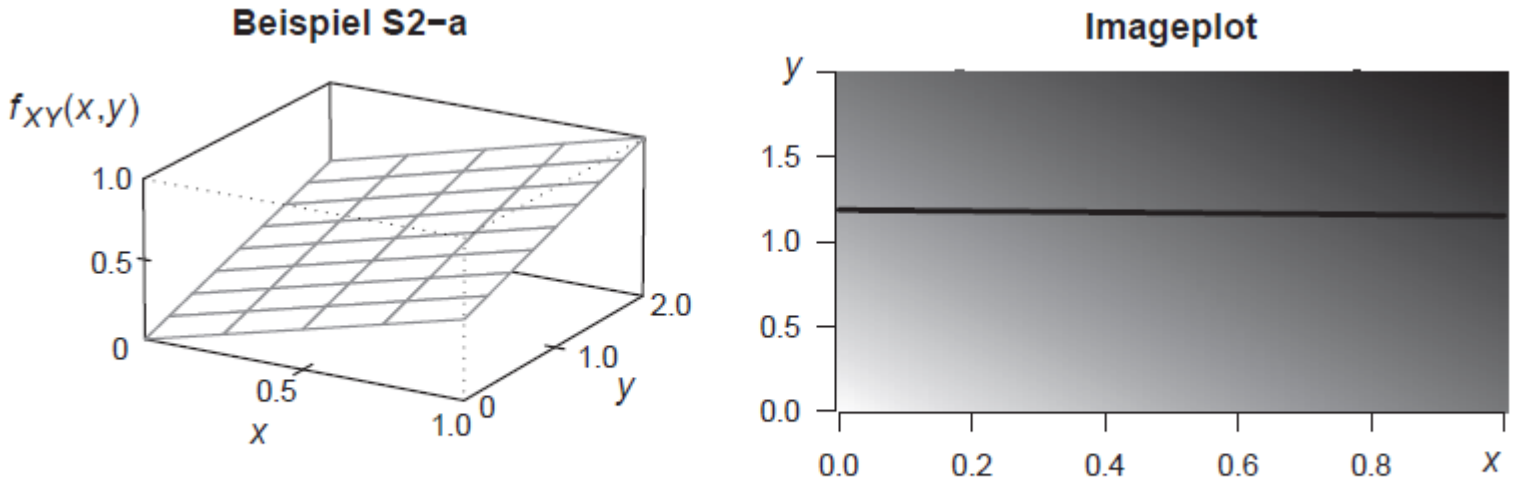


Abb. 7.2.7: Theoretischer Zusammenhang bei stetiger gemeinsamer Verteilung

Die Randdichten von X und Y lauten (vgl. Folie 374 bzw. 375):

$$f_X(x) = (x + 0.5)I_{[0,1]}(x) \quad \text{bzw.} \quad f_Y(y) = (0.25y + 0.25)I_{[0,2]}(y)$$

Daraus folgt dann für X:

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)dx = \int_{-\infty}^{\infty} (x^2 + 0.5x)dx = \left[\frac{1}{3}x^3 + \frac{1}{4}x^2\right]_0^1 = \frac{7}{12}$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)dx = \int_{-\infty}^{\infty} (x^3 + 0.5x^2)dx = \left[\frac{1}{4}x^4 + \frac{1}{6}x^3\right]_0^1 = \frac{5}{12}$$

$$\Rightarrow \quad Var(X) = E(X^2) - \big(E(X)\big)^2 = \frac{5}{12} - \left(\frac{7}{12}\right)^2 = \frac{11}{144}$$

Daraus berechnet man für Y:

$$E(Y) = \frac{7}{6}, \quad E(Y^2) = \frac{5}{3} \quad \text{und} \quad Var(Y) = \frac{11}{36}$$

Kovarianz und Korrelation ergeben sich dann aus:

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{XY}(x,y)\,dydx$$

$$= \int_{0}^{1}\int_{0}^{2} (0.5x^2y + 0.25xy^2)\,dydx$$

$$= \int_{0}^{1}\left(\left[\frac{1}{4}x^2y^2 + \frac{1}{12}xy^3\right]_{y=0}^{y=2}\right)dx$$

$$= \int_{0}^{1}\left(x^2 + \frac{2}{3}x\right)dx = \left[\frac{1}{3}x^3 + \frac{1}{3}x^2\right]_{0}^{1} = \frac{2}{3}$$

$$\Rightarrow \quad Cov(X,Y) = E(XY) - E(X)E(Y) = \frac{2}{3} - \frac{7}{12}\times\frac{7}{6} = -\frac{1}{72}$$

$$\Rightarrow \quad \rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{-1/72}{\sqrt{11/144\times 11/36}} = -\frac{1}{11} \approx -0.0909$$

Theoretische Regressionskoeffizienten

• Definition und Notation •

Definition 7.2.6: Theoretische Kleinste-Quadrate-Regression

Gegeben seien zwei Zufallsvariablen X und Y (diskret oder stetig) mit

$$E(X) = \mu_X,\ Var(X) = \sigma_X^2,\ E(Y) = \mu_Y,\ Cov(X, Y) = \sigma_{XY}.$$

Dann sind die **theoretischen KQ-Regressionskoeffizienten** definiert als

$$\overline{\hat{\beta}}_1 = \frac{Cov(X, Y)}{Var(X)} = \frac{\sigma_{XY}}{\sigma_X^2} \text{ und}$$

$$\overline{\hat{\beta}}_0 = E(Y) - \overline{\hat{\beta}}_1 E(X) = \mu_Y - \overline{\hat{\beta}}_1 \mu_X .$$

Die **theoretische KQ-Regressionsgerade** lautet dann

$$\overline{\hat{y}}(x) = \overline{\hat{\beta}}_0 + \overline{\hat{\beta}}_1 x.$$

• Interpretationen •

Interpretation der theoretischen KQ-Regression

Die theoretischen KQ-Regressionskoeffizienten sind die sich „langfristig“ einstellenden empirischen KQ-Regressionskoeffizienten, falls laufend Beobachtungswerte aus der gemeinsamen Wahrscheinlichkeitsverteilung zweier Zufallsvariablen X und Y gewonnen werden. Informal ausgedrückt gilt also:

$\hat{b}_0 \to \bar{\hat{\beta}}_0$ und $\hat{b}_1 \to \bar{\hat{\beta}}_1$ für wachsendes n oder

$\hat{b}_0 \approx \bar{\hat{\beta}}_0$ und $\hat{b}_1 \approx \bar{\hat{\beta}}_1$ für großes n.

• Beispiel D2-c fortgesetzt • (vgl. Folie 474 ff.)

$$\bar{\hat{\beta}}_1 = \frac{Cov(X,Y)}{Var(X)} = \frac{0.13}{0.51} \approx 0.2549$$

$$\bar{\hat{\beta}}_0 = E(Y) - \bar{\hat{\beta}}_1 E(X) = 1.2 - \frac{0.13}{0.51} \times 1.3 \approx 0.8686$$

$$\Rightarrow \bar{\hat{y}}(x) = 0.8686 + 0.2549x$$

- **Beispiel D2-d fortgesetzt** (vgl. Folie 478)

$$\bar{\hat{\beta}}_1 = 0, \quad \bar{\hat{\beta}}_0 = E(Y) = 1.4$$

- **Beispiel S2-a fortgesetzt** (vgl. Folie 479 ff.)

$$\bar{\hat{\beta}}_1 = \frac{Cov(X,Y)}{Var(X)} = \frac{-1/72}{11/144} = -\frac{2}{11} \approx -0.1818$$

$$\bar{\hat{\beta}}_0 = E(Y) - \bar{\hat{\beta}}_1 E(X) = \frac{7}{6} - \frac{2}{11} \times \frac{7}{12} \approx 1.0606$$

$$\Rightarrow \bar{\hat{y}}(x) = 1.0606 - 0.1818x$$

7.2.3 Spezifische Eigenschaften theoretischer Kennwerte

Minimumeigenschaften von Lagekennwerten

- **Minimumeigenschaft des Erwartungswertes** -

Satz 7.2.8: Minimumeigenschaft des Erwartungswerts

Gegeben sei eine Zufallsvariable X mit Erwartungswert μ_X und Varianz σ_X^2. Dann gilt für jedes $c \in \mathbb{R}$:

$$\sigma_X^2 = E[(X - \mu_X)^2] \leq E[(X - c)^2].$$

Empirisches Pendant: $$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \leq \frac{1}{n}\sum_{i=1}^{n}(x_i - c)^2$$

● Minimumeigenschaft des Medians ●

Satz 7.2.9: Minimumeigenschaft des theoretischen Medians

Gegeben sei eine Zufallsvariable X mit theoretischem Median $q_{0.5}$. Dann gilt für jedes $c \in \mathbb{R}$:

$$E[|X - q_{0.5}|] \leq E[|X - c|].$$

Empirisches Pendant: $$\frac{1}{n}\sum_{i=1}^{n}|x_i - \tilde{x}_{0.5}| \leq \frac{1}{n}\sum_{i=1}^{n}|x_i - c|$$

Wichtige Transformationseigenschaften

- **Transformationseigenschaften des Erwartungswertes**
- **Transformationseigenschaften der theoretischen Varianz**
- **Transformationseigenschaften der theoretischen Quantile**
- **Transformationseigenschaften von theoretischer Kovarianz und Korrelation**

Betrachte $U = c_X X$ und $V = c_Y Y$

Dann gilt:

$$Cov(U,V) = E\big[\big(U - E(U)\big)\big(V - E(V)\big)\big]$$

$$= E[(c_X X - c_X \mu_X)(c_Y Y - c_Y \mu_Y)]$$

$$= c_X c_Y E[(X - \mu_X)(Y - \mu_Y)]$$

$$= c_X c_Y Cov(X,Y)$$

7.2 Theoretische Kennwerte

Wichtige Transformationseigenschaften theoretischer Kennwerte

Für Erwartungswert, Varianz, Kovarianz und Korrelation von Zufallsvariablen gelten folgende Transformationsregeln:

(1) $E(X + c) = E(X) + c$ für jedes $c \in \mathbb{R}$.

(2) $E(cX) = cE(X)$ für jedes $c \in \mathbb{R}$.

(3) $Var(X + c) = Var(X)$ für jedes $c \in R$.

(4) $Var(cX) = c^2 Var(X)$ für jedes $c \in \mathbb{R}$.

(5) $\sqrt{Var(cX)} = c\sqrt{Var(X)}$ für jedes $c > 0$.

(6) $Cov(X + c_X, Y + c_Y) = Cov(X, Y)$ für alle $c_X, c_Y \in \mathbb{R}$.

(7) $Corr(X + c_X, Y + c_Y) = Corr(X, Y)$ für alle $c_X, c_Y \in \mathbb{R}$.

(8) $Cov(c_X X, c_Y Y) = c_X c_Y Cov(X, Y)$ für alle $cX, cY \in \mathbb{R}$.

(9) $Corr(c_X X, c_Y Y) = Corr(X, Y)$ für alle $c_X, c_Y > 0$.

Insbesondere folgen daraus die zu den entsprechenden empirischen Kennwerten völlig analogen Äquivarianz- und Invarianzeigenschaften in Bezug auf Verschiebungen und Umskalierungen.

• Erwartungswert und Varianz nach Standardisierung •

Analog zum empirischen Fall ist die „theoretische z-Standardisierung“ einer Zufallsvariable X mit Erwartungswert μ_X und Varianz σ_X^2 definiert als

$$Z = \frac{X - \mu_X}{\sigma_X}$$

Dabei gilt: (i) $$E(Z) = \frac{1}{\sigma_x} E(X - \mu_X) = \frac{1}{\sigma_x}\big(E(X) - E(\mu_X)\big) = 0$$

(ii) $$Var(Z) = \frac{1}{\sigma_X^2} Var(X - \mu_X) = \frac{1}{\sigma_X^2} Var(X) = 1$$

Kennwerte standardisierter Zufallsvariablen

Gegeben sei eine Zufallsvariable X mit Erwartungswert μ_X und Varianz $\sigma_X^2 > 0$. Für die standardisierte Größe

$$Z = \frac{X - \mu_X}{\sigma_X}$$

gilt stets: $E(Z) = 0$ und $Var(Z) = 1$.

Endliche und nicht endliche theoretische Momente

- **Momente**

$E(X^k)$ ... k-tes (theoretisches) Moment

- **Endlichkeit theoretischer Momente**
- **Momenterzeugende Funktionen**
- **Implikation und Deutung nicht endlicher Momente**
- **Beispiel S1-d**

Sei X stetig verteilt mit Dichte $f_X(x) = \frac{1}{x^2} I_{[1,\infty)}(x)$

Dann gilt: $E(X) = \int_{-\infty}^{\infty} x\, f_X(x) dx = \int_{1}^{\infty} \frac{1}{x} dx = [ln(x)]_1^{\infty}$

● Beispiel S1-e ●

Sei Y stetig verteilt mit Dichte $f_Y(y) = \dfrac{1}{\pi(1+y^2)}$

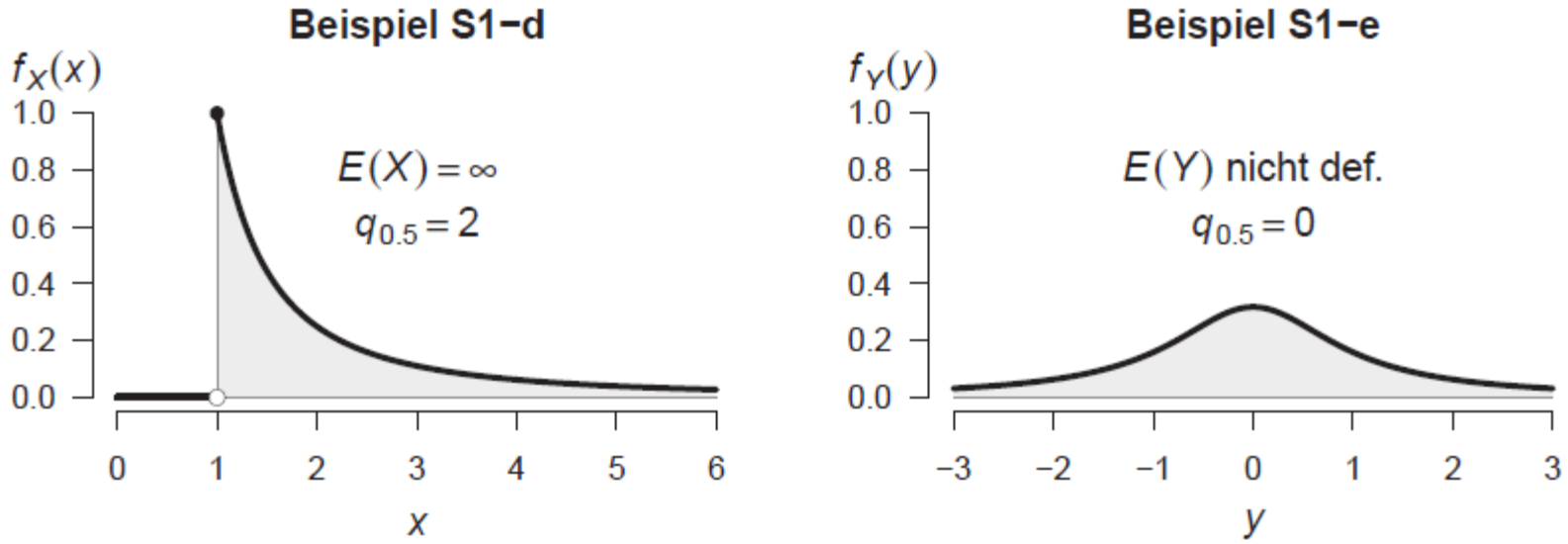


Abb. 7.2.8: Verteilungen mit nicht endlichen Erwartungswerten

● Weitere Beispiele ●

Literaturhinweis zu Beispielen mit nicht endlichen Momenten bei diskreten Verteilungen finden sich im Lehrbuch auf S. 379.

7.3 Spezielle eindimensionale Verteilungen

7.3.1 Spezielle diskrete Verteilungen

Elementare Kombinatorik

● Hintergrund ●

Bestimmung der Anzahl von Anordnungen und Auswahlmöglichkeiten von Objekten aus vorgegebenen Mengen

● Fakultät und Permutationen ●

Fakultät und Permutationen

Die **Fakultät** einer natürlichen Zahl n wird über das Symbol „!“ notiert und ist definiert als

$$n! = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 2 \cdot 1 \quad \text{(lies: } n \textit{ Fakultät}\text{)}.$$

Für $n = 1$ und $n = 0$ definiert man dabei $1! = 1$ und $0! = 1$.

Für $n \geq 2$ entspricht $n!$ der Anzahl der Anordnungsmöglichkeiten (**Permutationen**) von n verschiedenen Objekten.

● Auswahlmöglichkeiten bei Berücksichtigung der Reihenfolge ●

Werden k aus n Objekten ($k \leq n$) ausgewählt, so gibt es dafür

$$\frac{n!}{(n-k)!}$$

Möglichkeiten, falls die **Reihenfolge berücksichtigt** wird.

● Binomialkoeffizient und Auswahlmöglichkeiten ohne Berücksichtigung der Reihenfolge ●

Werden k aus n Objekten ($k \leq n$) ausgewählt, so gibt es dafür

$$\frac{n!}{(n-k)!\,k!}$$

Möglichkeiten, falls die **Reihenfolge nicht berücksichtigt** wird.
Dies definiert zugleich den **Binomialkoeffizienten**.

Binomialkoeffizient und Auswahlmöglichkeiten für „*k aus n*“

Für zwei natürliche Zahlen n und k mit $k \leq n$ ist der **Binomialkoeffizient**

$$\binom{n}{k} \quad \text{(lies: } n \text{ über } k\text{)}$$

definiert als

$$\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!} .$$

Der Binomialkoeffizient entspricht der Anzahl von Möglichkeiten, aus n Objekten k Objekte auszuwählen, wobei die Reihenfolge der Auswahl nicht berücksichtigt wird.

● Beispiel 7.3.1: Lotto „6 aus 49“ ●

Beim Lotto werden 6 aus 49 Kugeln ohne Zurücklegen gezogen. Die Reihenfolge der gezogenen Kugeln ist dabei irrelevant. Folglich gibt es

$$\binom{49}{6} = \frac{49!}{(49-6)!\,6!} = 13\,983\,816$$ verschiedene Ergebnisse

Gemäß Laplace-Kalkül beträgt die Wahrscheinlichkeit für „6 Richtige“

$$1/13\,983\,816 \approx 0.00000715\%$$

7.3 Spezielle eindimensionale Verteilungen

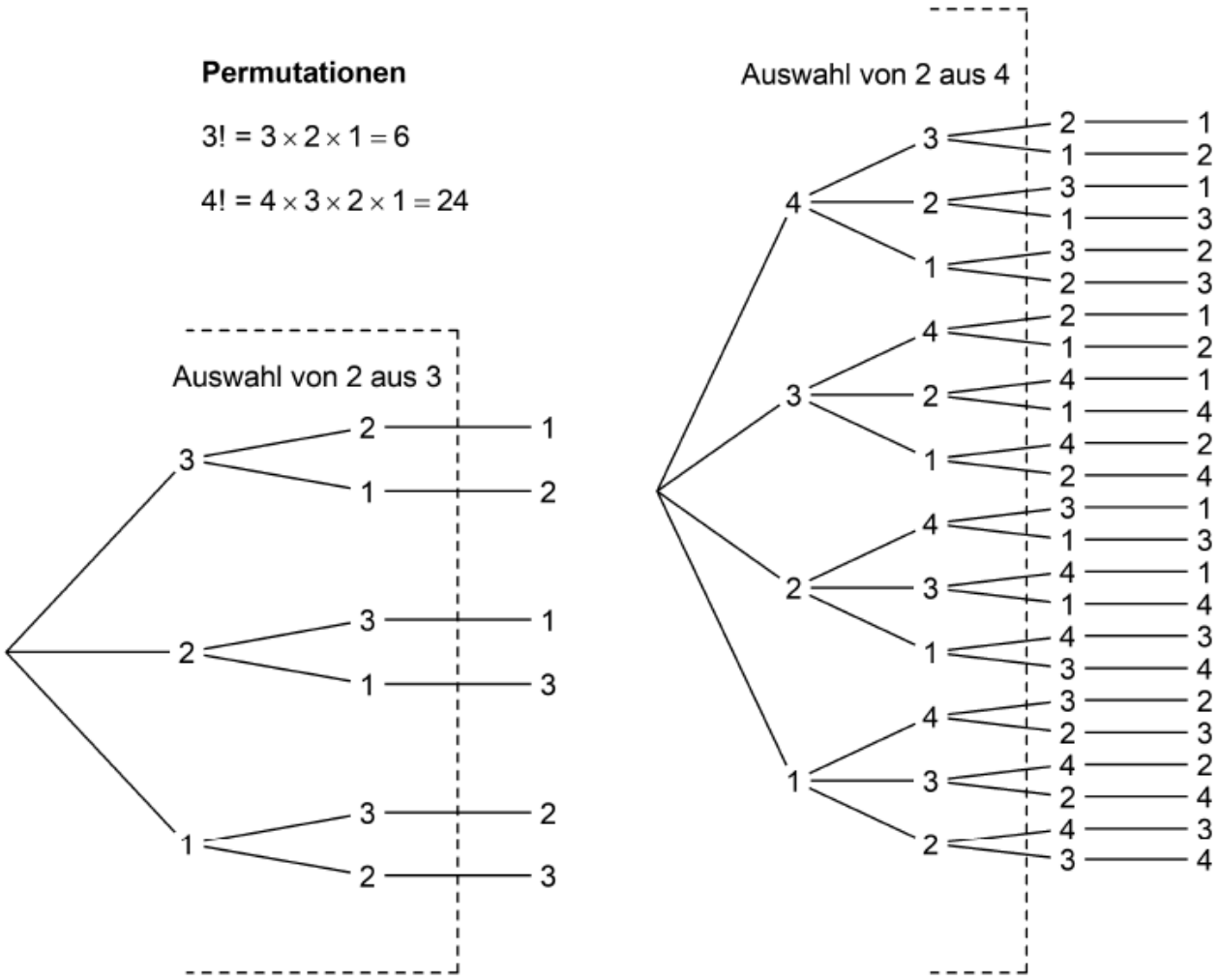


Abb. 7.3.1: Anzahl von Permutationen und Auswahlmöglichkeiten

Einpunktverteilung

• Modell und Definition •

Eine Zufallsvariable X, die nur einen bestimmten Wert a annehmen kann, für die also gilt:

$$P(X = a) = 1 \quad \text{und} \quad P(X = x) = 0 \quad \text{für} \quad x \neq a$$

heißt **einpunktverteilt auf *a***. Man spricht dann auch von einer **deterministischen** oder **degenerierten Verteilung**.

• Erwartungswert und Varianz •

Ist X einpunktverteilt auf a, so gilt:

$$E(X) = a \quad \text{und} \quad Var(X) = 0$$

7.3 Spezielle eindimensionale Verteilungen

Bernoulli-Verteilung

• Modell und Definition •

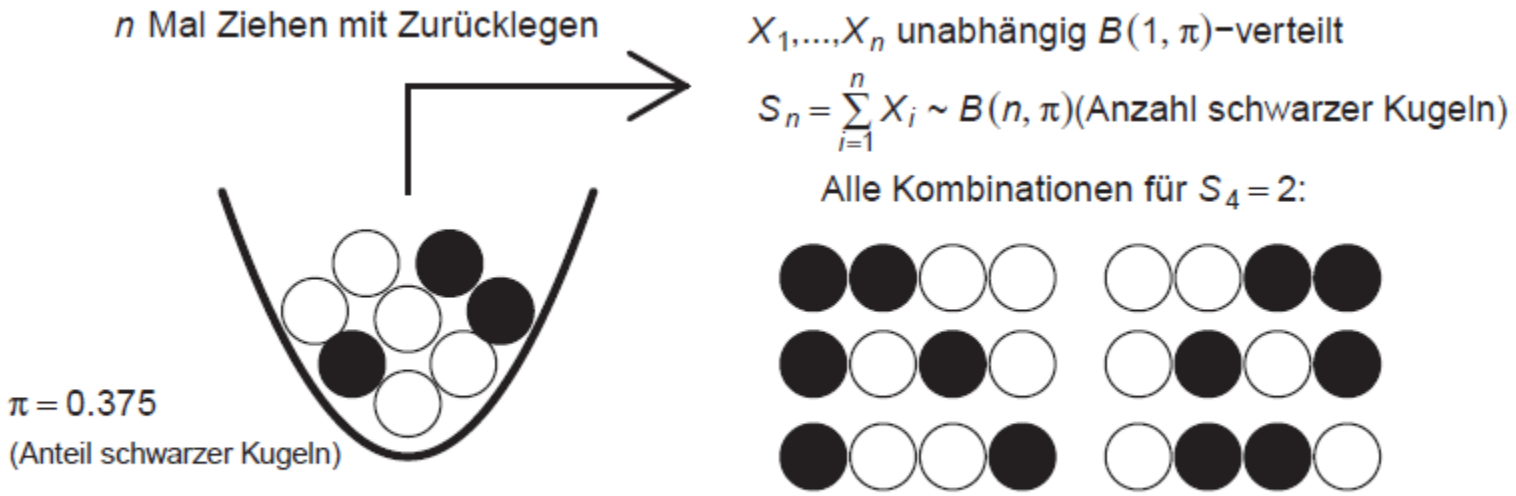


Abb. 7.3.2: Urnenmodell – Ziehen mit Zurücklegen

Aus einer Urne mit Schwarzkugelanteil π wird zufällig eine Kugel gezogen. Sei $X = 1$, falls schwarz und $X = 0$, falls weiß. Dann genügt X einer **Bernoulli-Verteilung** mit Parameter π.

7.3 Spezielle eindimensionale Verteilungen

Bernoulli-Verteilung

Eine Zufallsvariable X heißt **Bernoulli-verteilt** mit Parameter $\pi \in (0, 1)$, kurz $\boldsymbol{X \sim B(1, \pi)}$, wenn gilt:

$$P(X = 0) = 1 - \pi \quad \text{und} \quad P(X = 1) = \pi.$$

Die korrespondierende Verteilung heißt **Bernoulli-Verteilung**. Dabei gilt:

$$E(X) = \pi \quad \text{und} \quad Var(X) = \pi(1 - \pi).$$

- **Bernoulli-Verteilung als parametrische Verteilungsfamilie**
- **Erwartungswert und Varianz**

$$E(X) = 1 \times \pi + 0 \times (1 - \pi) = \pi$$

$$E(X^2) = 1^2 \times \pi + 0^2 \times (1 - \pi) = \pi$$

$$\Rightarrow \quad Var(X) = E(X^2) - E(X)^2 = \pi - \pi^2 = \pi(1 - \pi)$$

Binomialverteilung

• Modell und Definition •

Gegeben sei eine Situation wie auf Folie 497:
Aus einer Urne mit Schwarzkugelanteil π werden zufällig n Kugeln mit Zurücklegen gezogen. Dies definiere die Zufallsvariablen $X_1, X_2, \ldots, X_n$.
Dabei gelte:

$X_i = 0$, falls im i-ten Zug eine weiße und
$X_i = 1$, falls im i-ten Zug eine schwarze

Kugel gezogen wird. Damit sind $X_1, X_2, \ldots, X_n$ unabhängig $B(1, \pi)$-verteilt.
Die Summe

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

entspricht der Anzahl der schwarzen Kugeln in der Stichprobe und genügt dann einer sog. **Binomialverteilung** mit Parametern n und π.

Speziell in der Situation von Folie 497 würde dann gelten:

$$X_i \sim B(1,0.375) \text{ für } i = 1, \ldots, n$$

Die Summe S_4 genügt dann einer Binomialverteilung mit Parameter $n = 4$ und $\pi = 0.375$, kurz

$$S_4 \sim B(4,0.375).$$

Dabei gilt z. B.:

$$P(S_4 = 0) = \binom{4}{0} 0.375^0 (1 - 0.375)^{4-0} = 0.625^4 \approx 0.1526$$

$$P(S_4 = 1) = \binom{4}{1} 0.375^1 (1 - 0.375)^{4-1} = 4 \times 0.375 \times 0.625^3 \approx 0.3662$$

$$P(S_4 = 2) = \binom{4}{2} 0.375^2 (1 - 0.375)^{4-2} = 6 \times 0.375^2 \times 0.625^2 \approx 0.3296$$

Man beachte:

$$\binom{4}{0} = \binom{4}{4} = 1$$

7.3 Spezielle eindimensionale Verteilungen

- **Bernoulli-Verteilung als Spezialfall**
- **Binomial-Verteilung bei unterschiedlichen Parameterwerten**

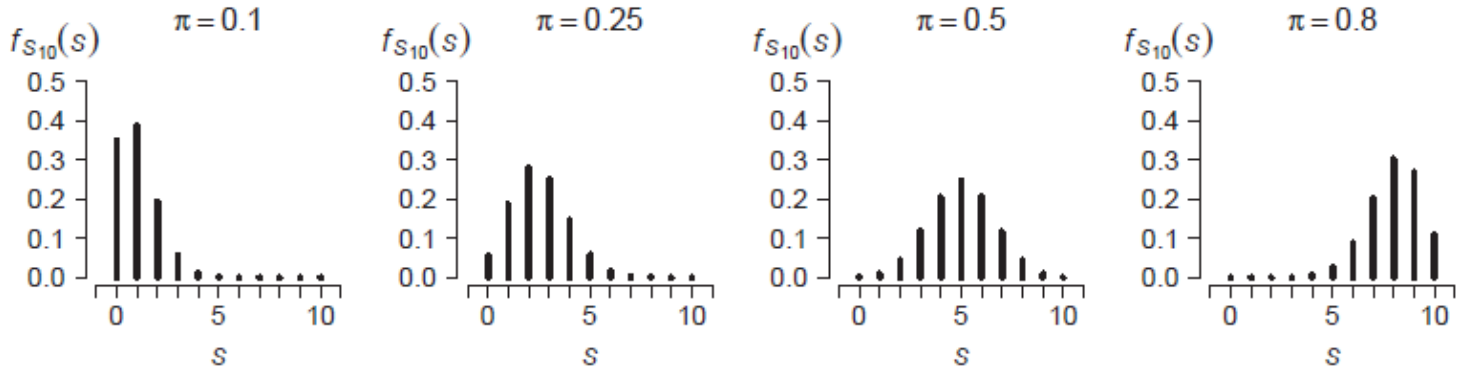


Abb. 7.3.3: $B(10, \pi)$-Verteilung für unterschiedliche Werte von π

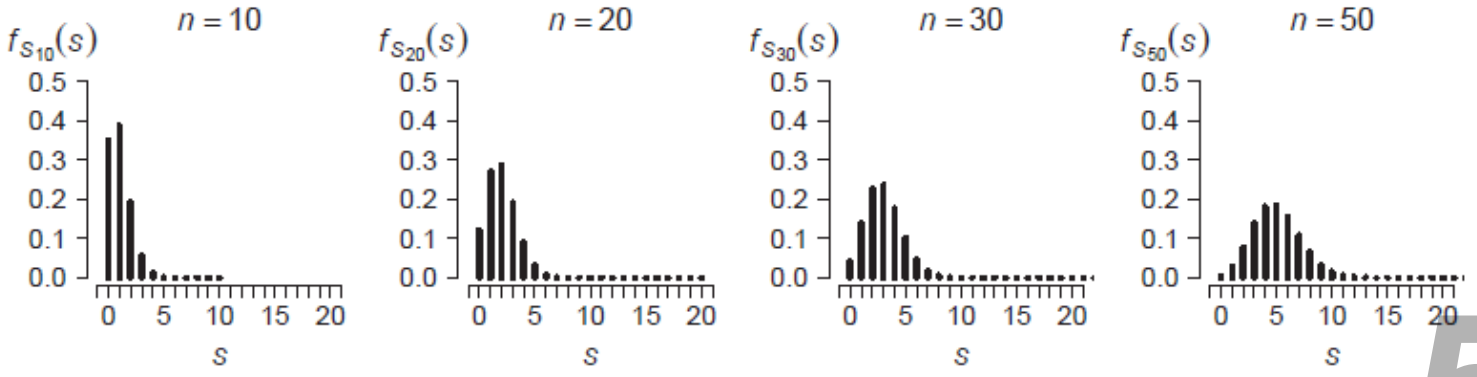


Abb. 7.3.4: $B(n, 0.1)$-Verteilung für unterschiedliche Werte von n

• Erwartungswert und Varianz •

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

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

Binomialverteilung

Eine Zufallsvariable S_n heißt **binomialverteilt** mit Parametern $n \in \mathbb{N}$ und $\pi \in (0, 1)$, kurz $\boldsymbol{S_n \sim B(n, \pi)}$, wenn gilt:

$$P(S_n = s) = \binom{n}{s} \pi^s (1-\pi)^{n-s} \quad \text{für } s = 0, 1, 2, \ldots, n.$$

Die korrespondierende Verteilung heißt **Binomialverteilung**. Dabei gilt:

$$E(S_n) = n\pi \quad \text{und} \quad Var(S_n) = n\pi(1-\pi).$$

● Anmerkung zur Notation „S_n“ ●

Stellt Interpretierbarkeit als Summe heraus

● Beispiel 7.3.2 ●

Angenommen eine Versicherung schließt Versicherungsverträge mit $n = 100$ Kunden ab. Aus früheren Datenbeständen weiß man, dass die Wahrscheinlichkeit für einen Versicherungsfall innerhalb eines Jahres pro Kunde bei etwa 0.5% liegt. Mit welcher Wahrscheinlichkeit wird dann im nächsten Jahr mindestens ein Versicherungsfall eintreten? Mit welcher Wahrscheinlichkeit werden höchstens zwei Versicherungsfälle eintreten? Solche und ähnliche Fragen lassen sich nun unter Umständen mithilfe einer Binomialverteilung beantworten.

Erscheint hier Modellierung über Binomialverteilung überhaupt adäquat? Unabhängigkeit für den Eintritt der einzelnen Todesfälle bei einer Lebensversicherung erscheint weitestgehend realitätsnah. Anders wäre es beispielsweise bei einer Versicherung für Hochwasserschäden. Warum?

Unter der Annahme einer $B(100, 0.005)$-Verteilung würde gelten:

$$P(S_{100} = 0) = \binom{100}{0} 0.005^0 (1 - 0.005)^{100-0} \approx 0.6058$$

$$P(S_{100} = 1) = \binom{100}{1} 0.005^1 (1 - 0.005)^{100-1} \approx 0.3044$$

$$P(S_{100} = 2) = \binom{100}{2} 0.005^2 (1 - 0.005)^{100-2} \approx 0.0757$$

Die Wahrscheinlichkeit, dass mindestens ein Versicherungsfall eintritt, beträgt

$$P(S_{100} > 0) = 1 - P(S_{100} = 0) \approx 1 - 0.6058 = 0.3942$$

und die Wahrscheinlichkeit, dass höchstens zwei Versicherungsfälle eintreten

$$P(S_{100} \leq 2) = P(S_{100} = 0) + P(S_{100} = 1) + P(S_{100} = 2) \approx 0.9859$$

Insbesondere die der Binomialverteilung innewohnende Unabhängigkeitsannahme ist bei statistischen Modellierungen kritisch zu prüfen.

Poisson-Verteilung

• Modell und Definition •

Sei X die Anzahl des Auftretens eines bestimmten Ereignisses in einem fest vorgegebenen Zeitraum. Dann kann X häufig als **Poisson-verteilt mit Parameter λ** angenommen werden. Die Wahrscheinlichkeitsfunktion lautet dann

$$f_X(x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad \text{für } x = 0, 1, 2, \ldots \text{ und } \lambda > 0.$$

• Zusammenhang zwischen Binomialverteilung und Poisson-Verteilung •

Zusammenhang zwischen Binomialverteilung und Poisson-Verteilung

Eine $B(n, \pi)$-verteite Zufallsvariable S_n ist für „kleines“ π und „großes“ n approximativ $Po(\lambda)$-verteit mit $\lambda = n\pi$. Es gilt dann also:

$$P(S_n = s) = \binom{n}{s} \pi^s (1-\pi)^{n-s} \approx \frac{\lambda^s}{s!} e^{-\lambda}.$$

Die Approximation ist für $\pi < 0.05$ und $n \geq 30$ hinreichend gut.

● Poisson-Verteilung bei unterschiedlichen Parameterwerten ●

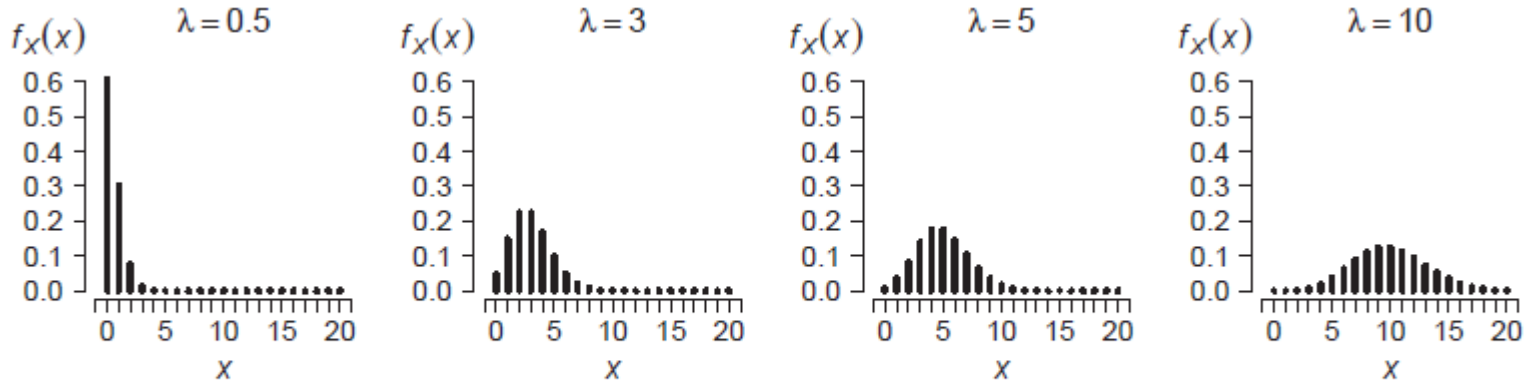


Abb. 7.3.5: $Po(\lambda)$-Verteilung für unterschiedliche Werte von λ

● Erwartungswert und Varianz ●

Für $X \sim Po(\lambda)$ kann dann mit etwas Aufwand gezeigt werden

$$E(X) = \sum_{j=0}^{\infty} \frac{j\lambda^j}{j!} e^{-\lambda} = 0 + \frac{\lambda^1}{1!} e^{-\lambda} + \frac{2\lambda^2}{2!} e^{-\lambda} + \cdots = \lambda$$

$$E(X^2) = \sum_{j=0}^{\infty} \frac{j^2\lambda^j}{j!} e^{-\lambda} = 0 + \frac{\lambda^1}{1!} e^{-\lambda} + \frac{4\lambda^2}{2!} e^{-\lambda} + \cdots = \lambda + \lambda^2$$

Poisson-Verteilung

Eine Zufallsvariable X heißt **poissonverteilt** mit Parameter $\lambda > 0$, kurz $\boldsymbol{X \sim Po(\lambda)}$, wenn gilt:

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad \text{für } x = 0, 1, 2, \ldots$$

Die korrespondierende Verteilung heißt ***Poisson-Verteilung***. Dabei gilt:

$$E(X) = \lambda \quad \text{und} \quad Var(X) = \lambda.$$

● Beispiel 7.3.3 ●

Betrachten wir nochmals Beispiel D1-b aus Abschnitt 7.1.1, in dem es um die Anzahl eingehender Notrufe an einem Rettungswagen-Stützpunkt während einer Stunde ging. Für die Anzahl X wurde dabei eine $Po(3)$-Verteilung unterstellt, so dass laut Modell im Durchschnitt theoretisch 3 Notrufe eingehen sollten. Die Wahl von $\lambda = 3$ kann andersherum argumentiert auf empirischen Erfahrungen beruhen. Weshalb erscheint eine Poisson-Verteilung in einem solchen Kontext jedoch überhaupt aus theoretischer Sicht gerechtfertigt?

> Argumente für eine Poisson-Verteilung:

– Anzahl von Anrufen (theoretisch) nicht nach oben beschränkt

– Anzahl potenzieller Anrufer „sehr groß“ bei gleichzeitig geringer Eintrittswahrscheinlichkeit und (weitgehender) Unabhängigkeit
$\Rightarrow \quad B(n,\pi) \approx Po(n\pi)$

– Empirische Bestätigung, dass eine Poisson-Verteilung gut passt

> Beispiel: In Beispiel 7.3.2 (Folie 503) könnte man approximieren

$$S_{100} \sim B(100,0.005) \approx Po(100 \times 0.005) = Po(0.5)$$

$$\Rightarrow \quad P(S_{100} = 0) \approx \frac{0.5^0}{0!} e^{-0.5} \approx 0.6065$$

$$P(S_{100} = 1) \approx \frac{0.5^1}{1!} e^{-0.5} \approx 0.3033$$

$$P(S_{100} = 2) \approx \frac{0.5^2}{2!} e^{-0.5} \approx 0.0758$$

7.3.2 Spezielle stetige Verteilungen

Stetige Gleichverteilung

- **Modell und Definition**
- **Stetige Gleichverteilung bei unterschiedlichen Parameterwerten**

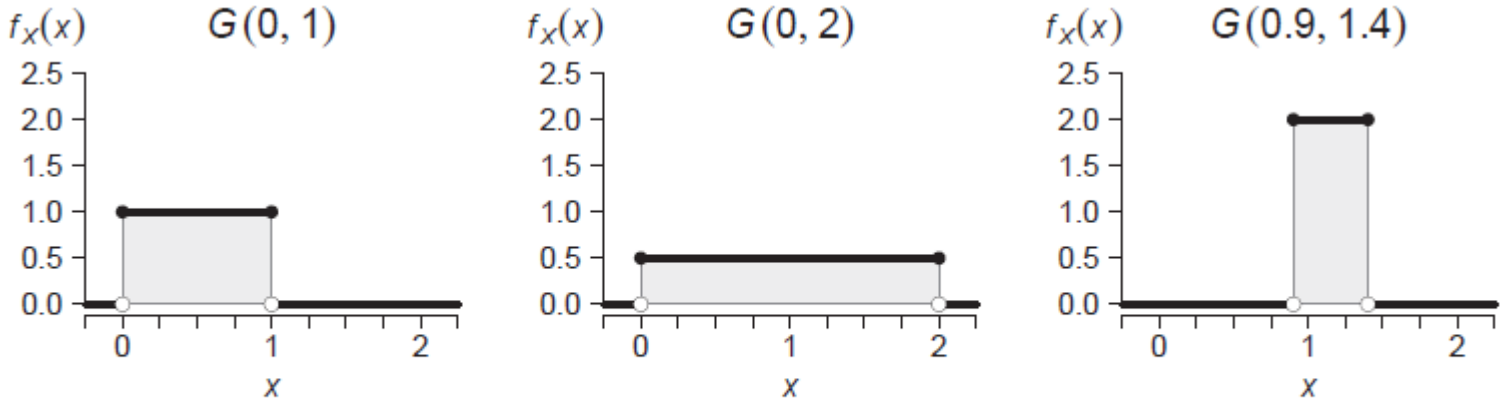


Abb. 7.3.6: $G(a, b)$-Verteilung für unterschiedliche Intervalle $[a, b]$

7.3 Spezielle eindimensionale Verteilungen

Stetige Gleichverteilung

Für $a < b$ heißt eine Zufallsvariable X **auf** $[a, b]$ **stetig gleichverteilt**, kurz $X \sim G(a, b)$, wenn sie die Dichte

$$f_X(x) = \frac{1}{b-a} I_{[a,b]}(x)$$

besitzt. Die korrespondierende Verteilung heißt **stetige Gleichverteilung** oder **Rechteckverteilung**. Dabei gilt:

$$E(X) = \frac{a+b}{2} \quad \text{und} \quad Var(X) = \frac{(b-a)^2}{12}.$$

• Erwartungswert und Varianz •

$$E(X) = \int_a^b \frac{x}{b-a}\,dx \ldots = \frac{a+b}{2}$$

$$E(X^2) = \int_a^b \frac{x^2}{b-a}\,dx = \cdots = \frac{(a+b)^2 - ab}{3}$$

$$\Rightarrow \quad Var(X) = E(X^2) - E(X)^2 = \frac{(b-a)^2}{12}$$

• Verteilungsfunktion •

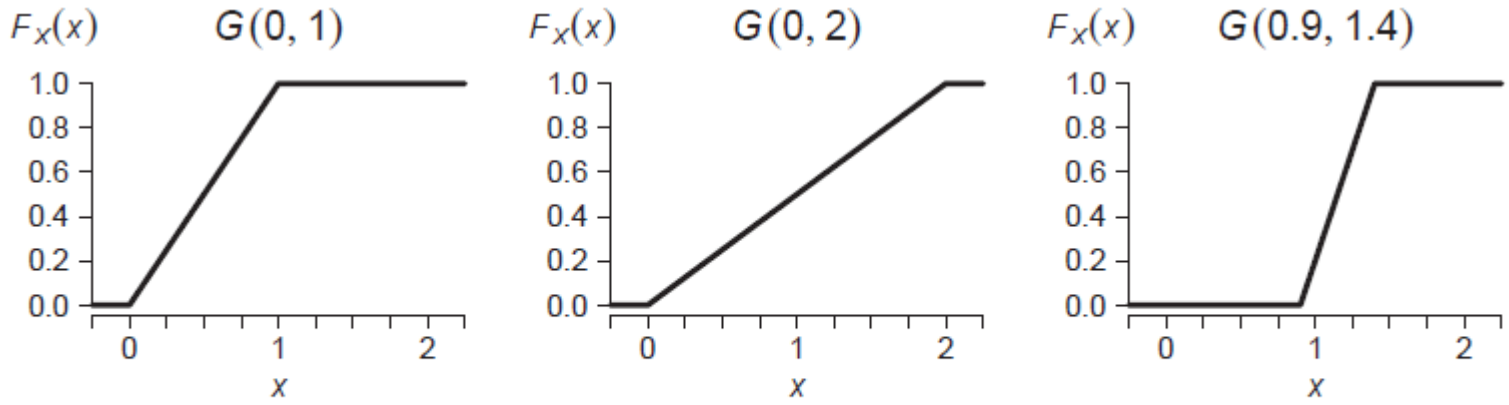


Abb. 7.3.7: Verteilungsfunktionen stetiger Gleichverteilungen

• Beispiel 7.3.4 •

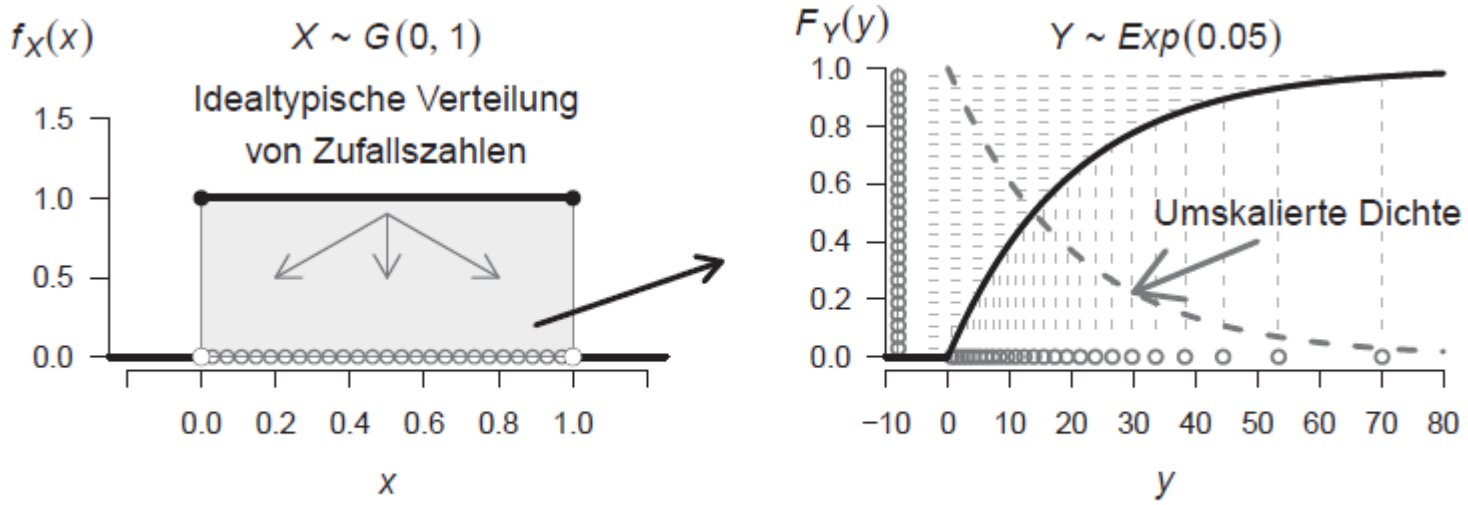


Abb. 7.3.8: Zufallszahlenerzeugung illustriert am Beispiel S1-b

Exponentialverteilung

• Modell und Definition •

Sei X eine Wartezeit (auf ein bestimmte Ereignis) oder eine bestimmte Lebensdauer (einer Person, eines Objektes). Dann kann X häufig über eine **Exponentialverteilung mit Parameter λ** modelliert werden. Die Dichtefunktion dieser stetigen Verteilung lautet dann

$$f_X(x) = \lambda e^{-\lambda x} I_{[0,\infty)}(x) \text{ für } \lambda > 0.$$

• Zusammenhang zwischen Poisson-Verteilung und Exponentialverteilung •

Zusammenhang zwischen Poisson-Verteilung und Exponentialverteilung

Sei X die Anzahl des Auftretens eines bestimmten Ereignisses in einem bestimmten Zeitfenster, dessen Länge in einer bestimmten Einheit gemessen wird. Ist X $Po(\lambda)$-verteilt, so ist die in der gleichen Einheit gemessene Wartezeit zwischen aufeinanderfolgenden Ereignissen $Exp(\lambda)$-verteilt.

● Eigenschaft der Gedächtnislosigkeit ●

Eine der Exponentialverteilung innewohnende Eigenschaft ist die der sog. **Gedächtnislosigkeit**. Dabei gilt, dass für jeden Zeitpunkt t eine noch verbleibende Wartezeit (bis zum nächsten Eintritt eines bestimmten Ereignisses) oder Lebensdauer (bis zum Verfall eines Objekts) nicht von der bereits bis t verstrichenen Wartezeit bzw. Lebensdauer abhängt. Formal ausgedrückt:

$$P(X \leq t + s | X > t) = P(X \leq s) \quad \text{für} \quad t, s \in [0, \infty).$$

Beispiel:

$$P(X \leq 2 | X > 1) = P(X \leq 1)$$

$$P(X \leq 4 | X > 3) = P(X \leq 1)$$

$$P(X \leq 28.5 | X > 27.5) = P(X \leq 1)$$

Die Wahrscheinlichkeit eine weitere Minute zu warten, wenn bereits 1, 3 oder 27.5 Minuten verstrichen sind, ist gleich der Wahrscheinlichkeit überhaupt höchstens 1 Minute warten zu müssen.

7.3 Spezielle eindimensionale Verteilungen

● Exponentialverteilung bei unterschiedlichen Parameterwerten ●

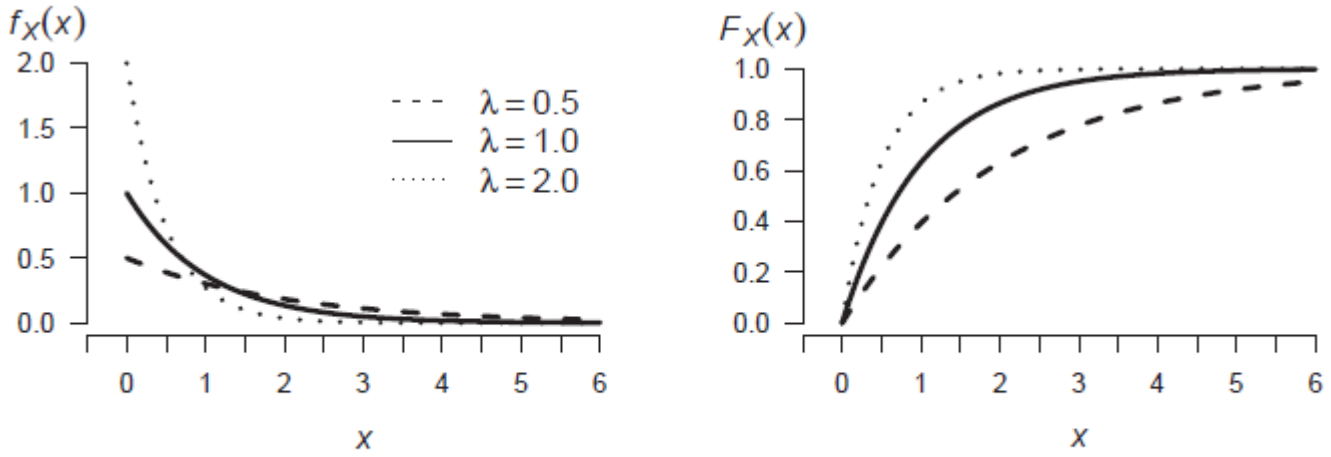


Abb. 7.3.9: *Exp*(λ)-Verteilung für unterschiedliche Werte von λ

● Erwartungswert und Varianz ●

Für $X \sim Exp(\lambda)$ kann dann mit etwas Aufwand gezeigt werden

$$E(X) = \int_0^\infty \lambda x e^{-\lambda x}\, dx = \cdots = \frac{1}{\lambda}$$

$$E(X^2) = \int_0^\infty \lambda x^2 e^{-\lambda x}\, dx = \cdots = \frac{2}{\lambda^2}$$

Exponentialverteilung

Eine stetige Zufallsvariable X heißt **exponentialverteilt** mit Parameter $\lambda > 0$, kurz $\boldsymbol{X \sim Exp(\lambda)}$, wenn sie die Dichte

$$f_X(x) = \lambda e^{-\lambda x} I_{[0,\infty)}(x)$$

besitzt. Die korrespondierende Verteilung heißt **Exponentialverteilung**. Dabei gilt:

$$E(X) = \frac{1}{\lambda} \quad \text{und} \quad Var(X) = \frac{1}{\lambda^2}.$$

• Verteilungsfunktion •

Vgl. Folie 354 (Beispiel S1-b) speziell für $\lambda = 0.05$

Allgemeine Lösung: $F_X(x) = \left(1 - e^{-\lambda x}\right) I_{[0,\infty)}(x)$

● Beispiel 7.3.5 ●

Beispiel 1
Betrachten wir nochmals Beispiel S1-b (Folie 347), in dem es um die Wartezeit zwischen aufeinanderfolgenden Notrufen ging. Die betreffende Wartezeit in Minuten wurde als $Exp(0.05)$-verteilt angenommen. Über den Zusammenhang zwischen Poisson-Verteilung und Exponentialverteilung korrespondiert das Beispiel direkt mit Beispiel D1-b, in dem es um die Anzahl eingehender Notrufe an einem Rettungswagen-Stützpunkt während einer Stunde geht. Diese wurde als $Po(3)$-verteilt angenommen. Insofern erscheint eine Modellierung der Wartezeit zwischen zwei Notrufen mit einer Exponentialverteilung dann adäquat, falls eine Modellierung der Anzahl während einer Stunde mit einer Poisson-Verteilung adäquat erscheint. Außerdem kann zusätzlich die Eigenschaft der Gedächtnislosigkeit separat auf Plausibilität geprüft werden.

Beispiel 2
Angenommen eine U-Bahnlinie verkehre im 20-Minuten-Takt. Man komme ohne Kenntnis des Fahrplans zufällig an den Bahnsteig. Dann könnte man die anstehende Wartezeit als auf dem Intervall [0, 20] stetig gleichverteilt annehmen. In diesem Fall würde aber die Wahrscheinlichkeit stetig zunehmen, dass die U-Bahn in den nächsten 5 Minuten eintritt, je länger man bereits warten würde. Nach 10 Minuten würde diese genau 50% betragen, da

$$P(X \leq 15 | X > 10) = \frac{P(10 < X \leq 15)}{1 - P(X \leq 10)} = \frac{0.25}{1 - 0.5} = 0.5$$

Insbesondere die der Exponentialverteilung innewohnende Eigenschaft der Gedächtnislosigkeit ist bei statistischen Modellierungen kritisch zu prüfen.

7.3 Spezielle eindimensionale Verteilungen

Normalverteilung

- **Modell und Definition** -

Dichte: $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$ Notation: $X \sim N(\mu, \sigma^2)$

- **Normalverteilung bei unterschiedlichen Parameterwerten** -

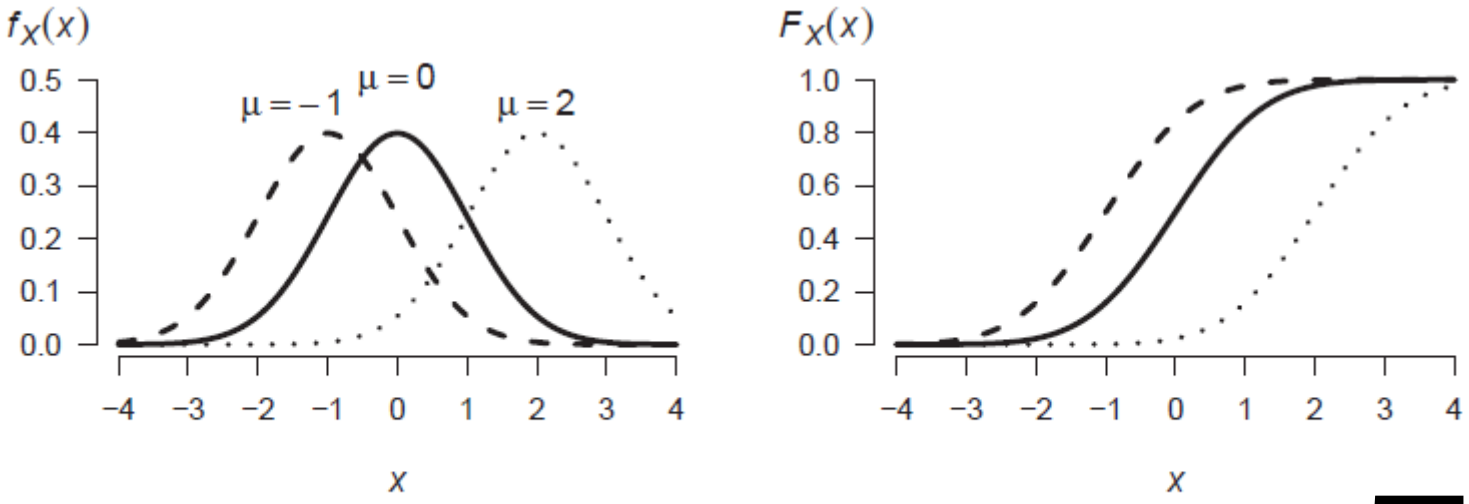


Abb. 7.3.10: $N(\mu, 1)$-Verteilung für unterschiedliche Werte von μ

7.3 Spezielle eindimensionale Verteilungen

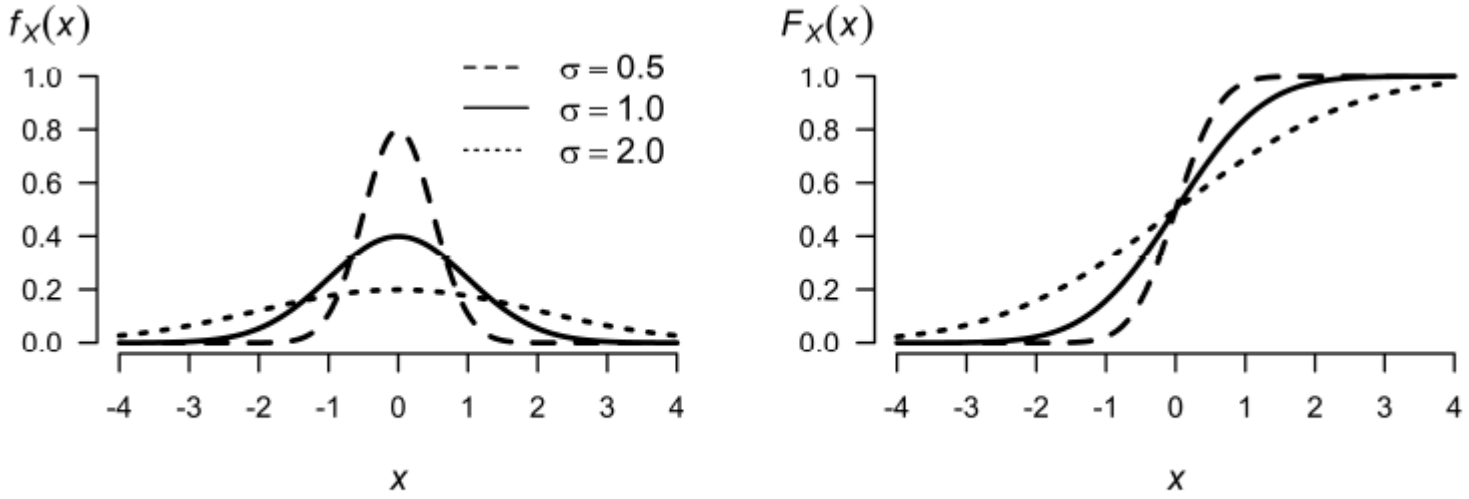


Abb. 7.3.11: $N(0, \sigma^2)$-Verteilung für unterschiedliche Werte von σ^2

- **Erwartungswert und Varianz**
- **Lineartransformationen bei Normalverteilungen**

Satz 7.3.1: Lineartransformation bei Normalverteilung

Falls $X \sim N(\mu, \sigma^2)$-verteilt ist, so gilt für $b \neq 0$:

$$Z = a + bX \sim N(a + b\mu, b^2\sigma^2).$$

7.3 Spezielle eindimensionale Verteilungen

• Standardisierung bei Normalverteilung •

Standardisierung bei Normalverteilung

Ist $X \sim N(\mu, \sigma^2)$-verteilt ist, so ist die standardisierte Größe

$$Z = \frac{X - \mu}{\sigma}$$

standardnormalverteilt, d.h. $N(0, 1)$-verteilt.

• Verteilungsfunktion •

> Verteilungsfunktion für $N(0, 1)$: $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} exp\left(-\frac{1}{2}t^2\right) dt$

> Verteilungsfunktion für $N(\mu, \sigma^2)$: $F_X(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$

Man beachte: $P(X \leq x) = P\left(\underbrace{\frac{X - \mu}{\sigma}}_{N(0,1)} \leq \frac{x - \mu}{\sigma}\right)$

● Theoretische Quantile ●

Sei q_α das theoretische α-Quantil einer $N(\mu, \sigma^2)$-verteilten Zufallsvariable X.

Dann gilt: $P(X \leq q_\alpha) = \alpha$

$$\Leftrightarrow \quad P\left(\underbrace{\frac{X-\mu}{\sigma}}_{N(0,1)} \leq \frac{q_\alpha - \mu}{\sigma}\right) = \alpha$$

$$\Leftrightarrow \quad \Phi\left(\frac{q_\alpha - \mu}{\sigma}\right) = \alpha$$

$$\Leftrightarrow \quad z_\alpha = \frac{q_\alpha - \mu}{\sigma}$$

$$\Leftrightarrow \quad q_\alpha = \mu + \sigma z_\alpha$$

• Zusammenfassung für die Normalverteilung •

Zusammenfassung für die Normalverteilung

Eine stetige Zufallsvariable X heißt **normalverteilt** mit Erwartungswert μ und Varianz $\sigma^2 > 0$, kurz $\boldsymbol{X \sim N(\mu, \sigma^2)}$, wenn sie die Dichte

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

besitzt. Speziell heißt die $N(0, 1)$-Verteilung auch **Standardnormalverteilung**. Mit φ bezeichnen wir die Dichte der Standardnormalverteilung. Für die Verteilungsfunktion F_X einer $N(\mu, \sigma^2)$-verteilten Zufallsvariable X gilt:

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

wobei Φ die Verteilungsfunktion der $N(0, 1)$-Verteilung bezeichnet, deren Werte aus Verteilungstafeln (vgl. A.1) abgelesen werden können. Außerdem gilt:

$$q_\alpha = \mu + \sigma \cdot z_\alpha,$$

wobei q_α das theoretische α-Quantil einer $N(\mu, \sigma^2)$-Verteilung und z_α das theoretische α-Quantil der $N(0, 1)$-Verteilung bezeichnet.

• Vertafelung der Standardnormalverteilung •

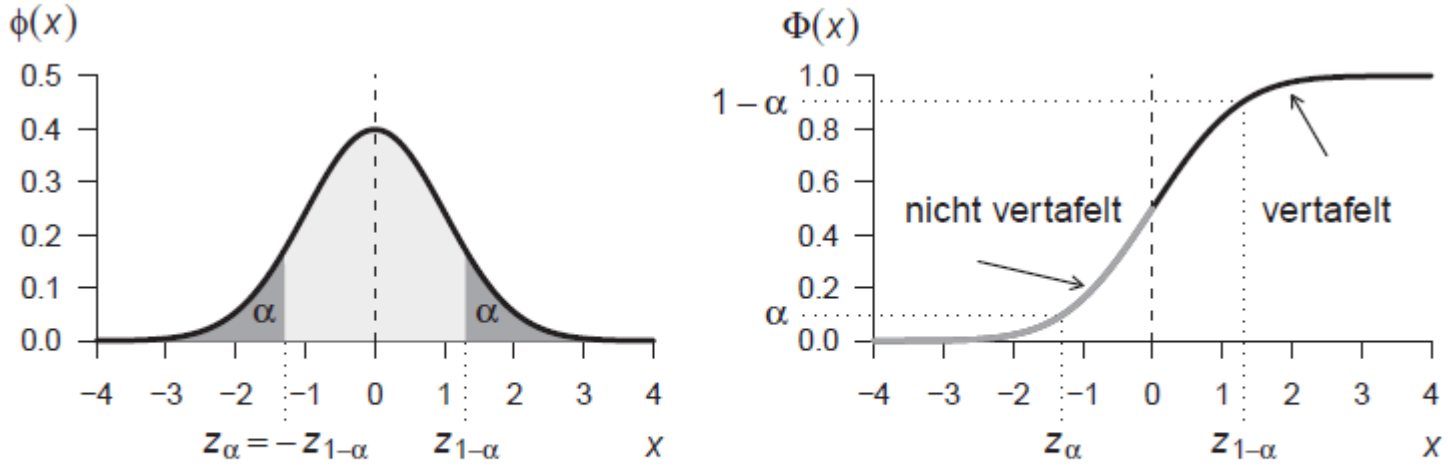


Abb. 7.3.12: Dichte, Verteilungsfunktion und Quantile der Standardnormalverteilung

• Zur Lesart der Tabelle •

Für eine $N(0, 1)$-verteilte Zufallsvariable X gilt:

$$P(X \leq 1.75) \approx 0.9599$$

$$P(X \leq -1.75) = \Phi(-1.75) = 1 - \Phi(1.75)$$
$$\approx 1 - 0.9599 = 0.0401$$

7.3 Spezielle eindimensionale Verteilungen

Tab. A.1: Wertetabelle der Standardnormalverteilung

Tabelliert sind die Werte der Verteilungsfunktion der Normalverteilung, $\Phi(z)$.
Ablesebeispiel: $\Phi(1.23) = \Phi(1.2 + 0.03) \approx 0.8907$.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

● Beispiel 7.3.6 ●

Gegeben sei eine $N(\mu, \sigma^2)$-verteilte Zufallsvariable X.

Wie berechnet man dann z. B.

(i) $P(X < 0)$, (ii) $P(0 < X < 1)$, (iii) $q_{0.05}$, (iv) $q_{0.95}$

ad (i):

$$P(X < 0) = P(X \leq 0) = P\left(\frac{X-1}{\sqrt{4}} \leq \frac{0-1}{\sqrt{4}}\right) = \Phi\left(\frac{0-1}{\sqrt{4}}\right)$$

(kann entfallen) vgl. Folie 521

$$= \Phi(-0.5) = 1 - \Phi(0.5)$$

$$\approx 1 - 0.6915$$

$$= 0.3085$$

ad (ii):

$$P(0 < X < 1) = P(0 < X \leq 1) = P(X \leq 1) - P(X \leq 0)$$

$$= P\left(\frac{X-1}{2} \leq \frac{1-1}{2}\right) - P\left(\frac{X-1}{2} \leq \frac{0-1}{2}\right)$$ (kann entfallen)

$$= \Phi(0) - \Phi(-0.5)$$

$$\approx 0.5 - 0.3085 = 0.1915$$

ad (iii) und (iv):

$$q_{0.95} = 1 + 2z_{0.95} \approx 1 + 2 \times 1.64 = 4.28$$

$$q_{0.05} = 1 + 2z_{0.05} = 1 - 2z_{0.95} \approx 1 - 2 \times 1.64 = -2.28$$

7.3 Spezielle eindimensionale Verteilungen

● Theoretische Schwankungsintervalle ●

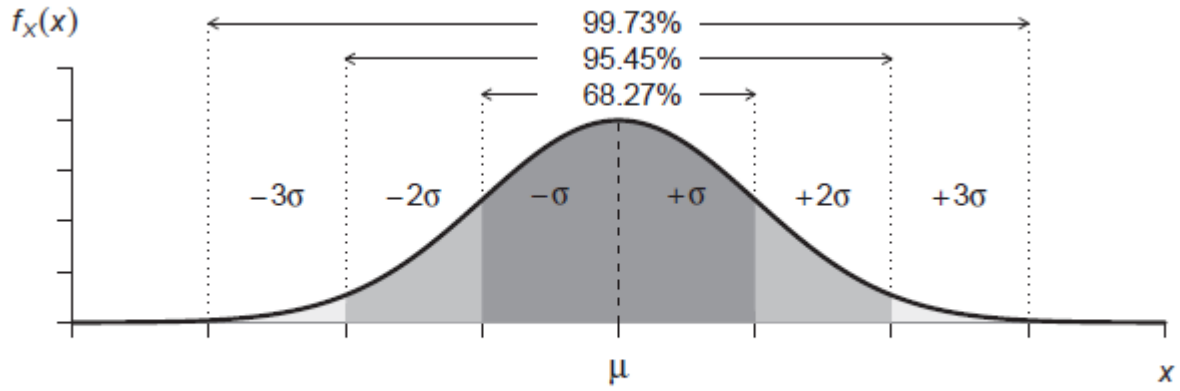


Abb. 7.3.13: Normalverteilung und theoretische Schwankungsintervalle

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = P\left(-2 < \frac{X-\mu}{\sigma} \leq 2\right)$$

$$= \Phi(2) - \Phi(-2) = \Phi(2) - \big(1 - \Phi(2)\big)$$

$$= 2 \times \Phi(2) - 1 \approx 2 \times 0.9772 - 1$$

$$= 0.9545$$

Analog die anderen Schwankungsintervalle...

7.4.1 Exakte Aussagen

Erwartungswerte und Varianzen

● **Grundrahmen und benötigte Resultate** ●

Wir betrachten n Zufallsvariablen $X_1, X_2, ..., X_n$ mit

$$\mu_1, \dots, \mu_n \quad \text{bzw.} \quad \sigma_1^2, \dots, \sigma_n^2$$

Dazu definieren wir folgende stochastische und theoretische Statistiken

$$S_n = \sum_{i=1}^{n} X_i\,, \qquad \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i\,, \qquad \bar{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \mu_i\,, \qquad \bar{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} \sigma_i^2$$

Man beachte im Folgenden folgendes Resultat:

$$Cov(X,X) = E[(X-\mu)(X-\mu)] = E[(X-\mu)^2] = Var(X)$$

7.4 Verteilung stochastischer Summen und Mittelwerte

- **Erwartungswerte stochastischer Summen** -

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

- **Erwartungswerte stochastischer Mittelwerte** -

$$E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu_i = \bar{\mu}_n$$

- **Varianz stochastischer Summen** -

> Für 2 Zufallsvariablen gilt:

$$Var(X_1 + X_2) = E[(X_1 + X_2 - \mu_1 - \mu_2)^2]$$

$$= E[(X_1 + X_2 - \mu_1 - \mu_2)(X_1 + X_2 - \mu_1 - \mu_2)]$$

$$= E\big[\big((X_1 - \mu_1) + (X_2 - \mu_2)\big)\big((X_1 - \mu_1) + (X_2 - \mu_2)\big)\big]$$

$$= E[(X_1 - \mu_1)^2 + (X_2 - \mu_2)^2 + 2(X_1 - \mu_1)(X_2 - \mu_2)]$$
$$= E[(X_1 - \mu_1)^2] + E[(X_2 - \mu_2)^2] + 2E[(X_1 - \mu_1)(X_2 - \mu_2)]$$
$$= Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$$

> Auf analoge Weise erhält man für drei Zufallsvariablen

$$Var(X_1 + X_2 + X_3) = Var(X_1) + Var(X_2) + Var(X_3)$$
$$+2Cov(X_1, X_2) + 2Cov(X_1, X_3) + 2Cov(X_2, X_3)$$

$$= Cov(X_1, X_1) + Cov(X_1, X_2) + Cov(X_1, X_3)$$
$$+Cov(X_2, X_1) + Cov(X_2, X_2) + Cov(X_2, X_3)$$
$$+Cov(X_3, X_1) + Cov(X_3, X_2) + Cov(X_3, X_3)$$
$$= \sum_{i=1}^{3} \sum_{j=1}^{3} Cov(X_i, X_j)$$

> Verallgemeinerung auf *n* Zufallsvariablen:

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j)$$

● Varianz stochastischer Mittel ●

$$Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j)$$

Die Varianz der Summe ist gleich der Summe aller wechselseitigen Kovarianzen

Bei Unkorreliertheit und identischen Varianzen folgt:

$$Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{\sigma^2}{n}$$

● Zusammenfassung ● (siehe folgende Folie)

7.4 Verteilung stochastischer Summen und Mittelwerte

Erwartungswerte und Varianzen von Summen und Mittelwerten

Für n Zufallsvariablen $X_1, \ldots, X_n$ mit $E(X_i) = \mu_i$ und $Var(X_i) = \sigma_i^2$ für $i = 1, \ldots, n$ gelten folgende Aussagen allgemein bzw. bei Vorliegen von identischen Erwartungswerten bzw. Varianzen:

$$E(S_n) = \sum_{i=1}^{n} \mu_i \quad \text{bzw.} \quad E(S_n) = n\mu, \tag{7.4.14}$$

$$E(\overline{X}_n) = \overline{\mu}_n \quad \text{bzw.} \quad E(\overline{X}_n) = \mu, \tag{7.4.15}$$

$$Var(S_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} Cov(X_i, X_j), \tag{7.4.16}$$

d.h. speziell für $n = 2$:

$$Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2).$$

Sind $X_1, \ldots, X_n$ paarweise unkorreliert, so gilt:

$$Var(S_n) = \sum_{i=1}^{n} \sigma_i^2 \quad \text{bzw.} \quad Var(S_n) = n\sigma^2, \tag{7.4.17}$$

$$Var(\overline{X}_n) = \overline{\sigma}_n^2/n \quad \text{bzw.} \quad Var(\overline{X}_n) = \sigma^2/n. \tag{7.4.18}$$

Insbesondere gelten die letzten Resultate in (7.4.14), (7.4.15), (7.4.17) und (7.4.18), falls $X_1, \ldots, X_n$ unabhängig und identisch verteilt (u.i.v.) sind.

Verteilungen unter bestimmten Ausgangsverteilungen

• Allgemeines •

Sofern die Dichte oder Verteilungsfunktion der stochastischen Summe vorliegt, lässt sich daraus die Verteilungsfunktion bzw. Dichte des stochastischen Mittels ableiten:

$$F_{\bar{X}_n}(x) = P\left(\frac{S_n}{n} \leq x\right) = P(S_n \leq nx) = F_{S_n}(nx)$$

$$\Rightarrow \quad f_{\bar{X}_n}(x) = n f_{S_n}(nx)$$

• Diskrete und stetige Gleichverteilung •

Sind X_1 und X_2 jeweils stetig gleichverteilt über [0, 1] und unabhängig, so ist die Summe $S_2 = X_1 + X_2$ dreiecksverteilt (ohne Beweis) mit Dichte

$$f_{S_2}(s) = sI_{[0,1]}(s) + (2 - s)I_{(1,2]}(s)$$

$$\Longrightarrow \quad f_{\bar{X}_2}(x) = 2f_{S_2}(2x) = 2 \times \left[2xI_{[0,1]}(2x) + (2-2x)I_{(1,2]}(2x)\right]$$

$$= 4xI_{[0,0.5]}(x) + (4-4x)I_{(0.5,1]}(x)$$

Die letzte Gleichung folgt aufgrund der Äquivalenzen

$$0 \leq 2x \leq 1 \Leftrightarrow 0 \leq x \leq 0.5 \quad \text{und} \quad 1 < 2x \leq 2 \Leftrightarrow 0.5 < x \leq 1$$

7.4 Verteilung stochastischer Summen und Mittelwerte

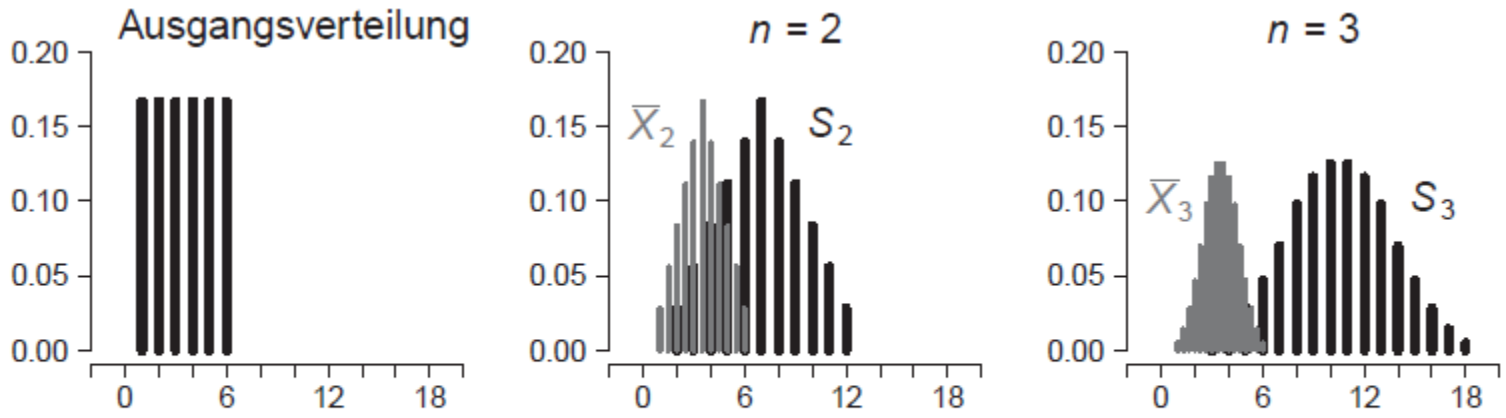


Abb. 7.4.1: Verteilung von Summe und Mittelwert bei einer diskreten Gleichverteilung

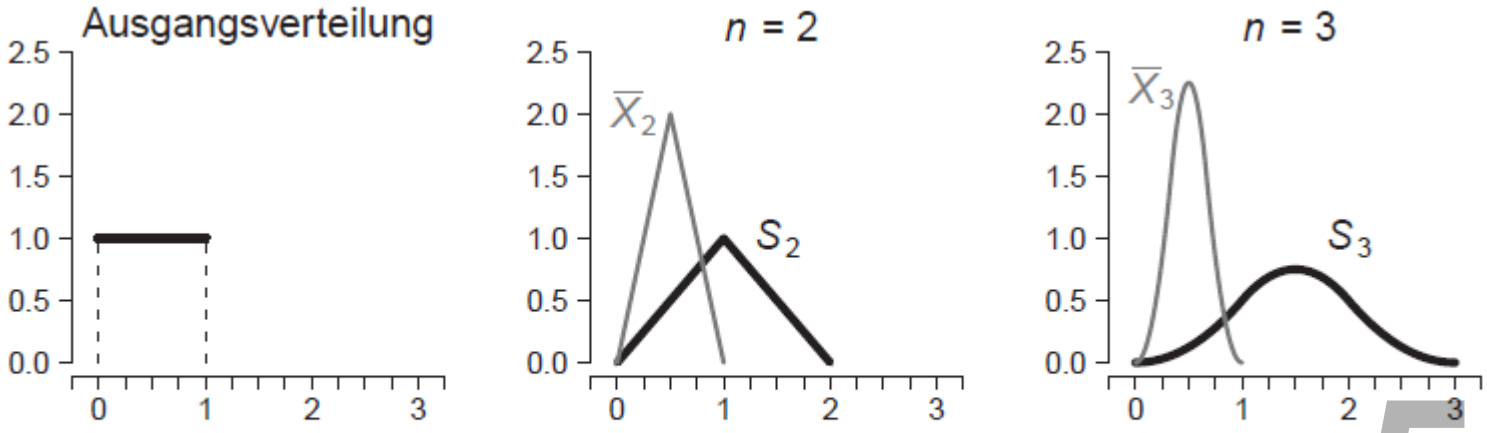


Abb. 7.4.2: Verteilung von Summe und Mittelwert bei einer stetigen Gleichverteilung

7.4 Verteilung stochastischer Summen und Mittelwerte

• Binomialverteilung •

$$B(n_1, \pi) + B(n_2, \pi) = B(n_1 + n_2, \pi).$$

Unabhängigkeit vorausgesetzt!

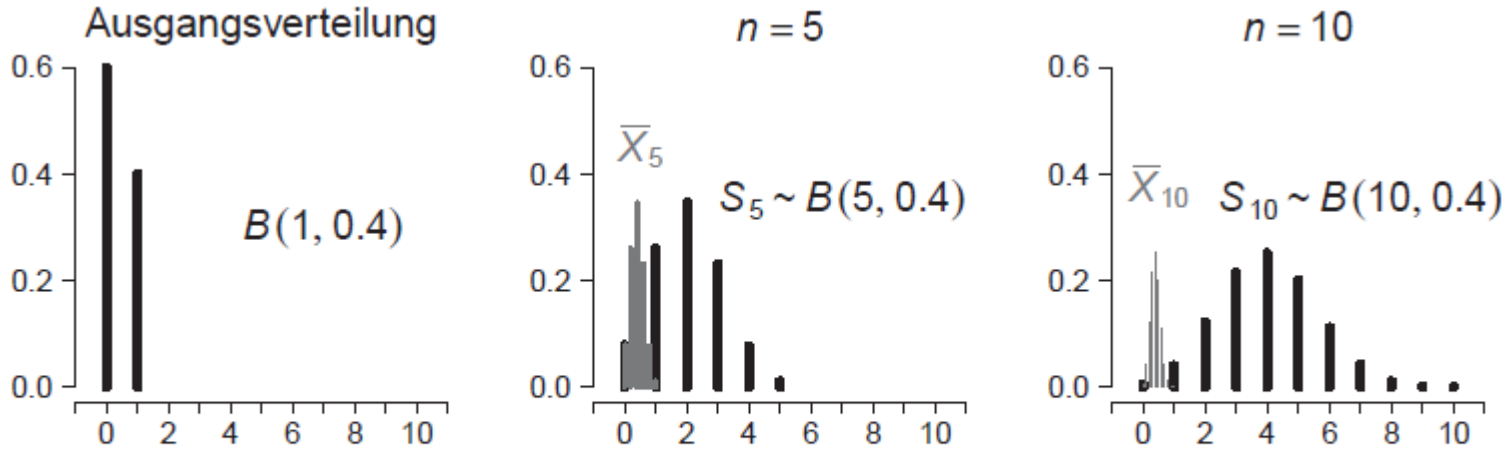


Abb. 7.4.3: Verteilung von Summe und Mittelwert bei einer Binomialverteilung

• Poisson-Verteilung •

$$Po(\lambda) + \cdots + Po(\lambda) = Po(n\lambda).$$

Unabhängigkeit vorausgesetzt!

7.4 Verteilung stochastischer Summen und Mittelwerte

• Normalverteilung •

$$N(\mu_1, \sigma_1^2) + N(\mu_2, \sigma_2^2) = N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Unabhängigkeit vorausgesetzt!

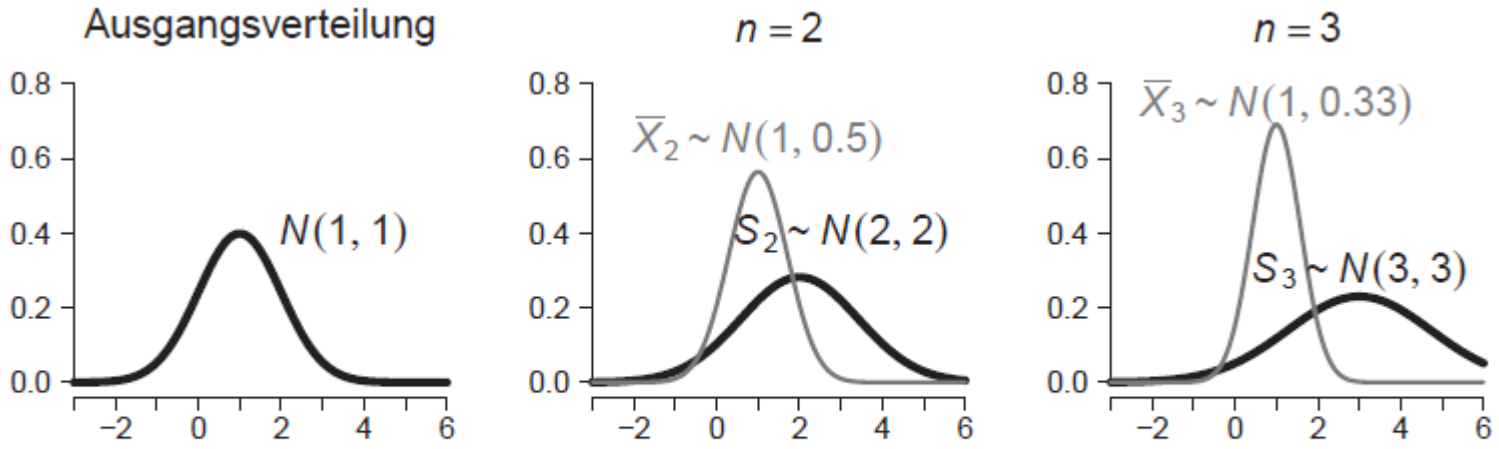


Abb. 7.4.4: Verteilung von Summe und Mittelwert bei einer Normalverteilung

7.4 Verteilung stochastischer Summen und Mittelwerte

Verteilung von Summen und Mittelwerten bei Normalverteilung

Sind $X_1, \ldots, X_n$ unabhängig mit $X_i \sim N(\mu_i, \sigma_i^2)$ für $i = 1, \ldots, n$, dann gilt:

$$c_0 + c_1 X_1 + \cdots + c_n X_n \sim N(c_0 + c_1\mu_1 + \cdots + c_n\mu_n, c_1^2\sigma_1^2 + \cdots + c_n^2\sigma_n^2). \quad (7.4.21)$$

für beliebige Konstanten $c_0, c_1, \ldots, c_n$, wobei mindestens ein $c_i \neq 0$ für $i > 0$ sei. Speziell folgt daraus:

$$S_n \sim N\Big(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\Big) \quad \text{bzw.} \quad (7.4.22)$$

$$\overline{X}_n \sim N(\overline{\mu}_n, \overline{\sigma}_n^2/n). \quad (7.4.23)$$

Sind $X_1, \ldots, X_n$ unabhängig und identisch verteilt mit $X_i \sim N(\mu, \sigma^2)$ gilt:

$$S_n \sim N(n\mu, n\sigma^2) \quad \text{bzw.} \quad (7.4.24)$$

$$\overline{X}_n \sim N(\mu, \sigma^2/n). \quad (7.4.25)$$

Die Resultate (7.4.21) bis (7.4.25) lassen sich für den Fall abhängiger (korrelierter) Zufallsvariablen verallgemeinern, sofern diese sofern diese **multivariat normalverteilt** sind (Abschnitt 8.3.3)

7.4 Verteilung stochastischer Summen und Mittelwerte

• Exponentialverteilung und Erlangverteilung •

$$Exp(\lambda) + \cdots + Exp(\lambda) = Erlang(n, \lambda).$$

Unabhängigkeit vorausgesetzt!

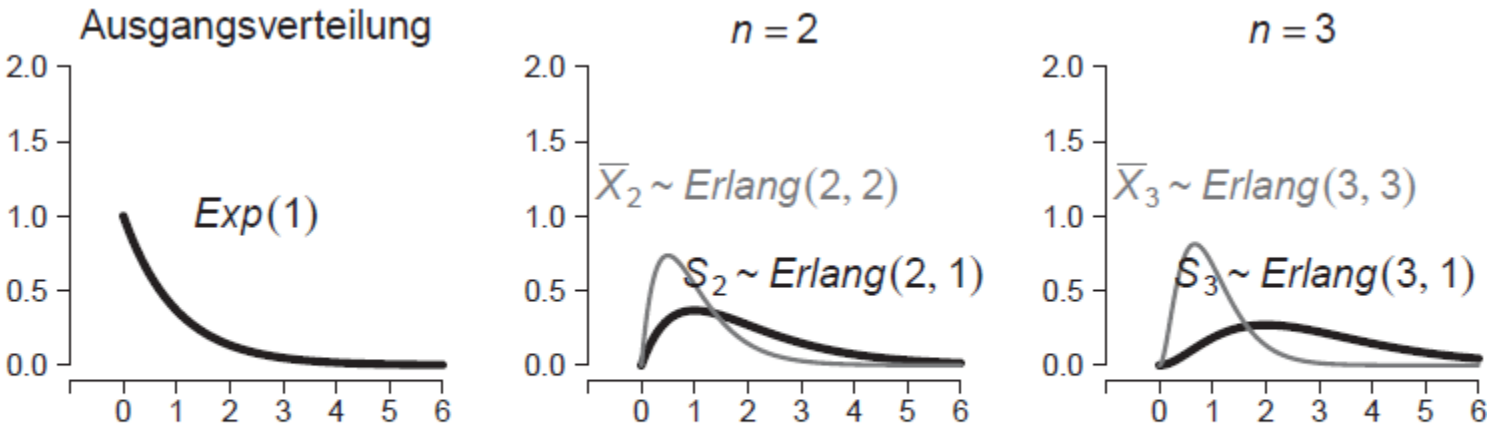


Abb. 7.4.5: Verteilung von Summe und Mittelwert bei einer *Exp*(1)-Verteilung

7.4 Verteilung stochastischer Summen und Mittelwerte

$$Erlang(n_1, \lambda) + Erlang(n_2, \lambda) = Erlang(n_1 + n_2, \lambda).$$

Unabhängigkeit vorausgesetzt!

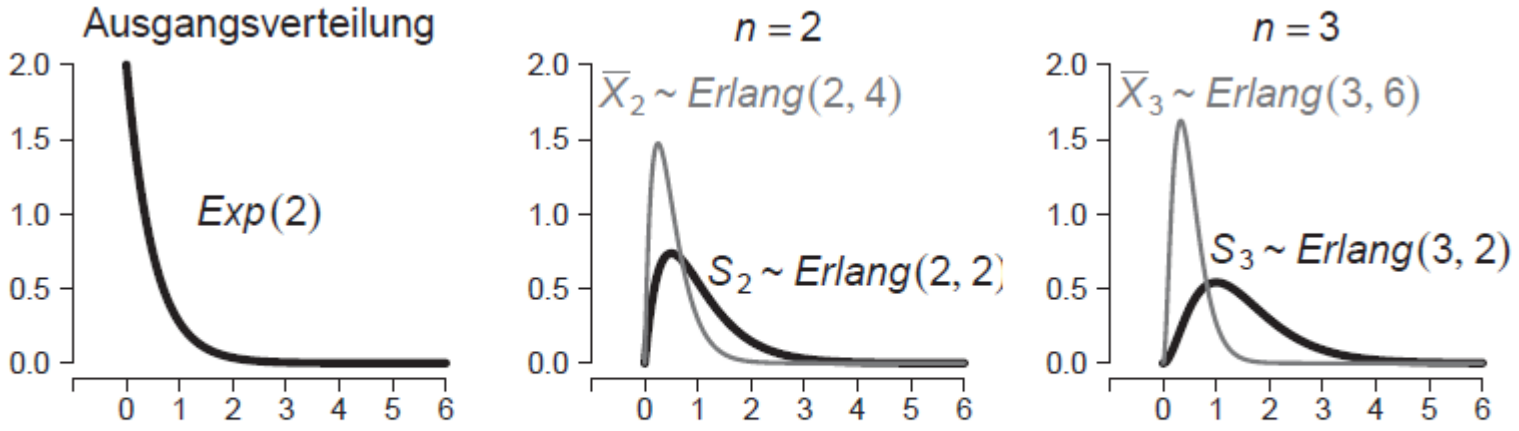


Abb. 7.4.6: Verteilung von Summe und Mittelwert bei einer *Exp*(2)-Verteilung

7.4.2 Asymptotische und approximative Aussagen

Gesetz der großen Zahlen (GGZ)

- Hintergrund
- Schwaches Gesetz der großen Zahlen

Satz 7.4.1: Schwaches GGZ für u.i.v. Zufallsvariablen

Seien $X_1, \ldots, X_n$ unabhängig und identisch verteilte Zufallsvariablen mit $E(X_i) = \mu$ für $i = 1, \ldots, n$. Dann gilt für jedes $c > 0$:

$$P(|\overline{X}_n - \mu| \leq c) \xrightarrow{n\to\infty} 1 \quad \text{bzw.} \quad P(|\overline{X}_n - \mu| > c) \xrightarrow{n\to\infty} 0.$$

Man sagt, das (stochastische) Mittel **konvergiert stochastisch (nach Wahrscheinlichkeit)** gegen μ. Dafür schreibt man auch kurz

$$\overline{X}_n \xrightarrow{p} \mu.$$

7.4 Verteilung stochastischer Summen und Mittelwerte

• Interpretation •

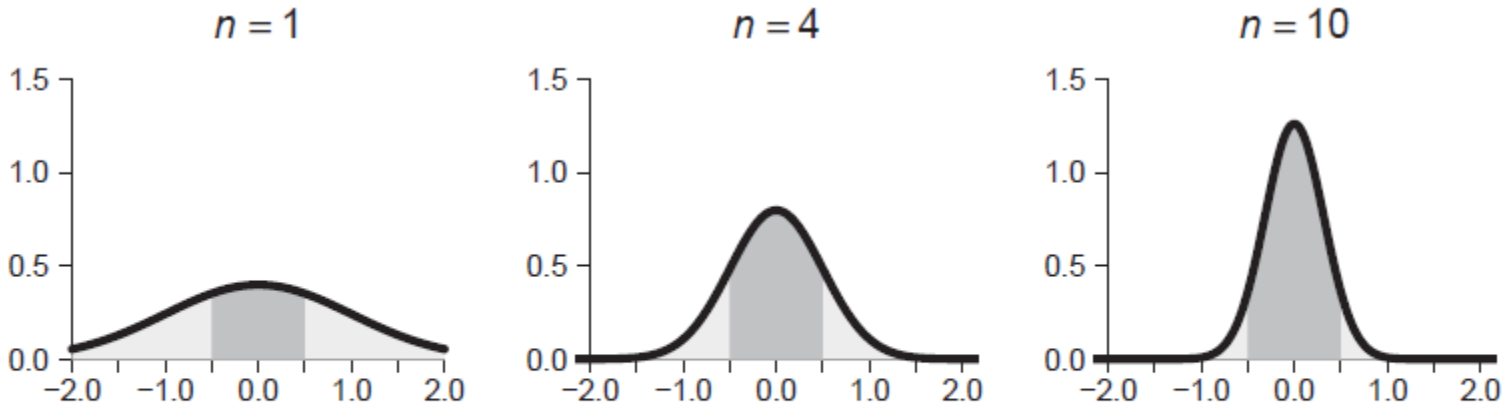


Abb. 7.4.7: Gesetz der großen Zahlen illustriert anhand eines Schwankungsintervalls

Die Wahrscheinlichkeit einer Realisation von $\bar{X}_n$ im Intervall

$$[\mu - c, \mu + c]$$

strebt (konvergiert) für wachsendes n gegen 1. Die Wahrschein lichkeit für eine Realisation außerhalb dieses Intervalls konver giert entsprechend gegen 0. Dies gilt für jeden positiven Wert c, insbesondere also auch für jeden noch so kleinen Wert $c > 0$.

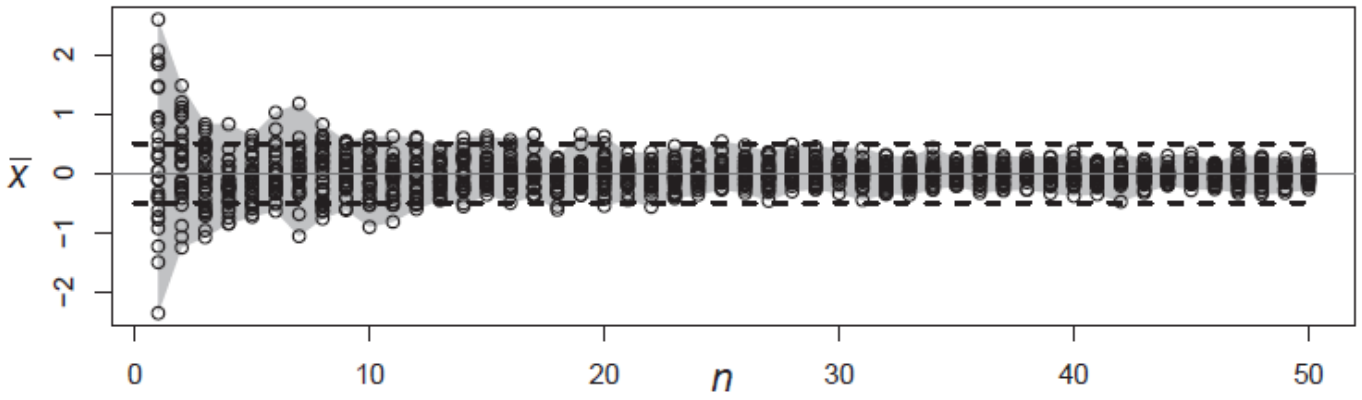


Abb. 7.4.8: Gesetz der großen Zahlen illustriert anhand einer Simulation

Zu jedem $n = 1, 2, ..., 50$ wurden jeweils 30 Stichproben aus einer $N(0, 1)$-Verteilung gezogen und die jeweiligen Realisationen des stochastischen Mittels berechnet. Für jedes n liegen also jeweils 30 Realisationen eines auf n Beobachtungen beruhenden Mittelwerts vor. Die schwarz gestrichelten Linien markieren das Intervall $[-0.5, 0.5]$. Wie man sieht, liegen mit wachsendem n zunehmend mehr Realisationen innerhalb dieses Intervalls. Jedoch verdeutlicht das Schaubild auch die Bedeutung der **stochastischen Konvergenz**. Zu jedem noch so großen n gibt es keine 100%-tige Sicherheit dafür, dass alle Realisationen innerhalb eines bestimmen Intervalls liegen werden.

• Satz von Bernoulli •

Satz 7.4.2: Satz von Bernoulli

Seien $X_1, \ldots, X_n$ unabhängig und identisch verteilt und sei $A \subset \mathbb{R}$ ein beliebiges (messbares) Ereignis. Definiere

$$Y_i = I_A(X_i) \quad \text{für } i = 1, \ldots, n.$$

Dann sind $Y_1, \ldots, Y_n$ unabhängig $B(1, \pi)$-verteilt mit $\pi = P(X_i \in A)$, und die (stochastische) relative Häufigkeit des Ereignisses A konvergiert stochastisch gegen die Wahrscheinlichkeit von A, d.h.

$$\overline{Y}_n \xrightarrow{p} P(A).$$

Das GGZ liefert eine formale Begründung für den frequentistischen Deutungsansatz von Wahrscheinlichkeiten (Abschnitt 6.1.1).

7.4 Verteilung stochastischer Summen und Mittelwerte

- **Spezialfall: Hauptsatz der Statistik** -

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}(X_i) \xrightarrow{p} F(x)$$

Die stochastische Verteilungsfunktion konvergiert gegen die theoretische.

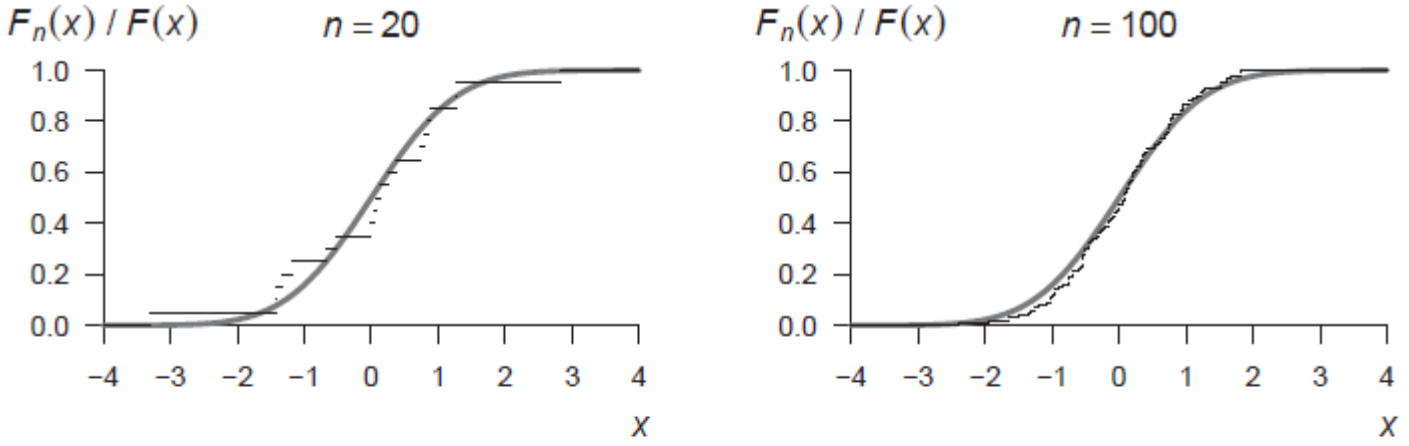


Abb. 7.4.9: Hauptsatz der Statistik illustriert anhand einer Simulation

7.4 Verteilung stochastischer Summen und Mittelwerte

Zentraler Grenzwertsatz

• Hintergrund •

(1) Sind $X_1, \ldots, X_n$ unabhängig $N(\mu, \sigma^2)$-verteilt, so gilt (vgl. Folie 540):

$$S_n \sim N(n\mu, n\sigma^2) \quad \text{und} \quad \bar{X}_n \sim N(\mu, \sigma^2/n).$$

(2) Werden Summe und Mittelwert standardisiert, erhält man

$$Z_n = \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \sim N(0,1) \quad \text{und} \quad Z_n = \frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$$

Der zentrale Grenzwertsatz (ZGWS) besagt, dass die Resultate in (1) und (2) näherungsweise auch für nicht normalverteilte unabhängige und identisch verteilte Zufallsvariablen gelten, falls *n* „hinreichend groß“ ist.

• ZGWS nach Lindeberg-Lévy •

„Einfachste ZGWS-Version“ in Bezug auf Ausgangsbedingungen

7.4 Verteilung stochastischer Summen und Mittelwerte

Satz 7.4.3: Zentraler Grenzwertsatz nach Lindeberg-Lévy

Seien $X_1, \ldots, X_n$ unabhängig und identisch verteilte Zufallsvariablen mit $E(X_i) = \mu$ und $Var(X_i) = \sigma^2$ mit $0 < \sigma^2 < \infty$ für $i = 1, \ldots, n$. Dann ist

$$Z_n = \frac{S_n - n\mu}{\sqrt{n\sigma^2}} = \frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} \quad \textbf{asymptotisch } \boldsymbol{N(0,1)}\textbf{-verteilt,} \tag{7.4.30}$$

d.h.

$$P(Z_n \leq x) = F_{Z_n}(x) \xrightarrow{n\to\infty} \Phi(x) \quad \text{für jedes } x \in \mathbb{R}, \tag{7.4.31}$$

wobei F_{Z_n} die Verteilungsfunktion von Z_n und Φ die Verteilungsfunktion der Standardnormalverteilung bezeichnet. Wir schreiben dafür auch kurz

$$Z_n \overset{a}{\sim} N(0,1). \tag{7.4.32}$$

Summe und Mittelwert sind damit für großes n **approximativ normalverteilt.** Konkret gelten folgende Approximationen:

$$S_n \overset{approx}{\sim} N(n\mu, n\sigma^2), \tag{7.4.33}$$

$$\overline{X}_n \overset{approx}{\sim} N(\mu, \sigma^2/n). \tag{7.4.34}$$

Als **Daumenregel** verwenden wir $n \geq 30$ gilt als „groß“.

• Interpretation •

7.4 Verteilung stochastischer Summen und Mittelwerte

• Approximationsgüte und Daumenregel •

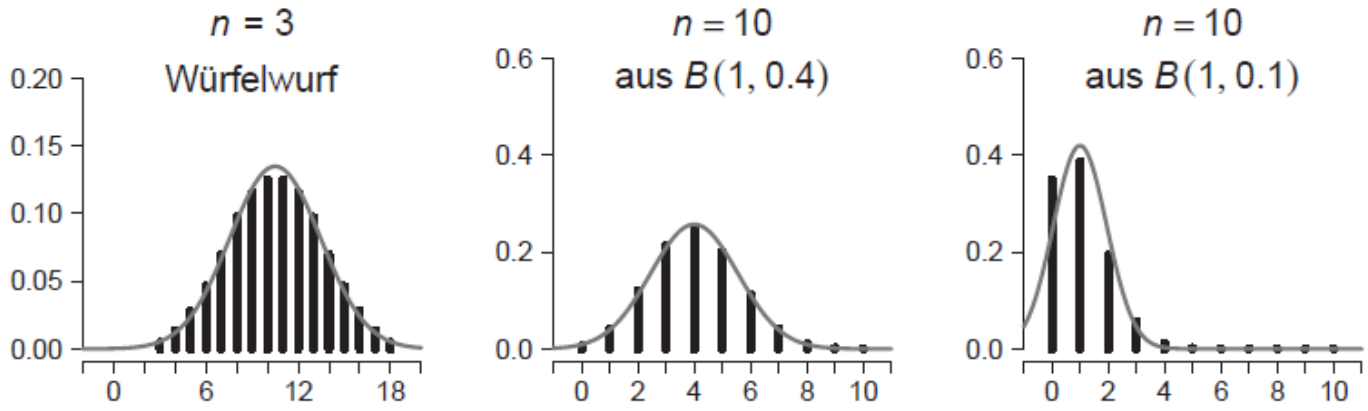


Abb. 7.4.10: Approximationsgüte des ZGWS bei verschiedenen diskreten Verteilungen

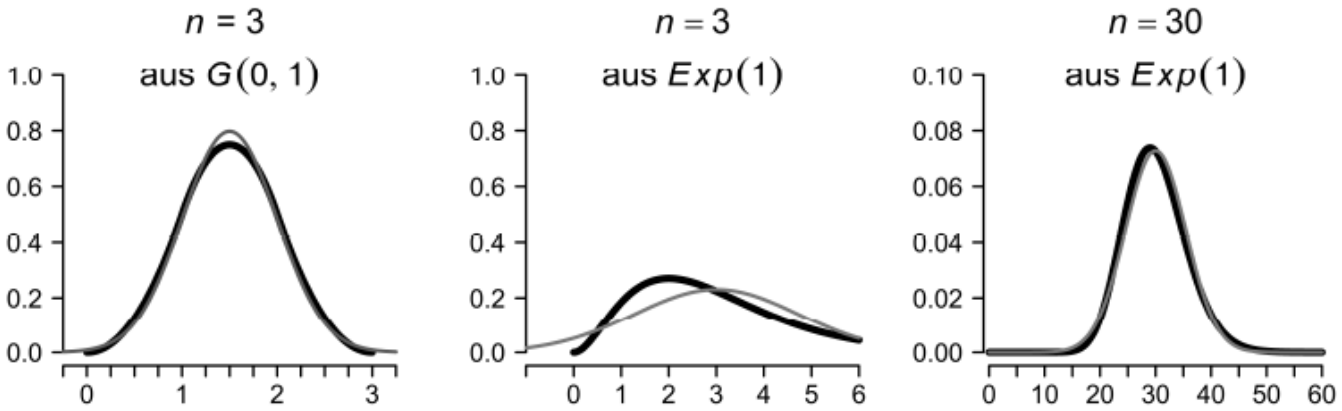


Abb. 7.4.11: Approximationsgüte des ZGWS bei verschiedenen stetigen Verteilungen

Unsere Konvention:

$n \geq 30$

• Spezialfall: Grenzwertsatz von de Moivre •

Satz 7.4.4: Grenzwertsatz nach de Moivre

Seien $X_1, \ldots, X_n$ unabhängig $B(1, \pi)$-verteilt. Dann gilt:

$$Z_n = \frac{S_n - n\pi}{\sqrt{n\pi(1-\pi)}} = \frac{\overline{X}_n - \pi}{\sqrt{\pi(1-\pi)/n}} \overset{a}{\sim} N(0,1), \tag{7.4.36}$$

$$S_n \overset{approx}{\sim} N(n\pi, n\pi(1-\pi)) \quad \text{für großes } n \text{ und} \tag{7.4.37}$$

$$\overline{X}_n \overset{approx}{\sim} N(\pi, \pi(1-\pi)/n) \quad \text{für großes } n. \tag{7.4.38}$$

• Anwendung •

• **Beispiel 7.4.1** • (vgl. Folien 339)

$$P(X=0) = 0.4, P(X=1) = 0.2, P(X=2) = 0.3, P(X=3) = 0.1$$

Wie bereits berechnet (vgl. Folien 424 und 450) gilt hier:

$$E(X) = 1.1 \text{ und } \mathrm{Var}(X) = 1.09.$$

7.4 Verteilung stochastischer Summen und Mittelwerte

Sind $X_1, \ldots, X_{50}$ u. i. v. wie X, so gilt:

$$E(S_{50}) = \sum_{i=1}^{50} E(X_i) = 50 \times 1.1 = 55$$

$$Var(S_{50}) = \sum_{i=1}^{50} Var(X_i) = 50 \times 1.09 = 54.5$$

Damit erhalten wir gemäß ZGWS: $S_{50} \overset{approx.}{\sim} N(55, 54.5)$

Daraus ergibt sich weiter (vgl. Folie 523 ff.)

$$q_{0.05} = 55 + \sqrt{54.5} \times z_{0.05} = 55 + \sqrt{54.5} \times (-1.64) \approx 42.9 \text{ bzw.}$$

$$q_{0.95} = 55 + \sqrt{54.5} \times z_{0.95} = 55 + \sqrt{54.5} \times 1.64 \approx 67.1$$

Mit einer Wahrscheinlichkeit von 90% sollte die Anzahl mitreisender Kinder bei 50 gebuchten Reisen etwa zwischen 43 und 67 liegen.

Angenommen, ab 70 Kindern erleide der Reisveranstalter bestimmte finanzielle Einbußen. Mit welcher Wahrscheinlichkeit tritt dieser Fall ein?

$$
\begin{aligned}
P(S_{50} \geq 70) &= 1 - P(S_{50} \leq 69) \\
&= 1 - P\left(\frac{S_{50} - 55}{\sqrt{54.5}} \leq \frac{69 - 55}{\sqrt{54.5}}\right) \\
&\approx 1 - \Phi(1.90) \approx 1 - 0.9713 = 0.0287
\end{aligned}
$$

Antwort: ca. 2.9%

● Beispiel 7.4.2 ●

Fluggesellschaften verkaufen manchmal mehr Tickets als eigentlich Plätze zur Verfügung stehen, da Buchungen regelmäßig kurzfristig storniert werden oder aus anderen Gründen der Flug nicht angetreten wird. Angenommen, die Wahrscheinlichkeit, dass ein Passagier auf einer bestimmten Strecke seien Reise tatsächlich antritt, betrage 90%. Auf dieser Strecke werde eine Maschine mit 200 Sitzplätzen eingesetzt.

Mit welcher Wahrscheinlichkeit wird diese Maschine dann entweder

(i) voll,

(ii) mit genau 180 Passagieren,

(iii) mit höchsten 190 Passagieren

besetzt abheben, sofern genau 200 Tickets verkauft wurden?

(1) Berechnung über exaktes Verteilungsresultat

Es gilt: $S_{200} = \sum_{i=1}^{200} X_i \sim B(200, 0.9)$

Daraus ergeben sich die exakten Wahrscheinlichkeiten:

(i) $P(S_{200} = 200) = 0.9^{200} \approx 0.000\,000\,000\,705\,5,$

(ii) $P(S_{200} = 180) = \binom{200}{180} \times 0.9^{180}(1 - 0.9)^{20} \approx 0.0936,$

(iii) $P(S_{200} \leq 190) = \sum_{s=0}^{190} \binom{200}{s} 0.9^{s}(1 - 0.9)^{200-s} \approx 0.9965.$

(1) Berechnung über approximatives Verteilungsresultat (ZGWS):

Es gilt: $S_{200} \overset{approx.}{\sim} N(180, 18)$

Daraus ergeben sich die approximativen Wahrscheinlichkeiten:

(i) $P(S_{200} = 200) \approx 0$ oder

$P(S_{200} = 200) \approx f_{S_{200}}(200) = 0.000\,001\,405,$

$f_{S_{200}}(200)$... Dichte der $N(180, 18)$

(ii) $P(S_{200} = 180) \approx 0$ oder

$P(S_{200} = 180) \approx f_{S_{200}}(180) = 0.0940,$

(iii) $$P(S_{200} \leq 190) = P\left(\frac{S_{200} - 180}{\sqrt{18}} \leq \frac{190 - 180}{\sqrt{18}}\right)$$

$$\approx \Phi(2.36) \approx 0.9909$$

In (i) und (ii) kann alternativ noch eine Stetigkeitskorrektur verwendet werden (vgl. Folie 558)

7.4 Verteilung stochastischer Summen und Mittelwerte

Nun könnte man fragen, wie viele Tickets über die eigentliche Sitzplatzanzahl hinaus verkauft werden könnten, sodass höchstens mit einer Wahrscheinlichkeit von 10% die Plätze nicht ausreichen würden. Im Rahmen eines solchen Optimierungskalküls erscheint die Verwendung der ZGWS-Approximation eleganter und praktikabler.

Für großes n (was hier angenommen werden kann) gilt:

$$S_n \overset{approx.}{\sim} N(n \times 0.9, n \times 0.9 \times 0.1) = N(0.9n, 0.09n)$$

Daraus erhält man dann folgenden Kalkül:

$$P(S_n \leq 200) = P\left(\frac{S_n - 0.9n}{\sqrt{0.09n}} \leq \frac{200 - 0.9n}{\sqrt{0.09n}}\right)$$

$$\approx \Phi\left(\frac{200 - 0.9n}{\sqrt{0.09n}}\right) \geq 0.9$$

Die letzte Ungleichung impliziert:

$$\frac{200 - 0.9n}{\sqrt{0.09n}} \geq z_{0.9} \approx 1.28 \qquad (*)$$

$$\Leftrightarrow \quad n^2 - 444.63 + 49382.7 \geq 0 \qquad (**)$$

wobei der Zähler von (*) nichtnegativ sein muss, d. h. wir erhalten für (**) die Restriktion: $200 - 0.9n \geq 0 \quad \Leftrightarrow \quad n \leq 222.22$

Für (**) erhalten wir zunächst die Lösungen:

$$n_1 \approx 215.9 \quad \text{und} \quad n_2 \approx 228.7.$$

Für $n \leq 215.9$ und $n \geq 228.7$ ist somit Ungleichung (**) erfüllt.

Aufgrund der Restriktion verbleibt als korrekte Lösung für (*), also für das eigentliche Problem, schließlich $n \leq 215.9$.

Folglich dürfen höchstens 216 Tickets verkauft werden.

7.4 Verteilung stochastischer Summen und Mittelwerte

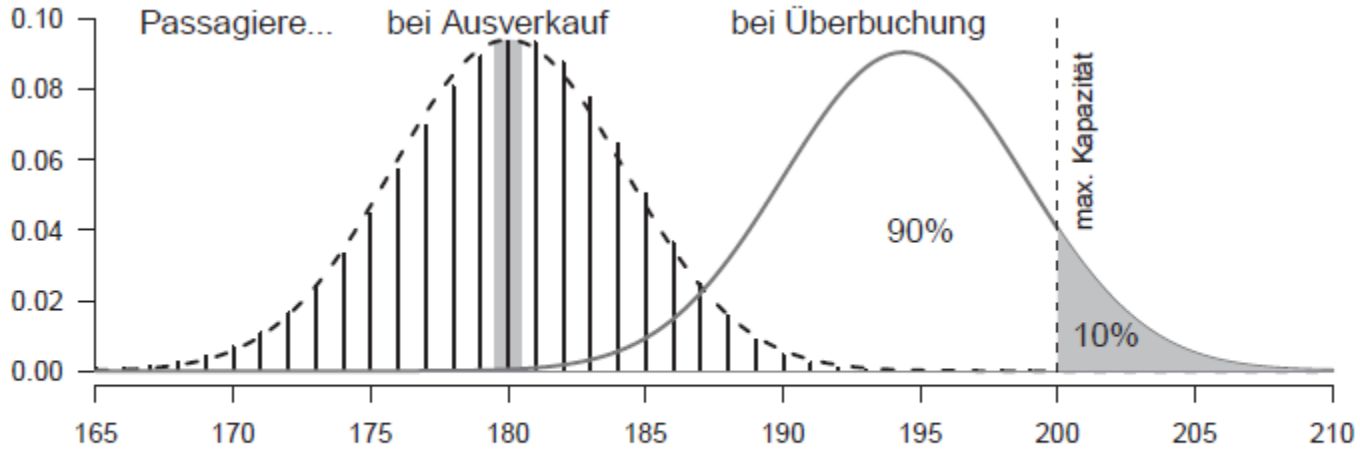


Abb. 7.4.12: Illustration zu Beispiel 7.4.2. ZGWS-Approximationen und Stetigkeitskorrektur

● Stetigkeitskorrektur für einzelne Trägerpunkte ●

Gegeben die Situation von Beispiel 7.4.2 (vgl. Folie 553).
Hier erhalten wir für (i) und (ii):

$$P(S_{200} = 200) \approx \int_{200-0.5}^{200+0.5} f_{S_{200}}(s)ds = \cdots = 0.000\,001\,5,$$

$$P(S_{200} = 180) \approx \int_{180-0.5}^{180+0.5} f_{S_{200}}(s)ds = \cdots = 0.093.$$

● Kritische Prüfung der Annahmen des ZGWS ●

> Identische Verteilungen: In Beispiel 7.4.2 beispielsweise verletzt, falls die Wahrscheinlichkeit einer Stornierung stark schwankt

> Unabhängigkeit: In Beispiel 7.4.1 beispielsweise verletzt, wenn Familien mit Kindern immer nur dann eine Reise buchen, wenn bereits genügend andere Familien mit Kindern bereits eine Reise gebucht haben.

8 Ergänzungen und Verallgemeinerungen

8 Ergänzungen und Verallgemeinerungen

8.1 Weitere eindimensionale Verteilungen 563

8.1.1 Weitere diskrete Verteilungen 563

8.1.2 Weitere stetige Verteilungen 566

8.1.3 Sonstige eindimensionale Verteilungen 572

8.2 Ergänzungen zu asymptotischen Aussagen 574

8.2.1 Alternative GGZ- und ZGWS-Varianten 574

8.2.2 Stochastische Konvergenz und Verteilungskonvergenz 579

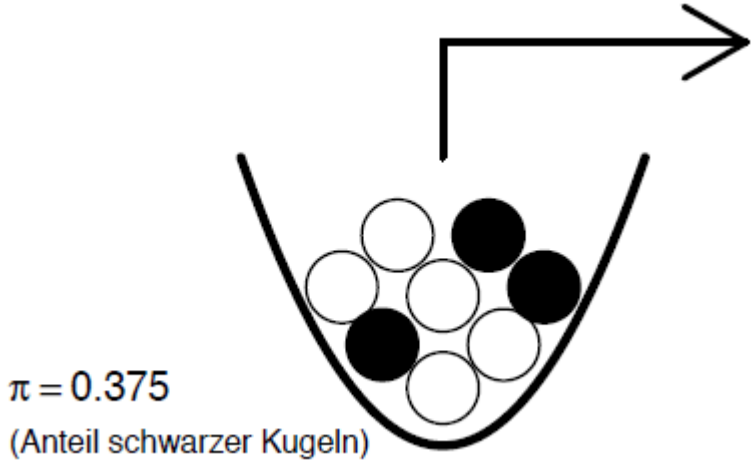
8 Ergänzungen und Verallgemeinerungen

8.3 Einige multivariate Verallgemeinerungen	586
8.3.1 Multivariate Abhängigkeiten	586
Abhängigkeit und Unabhängigkeit von Zufallsvektoren	586
Bedingte Abhängigkeit und Unabhängigkeit	591
8.3.2 Kenngrößen multivariater Abhängigkeiten	592
Theoretische Kovarianz- und Korrelationsmatrix	592
Bedingte Kennwerte für $n > 2$	595
8.3.3 Sonstige Verallgemeinerungen	600

8.1 Weitere eindimensionale Verteilungen

8.1.1 Weitere diskrete Verteilungen

• Hypergeometrische Verteilung •



n Mal Ziehen ohne Zurücklegen

$(n \leq N)$

Sei S_n die Anzahl schwarzer Kugeln in der Stichprobe. Dann gilt für

$S_n \sim H(n, M, N)$:

$$P(S_n = s) = \frac{\binom{M}{s}\binom{N-M}{n-s}}{\binom{N}{n}}$$

N ... Anzahl der Kugeln in der Urne

M ... Anzahl der schwarzen Kugeln

$\pi = M/N$ ... Anteil der schwarzen Kugeln

● Geometrische Verteilung ●

> Wahrscheinlichkeitsfunktion für $X \sim Geo(\pi)$ mit $\pi \in (0,1)$:

$$f_X(x) = P(X = x) = \begin{cases} (1-\pi)^x \pi, & x = 0,1,2,\dots, \\ 0, & sonst. \end{cases}$$

> Modellierung diskret gemessener Wartezeiten. Beispiel: Anzahl benötigter **Fehl**versuche beim Würfeln bis zur ersten „6“ $\Rightarrow X \sim Geo(1/6)$

> Diskretes Pendant zur stetigen **Exponentialverteilung**

● Negative Binomialverteilung ●

> Wahrscheinlichkeitsfunktion für $X \sim NB(n,\pi)$ mit $\pi \in (0,1)$ und $n \in \mathbb{N}_0$:

$$f_X(x) = P(X = x) = \begin{cases} \binom{n+x-1}{x} \pi^n (1-\pi)^x, & x = 0,1,2,\dots, \\ 0, & sonst. \end{cases}$$

> Diskrete Wartezeit bis zum n-ten Eintreten eines Ereignisses. Beispiel: Anzahl benötigter **Fehl**versuche beim Würfeln bis zur n-ten „6“ $\Rightarrow X \sim NB(n, 1/6)$

> Diskretes Pendant zur stetigen **Erlang-Verteilung**

> Es gilt:

- $NB(1,\pi) = Geo(\pi)$
- Sind $X_1, \ldots, X_n$ unabhängig $Geo(\pi)$-verteilt, so folgt

$$S_n = \sum_{i=1}^{n} X_i \sim NB(n,\pi)$$

Informal ausgedrückt:

$$Geo(\pi) + \cdots + Geo(\pi) = NB(n,\pi).$$

> Beispiel: Sei X die Anzahl von Fehlversuchen beim Würfeln bis zum dritten Mal die „6“ fällt. Dann gilt: $X \sim NB(3, 1/6)$

Beispielsweise folgt daraus dann

$$P(X = 18) = \binom{20}{18}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^{18} \approx 0.0330$$

8.1.2 Weitere stetige Verteilungen

● Gamma-Verteilung ●

> Dichte für $X \sim Gamma(\delta, \lambda)$ mit $\delta > 0$ und $\lambda > 0$:

$$f_X(x) = \frac{\lambda^\delta x^{\delta-1}}{\Gamma(\delta)} e^{-\lambda x} I_{[0,\infty)}(x) \quad \text{wobei} \quad \Gamma(\delta) = \int_0^\infty x^{\delta-1} e^{-x} dx$$

(Gammafunktion)

> Es gilt: $Gamma(\delta_1, \lambda) + Gamma(\delta_2, \lambda) = Gamma(\delta_1 + \delta_2, \lambda).$

Unabhängigkeit vorausgesetzt

Spezielle Gamma-Verteilungen

Für $\lambda > 0$ und $n \in \mathbb{N}$ gelten folgende Zusammenhänge:

$$Gamma(1, \lambda) = Exp(\lambda) = Erlang(1, \lambda), \tag{8.1.3}$$

$$Gamma(n, \lambda) = Erlang(n, \lambda), \tag{8.1.4}$$

8.1 Weitere eindimensionale Verteilungen

• χ^2-Verteilung •

> Seien $Z_1, \dots, Z_n$ unabhängig $N(0,1)$-verteilt. Dann gilt (ohne Beweis):

$$X := \sum_{i=1}^{n} Z_i^2 \sim \chi^2(n)$$ **χ^2-Verteilung mit *n* Freiheitsgraden**

> Für $Z \sim N(0,1)$ gilt: $E(Z) = 0, E(Z^2) = 1$ und $E(Z^4) = 3$

$$\Longrightarrow \quad E(X) = E(Z_1^2 + \cdots + Z_n^2) = n$$

$$\Longrightarrow \quad \mathrm{Var}(X) = Var(Z_1^2 + \cdots + Z_n^2) = Var(Z_1^2) + \cdots + Var(Z_n^2) = 2n$$

Beziehung der χ^2-Verteilung zu anderen Verteilungen

Für $n \in \mathbb{N}$ gelten folgende Zusammenhänge:

$$\chi^2(n) = Gamma(0.5n, 0.5), \tag{8.1.6}$$

$$\chi^2(2) = Gamma(1, 0.5) = Erlang(1, 0.5) = Exp(0.5). \tag{8.1.7}$$

● *t*-Verteilung ●

> Dichte für $X \sim t(n)$ mit $n \in \mathbb{N}$:

$$f_X(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\,\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

***t*-Verteilung mit *n* Freiheitsgraden**

> $E(X)$ existiert nicht für $n = 1$, $E(X) = 0$ für $n \geq 2$

> $Var(X)$ existiert nicht für $n < 3$, $Var(X) = \frac{n}{n-2}$ für $n \geq 3$

> Sind $Z \sim N(0,1)$ und $Y_n \sim \chi^2(n)$ unabhängig, so gilt (ohne Beweis):

$$X_n := \frac{Z}{\sqrt{Y_n/n}} \sim t(n)$$

> Die $t(n)$-Verteilung „konvergiert“ für $n \to \infty$ gegen die $N(0,1)$-Verteilung, d. h. für großes *n* gilt:

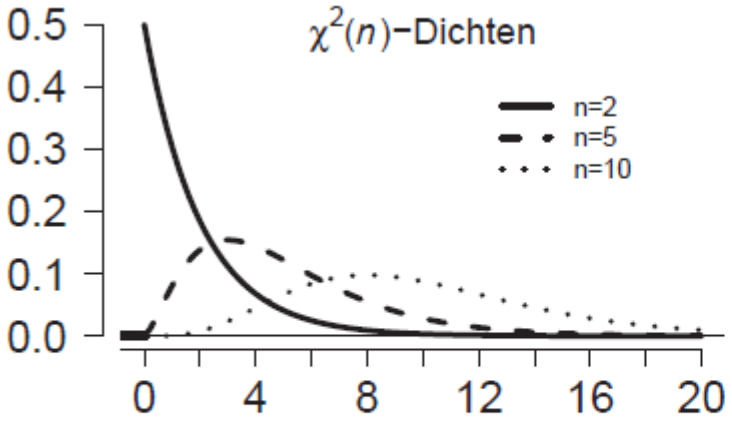
$$t_{n,\alpha} \approx z_\alpha$$

8.1 Weitere eindimensionale Verteilungen

Hierzu beachte man:

$$Y_n/n = \frac{1}{n}\sum_{i=1}^{n} Z_i^2 \xrightarrow{p} E\left(Z_i^2\right) = 1 \quad \text{und } Z \sim N(0,1)$$

$$\Rightarrow \quad X_n \overset{a}{\sim} N(0,1) \qquad (\textbf{Slutsky's Theorem}, \text{ Folie } 583)$$



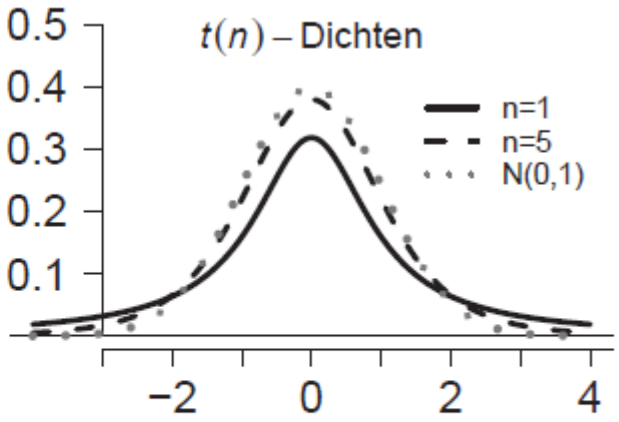


Abb. 8.1.1: Dichten der χ^2-Verteilung und t-Verteilung

● Cauchy-Verteilung ●

> Dichte für $X \sim C(\alpha, \beta)$ mit $\alpha \in \mathbb{R}$ und $\beta > 0$:

$$f_X(x) = \frac{1}{\pi\beta\left(1 + \frac{(x-\alpha)^2}{\beta^2}\right)}$$ (π ... hier Kreiszahl)

> Es gilt: $C(0,1) = t(1)$ (**Standard-Cauchy-Verteilung** = $t(1)$-Verteilung)

> Es gilt: $C(\alpha_1, \beta_1) + C(\alpha_2, \beta_2) = C(\alpha_1 + \alpha_2, \beta_1 + \beta_2).$

Unabhängigkeit vorausgesetzt

> Sind $X_1, \ldots, X_n$ unabhängig $C(\alpha, \beta)$-verteilt, so gilt:

$$\bar{X}_n \sim C(\alpha, \beta)$$

Das stochastische Mittel ist identisch verteilt wie die Ausgangsvariablen selbst, d. h. keine stochastische Konvergenz gegen 0!

8.1 Weitere eindimensionale Verteilungen

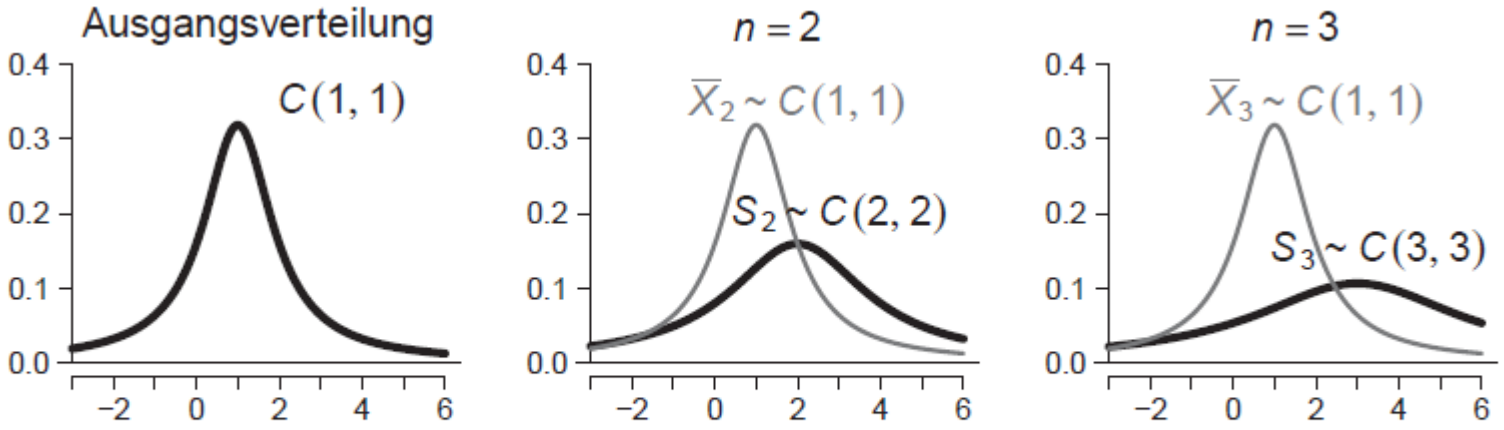


Abb. 8.1.2: Verteilung von Summe und Mittelwert bei einer Cauchy-Verteilung

$$S_n \sim C(n\alpha, n\beta)$$

$$\bar{X}_n \sim C(\alpha, \beta)$$

8.1.2 Sonstige eindimensionale Verteilungen

- Weder diskret noch stetig
- Mixturverteilungen

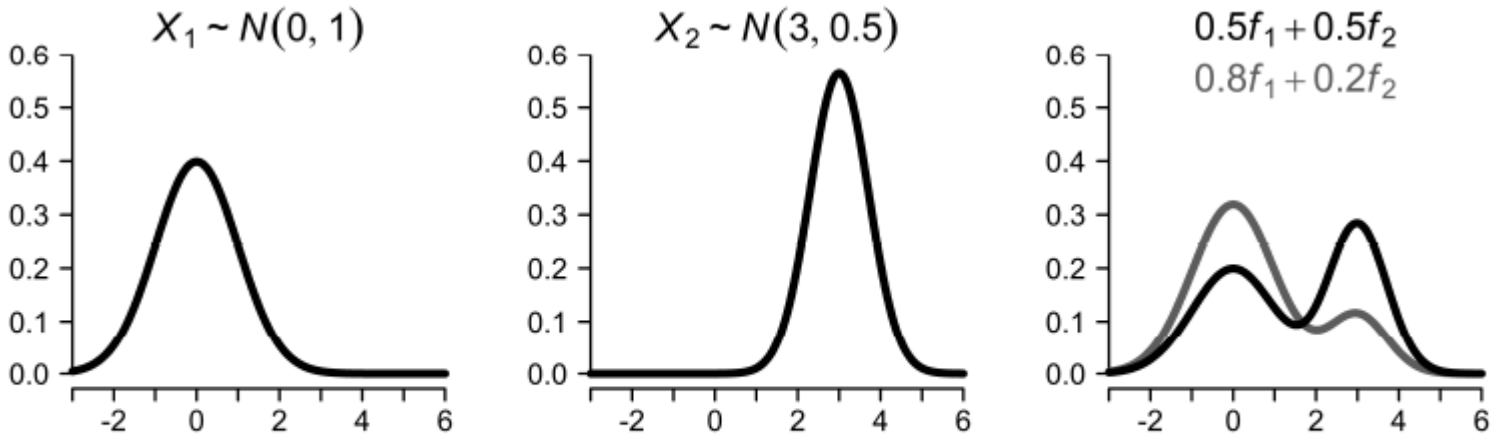


Abb. 8.1.3: Mixturverteilung aus zwei Normalverteilungen

8.1 Weitere eindimensionale Verteilungen

● Gestutzte und zensierte Verteilungen ●

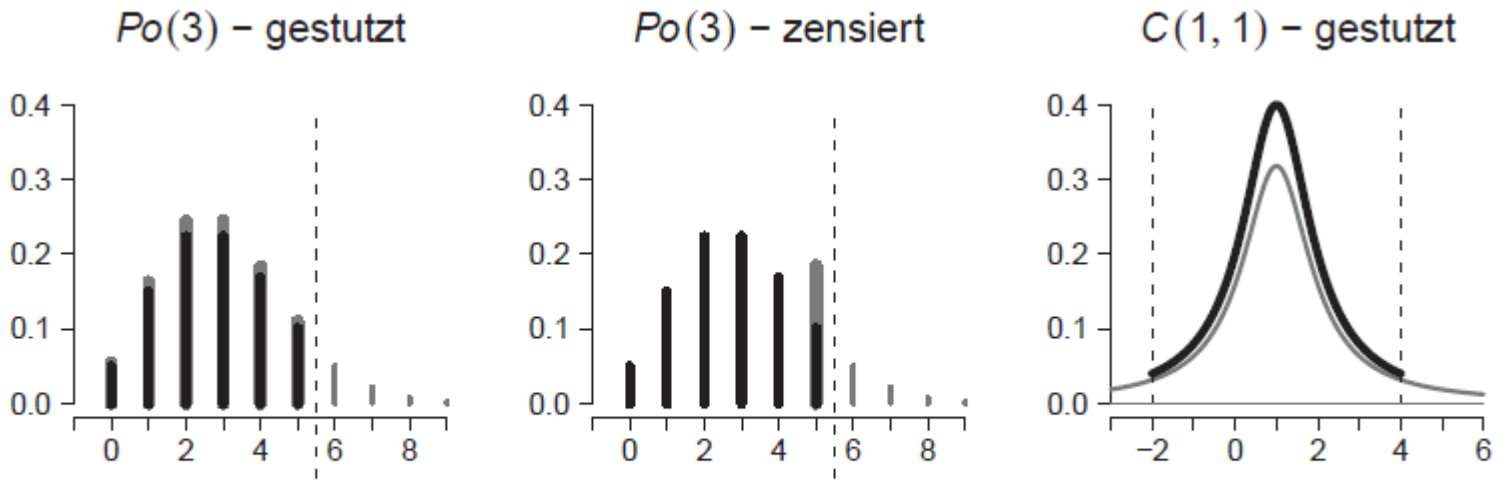


Abb. 8.2.1: Gestutzte und zensierte Verteilungen

> **Stutzung:** Verbleibende Wahrscheinlichkeitsmasse wird proportional auf die Ziel-Trägermenge aufgeteilt

> **Zensierung:** Verbleibende Wahrscheinlichkeitsmasse wird vollständig letztem Trägerpunkt der Ziel-Trägermenge zugeschlagen

8.2 Ergänzungen zu asymptotischen Aussagen

8.2.1 Alternative GGZ- und ZGWS-Varianten

• Vereinfachende Annahmen •

Definition 8.2.1: Annahmen (M) und (B)

Für eine gegebene Folge von Zufallsvariablen $X_1, \ldots, X_n, \ldots$ gelte

(i) **Annahme (M)**, falls für alle Zufallsvariablen jeweils alle Momente endlich sind, d.h. falls gilt:

$$E(X_i^k) < \infty \quad \text{für alle } i = 1, \ldots, n, \ldots \text{ und jedes } k \in \mathbb{N}.$$

(ii) **Annahme (B)**, falls die Trägermengen aller Zufallsvariablen gleichmäßig beschränkt sind, d.h. falls es einen festen Wert c mit $0 < c < \infty$ gibt, sodass

$$P(|X_i| < c) = 1 \quad \text{für alle } i = 1, \ldots, n, \ldots$$

> **Annahme (M)** bei allen „bekannten Verteilungen“ sowie Verteilungen mit beschränkten Trägermengen erfüllt außer *t*-, *F*- und *Cauchy*-Verteilung

> **Annahme (B)** schließt Verteilungen mit unbeschränkter Trägermenge (z. B. Normalverteilung) aus; impliziert (M)

8.2 Ergänzungen zu asymptotischen Aussagen

• Konvergenzformen und alternative GGZ-Varianten •

Satz 8.2.1: Schwaches GGZ für u.h.v. Zufallsvariablen

Seien $X_1, \ldots, X_n$ unabhängig. Dabei gelte Annahme (M) sowie

$$\overline{\mu}_n \xrightarrow{n\to\infty} \mu \quad \text{und} \quad \overline{\sigma}_n^2 \xrightarrow{n\to\infty} \sigma^2.$$

Dann folgt daraus

$$\overline{X}_n \xrightarrow{p} \mu.$$

Hinweis: Es gelte die Notation der Abschnitte 7.4.1 und 7.4.2.

> **u. h. v.** steht für **unabhängig heterogen verteilt**

> GGZ-Variante erlaubt unterschiedliche Verteilungen der einzelnen Zufallsvariablen. Durchschnittlicher Erwartungswert und durchschnittliche theoretische Varianz müssen jedoch gegen feste Werte konvergieren.

● **Beispiel 8.2.1** ●

$X_1 \sim N(0,1), X_1 \sim N(1,2), X_1 \sim N(0,1), X_1 \sim N(1,2), X_1 \sim N(0,1), \ldots$

$\Rightarrow \bar{\mu}_n \xrightarrow{n\to\infty} 0.5$ und $\bar{\sigma}_n^2 \xrightarrow{n\to\infty} 1.5$

$\Rightarrow \bar{X}_n \xrightarrow{p} 0.5$

● **Beispiel 8.2.2** ●

Seien $X_1, \ldots, X_n$ unabhängig mit identischem Erwartungswert μ und heterogenen Varianzen gemäß folgendem Schema:

1, 3, 1, 1, 3, 3, 1, 1, 1, 1, 1, 1, 3, 3, 3, 3, 3, 3,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,
1, 1, ... Wie viele Einsen folgen jetzt? Antwort: 32

Hier konvergiert $\bar{\sigma}_n^2$ gegen keinen festen Wert, sondern oszilliert ständig zwischen 2 und 1.5. Es lässt sich zeigen, dass $\bar{X}_n \xrightarrow{p} \mu$ dennoch gilt, sofern **Annahme (B)** erfüllt ist (über allgemeinere GGZ-Variante).

8.2 Ergänzungen zu asymptotischen Aussagen

• Normalverteilung und ZGWS-Varianten •

Satz 8.2.2: ZGWS für u.h.v. Zufallsvariablen

Seien $X_1, \ldots, X_n$ unabhängig. Dabei gelte Annahme (B) sowie

$$\overline{\mu}_n \xrightarrow{n\to\infty} \mu \quad \text{und} \quad \overline{\sigma}_n^2 \xrightarrow{n\to\infty} \sigma^2 > 0.$$

Dann folgt daraus

$$Z_n = \frac{S_n - n\overline{\mu}_n}{\sqrt{n\overline{\sigma}_n^2}} = \frac{\overline{X}_n - \overline{\mu}_n}{\sqrt{\overline{\sigma}_n^2/n}} \overset{a}{\sim} N(0,1) \tag{8.2.1}$$

sowie für großes n, sofern $\overline{\sigma}_n^2 > 0$:

$$S_n \overset{approx}{\sim} N\Big(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\Big) = N(n\overline{\mu}_n, n\overline{\sigma}_n^2), \tag{8.2.2}$$

$$\overline{X}_n \overset{approx}{\sim} N(\overline{\mu}_n, \overline{\sigma}_n^2/n) \quad \text{und} \tag{8.2.3}$$

$$Z_n \overset{approx}{\sim} N(0,1). \tag{8.2.4}$$

Konvergenzannahmen hinreichend aber nicht notwendig. Theoretisches Mittel dürfte z. B. auch oszillieren wie die Varianz in Beispiel 8.2.2.

● Beispiel 8.2.3 ●

Ein Taxi-Unternehmen verfüge über 50 Fahrzeuge, die teils unterschiedlich eingesetzt werden und unterschiedlich viel Kraftstoff verbrauchen. Der durchschnittliche tägliche Verbrauch betrage 15 Liter bei einer durchschnittlichen Varianz von 2 (Liter^2). Dann sollte sich der tägliche Gesamtverbrauch durch eine

$N(750, 100)$-Verteilung

approximieren lassen, sofern die einzelnen Verbrauchswerte (weitgehend) unabhängig voneinander sind.

8.2.2 Stochastische Konvergenz und Verteilungskonvergenz

- **Hintergrund und Überblick**
- **Stochastische Konvergenz**

Definition 8.2.2: Stochastische Konvergenz (gegen feste Werte)

Sei (Y_n) eine Folge von Zufallsvariablen. Sofern für einen festen Wert θ und jedes $c > 0$

$$P(|Y_n - \theta| \leq c) \xrightarrow{n\to\infty} 1 \quad \text{bzw.} \quad P(|Y_n - \theta| > c) \xrightarrow{n\to\infty} 0$$

gilt, sagen wir, dass Y_n **stochastisch (nach Wahrscheinlichkeit)** gegen θ **konvergiert** und schreiben dafür auch kurz

$$Y_n \xrightarrow{p} \theta.$$

> Bisher nur Spezialfall GGZ mit$(\bar{X}_n)$, d. h.

$\bar{X}_1 = X_1, \bar{X}_2 = \frac{1}{2}(X_1 + X_2), \bar{X}_3 = \frac{1}{3}(X_1 + X_2 + X_3), \ldots$ und

$$\bar{X}_n \xrightarrow{p} \mu$$

> Jetzt allgemeiner: „Beliebige“ Folgen von Zufallsvariablen

8.2 Ergänzungen zu asymptotischen Aussagen

• Deterministische Konvergenz •

Deterministische Konvergenz kann als Spezialfall stochastischer Konvergenz aufgefasst werden. Definiere dazu **einpunktverteilte** Zufallsvariablen $X_1, \ldots, X_n$ mit

$$P(X_i = c_i) = 1 \quad \text{mit } c_i \in \mathbb{R} \quad \text{für } i = 1, \ldots, n$$

Beispiel: $c_i = 1/i$ für Folge $(1/n)$ mit $1/n \xrightarrow{n\to\infty} 0$

• Transformation stochastisch konvergenter Folgen •

Satz 8.2.3: Stetigkeitssatz

Sei (Y_n) eine Folge von Zufallsvariablen, $h : \mathbb{R} \to \mathbb{R}$ eine stetige Funktion und θ ein fester Wert. Sofern $h(\theta)$ definiert ist und gilt

$$Y_n \xrightarrow{p} \theta,$$

folgt daraus für $U_n = h(Y_n)$:

$$U_n \xrightarrow{p} h(\theta) \quad \text{bzw.} \quad h(Y_n) \xrightarrow{p} h(\theta).$$

● Beispiel 8.2.4 ●

Aus $\bar{X}_n \xrightarrow{p} \mu$ folgt beispielsweise

(i) $\bar{X}_n^2 \xrightarrow{p} \mu^2$

(ii) $\bar{X}_n^{-1} \xrightarrow{p} 1/\mu$ für $\mu \neq 0$

Wie lässt sich das begründen? Antwort: Dies folgt mit der Stetigkeit von

$$h_1(t) = t^2 \text{ bzw. } h_2(t) = 1/t$$

● GGZ für *k*-te Momente ●

Satz 8.2.4: GGZ für k-te Momente

Seien $X_1, \ldots, X_n$ u.i.v. mit $E(X_i^k) < \infty$. Dann folgt daraus

$$\frac{1}{n}\sum_{i=1}^{n} X_i^m \xrightarrow{p} E(X_i^m) \quad \text{für } m \leq k, \text{ wobei } k, m \in \mathbb{N}.$$

● Konvergenz nach Verteilung ●

Definition 8.2.3: Verteilungskonvergenz gegen eine Normalverteilung

Sei (Y_n) eine Folge von Zufallsvariablen. Sofern gilt:

$$P(Y_n \leq x) = F_{Y_n}(x) \xrightarrow{n\to\infty} \Phi((x-\mu)/\sigma) \quad \text{für jedes } x \in \mathbb{R},$$

sagen wir, dass Y_n **asymptotisch** $N(\mu, \sigma^2)$**-verteilt** ist oder dass Y_n **gegen eine** $N(\mu, \sigma^2)$**-Verteilung konvergiert** und schreiben dafür auch kurz

$$Y_n \overset{a}{\sim} N(\mu, \sigma^2).$$

> Bisher nur Spezialfall ZGWS, wobei (Y_n) eine standardisierte(r) Summe (Mittelwert) ist und Grenzverteilung die Standardnormalverteilung

> Jetzt allgemeiner: „Beliebige“ Folgen von Zufallsvariablen und beliebige Normalverteilungen als Grenzverteilung

> Weitere Verallgemeinerungen möglich wie etwa Konvergenz gegen andere Grenzverteilungen als Normalverteilung (wird nicht benötigt)

8.2 Ergänzungen zu asymptotischen Aussagen

• Alternative Notation •

Anstelle von $F_{Y_n}(x) \xrightarrow{n\to\infty} \Phi((x-\mu)/\sigma)$ kann man auch äquivalent schreiben: $Y_n \xrightarrow{d} X$, wobei $X \sim N(\mu, \sigma^2)$

• Stochastische Konvergenz und Verteilungskonvergenz parallel •

Satz 8.2.5: Slutsky's Theorem für Normalverteilungen

Seien (A_n), (B_n) und (Y_n) jeweils Folgen von Zufallsvariablen mit

$$A_n \xrightarrow{p} a,\ B_n \xrightarrow{p} b \quad \text{und} \quad Y_n \overset{a}{\sim} N(\mu, \sigma^2).$$

Dann folgt daraus:

$$A_n + B_n Y_n \overset{a}{\sim} N(a + b\mu, b^2\sigma^2), \quad \text{falls } b \neq 0, \tag{8.2.6}$$

$$A_n + B_n Y_n \xrightarrow{p} a, \quad \text{falls } b = 0, \tag{8.2.7}$$

$$Y_n/B_n \overset{a}{\sim} N(\mu/b, \sigma^2/b^2), \quad \text{falls } b \neq 0. \tag{8.2.8}$$

● Beispiel 8.2.5 ●

Sei Z_n wie auf Folie 577 (ZGWS für u. h. v. Zufallsvariablen), d. h.

$$Z_n = \frac{\bar{X}_n - \bar{\mu}_n}{\sqrt{\bar{\sigma}_n^2/n}} \overset{a}{\sim} N(0,1)$$

Dann gilt auch: $\tilde{Z}_n = \frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}} \overset{a}{\sim} N(0,1)$

Dies folgt mit Slutsky‘s Theorem. Definiere dazu

$\tilde{Z}_n = A_n + B_n Z_n$, wobei

$$Z_n = \frac{\bar{X}_n - \bar{\mu}_n}{\sqrt{\bar{\sigma}_n^2/n}} = \frac{\bar{X}_n}{\sqrt{\bar{\sigma}_n^2/n}} - \frac{\bar{\mu}_n}{\sqrt{\bar{\sigma}_n^2/n}} \overset{a}{\sim} N(0,1)$$

$$A_n = \frac{\bar{\mu}_n}{\sqrt{\bar{\sigma}_n^2/n}} - \frac{\mu}{\sqrt{\bar{\sigma}_n^2/n}} \xrightarrow{n\to\infty} 0 \text{ und } B_n = \sqrt{\bar{\sigma}_n^2/\sigma^2} \xrightarrow{n\to\infty} 1$$

Weiter gilt auch: $\check{Z}_n = \sqrt{n}(\bar{X}_n - \mu) \overset{a}{\sim} N(0, \sigma^2)$

Definiere dazu

$\check{Z}_n = A_n^* + B_n^* \tilde{Z}_n$, wobei

$$\tilde{Z}_n = \frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}} \overset{a}{\sim} N(0,1)$$

$A_n^* = 0$ und $B_n^* = \sigma$

8.3.1 Multivariate Abhängigkeiten

Abhängigkeit und Unabhängigkeit von Zufallsvektoren

● Hintergrund und Definition ●

Definition 8.3.1: Abhängigkeit und Unabhängigkeit von Zufallsvektoren

Zwei Zufallsvektoren (diskret oder stetig)

$$\mathbf{X} = (X_1, X_2, \ldots, X_p)^T \quad \text{und} \quad \mathbf{Y} = (Y_1, Y_2, \ldots, Y_q)^T$$

sind **stochastisch unabhängig**, falls sich deren gemeinsame Verteilung aus dem Produkt der Randverteilungen ergibt, d.h. falls gilt

$$f_{\mathbf{X},\mathbf{Y}} = f_{\mathbf{X}} f_{\mathbf{Y}}.$$

Anderenfalls sind sie **stochastisch abhängig**.

● Interpretation ●

● Beispiel D3-a fortgesetzt ●

$X = 0$ für weiblich $Y = 0$ für Raucher $Z = 0$ für nichtvegetarisch

$X = 1$ für männlich $Y = 1$ für Gelegenheitsraucher $Z = 1$ für vegetarisch

$Y = 2$ für Nichtraucher

Tab. 7.1.6: Gemeinsame Verteilung von Beispiel D3-a

$Z = 0$

X \ Y	0	1	2	Summe
0	0.04	0.07	0.26	0.37
1	0.11	0.10	0.32	0.53
Summe	0.15	0.17	0.58	**0.90**

$Z = 1$

X \ Y	0	1	2	Summe
0	0.00	0.01	0.02	0.03
1	0.01	0.02	0.04	0.07
Summe	0.01	0.03	0.06	**0.10**

8.3 Einige multivariate Verallgemeinerungen

Tab. 7.1.9: Bedingte Verteilung von Z gegeben $X = x$ und $Y = y$

Z \ (x, y)	(0,0)	(0,1)	(0,2)	(1,0)	(1,1)	(1,2)
0	1.0	0.875	0.93	0.92	0.83	0.89
1	0.0	0.125	0.07	0.08	0.17	0.11
Summe	1.0	1.00	1.00	1.00	1.00	1.00

Tab. 7.1.10: Bedingte Verteilung von $(X, Y)^T$ gegeben $Z = z$ von Beispiel D3-a

z=0 Y / X	0	1	2	Σ
0	0.04	0.08	0.29	0.41
1	0.12	0.11	0.36	0.59
Σ	0.16	0.19	0.65	1.00

z=1 Y / X	0	1	2	Σ
0	0.00	0.10	0.20	0.30
1	0.10	0.20	0.40	0.70
Σ	0.10	0.30	0.60	1.00

Fazit:

$$\left.\begin{aligned} & f_{Z|XY} \neq f_Z \\ \Leftrightarrow\ & f_{XY|Z} \neq f_{XY} \\ \Leftrightarrow\ & f_{XYZ} \neq f_{XY} f_Z \end{aligned}\right\} \quad (X, Y)^T \text{ und } Z \text{ sind abhängig}$$

8.3 Einige multivariate Verallgemeinerungen

• Beispiel D3-b •

Tab. 8.3.1: Gemeinsame Verteilung von $(X, Y, Z)^T$ von Beispiel D3-b

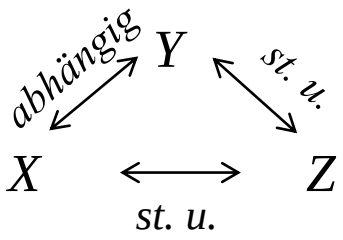
$Z = 0$

X \ Y	0	1	2	Summe
0	0.06	0.03	0.00	0.09
1	0.15	0.21	0.15	0.51
Summe	0.21	0.24	0.15	**0.60**

$Z = 1$

X \ Y	0	1	2	Summe
0	0.04	0.02	0.00	0.06
1	0.10	0.14	0.10	0.34
Summe	0.14	0.16	0.10	**0.40**

Hier:



?

$\Longrightarrow$ $(X, Y)^T$ und Z unabhängig

Antwort: Nein, nicht zwingend.
Es ist zu überprüfen.

8.3 Einige multivariate Verallgemeinerungen

Tab. 8.3.2: Randverteilung von $(X, Y)^T$ von Beispiel D3-b

X \ Y	0	1	2	$P(X = x)$
0	0.10	0.05	0.00	0.15
1	0.25	0.35	0.25	0.85
$P(Y = y)$	0.35	0.40	0.25	**1.00**

Tab. 8.3.3: Bedingte Verteilung von $(X, Y)^T$ gegeben $Z = z$ von Beispiel D3-b

z=0 X \ Y	0	1	2	Σ
0	0.10	0.05	0.00	0.15
1	0.25	0.35	0.25	0.85
Σ	0.35	0.40	0.25	1.00

z=1 X \ Y	0	1	2	Σ
0	0.10	0.05	0.00	0.15
1	0.25	0.35	0.25	0.85
Σ	0.35	0.40	0.25	1.00

Fazit:

$$\left.\begin{aligned} & f_{XYZ} = f_{XY} f_Z \\ \Leftrightarrow\ & f_{XY|Z} = f_{XY} \\ \Leftrightarrow\ & f_{Z|XY} = f_Z \end{aligned}\right\} \quad (X, Y)^T \text{ und } Z \text{ sind unabhängig}$$

Bedingte Abhängigkeit und Unabhängigkeit

● Definition und Hintergrund ●

Im vorhergehenden Beispiel D3-b wären X und Y unter Z bedingt unabhängig, falls gelten würde:

$$f_{XY|Z} = f_{X|Z}\, f_{Y|Z}$$

Das ist jedoch nicht der Fall.

● Beispiel 8.3.1 ●

Tab. 8.3.4: Beispiel einer 3-dimensionalen Verteilung mit bedingter Unabhängigkeit

	$Z=0$				$Z=1$		
X \ Y	0	1	Summe	X \ Y	0	1	Summe
0	0.05	0.05	0.10	0	0.05	0.20	0.25
1	0.20	0.20	0.40	1	0.05	0.20	0.25
Summe	0.25	0.25	**0.50**	Summe	0.10	0.40	**0.50**

Hier sind X und Y abhängig, bedingt unter Z jedoch unabhängig.

8.3.2 Kenngrößen multivariater Abhängigkeiten

Theoretische Kovarianz- und Korrelationsmatrix

- **Definition** -

Definition 8.3.2: Theoretische Kovarianz- und Korrelationsmatrizen

Die (theoretische) **Varianz-Kovarianzmatrix**, kurz **Kovarianzmatrix**, eines Zufallsvektors $\mathbf{X} = (X_1, X_2, \ldots, X_p)^T$ ist definiert als

$$\Sigma_X = Cov(\mathbf{X}) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \ldots & \sigma_{2p} \\ \vdots & & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \ldots & \sigma_{pp} \end{pmatrix}, \text{ wobei}$$

$\sigma_{ij} = Cov(X_i, X_j)$ für $i, j = 1, \ldots, p$ sind.

Die **(theoretische) Korrelationsmatrix von X** ist die zu Σ_X korrespondierende Matrix der Korrelationen (sofern $\sigma_{ii} > 0$ für $i = 1, \ldots, p$)

$$\varrho_{ij} = Corr(X_i, X_j) = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} \quad \text{für } i, j = 1, \ldots, p.$$

8.3 Einige multivariate Verallgemeinerungen

Für gegebene Zufallsvektoren

$$\mathbf{X} = (X_1, X_2, \ldots, X_p)^T \text{ und } \mathbf{Y} = (Y_1, Y_2, \ldots, Y_q)^T$$

ist der aus $\mathbf{X}$ und $\mathbf{Y}$ gestapelte Vektor $\mathbf{Z}$ definiert als

$$\mathbf{Z} = \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = (X_1, \ldots, X_p, Y_1, \ldots, Y_q)^T.$$

Dabei gelte folgende Partition:

$$Cov(\mathbf{Z}) = Cov\left(\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}\right) = \begin{pmatrix} \Sigma_{\mathbf{X}} & \Sigma_{\mathbf{XY}} \\ \Sigma_{\mathbf{YX}} & \Sigma_{\mathbf{Y}} \end{pmatrix} \tag{8.3.3}$$

Dann ist die **Kovarianzmatrix zwischen X und Y** gegeben durch

$$\Sigma_{\mathbf{XY}} = Cov(\mathbf{X}, \mathbf{Y})$$

und die **Kovarianzmatrix zwischen Y und X** durch

$$\Sigma_{\mathbf{YX}} = Cov(\mathbf{Y}, \mathbf{X}) = \Sigma_{\mathbf{XY}}^T.$$

Die **Korrelationsmatrix zwischen X und Y** bzw. **zwischen Y und X** ist die zu $\Sigma_{\mathbf{XY}}$ bzw. $\Sigma_{\mathbf{YX}}$ korrespondierende Matrix der Korrelationen.

● Beispiel 8.3.2 ●

Für das vorhergehende Beispiel D3-b erhält man mit den Regeln zur Berechnung von Kovarianzen als Varianz-Kovarianzmatrix des Vektors $(X, Y, Z)^T$:

$$Cov\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} Cov(X,X) & Cov(X,Y) & Cov(X,Z) \\ Cov(Y,X) & Cov(Y,Y) & Cov(Y,Z) \\ Cov(Z,X) & Cov(Z,Y) & Cov(Z,Z) \end{pmatrix}$$

$$= \begin{pmatrix} 0.13 & 0.09 & 0 \\ 0.09 & 0.59 & 0 \\ 0 & 0 & 0.24 \end{pmatrix}$$

Daraus erhält man als Kovarianzmatrizen zwischen $(X, Y)^T$ und Z und zwischen Z und $(X, Y)^T$:

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ bzw. } (0 \quad 0)$$

Man beachten, dass dies bereits aus der Unabhängigkeit von $(X, Y)^T$ und Z folgt.

Bedingte Kennwerte für $n > 2$

- **Hintergrund**
- **Allgemeine Vorgehensweise**

> Berechnung bedingter Kennwerte setzt an bedingten Verteilungen an

> Beispiele:

Für $E(Y|X = x, Z = z)$ und $E(XY|Z = z)$

benötigt man $f_{Y|XZ} = \frac{f_{XYZ}}{f_{XZ}}$ bzw. $f_{XY|Z} = \frac{f_{XYZ}}{f_Z}$

> Vereinfachungsmöglichkeiten bei Vorliegen von Unabhängigkeiten:

– X, Y und Z unabhängig $\Rightarrow$ $f_{Y|XZ} = \frac{f_{XYZ}}{f_{XZ}} = \frac{f_X f_Y f_Z}{f_X f_Z} = f_Y$

$\Rightarrow$ $E(Y|X = x, Z = z) = E(Y)$

– Z und $(X, Y)^T$ unabhängig $\Rightarrow \quad f_{XY|Z} = \frac{f_{XYZ}}{f_Z} = \frac{f_{XY} f_Z}{f_Z} = f_{XY}$

$$\Rightarrow \quad E(XY|Z = z) = E(XY)$$

> Umformung bei Ausdrücken mit Variablen, die „vorne“ und „hinten“ stehen:

$$E(XYZ|Z = z) = zE(XY|Z = z)$$

● Beispiel D3-a fortgesetzt ●

Beispielhafte Berechnungen:

$$E(Z|X = 0, Y = 0) = \quad 0$$

$$E(Z|X = 0, Y = 1) = \quad 0.125$$

$$E(Z|X = 1, Y = 2) = \quad 0.11$$

$$E(XY|Z=0) = 1 \times 1 \times 0.11 + 1 \times 2 \times 0.36 = 0.83$$

$$E(XY|Z=1) = 1 \times 1 \times 0.2 + 1 \times 2 \times 0.4 = 1$$

$$\begin{aligned} Cov(X,Y|Z=0) &= E(XY|Z=0) - E(X|Z=0)E(Y|Z=0) \\ &= 0.83 - 0.59 \times (0.19 + 2 \times 0.65) \\ &= -0.0491 \end{aligned}$$

$$\begin{aligned} Cov(X,Y|Z=1) &= E(XY|Z=1) - E(X|Z=1)E(Y|Z=1) \\ &= 1 - 0.7 \times (0.3 + 2 \times 0.6) \\ &= -0.05 \end{aligned}$$

● **Beispiel D3-b fortgesetzt** ●

Hier sind X und Z, Y und Z und $(X,Y)^T$ und Z unabhängig

$$\Rightarrow \quad f_{Z|XY} = f_Z$$

$$\Rightarrow \quad f_{Y|XZ} = \frac{f_{XYZ}}{f_{XZ}} = \frac{f_{XY} f_Z}{f_X f_Z} = \frac{f_{XY}}{f_X} = f_{Y|X}$$

Beispielhafte Berechnungen:

$$E(Z|X = x, Y = y) = E(Z) = 0.4$$

$$E(Y|X = 0, Z = 0) = E(Y|X = 0) = 0.05/0.15 \approx 0.33$$

● **Beispiel 8.3.3** ●

In Tabelle 8.3.4 (Folie 591) gilt für jedes $z = 0{,}1$:

$$Cov(X, Y|Z = z) = 0$$

Andererseits gilt:

$$Cov(X, Y) = -0.0225$$

Somit sind X und Y korreliert und bedingt unter Z unkorreliert.

8.3.3 Sonstige Verallgemeinerungen

- **Funktionen unabhängiger und identisch verteilter Zufallsvektoren** -

Satz 8.3.1: Funktionen von unabhängigen Zufallsvektoren

Für gegebene Zufallsvektoren (diskret oder stetig)

$$\mathbf{X} = (X_1, X_2, \ldots, X_p)^T \quad \text{und} \quad \mathbf{Y} = (Y_1, Y_2, \ldots, Y_q)^T$$

und beliebige (messbare) Funktionen

$$g_1 : \mathbb{R}^p \to \mathbb{R}^r \quad \text{und} \quad g_2 : \mathbb{R}^q \to \mathbb{R}^s$$

gilt: Sind $\mathbf{X}$ und $\mathbf{Y}$

(i) stochastisch unabhängig, so sind auch $g_1(\mathbf{X})$ und $g_2(\mathbf{Y})$ stochastisch unabhängig.

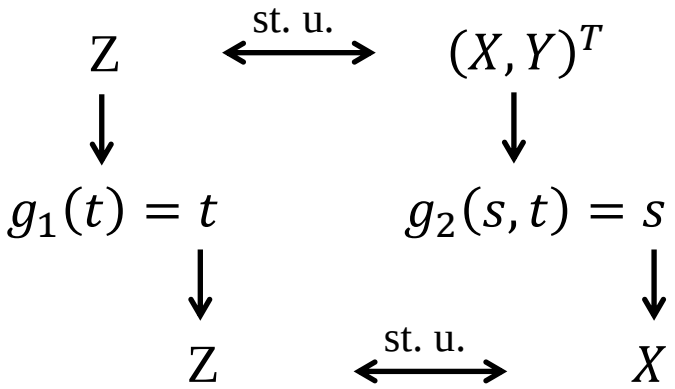
(ii) identisch verteilt, so sind auch $g_1(\mathbf{X})$ und $g_1(\mathbf{Y})$ identisch verteilt.

(iii) stochastisch unabhängig und identisch verteilt, so sind auch $g_1(\mathbf{X})$ und $g_1(\mathbf{Y})$ unabhängig und identisch verteilt.

Erweiterbar auf n unabhängige Vektoren $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$

● Beispiel D3-b fortgesetzt ●

Wie bereits festgestellt folgt aus der Unabhängigkeit von X und Z und Y und Z nicht diejenige von $(X, Y)^T$ und Z (vgl. Folie 589). Die Umkehrung gilt jedoch. Warum?



● Anwendungsbeispiel: Multivariate Stichproben ●

Stichprobe von Personen mit X_i als Größe und Y_i als Gewicht der i-ten Person (i = 1, ..., n), modelliert über unabhängige Vektoren (X_i, Y_i):

$$(X_1, Y_1), (X_2, Y_2), \ldots (X_n, Y_n)$$

8.3 Einige multivariate Verallgemeinerungen

● Multivariate Normalverteilung ●

> $\mathbf{X} = (X_1, X_2, \ldots, X_p)^T$ ist **multivariat normalverteilt** bzw. $X_1, \ldots, X_p$ sind **gemeinsam normalverteilt**, falls die gemeinsame Dichte gegeben ist durch

$$f_{\mathbf{X}}(x_1, \ldots, x_p) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} exp\left(-0.5(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

Kurzschreibweise: $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$

Voraussetzung: $|\Sigma| > 0$ (Invertierbarkeit)

Notation: $\mathbf{x} = (x_1, x_2, \ldots, x_p)^T$, $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_p)^T$,

$\Sigma = Cov(\mathbf{X})$, $|\Sigma|$ ... Determinante von Σ

> Spezialfall: Dichte einer normalverteilten Zufallsvariable X:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma}} exp\left(-0.5\frac{(x-\mu)^2}{\sigma^2}\right)$$

Voraussetzung: $\sigma^2 > 0$

Kurzschreibweise: $X \sim N(\mu, \sigma^2)$

8.3 Einige multivariate Verallgemeinerungen

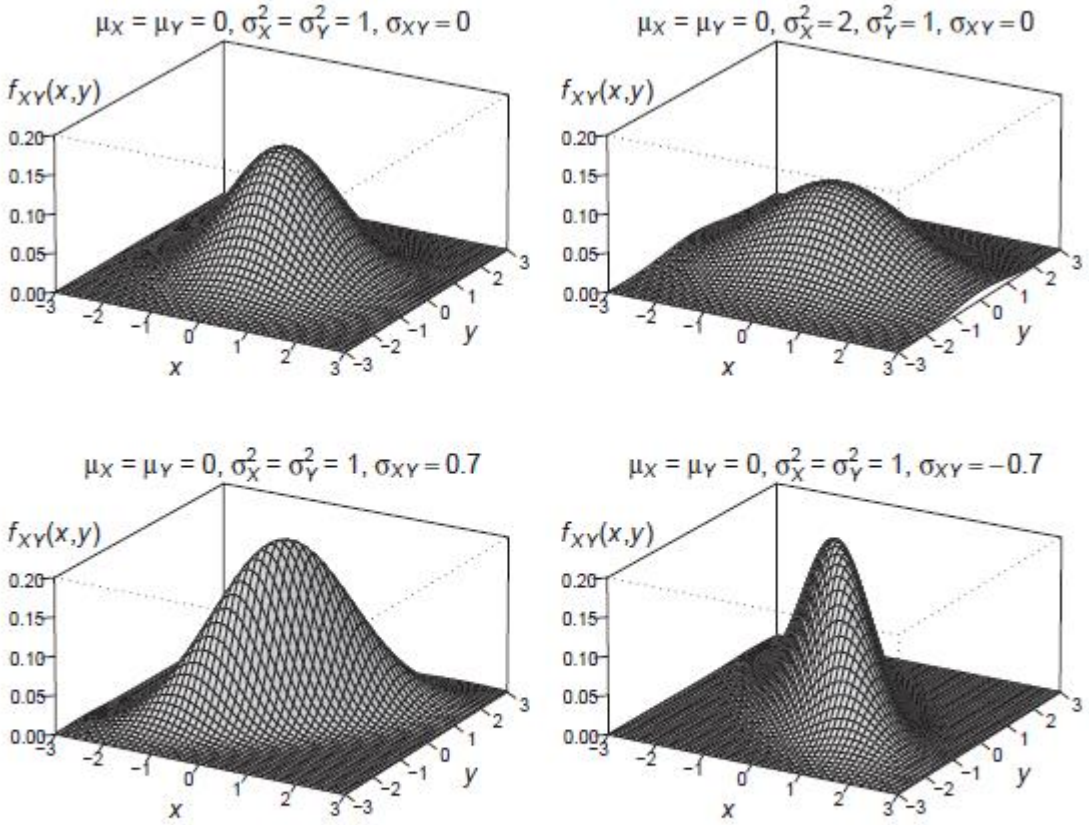


Abb. 8.3.1: Beispiele von Dichten der bivariaten Normalverteilung

8.3 Einige multivariate Verallgemeinerungen

Satz 8.3.2: Ableitungen aus der multivariaten Normalverteilung

Für einen r-dimensionalen Zufallsvektor $\mathbf{Z}$ gelte

$$Z \sim N_r(\mu_{\mathbf{Z}}, \Sigma_{\mathbf{ZZ}}).$$

Sei weiter $\mathbf{C}$ eine feste (nichtstochastische) $(s \times r)$-Matrix und $\mathbf{c}_0$ ein fester (nichtstochastischer) s-dimensionaler Vektor. Dann gilt:

$$\mathbf{c}_0 + \mathbf{CZ} \sim N_s(\mathbf{C}\mu_{\mathbf{Z}}, \mathbf{C}\Sigma_{\mathbf{ZZ}}\mathbf{C}^T) \text{ und} \tag{8.3.6}$$

$$(\mathbf{Z} - \mu_{\mathbf{Z}})^T \Sigma_{\mathbf{ZZ}}^{-1} (\mathbf{Z} - \mu_{\mathbf{Z}}) \sim \chi^2(r). \tag{8.3.7}$$

In Anlehnung an die Notation aus Abschnitt 8.3.2 gelte

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim N_{p+q}\left(\begin{pmatrix} \mu_{\mathbf{X}} \\ \mu_{\mathbf{Y}} \end{pmatrix}, \begin{pmatrix} \Sigma_{\mathbf{XX}} & \Sigma_{\mathbf{XY}} \\ \Sigma_{\mathbf{YX}} & \Sigma_{\mathbf{YY}} \end{pmatrix}\right).$$

Dann folgt daraus:

$$\mathbf{X} \sim N_p(\mu_{\mathbf{X}}, \Sigma_{\mathbf{XX}}), \tag{8.3.8}$$

$$\mathbf{Y} \sim N_q(\mu_{\mathbf{Y}}, \Sigma_{\mathbf{YY}}) \text{ und} \tag{8.3.9}$$

$$\mathbf{Y}|\mathbf{X} = \mathbf{x} \sim N_q(\mu_{\mathbf{Y}} + \Sigma_{\mathbf{YX}}\Sigma_{\mathbf{XX}}^{-1}(\mathbf{x} - \mu_{\mathbf{X}}), \Sigma_{\mathbf{YY}} - \Sigma_{\mathbf{YX}}\Sigma_{\mathbf{XX}}^{-1}\Sigma_{\mathbf{XY}}), \tag{8.3.10}$$

$$\mathbf{X}|\mathbf{Y} = \mathbf{y} \sim N_p(\mu_{\mathbf{X}} + \Sigma_{\mathbf{XY}}\Sigma_{\mathbf{YY}}^{-1}(\mathbf{y} - \mu_{\mathbf{Y}}), \Sigma_{\mathbf{XX}} - \Sigma_{\mathbf{XY}}\Sigma_{\mathbf{YY}}^{-1}\Sigma_{\mathbf{YX}}), \tag{8.3.11}$$

wobei $\mathbf{Y}|\mathbf{X} = \mathbf{x}$ die bedingte Verteilung von $\mathbf{Y}$ unter $\mathbf{X} = \mathbf{x}$ und $\mathbf{X}|\mathbf{Y} = \mathbf{y}$ die bedingte Verteilung von $\mathbf{X}$ unter $\mathbf{Y} = \mathbf{y}$ bezeichnet.

Anhand von (8.3.10) bzw. (8.3.11) lässt sich feststellen:
Sofern $\mathbf{X}$ und $\mathbf{Y}$ unkorreliert sind, also $\Sigma_{\mathbf{XY}} = \mathbf{0}$ bzw. $\Sigma_{\mathbf{YX}} = \mathbf{0}$ gilt, folgt

$$\mathbf{Y}|\mathbf{X} = \mathbf{x} \sim N_q(\boldsymbol{\mu}_{\mathbf{Y}}, \Sigma_{\mathbf{YY}}) \text{ bzw. } \mathbf{X}|\mathbf{Y} = \mathbf{y} \sim N_q(\boldsymbol{\mu}_{\mathbf{X}}, \Sigma_{\mathbf{XX}})$$

$\Rightarrow$ die bedingten Verteilungen entsprechen dann den Randverteilungen

$\Rightarrow$ $\mathbf{X}$ und $\mathbf{Y}$ sind unabhängig

Fazit: Aus Unabhängigkeit folgt stets Unkorreliertheit jedoch nicht umgekehrt. Bei gemeinsamer Normalverteilung gilt jedoch auch die Umkehrung, d. h. Unkorreliertheit impliziert Unabhängigkeit.

Sind $\mathbf{X}$ und $\mathbf{Y}$ gemeinsam normalverteilt, so sind Unkorreliertheit und Unabhängigkeit äquivalent.

● **Beispiel 8.3.4** ●

Für $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix}\right)$

lassen sich mit (8.3.8), (8.3.9), (8.3.6) und (8.3.10) folgende Resultate ableiten:

– $X \sim N(\mu_X, \sigma_X^2)$ bzw. $Y \sim N(\mu_Y, \sigma_Y^2)$

(Randverteilungen)

– $c_0 + c_1 X + c_2 Y \sim N(c_0 + c_1\mu_X + c_2\mu_Y, c_1^2\sigma_X^2 + c_2^2\sigma_Y^2 + 2c_1c_2\sigma_{XY})$

(Summen normalverteilt auch bei Abhängigkeit)

– $Y|X = x \sim N\left(\mu_Y + \frac{\sigma_{XY}}{\sigma_X^2}(x - \mu_X), \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2}\right)$

(bedingte Verteilung)

● **Beispiel 8.3.5** ●

Für $\begin{pmatrix} Z \\ X \\ Y \end{pmatrix} \sim N_3\left(\begin{pmatrix} \mu_Z \\ \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_Z^2 & \sigma_{ZX} & \sigma_{ZY} \\ \sigma_{ZX} & \sigma_X^2 & \sigma_{XY} \\ \sigma_{YZ} & \sigma_{XY} & \sigma_Y^2 \end{pmatrix}\right)$

lässt sich mit (8.3.10) folgendes Resultat ableiten:

$$\Sigma_{XY|Z} = \begin{pmatrix} \sigma_{X|Z}^2 & \sigma_{XY|Z} \\ \sigma_{XY|Z} & \sigma_{Y|Z}^2 \end{pmatrix} = \cdots$$

$$= \begin{pmatrix} \sigma_X^2 - \sigma_{XZ}^2/\sigma_Z^2 & \sigma_{XY} - \sigma_{XZ}\sigma_{YZ}/\sigma_Z^2 \\ \sigma_{XY} - \sigma_{XZ}\sigma_{YZ}/\sigma_Z^2 & \sigma_Y^2 - \sigma_{YZ}^2/\sigma_Z^2 \end{pmatrix}$$

$$\Rightarrow \quad \rho_{XY|Z} = Corr(X, Y|Z) = \sigma_{XY|Z} / \sqrt{\sigma_{X|Z}^2 \, \sigma_{Y|Z}^2} = \cdots$$

$$= \frac{\rho_{XY} - \rho_{XZ}\rho_{YZ}}{\sqrt{(1-\rho_{XZ}^2)(1-\rho_{YZ}^2)}}$$

Bedingte Korrelation
(= partielle Korrelation)

8.3 Einige multivariate Verallgemeinerungen

• Multivariate asymptotische Aussagen •

> Es ist möglich, alle Definitionen und Sätze in Bezug auf stochastische Konvergenzen und Verteilungskonvergenzen auf den Vektorfall zu verallgemeiner (Multivariate zentrale Grenzwertsätze)

> Beschränkung hier auf multivariate Version des Stetigkeitssatzes (vgl. Satz 8.2.3 auf Folie 580):

Satz 8.3.3: Multivariater Stetigkeitssatz

Sei $(\mathbf{Y}_n)$ eine Folge p-dimensionaler Zufallsvektoren, $h : \mathbb{R}^p \to \mathbb{R}^q$ eine stetige Funktion und θ ein Vektor fester Werte. Sofern $h(\theta)$ definiert ist und gilt

$$\mathbf{Y}_n \xrightarrow{p} \theta,$$

folgt daraus für $\mathbf{U}_n = h(\mathbf{Y}_n)$:

$$\mathbf{U}_n \xrightarrow{p} h(\theta) \quad \text{bzw.} \quad h(\mathbf{Y}_n) \xrightarrow{p} h(\theta).$$

● Beispiel 8.3.6 ●

> Seien $(X_1, Y_1)^T, (X_2, Y_2)^T, \ldots, (X_n, Y_n)^T$ u. i. v. wie $(X, Y)^T$

> Mit Satz 8.3.1 (Folie 600) folgt dann:

$\Rightarrow \quad X_1, \ldots, X_n$ u. i. v. wie X

$\Rightarrow \quad Y_1, \ldots, Y_n$ u. i. v. wie Y

$\Rightarrow \quad X_1^2, \ldots, X_n^2$ u. i. v. wie X^2

$\Rightarrow \quad X_1 Y_1, \ldots, X_n Y_n$ u. i. v. wie XY

> Mit dem **GGZ für *k*-te Momente** (Folie 581) folgt dann weiter:

$$\Rightarrow \quad \bar{X}_n \xrightarrow{p} E(X)$$

$$\Rightarrow \quad \bar{Y}_n \xrightarrow{p} E(Y)$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{p} E(X^2)$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n} X_i Y_i \xrightarrow{p} E(XY)$$

> Vektoriell zusammengefasst:

$$\left(\bar{X}_n, \bar{Y}_n, \frac{1}{n}\sum_{i=1}^{n} X_i^2, \frac{1}{n}\sum_{i=1}^{n} X_i Y_i\right)^T \xrightarrow{p} \left(E(X), E(Y), E(X^2), E(XY)\right)^T$$

> Mit dem multivariaten Stetigkeitssatz folgt dann:

$$\Rightarrow \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2}_{} \xrightarrow{p} E(X^2) - E(X)^2 = Var(X)$$

$= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ ... Stichprobenvarianz

$$\Rightarrow \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \bar{X}_n \bar{Y}_n}_{} \xrightarrow{p} E(XY) - E(X)E(Y) = Cov(X, Y)$$

$= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)(Y_i - \bar{Y}_n)$ ... Stichprobenkovarianz

● Gemischt skalierte Vektoren ●

Zufallsvektor enthält sowohl diskrete als auch stetige Zufallsvariablen

● Beispiel 8.3.7 ●

Analysiert man den Zusammenhang zwischen Haushaltsgröße (Anzahl von Personen im Haushalt) und Haushaltsbruttoeinkommen (in Euro), so erscheint auf theoretischer Ebene eine Modellierung mit einer diskreten Zufallsvariablen X (Haushaltsgröße) und einer stetigen Variablen Y (Einkommen) adäquat.
Bei theoretischen Berechnungen werden nun Additions- und Integrationskalkül gemischt (siehe LB, S. 473-474 für weitere Ausführungen hierzu).

Teil 3:
Induktive Statistik - Verbindung von Empirie und Theorie

9 Einführung in die induktive Statistik

9 Einführung in die induktive Statistik

9.1 Modellierung und Handhabung von Schätz- und Testproblemen	616
9.2 Was versteht man unter einer Stichprobe	622
9.3 Formale Unterscheidungsebenen	631

• Hintergrund und Überblick •

	Induktive Statistik	
Statistisches Problem:	Schätzproblem	Entscheidungs−/Testproblem
Methodischer Ansatz:	Statistisches Schätzen	Statistisches Testen
Quantifizierung von...	Sicherheitswahrscheinlichkeiten	Irrtumswahrscheinlichkeiten

Abb. 9.1.1: Methodik der induktiven Statistik

• Beispiel 9.1.1: Mensabefragung •

Von 50 zufällig befragten Studierenden (Stichprobe) präferieren 30 die Mensa und 20 die Stadt zum Essengehen.

Statistisches Entscheidungsproblem (später: Testproblem):
Welches Mehrheitsverhältnis innerhalb einer Stichprobe vom Umfang $n = 50$ spricht „evident“ dafür, dass von einer mehrheitlichen Präferenz unter allen Studierenden ausgegangen werden kann?

Statistisches Schätzproblem:
Was ist der tatsächliche Anteil von Befürwortern unter allen Studierenden?
Wie lässt sich ein wohlbegründetes (Sicherheit-)Intervall dafür konstruieren?

Statistisches Modell:

Sei π der tatsächliche Anteil von Mensabefürwortern unter allen Studierenden. Dann lässt sich die Präferenz eines einzelnen zufällig ausgewählten Studierenden über $X \sim B(1, \pi)$ modellieren:

$$P(X = 1) = \pi \quad \text{und} \quad P(X = 0) = 1 - \pi$$

Eine Stichprobe vom Umfang $n = 50$ wird dann modelliert über unabhängige und identisch wie X verteilte Zufallsvariablen $X_1, \ldots, X_{50}$

Anzahl von Mensabefürwortern:

$$S_{50} = \sum_{i=1}^{50} X_i \sim B(50, \pi)$$

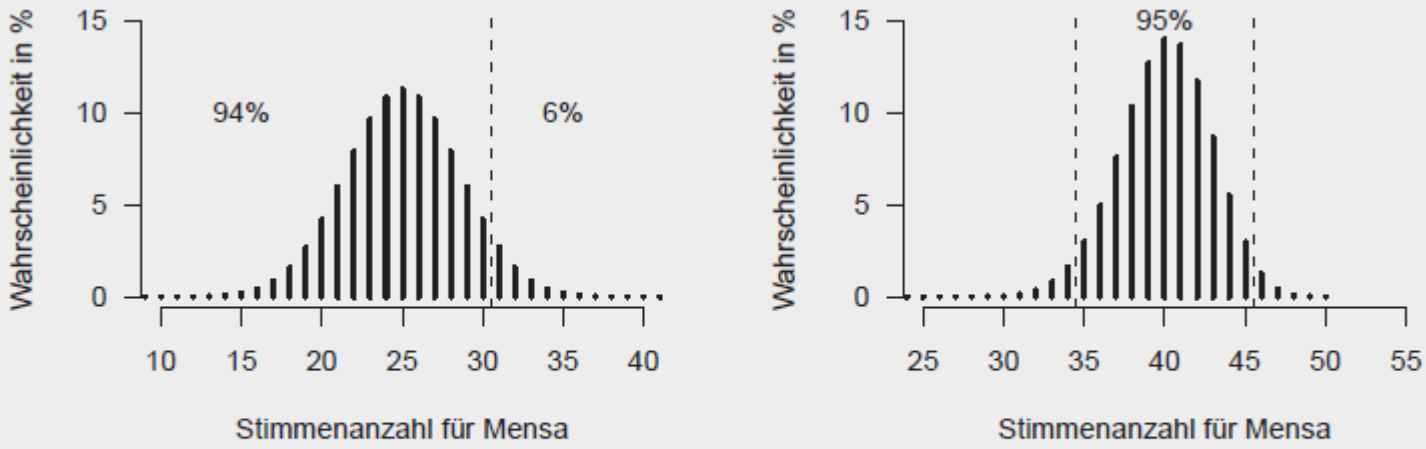


Abb. 1.2.3: 50% Zustimmung zur Mensa bzw. 80% Zustimmung zur Mensa

Vorgehensweise...

... bei Entscheidungsproblem (später: Testproblem)
Festlegung von **kritischen Werten** (Entscheidungsgrenzen) bzw. **Irrtumswahrscheinlichkeiten**

... bei Schätzproblem:
Festlegung eines **Konfidenzintervalls** (Sicherheitsintervalls) bzw. eines **Konfidenzniveaus** (Sicherheitsniveaus)

● Beispiel 9.1.2: Gepäckabfertigungszeiten ●

Zur Bedarfsermittlung von Personal bei der Gepäckannahme an einem Flughafen soll die durchschnittliche Abfertigungszeit geschätzt werden. Hierzu werden stichprobenhaft 10 Abfertigungsvorgänge zeitlich gemessen.

Statistisches Schätzproblem:
Wie hoch ist die tatsächliche durchschnittliche Abfertigungszeit?
Wie lässt sich ein wohlbegründetes (Sicherheit-)Intervall dafür konstruieren?

Statistisches Modell:

Die gemessenen Zeiten können beispielsweise über unabhängige $N(\mu, \sigma^2)$-verteilte Zufallsvariablen $X_1, \ldots, X_{10}$ modelliert werden, wobei dann der Parameter μ die interessierende „Zielgröße“ darstellt. Für das stochastische Mittel der Stichprobe gilt dann (vgl. Kap. 7, Folie 540):

$$\bar{X}_n = \frac{1}{10}\sum_{i=1}^{10} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

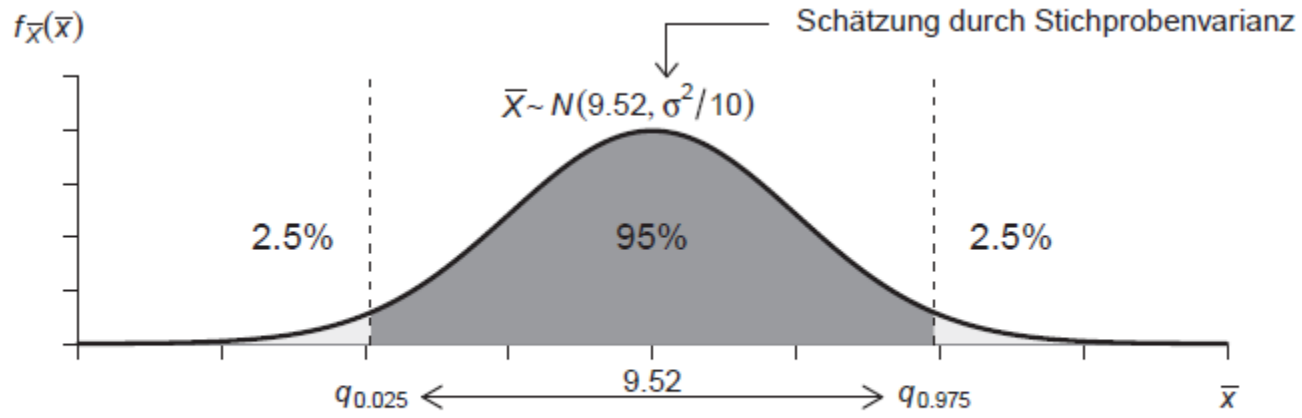


Abb. 9.1.2: Schwankungsverhalten des Stichprobenmittels für $\mu = 9.52$

Vorgehensweise:

Festlegung eines **Konfidenzintervalls** (Sicherheitsintervalls) bzw. eines **Konfidenzniveaus** (Sicherheitsniveaus)

Spezielles Problem hier:
$q_{0.025}$ und $q_{0.975}$ können ohne Kenntnis von σ^2 nicht ohne Weiteres ermittelt werden ($\Rightarrow$ Schätzung)

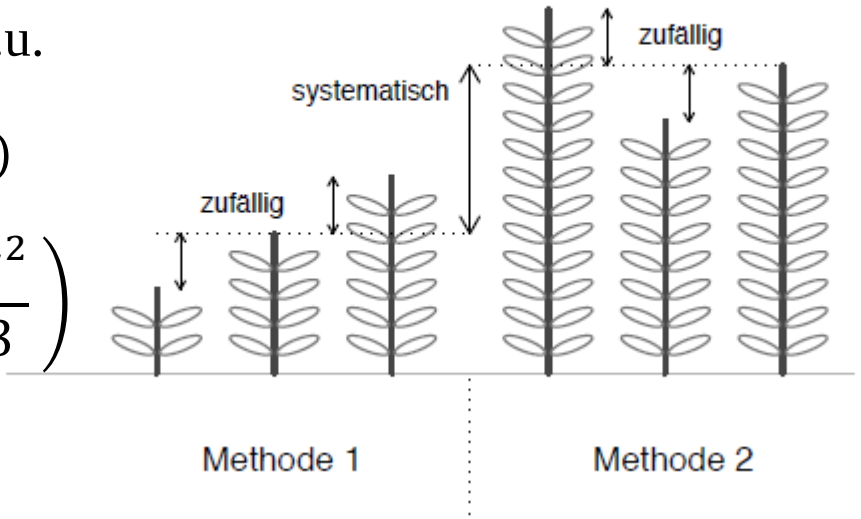
• Beispiel 9.1.3: Pflanzenexperiment •

Abb. 1.2.1: Streuungszerlegung – Anbaumethoden und Ertrag von Pflanzen

Y_{01}, Y_{02}, Y_{03} st.u.

$Y_{0i} \sim N(\mu_0, \sigma^2)$

$\Rightarrow \quad \bar{Y}_0 \sim N\left(\mu_0, \frac{\sigma^2}{3}\right)$



Y_{11}, Y_{12}, Y_{13} st.u.

$Y_{1i} \sim N(\mu_1, \sigma^2)$

$\Rightarrow \quad \bar{Y}_1 \sim N\left(\mu_1, \frac{\sigma^2}{3}\right)$

$$\Rightarrow \quad \bar{Y}_1 - \bar{Y}_0 \sim N\left(\mu_1 - \mu_0, \frac{2\sigma^2}{3}\right)$$

• Zusammenfassung •

9.2 Was versteht man unter einer Stichprobe?

• Stichproben und Stichprobenvariablen •

Definition 9.2.1: Stichprobe und Stichprobenvariablen

In der induktiven Statistik versteht man unter einer ***Stichprobe*** ein Ensemble von Zufallsvariablen, das bei der statistischen Modellierung eines Schätz- oder Testproblems jeweils relevant ist. Die Zufallsvariablen heißen dann auch ***Stichprobenvariablen***.

• Realisation einer Stichprobe •

Definition 9.2.2: Realisation einer Stichprobe

Unter der ***Realisation einer Stichprobe*** versteht man die korrespondierend zu einer Stichprobe $X_1, \ldots, X_n$ konkret vorliegenden Beobachtungswerte (Realisationen), welche mit Kleinbuchstaben $x_1, \ldots, x_n$ notiert werden.

Beispiel: Realisation einer Stichprobe in Beispiel 9.1.1 (Mensa vs. Stadt)

0, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 0, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1, 1, 0,
0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0.

9.2 Was versteht man unter einer Stichprobe

- **Induktion**
- **Reale Grundgesamtheit und reale Stichprobe**
- **Urnenmodell und Ziehungsschemata**

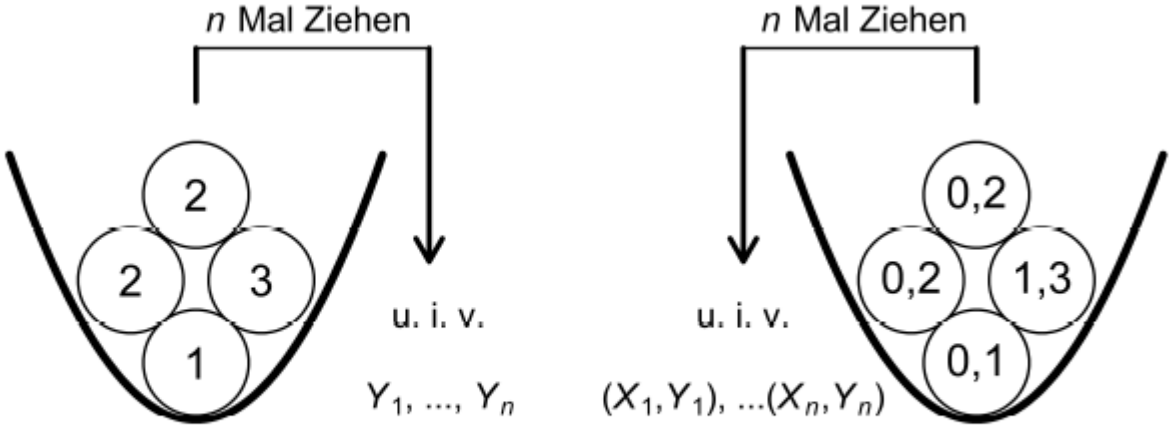


Abb. 9.2.1: Illustration realer Stichproben

9.2 Was versteht man unter einer Stichprobe

Tab. 9.2.1: Gemeinsame Verteilung von Y_1 und Y_2 (mit bzw. ohne Zurücklegen)

Y_1 \ Y_2	1	2	3	Σ
1	1/16	1/8	1/16	0.25
2	1/8	1/4	1/8	0.50
3	1/16	1/8	1/16	0.25
Σ	0.25	0.50	0.25	1.00

Y_1 \ Y_2	1	2	3	Σ
1	0	1/6	1/12	0.25
2	1/6	1/6	1/6	0.50
3	1/12	1/6	0	0.25
Σ	0.25	0.50	0.25	1.00

> Ziehen mit Zurücklegen

$$P(Y_1 = y_1 | Y_2 = y_2) = P(Y_1 = y_1) \text{ bzw.}$$

$$P(Y_1 = y_1, Y_2 = y_2) = P(Y_1 = y_1)P(Y_2 = y_2)$$

$\Rightarrow$ Y_1 und Y_2 sind unabhängig und identisch verteilt (u. i. v.)

> Ziehen ohne Zurücklegen

$$P(Y_1 = y_1 | Y_2 = y_2) \neq P(Y_1 = y_1) \text{ bzw.}$$

$$P(Y_1 = y_1, Y_2 = y_2) = P(Y_2 = y_2 | Y_1 = y_1)P(Y_1 = y_1)$$

$\Rightarrow$ Y_1 und Y_2 sind abhängig, jedoch identisch verteilt

Tab. 9.2.2: Gemeinsame Verteilung von $(X_1, Y_1)^T$ und $(X_2, Y_2)^T$ beim Ziehen „mit Zurücklegen“

Tupel	Wahrschein.	Tupel	Wahrschein.
(0, 1, 0, 1)	1/16	(1, 3, 0, 1)	1/16
(0, 1, 0, 2)	1/8	(0, 2, 1, 3)	1/8
(0, 2, 0, 1)	1/8	(1, 3, 0, 2)	1/8
(0, 1, 1, 3)	1/16	(1, 3, 1, 3)	1/16
(0, 2, 0, 2)	1/4		

Beim Ziehen mit Zurücklegen gilt:

$$f_{X_1Y_1X_2Y_2} = f_{X_1Y_1}\, f_{X_2Y_2}$$

bzw.

$$P(X_1 = x_1, Y_1 = y_1, X_2 = x_2, Y_2 = y_2) = P(X_1 = x_1, Y_1 = y_1)P(X_2 = x_2, Y_2 = y_2)$$

$\Rightarrow$ $(X_1, Y_1)^T$ und $(X_2, Y_2)^T$ sind unabhängig und identisch verteilt (u. i. v.)

9.2 Was versteht man unter einer Stichprobe

Tab. 9.2.3: Gemeinsame Verteilung von $(X_1, Y_1)^T$ und $(X_2, Y_2)^T$ bei Ziehen „ohne Zurücklegen“

Tupel	Wahrschein.	Tupel	Wahrschein.
(0, 1, 0, 1)	0	(1, 3, 0, 1)	1/12
(0, 1, 0, 2)	1/6	(0, 2, 1, 3)	1/6
(0, 2, 0, 1)	1/6	(1, 3, 0, 2)	1/6
(0, 1, 1, 3)	1/12	(1, 3, 1, 3)	0
(0, 2, 0, 2)	1/12		

Beim Ziehen ohne Zurücklegen gilt:

$$f_{X_1Y_1X_2Y_2} \neq f_{X_1Y_1}\, f_{X_2Y_2}$$

Stattdessen:

$$P(X_1 = x_1, Y_1 = y_1, X_2 = x_2, Y_2 = y_2)$$

$$= P(X_2 = x_2, Y_2 = y_2 | X_1 = x_1, Y_1 = y_1) P(X_1 = x_1, Y_1 = y_1)$$

$\Rightarrow$ $(X_1, Y_1)^T$ und $(X_2, Y_2)^T$ sind abhängig, jedoch identisch verteilt

9.2 Was versteht man unter einer Stichprobe

Satz 9.2.1: Urnenmodell und Ziehungsschemata

Gegeben sei eine Urne mit $N \geq 1$ Kugeln, die jeweils mit $p \geq 1$ Zahlen beschriftet sind. Es werden n Kugeln zufällig gezogen, wobei $1 \leq n \leq N$. Die gezogenen p-dimensionalen Zahlentupel werden durch die Zufallsvektoren $\mathbf{X}_1, \ldots, \mathbf{X}_n$ modelliert. Dann sind $\mathbf{X}_1, \ldots, \mathbf{X}_n$

(i) unabhängig und identisch verteilt (**u.i.v.**) wie $\mathbf{X}$, falls mit Zurücklegen gezogen wird,
(ii) abhängig identisch verteilt (**a.i.v.**) wie $\mathbf{X}$, falls ohne Zurücklegen gezogen wird,
(iii) annähernd unabhängig identisch verteilt wie $\mathbf{X}$, falls ohne Zurücklegen gezogen wird und der Auswahlsatz n/N sehr klein ist.

Dabei steht $\mathbf{X}$ für ein einmaliges zufälliges Ziehen. Die Verteilung von $\mathbf{X}$ deckt sich dann mit der empirischen Verteilung der Zahlen in der Urne.

9.2 Was versteht man unter einer Stichprobe

- **Hypothetische Grundgesamtheit und hypothetische Stichproben** -

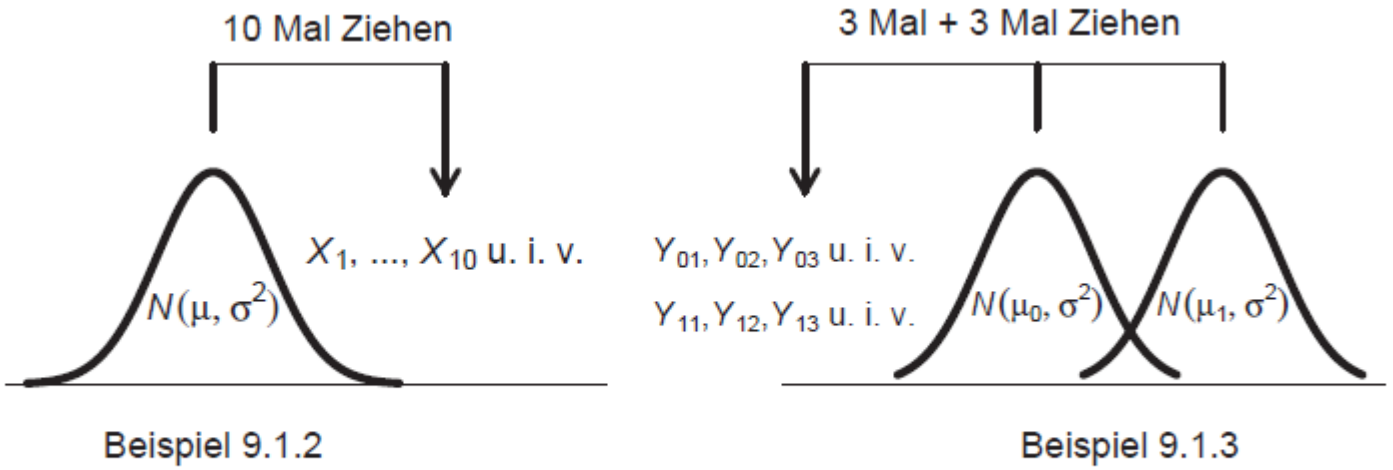


Abb. 9.2.2: Illustration hypothetischer Stichproben

vgl. Folie 618

vgl. Folie 620

> Urnenmodell nicht immer operabel; Stichprobe wird vielmehr aus einem hypothetisch unterstellen Verteilungsmuster gezogen

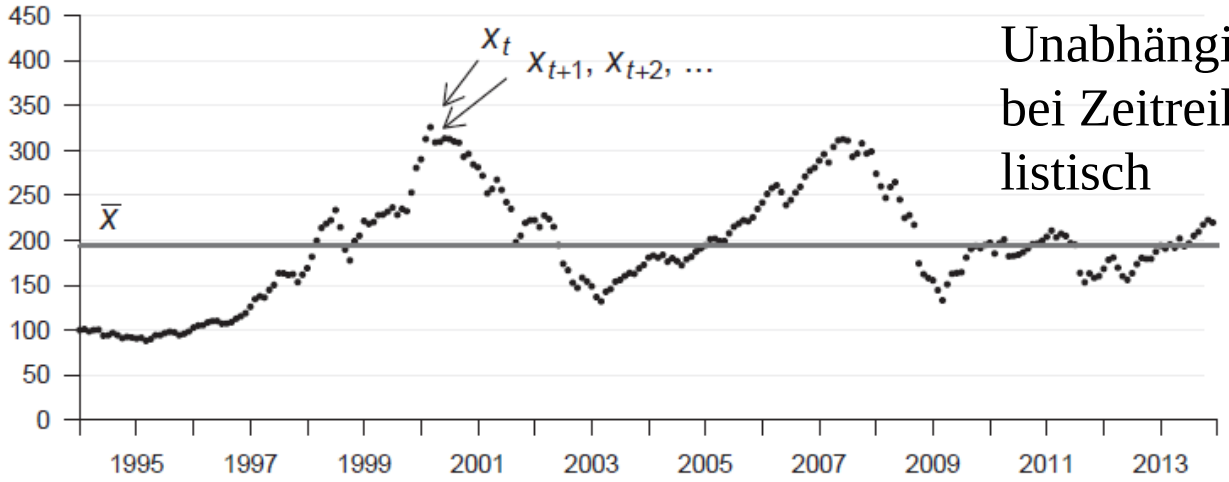
> Allgemein-typisch unterstelltes Potential (an Möglichkeiten) oder langfristig sich einstellende Verteilung

9.2 Was versteht man unter einer Stichprobe

- **Real vs. hypothetisch**
- **Spezifikation des Ziehungsschemas**

> Typische Schemata wie **u. i. v.**, **u. h. v.**, **a. i. v.** oder **a. h. v.**

> Typische Verteilungsannahmen: Bernoulli-Verteilung, Normalverteilung, ...



Unabhängigkeitsannahme bei Zeitreihen meist unrealistisch

Abb. 9.2.3: Positiv autokorrelierte Zeitreihe

- **Quasi-Stichproben**

9.2 Was versteht man unter einer Stichprobe

- **Repräsentativität**

Definition 9.2.3: Repräsentative Stichprobe

Eine **repräsentative Stichprobe** ist eine solche, bei der sich mit wachsendem Stichprobenumfang die Verteilung der Realisationen der Verteilung der interessierenden Merkmale (Variablen) in der Grundgesamtheit annähert.

Repräsentativität bezieht sich auf das stochastische Grenzverhalten einer Stichprobe, nicht auf konkrete Realisationen (Beobachtungen).

- **Prüfung von Repräsentativität**
- **Repräsentativität, Ziehungsschema und Induktion**
- **Ausblick: Beschränkung auf u.i.v.- und u.h.v.-Schemata**

9.3 Formale Unterscheidungsebenen

- **Zufallsvariablen und Realisationen**

  X, Y, Z oder $X_1, \ldots, X_n$ vs. x, y, z oder $x_1, \ldots, x_n$

- **Stochastische und empirische Statistiken**

  Beispiel: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ vs. $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

- **Empirischer vs. theoretischer Kalkül**

- **Empirische vs. theoretische Kennwerte**

  Beispiel: $\bar{x}$ vs. μ

  Bemerkung: Ein stochastische Kennwert wäre etwa $\bar{X}$

- **Theoretische Statistiken**

  Beispiel: $\bar{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \mu_i$ in Bezug auf $X_1, \ldots, X_n$ mit $E(X_i) = \mu_i$

9.3 Formale Unterscheidungsebenen

Tab. 9.3.1: Empirische vs. stochastische Statistiken und ihre Bezeichnungen

Empirische Statistiken (empirische Kennwerte)	Stochastische Statistiken (stochastische Kennwerte)
Merkmalssumme $\sum_{i=1}^{n} x_i$	Stochastische Summe/Stichprobensumme $\sum_{i=1}^{n} X_i$
arithmetisches Mittel $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$	Stochastisches Mittel/Stichprobenmittel $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$
empirische Varianz $\tilde{s}_X^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2$	Stochastische Varianz/Stichprobenvarianz $\tilde{S}_X^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$
empirische Kovarianz $\tilde{s}_{XY} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$	Stochastische Kovarianz/Stichprobenkovarianz $\tilde{S}_{XY} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})$
empirische Korrelation $r_{XY} = \frac{\tilde{s}_{XY}}{\tilde{s}_X \tilde{s}_Y}$	Stochastische Korrelation/Stichprobenkorrelation $R_{XY} = \frac{\tilde{S}_{XY}}{\tilde{S}_X \tilde{S}_Y}$
größter Wert (Maximum) $x_{(n)}$	stochastisches Maximum/Stichprobenmaximum $X_{(n)}$

9.3 Formale Unterscheidungsebenen

Tab. 9.3.2: Empirische vs. theoretische Kennwerte

Empirische Kennwerte		Theoretische Kennwerte	
arithmetisches Mittel:	$\bar{x}$	Erwartungswert:	$\mu_X = E(X)$
empirische Varianz:	$\tilde{s}_X^2$	theoretische Varianz:	$\sigma_X^2 = Var(X)$
empirische Kovarianz:	$\tilde{s}_{XY}$	theoretische Kovarianz:	$\sigma_{XY} = Cov(X, Y)$
empirische Korrelation:	r_{XY}	theoretische Korrelation:	$\rho_{XY} = Corr(X, Y)$
usw.		usw.	

- **Zusammenhang stochastischer, theoretischer und empirischer Kennwerte** -

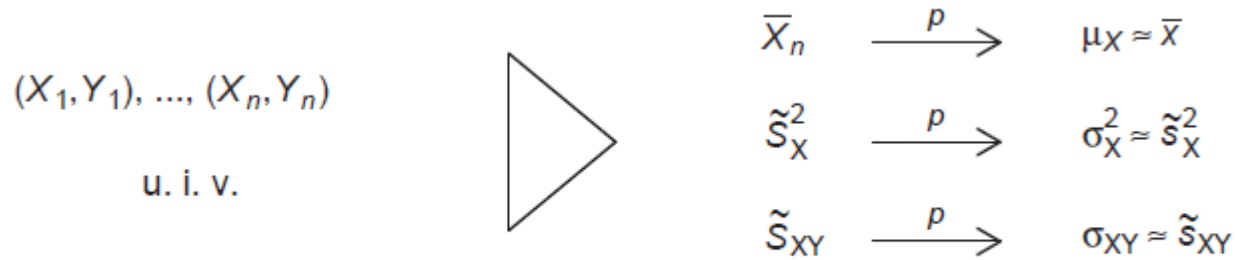


Abb. 9.3.1: Zusammenhang stochastischer, theoretischer und empirischer Kennwerte

10 Statistisches Schätzen

10 Statistisches Schätzen

10.1 Punktschätzung	636
10.1.1 Schätzer und ihre Gütekriterien	636
10.1.2 Erwartungstreue	641
10.1.3 Erwartete quadratische Abweichung (MSE)	651
10.1.4 Konsistenz	654
10.1.5 Effizienz	663
10.2 Intervallschätzung	668
10.2.1 Was versteht man unter einem Konfidenzintervall?	668
10.2.2 Konfidenzintervalle für Erwartungswerte	677
10.2.3 Konfidenzintervalle für Erwartungswertdifferenzen	692
10.2.4 Weitere Konfidenzintervalle	706
10.2.5 Adäquatheit bestimmter Modellannahmen	712

10 Statistisches Schätzen

10.3 Schätzmethoden	714
10.3.1 Momentenmethode	714
10.3.2 Maximum-Likelihood-Methode	717
10.3.3 Weitere Schätzmethoden	730

10.1.1 Schätzer und ihre Gütekriterien

- **Schätzprobleme**
- **Schätzer**

Schätzer sind stochastische Statistiken (Zufallsvariablen) deren Realisationen als Schätzwerte verwendet werden.

Statistische Modellierung	zu schätzende(r) Parameter(funktion)	Schätzfunktion
$S_{50} = \sum_{i=1}^{50} X_i \sim B(50, \pi)$	π	$\hat{\pi} = \overline{X} = S_{50}/50$
$\overline{X} \sim N(\mu, \sigma^2/10)$	μ, σ^2	$\hat{\mu} = \overline{X},\ \hat{\sigma}^2 = \tilde{S}_X^2$
$\overline{Y}_1 - \overline{Y}_0 \sim N(\mu_1 - \mu_0, 2\sigma^2/3)$	$\mu_1 - \mu_0, \sigma^2$	$\widehat{\mu_1 - \mu_0} = \overline{Y}_1 - \overline{Y}_0,\ \hat{\sigma}^2 = ?$

Abb. 10.1.1: Parametrisches Schätzen in den Beispielen 9.1.1–9.1.3

• Notation für Schätzer •

Beispiele: $\hat{\pi} = \bar{X}$ $\hat{\mu} = \bar{X}$ $\hat{\sigma}^2 = \tilde{S}^2$

Auf notationelle Unterscheidung stochastischer und empirischer Größen wie etwa $\widehat{\mathrm{M}}$ und $\hat{\mu}$ wird (i. A.) verzichtet (obwohl eigentlich „richtiger“)

• Statistische Schätztheorie •

Die Frage „Was soll wie geschätzt“ werden, lässt sich häufig nicht so einfach beantworten wie man zunächst möglicherweise meinen könnte

• Beispiel 10.1.1: Eine Simulation zur Illustration der Schätzproblematik •

Seien $X_1, \ldots, X_n$ unabhängig $N(0,1)$-verteilt. Abb. 10.1.2 zeigt realisierte Verteilungen von

(i) $\bar{X}$ Stichprobenmittel

(ii) $\tilde{X}_{0.5}$ Stichprobenmedian

(iii) $SMR = 0.5\left(X_{(1)} + X_{(n)}\right)$ Stichproben-Midrange

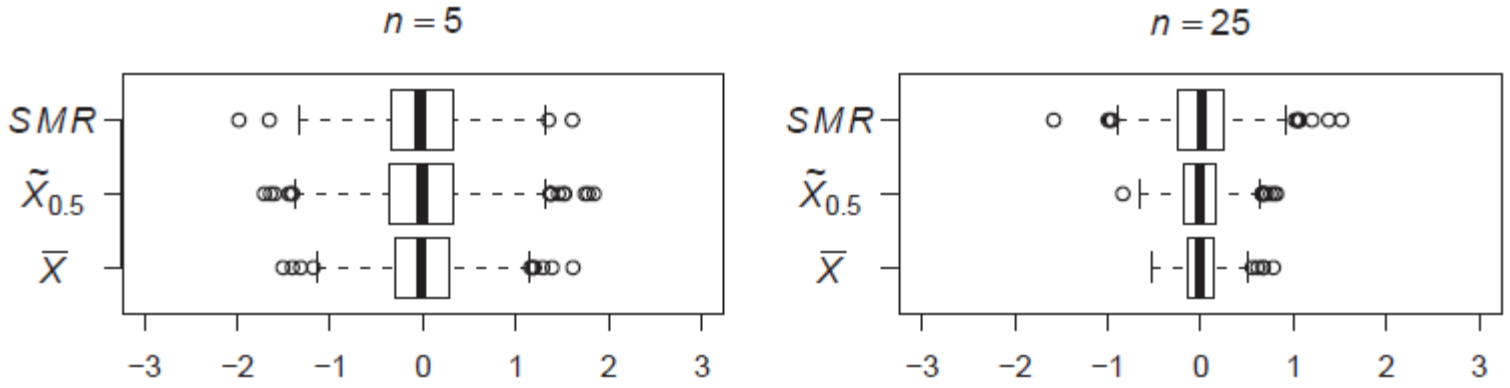


Abb. 10.1.2: Stichproben aus einer $N(0, 1)$-Verteilung

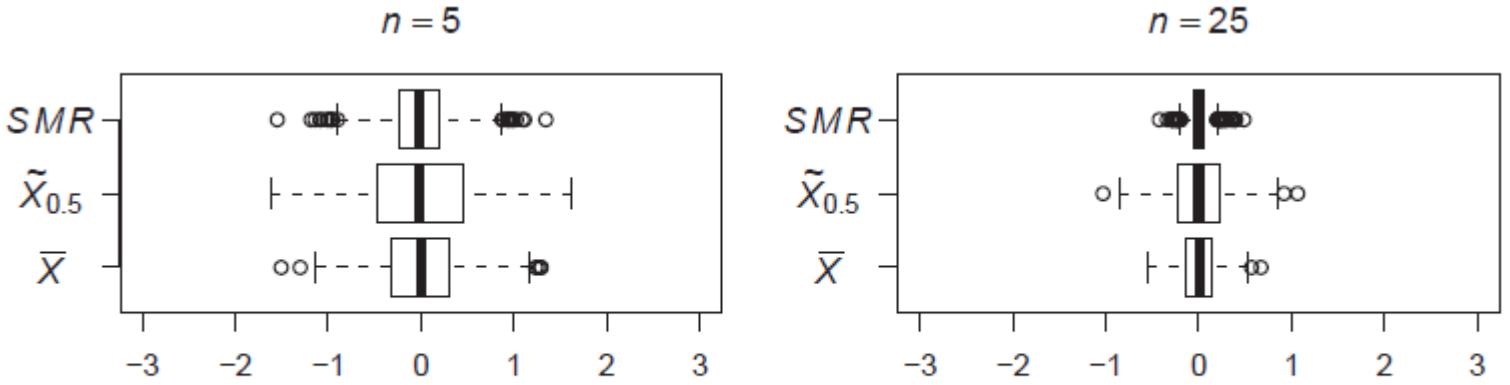


Abb. 10.1.3: Stichproben aus einer $G(-1.73, 1.73)$-Verteilung

● Verteilung von Schätzern ●

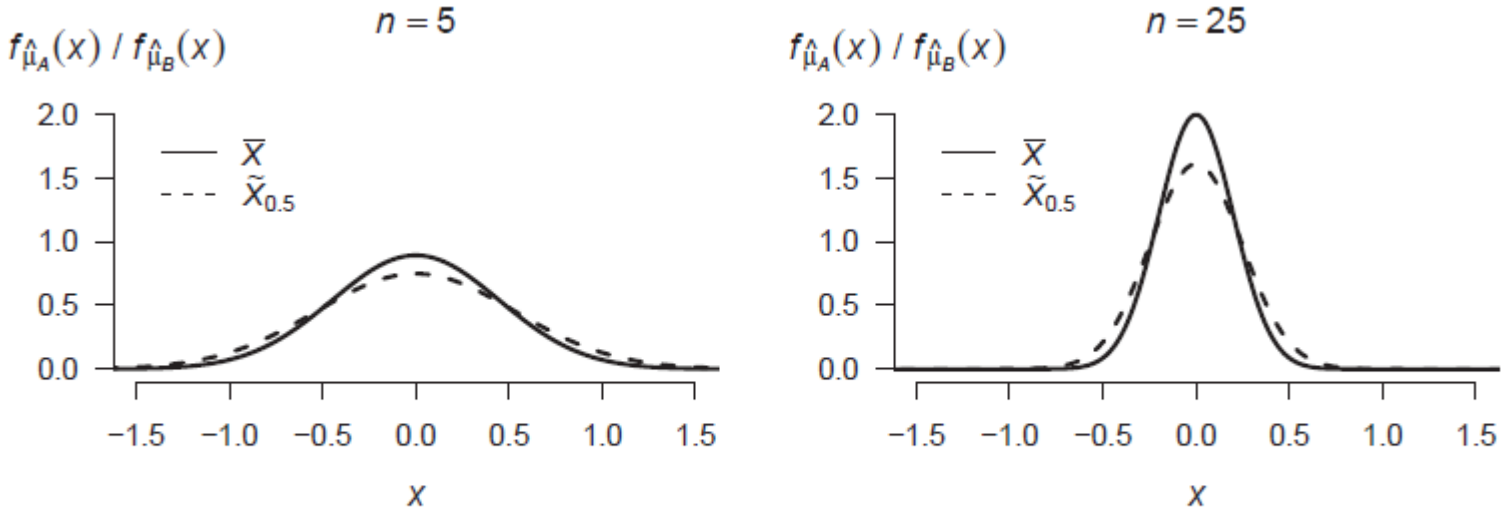


Abb. 10.1.4: Stichprobenmittel und Stichprobenmedian bei einer Normalverteilung

Hier kann die theoretische Verteilung von $\hat{\mu}_A = \bar{X}$ „ohne größere Schwierigkeiten“ angegeben werden. So gilt (vgl. Kap. 7, Folie 540):

$$\hat{\mu}_A = \bar{X} \sim N(\mu, \sigma^2/n)$$

Die Bestimmung der theoretischen Verteilung von $\hat{\mu}_B = \tilde{X}_{0.5}$ und $\hat{\mu}_C = 0.5\left(X_{(1)} + X_{(n)}\right)$ ist hier deutlich schwieriger (siehe LB, S. 501-503)

● Gütekriterien ●

(i) **Erwartungstreue**

Schätzer liefert „im Mittel“ den zu schätzenden Wert (liegt im Durchschnitt richtig)

(ii) **Erwartete quadratische Abweichung (MSE)**

Erwartete quadratische Abweichung vom zu schätzenden Wert (Genauigkeit des Schätzers) = Varianz bei E-Treue

(iii) **Konsistenz**

Schätzer „konvergiert“ für wachsendes *n* gegen zu schätzenden Wert

(iv) **Effizienz**

Frage nach dem besseren bzw. besten Schätzer innerhalb eines bestimmten Modellrahmens

10.1.2 Erwartungstreue

● Definition ●

Definition 10.1.1: Erwartungstreue und Verzerrung

Ein Schätzer $\hat{\theta}$ heißt **erwartungstreu** für θ, falls

$$E(\hat{\theta}) = \theta$$

für jeden zulässigen Wert θ gilt. Anderenfalls spricht man von einem **verzerrten** Schätzer. Die **Verzerrung** (engl. *Bias*) ist dabei definiert als

$$Bias(\hat{\theta}) = E(\hat{\theta}) - \theta.$$

Ein Schätzer heißt **asymptotisch erwartungstreu**, falls gilt:

$$\lim_{n \to \infty} E(\hat{\theta}) = \theta.$$

● Beispiel 10.1.2: Schätzung des Erwartungswerts ●

Sind $X_1, \ldots, X_n$ identisch verteilt mit $E(X_i) = \mu$ für $i = 1, \ldots, n$, dann gilt

$$E(\bar{X}) = \mu.$$

Man beachte:

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

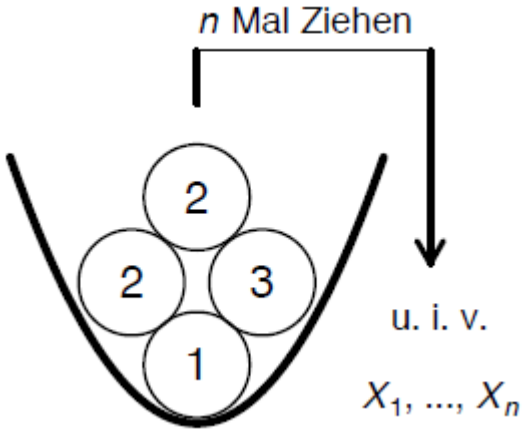
$$= \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n} \times n \times \mu$$

$$= \mu$$

Allgemein wurde dies auch schon in Kap. 7 gezeigt (Verteilung stochastischer Summen und Mittelwerte (vgl. Kap. 7, Folie 534)

Damit ist $\hat{\mu} = \bar{X}$ erwartungstreu für μ.

● Beispiel 10.1.3: Schätzung des Mittelwerts einer realen Grundgesamtheit ●



$\bar{z}$... Mittelwert der Zahlen in der Urne

Hier: $\bar{z} = 2$

X... Ergebnis bei 1 Mal zufällig ziehen

$$\Rightarrow \quad E(X) = 1 \times 0.25 + 2 \times 0.5 + 3 \times 0.25$$

$$= 2 = \bar{z}$$

Stichprobe: $X_1, \ldots, X_n$ unabhängig (mit Zurücklegen!) und identisch verteilt wie X

$$\Rightarrow \quad E(X_i) = 2 \quad \text{für } i = 1, \ldots, n$$

$$\Rightarrow \quad E(\bar{X}) = 2$$

Fazit: Das Stichprobenmittel ist (bei u.i.v.-Stichprobe) e-treu für den Mittelwert der Zahlen in der Urne

● Beispiel 10.1.4: Verzerrte und asymptotisch erwartungstreue Schätzung ●

Gegeben sei eine quadratische Metallplatte, deren Fläche zu bestimmen ist. Angenommen, eine Person A misst eine Kantenlänge von 32.1 cm und schließt daraus auf eine Fläche von 32.1×32.1 cm^2 = 1030.41 cm^2. Eine andere Person B misst dagegen eine Kantenlänge von 31.8 cm und schließt daraus auf eine Fläche von nur 1011.24 cm^2. Ist eine Mittelung der beiden Ergebnisse sinnvoll?

> Modellrahmen (Annahmen):

X... gemessene Kantenlänge

μ... tatsächliche Kantenlänge, d. h. μ^2 tatsächliche Fläche

Annahme: $X \sim N(\mu, \sigma^2)$ (häufige Annahme bei Messfehler-Problemen

$\Rightarrow$ $E(X) = \mu$

$\Rightarrow$ im Durchschnitt wird Kantenlänge richtig gemessen

$\Rightarrow \quad E(X^2) = Var(X) + E(X)^2 = \sigma^2 + \mu^2 > \mu^2$

$\Rightarrow$ im Durchschnitt wird Fläche überschätzt, sofern gemessene Kantenlänge quadriert wird

> Die gleiche Überschätzung tritt auch ein, wenn zwei oder mehrere Einzelschätzungen für die Fläche gemittelt würden:

Annahme: $X_1, \dots, X_n$ unabhängig $N(\mu, \sigma^2)$-verteilt

$\Rightarrow$ $X_1^2, \dots, X_n^2$ unabhängig und identisch verteilt (warum?) mit

$$E\left(X_i^2\right) = \sigma^2 + \mu^2$$

$$\Rightarrow \quad E\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(X_i^2\right) = \frac{1}{n} \times n \times (\sigma^2 + \mu^2) = \sigma^2 + \mu^2$$

$\Rightarrow$ Die Mittelung der beiden Flächenschätzungen führt tendenziell also zu einer Überschätzung, die umso größer ist desto ungenauer die Schätzung der Kantenlänge ist

> Eine alternative Vorgehensweise würde darin bestehen, den Mittelwert der gemessenen Kantenlängen zu nehmen und diesen dann zu quadrieren, d. h.

$\bar{X}^2$

Gilt dann $E(\bar{X}^2) = \mu^2$?

Es gilt zunächst einmal

$$E(\bar{X}) = \mu \quad \text{und} \quad Var(\bar{X}) = \sigma^2/n$$

Mit der Verschiebungsformel für die theoretische Varianz bezogen nun auf das Stichprobenmittel folgt dann:

$$Var(\bar{X}) = E(\bar{X}^2) - E(\bar{X})^2$$

$$\Rightarrow E(\bar{X}^2) = Var(\bar{X}) + E(\bar{X})^2$$

$$= \underbrace{\sigma^2/n + \mu^2}_{n\to\infty} \longrightarrow \mu^2$$

Damit ist $\bar{X}^2$ zwar nicht e-treu, jedoch asymptotisch e-treu.

● **Beispiel 10.1.5: Schätzung der theoretischen Varianz** ●

Seien $X_1, \ldots, X_n$ u. i. v. mit $E(X_i) = \mu$ und $Var(X_i) = \sigma^2$ für $i = 1, \ldots, n$.

Dann lässt sich zeigen (Ausführungen dazu im LB, S. 506-507):

$$E(\tilde{S}^2) = E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) = \frac{n-1}{n}\sigma^2$$

Die (nichtkorrigierte) **Stichprobenvarianz** $\tilde{S}^2$ ist folglich nicht e-treu, jedoch asymptotisch e-treu für die theoretische Varianz σ^2.

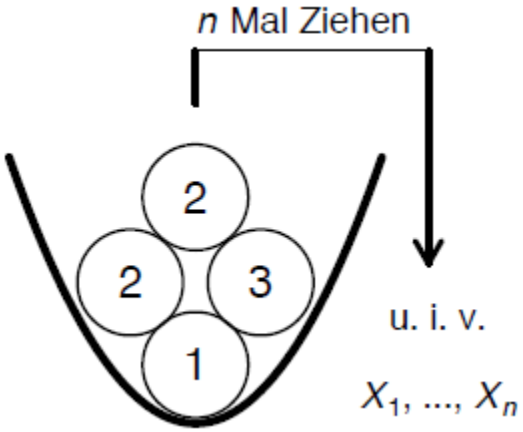
Definiere nun die **korrigierte Stichprobenvarianz**:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Diese ist dann e-treu:

$$E(S^2) = E\left(\frac{n}{n-1}\tilde{S}^2\right) = \frac{n}{n-1}E(\tilde{S}^2) = \frac{n}{n-1} \times \frac{n-1}{n}\sigma^2 = \sigma^2$$

- **Beispiel 10.1.6: Schätzung der empirischen Varianz einer realen Grundgesamtheit** ●



$\tilde{s}_Z^2$... empirische Varianz der Zahlen in der Urne

Hier:

$\tilde{s}_Z^2 = \frac{1}{n}\sum_{i=1}^{4}(z_i - \bar{z})^2$

$= 0.5$

X... Ergebnis bei 1 Mal zufällig ziehen

$\Rightarrow \quad E(X) = 2 = \bar{z}$

$\Rightarrow \quad E(X^2) = 1^2 \times 0.25 + 2^2 \times 0.5 + 3^2 \times 0.25$

$= 4.5$

$\Rightarrow \quad Var(X) = E(X^2) - E(X)^2 = 4.5 - 2^2$

$= 0.5 = \tilde{s}_Z^2$

Stichprobe: $X_1, \ldots, X_n$ u. i. v. wie X:

$\Rightarrow \quad E(S^2) = 0.5$ für $i = 1, \ldots, n$

Fazit: Die korrigierte Stichprobenvarianz ist (bei u.i.v.-Stichprobe) e-treu für die nichtkorrigierte Varianz der Zahlen in der Urne

• Zusammenfassung •

Satz 10.1.1: Erwartungstreue Schätzung von Erwartungswert und theoretischer Varianz

Seien $X_1, \ldots, X_n$ identisch verteilte Stichprobenvariablen mit Erwartungswert μ und Varianz σ^2. Weiter seien das Stichprobenmittel bzw. die korrigierte Stichprobenvarianz

$$\hat{\mu} = \overline{X} \quad \text{bzw.} \quad \hat{\sigma}^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

Dann gilt:

$$E(\hat{\mu}) = \mu, \tag{10.1.11}$$

$$E(\hat{\sigma}^2) = \sigma^2, \quad \text{falls } X_1, \ldots, X_n \text{ unabhängig sind.} \tag{10.1.12}$$

Satz 10.1.2: Erwartungstreue Schätzung von Mittelwert und empirischer Varianz bei realer Grundgesamtheit (Urnenmodell)

Gegeben sei eine Urne mit $N \geq 1$ Kugeln, die mit Zahlen $z_1, \ldots, z_N$ beschriftet sind. Sei $\mu = \overline{z}$ das arithmetische Mittel aller Zahlen und $\sigma^2 = \tilde{s}_Z^2$ die empirische Varianz. Es werden $n \geq 1$ Kugeln zufällig gezogen und die gezogenen Zahlen mit den Stichprobenvariablen $X_1, \ldots, X_n$ modelliert. Dann gilt:

$$E(\overline{X}) = \mu = \overline{z}, \quad \text{falls mit oder ohne Zurücklegen gezogen wird,} \tag{10.1.13}$$

$$E(S^2) = \sigma^2 = \tilde{s}_Z^2, \quad \text{falls mit Zurücklegen gezogen wird.} \tag{10.1.14}$$

Bemerkung: Für das Ziehen ohne Zurücklegen in (10.1.13) wird implizit $N \geq n$ vorausgesetzt.

10.1.3 Erwartete quadratische Abweichung (MSE)

● Hintergrund und Definition ●

Definition 10.1.2: Erwartete quadratische Abweichung (MSE)

Gegeben sei ein Schätzer $\hat{\theta}$ für einen Parameter θ. Die **erwartete quadratische Abweichung**, die auch als **MSE** (engl. *Mean Squared Error*) bezeichnet wird, ist definiert als

$$MSE(\hat{\theta}) = E([\hat{\theta} - \theta]^2) = Var(\hat{\theta}) + (Bias(\hat{\theta}))^2.$$

Die letzte Gleichung ergibt sich wie folgt:

$$E\left(\left[\hat{\theta} - \theta\right]^2\right) = E\left(\left[\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta\right]^2\right)$$

$$= E\left(\left[\hat{\theta} - E(\hat{\theta})\right]^2\right) + 2E\left(\left[\hat{\theta} - E(\hat{\theta})\right]\left[E(\hat{\theta}) - \theta\right]\right) + E\left(\left[E(\hat{\theta}) - \theta\right]^2\right)$$

$$= E\left(\left[\hat{\theta} - E(\hat{\theta})\right]^2\right) + 2\left[E(\hat{\theta}) - \theta\right] \times E\left(\left[\hat{\theta} - E(\hat{\theta})\right]\right) + \left(\left[E(\hat{\theta}) - \theta\right]^2\right)$$

$$= Var(\hat{\theta}) + 2\left[E(\hat{\theta}) - \theta\right] \times 0 + Bias(\hat{\theta})^2$$

● Beispiel 10.1.7 ●

Seien $X_1, \ldots, X_n$ unabhängig $N(\mu, \sigma^2)$-verteilt. Drei Schätzer für μ:

$$\hat{\mu}_A = \bar{X}, \quad \hat{\mu}_B = X_1 \quad \text{und} \quad \hat{\mu}_C = \bar{X} + 1/n$$

Dann gilt:

$$\hat{\mu}_A \sim N(\mu, \sigma^2/n), \quad \hat{\mu}_B \sim N(\mu, \sigma^2), \quad \hat{\mu}_C \sim N(\mu + 1/n, \sigma^2/n)$$

Für $\mu = 0, \sigma^2 = 1$

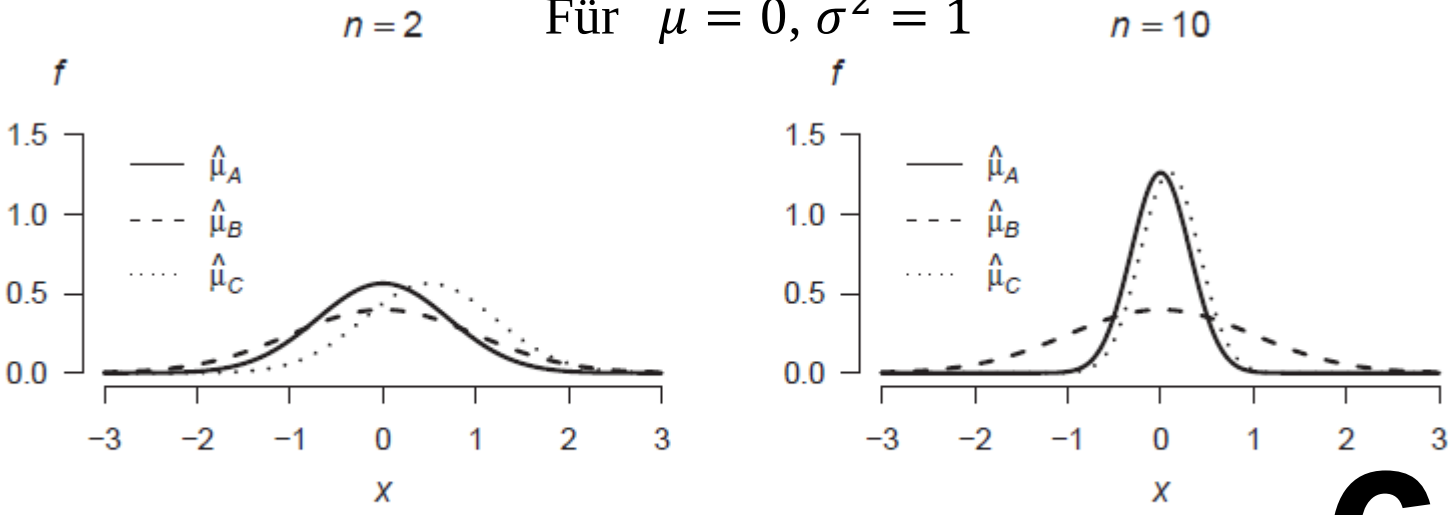


Abb. 10.1.5: Verteilung der drei Schätzer von Beispiel 10.1.7

$$\Rightarrow \quad MSE(\hat{\mu}_A) = MSE(\bar{X}) = Var(\bar{X}) = \sigma^2/n$$

$$\Rightarrow \quad MSE(\hat{\mu}_B) = MSE(X_1) = Var(X_1) = \sigma^2$$

$$\Rightarrow \quad MSE(\hat{\mu}_C) = MSE(\bar{X} + 1/n) = Var(\bar{X} + 1/n) + Bias(\bar{X} + 1/n)^2$$

$$= \sigma^2/n + 1/n^2$$

Beispielkonstellationen:

$\sigma^2 = 1, n = 2$: $MSE(\hat{\mu}_A) = 0.5$, $MSE(\hat{\mu}_B) = 1$ und $MSE(\hat{\mu}_C) = 0.75$

$\Rightarrow$ $\hat{\mu}_A$ besser als $\hat{\mu}_C$ besser als $\hat{\mu}_B$

$\sigma^2 = 0.1, n = 2$: $MSE(\hat{\mu}_A) = 0.05$, $MSE(\hat{\mu}_B) = 0.1$ und $MSE(\hat{\mu}_C) = 0.3$

$\Rightarrow$ $\hat{\mu}_A$ besser als $\hat{\mu}_B$ besser als $\hat{\mu}_C$

$\Rightarrow$ Hier $\hat{\mu}_A$ für $n \geq 2$ stets am besten

10.1.4 Konsistenz

- **Hintergrund und Definition**

Definition 10.1.3: MSE-Konsistenz und schwache Konsistenz

Ein Schätzer $\hat{\theta}_n$ für einen Parameter θ heißt

(i) **MSE-konsistent**, falls gilt:

$$MSE(\hat{\theta}_n) \xrightarrow{n\to\infty} 0.$$

(ii) **schwach konsistent**, falls für jedes $\varepsilon > 0$ gilt:

$$P(|\hat{\theta}_n - \theta| < \varepsilon) \xrightarrow{n\to\infty} 1 \quad \text{bzw.} \quad P(|\hat{\theta}_n - \theta| \geq \varepsilon) \xrightarrow{n\to\infty} 0,$$

in Kurzschreibweise: $\hat{\theta}_n \xrightarrow{p} \theta$.

- **Bemerkung zur Notation**

● MSE-Konsistenz impliziert schwache Konsistenz ●

MSE-Konsistenz impliziert schwache Konsistenz.
Die Umkehrung gilt nicht.

Beweis siehe LB, S. 513

● Konsistente Schätzung von Funktionen eines Parameters ●

Satz 10.1.3: Konsistente Schätzung von Funktionen eines Parameters

Ist $\widehat{\theta}_n$ ein schwach konsistenter Schätzer für einen Parameter θ und g eine stetige Funktion, dann ist $g(\widehat{\theta}_n)$ ein schwach konsistenter Schätzer für $g(\theta)$, sofern $g(\theta)$ definiert ist.

Man beachte hierzu das spätere Beispiel 10.1.9 auf Folie 658

• Nachweis von MSE-Konsistenz •

Ein erwartungstreuer oder asymptotisch erwartungstreuer Schätzer ist genau dann MSE-konsistent, falls dessen Varianz für wachsendes n gegen Null konvergiert.

• Nachweis von schwacher Konsistenz •

Schwache Konsistenz lässt sich, sofern vorhanden,

1. über die stärkere Form der MSE-Konsistenz,
2. anhand asymptotischer Resultate (GGZ + Stetigkeitssatz) oder
3. direkt anhand der Definition

nachweisen.

● Beispiel 10.1.8: Konsistente Schätzung von Erwartungswert und Varianz ●

Satz 10.1.4: Konsistente Schätzung von Erwartungswert und Varianz

Seien $X_1, \ldots, X_n$ u.i.v. mit Erwartungswert μ und Varianz σ^2 und $E(X_i^4) < \infty$. Dann gilt:

(i) Das Stichprobenmittel $\overline{X}$ ist ein MSE-konsistenter und schwach konsistenter Schätzer für μ.

(ii) Die korrigierte (nichtkorrigierte) Stichprobenvarianz S^2 ($\tilde{S}^2$) ist ein MSE-konsistenter und schwach konsistenter Schätzer für σ^2.

> Zu (i): $MSE(\bar{X}) = Var(\bar{X}) = \sigma^2/n \xrightarrow{n\to\infty} 0$

> Zu (ii): $S^2 \xrightarrow{p} \sigma^2$ folgt analog, nur komplizierter (siehe LB, S. 515)

Aus $S^2 \xrightarrow{p} \sigma^2$ und $n/(n-1) \xrightarrow{n\to\infty} 1$ folgt

$\tilde{S}^2 = \frac{n}{n-1} S^2 \xrightarrow{p} \sigma^2$ (Multivariater Stetigkeitssatz, Folie 608)

● Beispiel 10.1.9: Konsistente Schätzung einer Funktion eines Parameters ●

> 1. Beispiel: Gegeben sei die Situation aus Beispiel 10.1.4 (Folie 644)

Da $X_1, \dots, X_n$ u. i. v. mit $E(X_i) = \mu$ folgt mit GGZ und Stetigkeitssatz

$$\bar{X} \xrightarrow{p} \mu \quad \overset{\text{Satz 10.1.3}}{\Rightarrow} \quad \bar{X}^2 \xrightarrow{p} \mu^2$$

Stetige Funktion in diesem Fall: $g(t) = t^2$

> 2. Beispiel: Seien $X_1, \dots, X_n$ unabhängig $Exp(\lambda)$-verteilt

$$\Rightarrow \quad E(X_i) = 1/\lambda$$

Mit GGZ und Stetigkeitssatz folgt dann

$$\bar{X} \xrightarrow{p} 1/\lambda \quad \overset{\text{Satz 10.1.3}}{\Rightarrow} \quad 1/\bar{X} \xrightarrow{p} \lambda$$

Stetige Funktion in diesem Fall: $g(t) = 1/t$

● Beispiel 10.1.10: Konsistente Schätzung eines theoretischen Maximums ●

> Seien $X_1, \ldots, X_n$ unabhängig gleichverteilt über $[0, \theta]$, d. h. $E(X_i) = \theta/2$

Mit GGZ und Stetigkeitssatz folgt

$$\bar{X} \xrightarrow{p} \theta/2 \quad \Rightarrow \quad 2\bar{X} \xrightarrow{p} \theta$$

Somit wäre $2\bar{X}$ ein konsistenter Schätzer für das theoretische Maximum θ (die Intervallobergrenze).

> Ein alternativer Schätzer für θ wäre das Stichprobenmaximum $X_{(n)}$.
Auch dieses liefert eine konsistente Schätzung, d. h. es gilt:

$$X_{(n)} \xrightarrow{p} \theta$$

Der Nachweis der Konsistenz erfolgt in diesem Fall über die Definition der Konsistenzeigenschaft (Möglichkeit 3 von Folie 656).

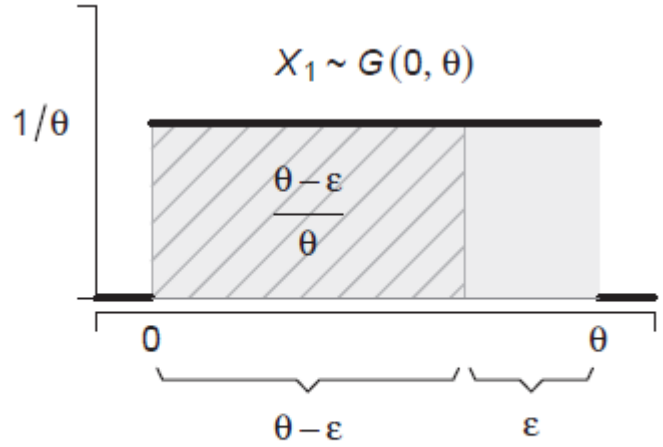


Abb. 10.1.6: Illustration zu Gleichung (10.1.22)

Nachweis der Konsistenzeigenschaft

> Für jedes $\varepsilon > 0$ gilt:

$$P\left(\left|X_{(n)} - \theta\right| \leq \varepsilon\right) = P\left(\theta - \varepsilon \leq X_{(n)} \leq \theta + \varepsilon\right)$$

$$= P\left(X_{(n)} \leq \theta + \varepsilon\right) - P\left(X_{(n)} < \theta - \varepsilon\right)$$

$$= 1 - P\left(X_{(n)} < \theta - \varepsilon\right)$$

> Ist die größte Beobachtung kleiner als $\theta - \varepsilon$, so trifft dies auf alle Beobachtungen zu, d. h. man erhält weiter

$$P\left(X_{(n)} < \theta - \varepsilon\right) = P(X_1 < \theta - \varepsilon, X_2 < \theta - \varepsilon, \ldots, X_n < \theta - \varepsilon)$$
$$= [P(X_1 < \theta - \varepsilon)]^n$$

> Für jedes $\varepsilon \geq \theta$ erhält man daraus dann

$$[P(X_1 < \theta - \varepsilon)]^n = 0^n = 0.$$

> Für $0 < \varepsilon < \theta$ erhält man hingegen (vgl. Abb. 10.1.6)

$$[P(X_1 < \theta - \varepsilon)]^n = \left(\frac{\theta - \varepsilon}{\theta}\right)^n$$

> Insgesamt erhält man damit für jedes $\epsilon \geq \theta$

$$(*) \quad P\left(\left|X_{(n)} - \theta\right| \leq \varepsilon\right) = 1 - 0 = 1$$

und für jedes $0 < \epsilon < \theta$

$$(**) \quad P\left(\left|X_{(n)} - \theta\right| \leq \varepsilon\right) = 1 - \left(\frac{\theta - \varepsilon}{\theta}\right)^n$$

> Wegen $(\theta - \varepsilon)/\theta < 1$ für $0 < \epsilon < \theta$ konvergiert die rechte Seite von $(**)$ für wachsendes n gegen 1.

> Insgesamt erhält man aus $(*)$ und $(**)$ die schwache Konsistenz des Stichprobenmaximums für die theoretische Obergrenze θ.

● Beispiel 10.1.11: Schwache Konsistenz impliziert keine MSE-Konsistenz ●

Nachweis über ein Beispiel mit einem Schätzer, der schwach konsistent jedoch nicht MSE-konsistent ist. Beispiel ist etwas schwierig zu verstehen und wird hier deshalb nicht präsentiert (vgl. LB, S. 518).

● Konsistenz ist Mindestanforderung ●

Die Konsistenz ist die wichtigste Eigenschaft eines Schätzers und zugleich auch eine Mindestanforderung, die man an einen Schätzer stellt.

10.1.5 Effizienz

● **Hintergrund** ●

● **Beispiel 10.1.12: Schätzung eines theoretischen Anteilswerts** ●

> Seien $X_1, \ldots, X_n$ unabhängig $B(1, \pi)$-verteilt, wobei π geschätzt werden soll. Für gerades n betrachten wir folgende 4 Schätzer:

$$\hat{\pi}_A = \frac{1}{n}\sum_{i=1}^{n} X_i\,, \quad \hat{\pi}_B = \frac{2}{n}\sum_{i=1}^{n/2} X_i\,,$$

$$\hat{\pi}_C = 0.5, \qquad \hat{\pi}_D = \frac{1}{n+2}\sum_{i=1}^{n} X_i + \frac{1}{n+2}$$

> Erwartungswerte der Schätzer:

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

$$E(\hat{\pi}_B) = E\left(\frac{2}{n}\sum_{i=1}^{n/2} X_i\right) = \frac{2}{n}\sum_{i=1}^{n/2} E(X_i) = \frac{2}{n}\sum_{i=1}^{n/2} \pi = \frac{2}{n} \times \frac{n\pi}{2} = \pi$$

$$E(\hat{\pi}_C) = E(0.5) = 0.5$$

$$E(\hat{\pi}_D) = E\left(\frac{1}{n+2}\sum_{i=1}^{n} X_i + \frac{1}{n+2}\right) = \frac{n\pi + 1}{n+2}$$

> Verzerrungen der Schätzer:

$$Bias(\hat{\pi}_A) = E(\hat{\pi}_A) - \pi = 0$$

$$Bias(\hat{\pi}_B) = E(\hat{\pi}_B) - \pi = 0$$

$$Bias(\hat{\pi}_C) = E(\hat{\pi}_C) - \pi = 0.5 - \pi$$

$$Bias(\hat{\pi}_D) = E(\hat{\pi}_D) - \pi = \frac{1 - 2\pi}{n+2}$$

> Varianzen der Schätzer:

$$Var(\hat{\pi}_A) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} \pi(1-\pi) = \frac{n\pi(1-\pi)}{n^2} = \frac{\pi(1-\pi)}{n}$$

$$Var(\hat{\pi}_B) = Var\left(\frac{2}{n}\sum_{i=1}^{n/2} X_i\right) = \frac{4}{n^2}\sum_{i=1}^{n/2} Var(X_i)$$

$$= \frac{4}{n^2}\sum_{i=1}^{n/2} \pi(1-\pi) = \frac{4n\pi(1-\pi)}{2n^2} = \frac{2\pi(1-\pi)}{n}$$

$$Var(\hat{\pi}_C) = Var(0.5) = 0$$

$$Var(\hat{\pi}_D) = Var\left(\frac{1}{n+2}\sum_{i=1}^{n} X_i + \frac{1}{n+2}\right) = Var\left(\frac{1}{n+2}\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{(n+2)^2}\sum_{i=1}^{n} Var(X_i) = \frac{n\pi(1-\pi)}{(n+2)^2}$$

> MSE‘s der Schätzer:

$$MSE(\hat{\pi}_A) = Var(\hat{\pi}_A) + Bias(\hat{\pi}_A)^2 = \frac{\pi(1-\pi)}{n}$$

$$MSE(\hat{\pi}_B) = Var(\hat{\pi}_B) + Bias(\hat{\pi}_B)^2 = \frac{2\pi(1-\pi)}{n}$$

$$MSE(\hat{\pi}_C) = Var(\hat{\pi}_C) + Bias(\hat{\pi}_C)^2 = (0.5-\pi)^2$$

$$MSE(\hat{\pi}_D) = Var(\hat{\pi}_D) + Bias(\hat{\pi}_D)^2 = \frac{n\pi(1-\pi)}{(n+2)^2} + \left(\frac{1-2\pi}{n+2}\right)^2$$

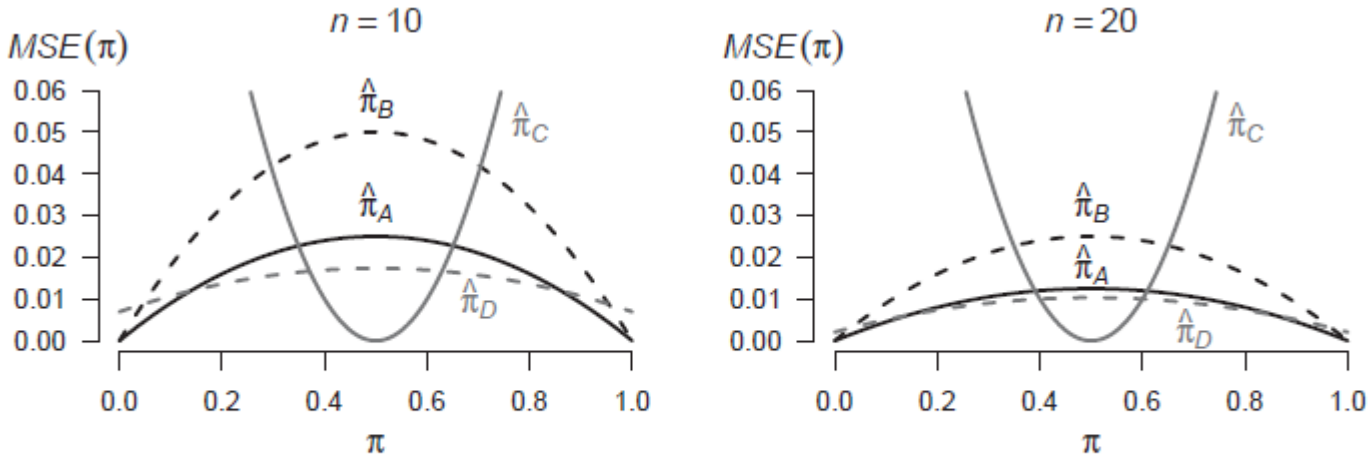


Abb. 10.1.7: MSE-Kurven der Schätzer von Beispiel 10.1.12

● Fazit ●

> Im Allgemeinen gibt es selten einen „uneingeschränkt besten“ Schätzer.

> Manchmal Einschränkung der in Betracht kommenden Schätzer auf die Klasse der erwartungstreuen Schätzer

> Suche nach einem **gleichmäßig besten erwartungstreuen Schätzer**

10.2.1 Was versteht man unter einem Konfidenzintervall?

- **Hintergrund und Überblick**
- **Herleitung eines Konfidenzintervalls für μ**

Seien $X_1, \dots, X_n$ unabhängig $N(\mu, \sigma^2)$-verteilt, wobei σ^2 bekannt und μ unbekannt ist. Dann gilt zunächst einmal

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

Man beachte, dass dies zugleich eine Aussage für die Verteilung des Schätzers $\hat{\mu} = \bar{X}$ ist. Weiter folgt dann

$$Z = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$$

Notiert man die Quantile der Standardnormalverteilung mit z_α erhält man...

10.2 Intervallschätzung

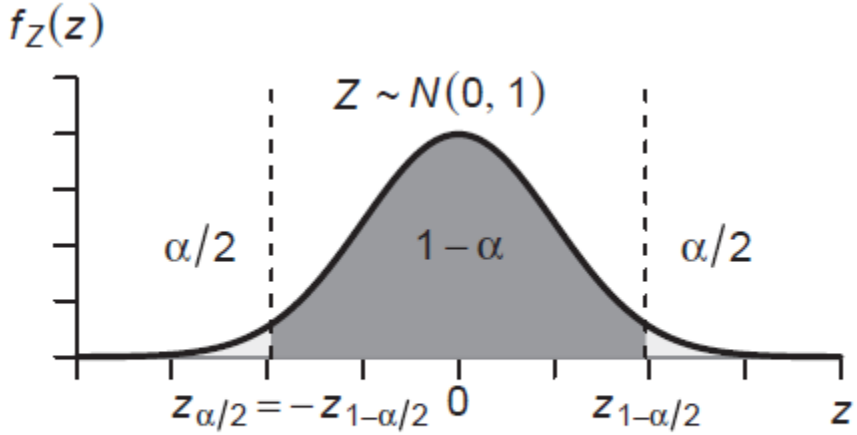


Abb. 10.2.1: Schwankungsverhalten des standarisierten Stichprobenmittels

$$1-\alpha = P\left(z_{\alpha/2} \leq Z \leq z_{1-\alpha/2}\right) = P\left(z_{\alpha/2} \leq \frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \leq z_{1-\alpha/2}\right)$$

$$= P\left(\underbrace{-z_{1-\alpha/2}\sqrt{\sigma^2/n} \leq \bar{X}-\mu}_{\text{Linke Ungleichung}} \leq \underbrace{z_{1-\alpha/2}\sqrt{\sigma^2/n}}_{\text{Rechte Ungleichung}}\right)$$

Linke Ungleichung umgeformt:

$$\mu \leq \bar{X} + z_{1-\alpha/2}\sqrt{\sigma^2/n}$$

Rechte Ungleichung umgeformt:

$$\mu \geq \bar{X} - z_{1-\alpha/2}\sqrt{\sigma^2/n}$$

$$= P\left(\bar{X} - z_{1-\alpha/2}\sqrt{\sigma^2/n} \leq \mu \leq \bar{X} + z_{1-\alpha/2}\sqrt{\sigma^2/n}\right)$$

Dies impliziert das $(1-\alpha)$-Konfidenzintervall

$$\left[\bar{X} - z_{1-\alpha/2}\sqrt{\sigma^2/n}\,, \bar{X} + z_{1-\alpha/2}\sqrt{\sigma^2/n}\right]$$

● Beispiel 10.2.1: Gepäckabfertigungszeiten ●

> Angenommen $X_1, \ldots, X_{10}$ seien unabhängig $N(\mu, 2)$-verteilt. Es liegen nun folgende Beobachtungswerte (Realisationen) vor:

7.8, 10.1, 9.0, 8.0, 11.6, 10.7, 8.1, 8.6, 9.4, 11.9.

Wie lautet dann das 0.95-Konfidenzintervall für μ?

$$\left[\bar{x} - z_{1-\alpha/2}\sqrt{\sigma^2/n}\,, \bar{x} + z_{1-\alpha/2}\sqrt{\sigma^2/n}\right]$$

9.52 | 2 | 10 | 1.96 ← Nachzuschlagendes Quantil:

$$0.95 = 1 - \alpha \quad \Rightarrow \quad \alpha = 0.05$$

$$\Rightarrow \quad z_{1-\alpha/2} = z_{0.975} = 1.96$$

$$= \left[9.52 - 1.96\sqrt{2/10}\,, 9.52 + 1.96\sqrt{2/10}\right]$$

$$= [8.64, 10.40]$$

10.2 Intervallschätzung

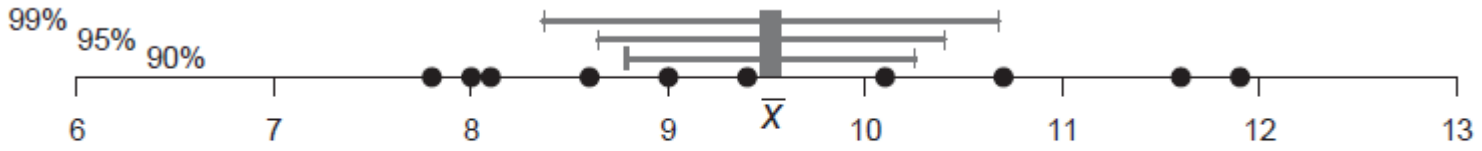


Abb. 10.2.2: Realisierte Konfidenzintervalle aus einer Stichprobe ($n = 10$)

Länge eines KI‘s: $L_Z = 2z_{1-\alpha/2}\sqrt{\sigma^2/n}$

> Alternative Sichtweise:

Angenommen $X_1, \ldots, X_{10}$ seien unabhängig $N(9.52, 2)$-verteilt. Dann würde mit $\sigma^2 = 2$ und $n = 10$ gelten:

$$\bar{X} \sim N(\mu, \sigma^2/n) = N(9.52, 2/10)$$

Das 0.025- und das 0.975-Quantil dieser Verteilung sind gegeben durch

$$q_{0.025} = 9.52 + \sqrt{2/10} \times z_{0.025}$$

$$= 9.52 - \sqrt{2/10} \times z_{0.975}$$

$$= 9.52 - \sqrt{2/10} \times 1.96 = 8.64$$

bzw. $q_{0.975} = 9.52 + \sqrt{2/10} \times z_{0.975}$

$= 9.52 + \sqrt{2/10} \times 1.96 = 10.40$

Diese Sichtweise führt uns zum gleichen Intervall.

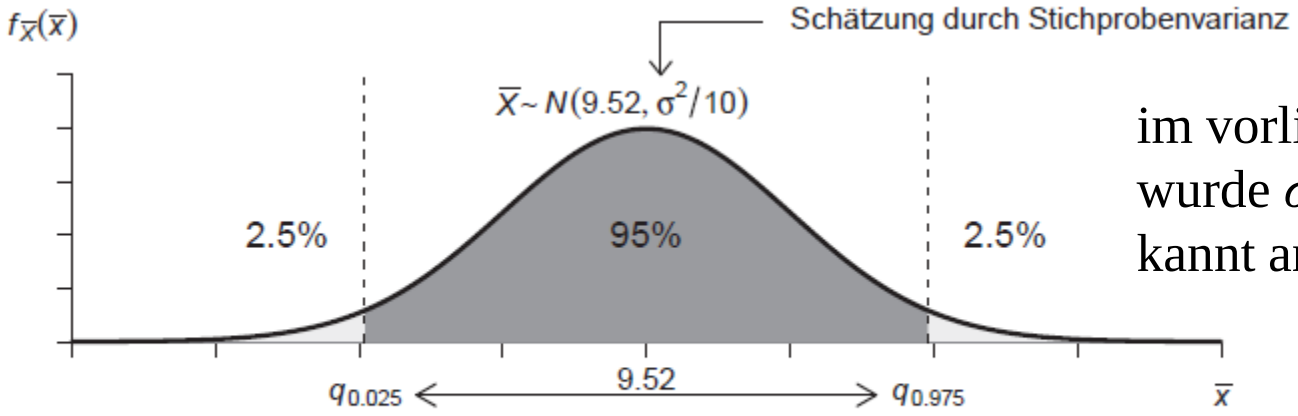


Abb. 9.1.2: Schwankungsverhalten des Stichprobenmittels für $\mu = 9.52$

im vorliegenden Fall wurde $\sigma^2 = 2$ als bekannt angenommen

> Problem in der Praxis: σ^2 i. d. R. unbekannt
Dann muss σ^2 geschätzt werden (vgl. Folie 677 ff.)

10.2 Intervallschätzung

• Interpretation •

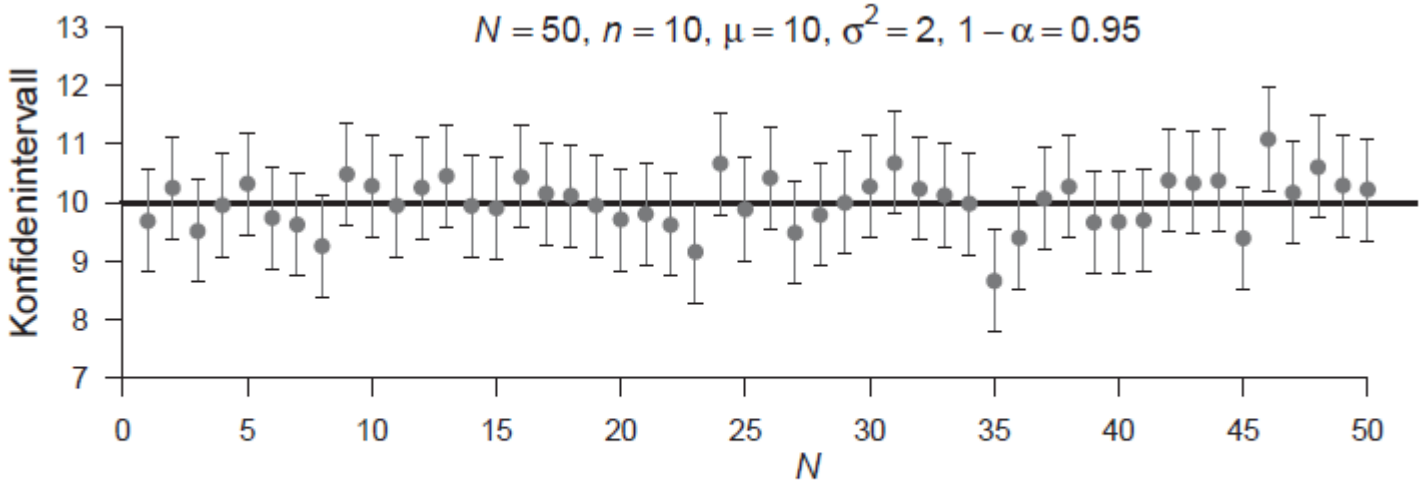


Abb. 10.2.3: Realisierte Konfidenzintervalle aus 50 Stichproben bei bekannter Varianz

Ein Konfidenzintervall ist ein Zufallsintervall, das mit einer vorgegebenen Wahrscheinlichkeit einen zu schätzenden Parameter überdeckt.

• Definition •

Definition 10.2.1: Konfidenzintervall

Seien $X_1, \ldots, X_n$ Stichprobenvariablen und θ ein unbekannter Parameter einer zugrunde liegenden Grundgesamtheit. Dann liefern die stochastischen Intervallgrenzen

$$G_u(X_1, \ldots, X_n) \quad \text{und} \quad G_o(X_1, \ldots, X_n)$$

ein $(1-\alpha)$-**Konfidenzintervall** für θ, falls gilt:

(i) $P(G_u \leq G_o) = 1$ und
(ii) $P(G_u \leq \theta \leq G_o) = 1 - \alpha$.

Der Wert $1-\alpha$ wird dabei als **Konfidenzniveau** oder als **Sicherheitswahrscheinlichkeit** bezeichnet.

10.2 Intervallschätzung

- **Grundlegende Eigenschaften von Konfidenzintervallen**
- **Übertragung auf andere Konfidenzintervalle**

Ein Konfidenzintervall wird i.d.R. länger, falls

(i) die Sicherheitswahrscheinlichkeit zunimmt,
(ii) die Streuung in der Grundgesamtheit zunimmt,
(iii) der Stichprobenumfang abnimmt.

- **Fazit**

Die Kunst der Konfidenzschätzung besteht darin, möglichst kurze Intervalle mit zugleich möglichst hoher Sicherheitswahrscheinlichkeit zu konstruieren.

10.2.2 Konfidenzintervalle für Erwartungswerte

● Modellrahmen und Überblick ●

Im Folgenden gilt stets: $X_1, \dots, X_n$ u. i. v. mit $E(X_i) = \mu$ und $Var(X_i) = \sigma^2$

Bei Gültigkeit weiterer Annahmen können dann spezifische KI‘s für μ konstruiert werden.

● Schätzung von μ bei Normalverteilung und bekannter Varianz ●

$X_1, \dots, X_n$ unabhängig $N(\mu, \sigma^2)$-verteilt, μ unbekannt und σ^2 bekannt:

Vgl. Folie 670: $\left[\bar{X} - z_{1-\alpha/2}\sqrt{\sigma^2/n}\,, \bar{X} + z_{1-\alpha/2}\sqrt{\sigma^2/n}\right]$

● Schätzung von μ bei Normalverteilung und unbekannter Varianz ●

> $X_1, \dots, X_n$ unabhängig $N(\mu, \sigma^2)$-verteilt, μ und σ^2 beide unbekannt

> Schätzung von σ^2: $\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

10.2 Intervallschätzung

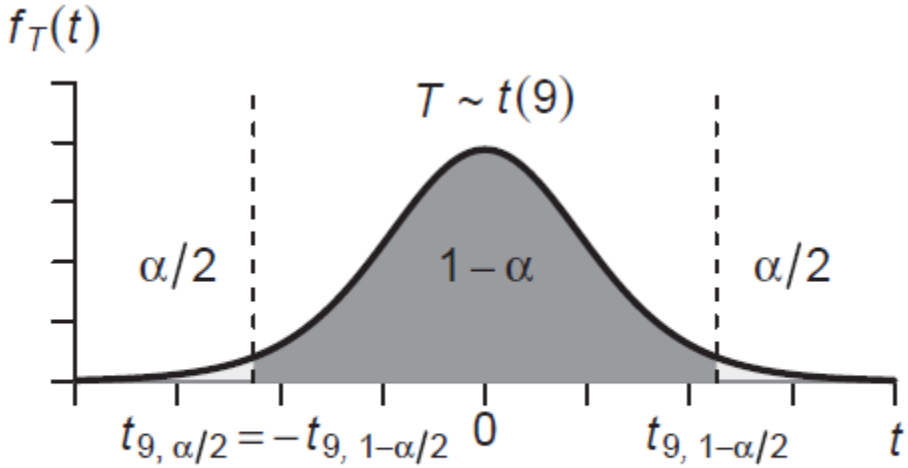


Abb. 10.2.4: Schwankungsverhalten der T-Statistik für $n = 10$

> Es gilt dann: $T = \dfrac{\bar{X} - \mu}{\sqrt{S^2/n}} \sim t(n-1)$ t-Verteilung mit n–1 Freiheitsgraden

> Notiert man die Quantile einer $t(n)$-Verteilung mit $t_{n,\alpha}$, erhält man

$$1 - \alpha = P\left(t_{n-1,\alpha/2} \leq T \leq t_{n-1,1-\alpha/2}\right)$$

$$= P\left(t_{n-1,\alpha/2} \leq \frac{\bar{X}-\mu}{\sqrt{S^2/n}} \leq t_{n-1,1-\alpha/2}\right)$$

$$= P\left(-t_{n-1,1-\alpha/2}\sqrt{S^2/n} \leq \bar{X}-\mu \leq t_{n-1,1-\alpha/2}\sqrt{S^2/n}\right)$$

Die linke Ungleichung umgeformt:

$$\mu \leq \bar{X} + t_{n-1,1-\alpha/2}\sqrt{S^2/n}$$

Die rechte Ungleichung umgeformt:

$$\mu \geq \bar{X} - t_{n-1,1-\alpha/2}\sqrt{S^2/n}$$

$$= P\left(\bar{X} - t_{n-1,1-\alpha/2}\sqrt{S^2/n} \leq \mu \leq \bar{X} + t_{n-1,1-\alpha/2}\sqrt{S^2/n}\right)$$

Dies impliziert das $(1-\alpha)$-Konfidenzintervall

$$\left[\bar{X} - t_{n-1,1-\alpha/2}\sqrt{S^2/n}\,, \bar{X} + t_{n-1,1-\alpha/2}\sqrt{S^2/n}\right]$$

10.2 Intervallschätzung

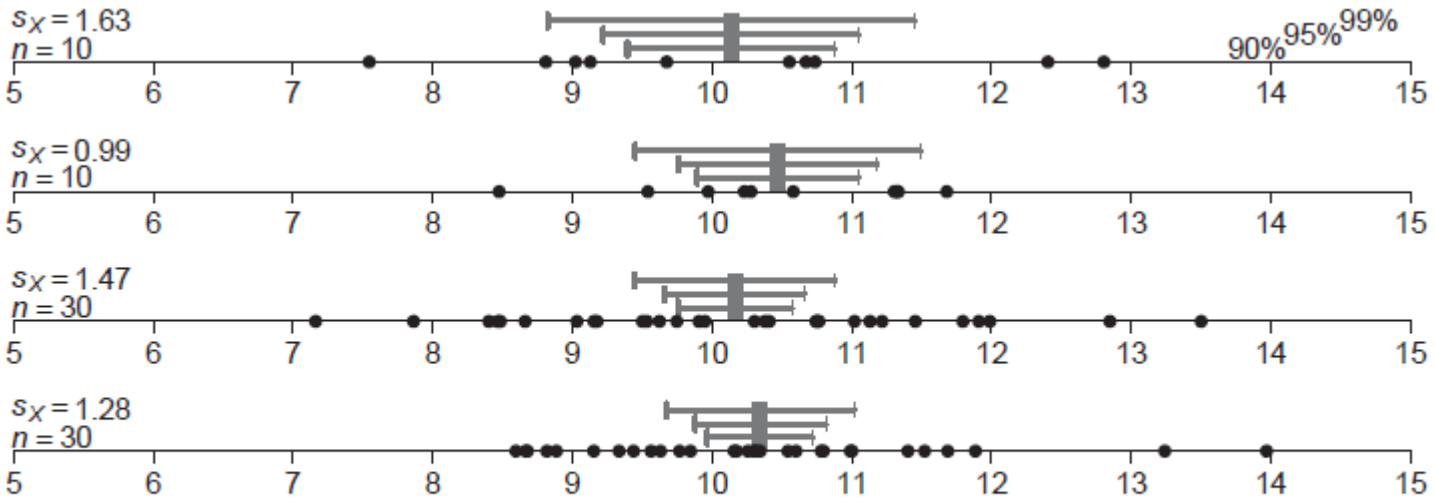


Abb. 10.2.5: Konfidenzintervalle für μ bei unbekannter Varianz ($n = 10$ und $n = 30$)

Länge eines KI's: $L_T = 2t_{n-1,1-\alpha/2}\sqrt{S^2/n}$

Im Unterschied zum Fall einer bekannten Varianz σ^2 ist die Länge jetzt stochastisch!

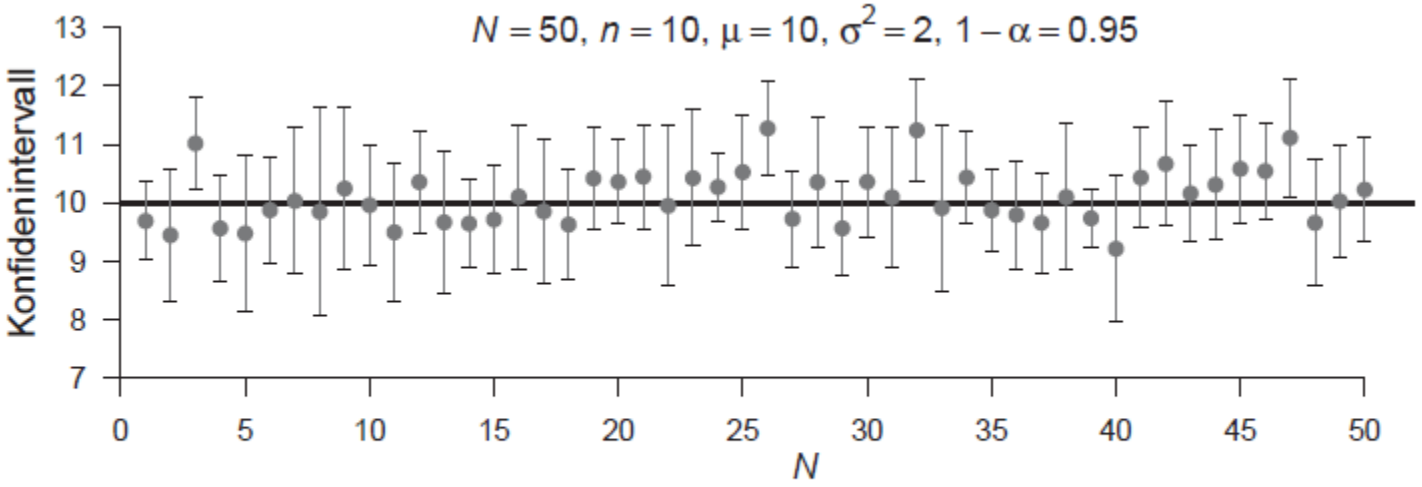


Abb. 10.2.6: Realisierte Konfidenzintervalle aus 50 Stichproben bei unbekannter Varianz

Grundsätzlich gilt: $t_{n-1,1-\alpha/2} > z_{1-\alpha/2}$

Daraus folgt jedoch nicht: $L_T \geq L_Z$

Jedoch gilt: $E(L_T) \geq E(L_Z)$

• Schätzung von μ bei beliebiger Ausgangsverteilung •

> Liegt keine Normalverteilung vor, so können nur für großes n ($n \geq 30$) noch „auf einfache Weise“ Konfidenzintervalle konstruiert werden.

> Mit dem *ZGWS* gilt dann:

$$\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \overset{a}{\sim} N(0,1)$$

> Es kann gezeigt werden (erscheint ohnehin schon plausibel), dass auch

$$\frac{\bar{X} - \mu}{\sqrt{S^2/n}} \overset{a}{\sim} N(0,1) \quad \text{bzw.} \quad \frac{\bar{X} - \mu}{\sqrt{\tilde{S}^2/n}} \overset{a}{\sim} N(0,1)$$

gilt. Diese Resultate sind für den Fall unbekannter Varianz nützlich.

> Auf Basis der obigen asymptotischen Resultate lassen sich dann approximative Konfidenzintervalle ableiten:
Dies sind die Fälle (iii) und (iv) auf Folie 684.

● Notwendiger Stichprobenumfang zur Erzielung bestimmter Intervalllängen ●

> Ziel: Konfidenzintervall soll bestimmte Länge (Genauigkeit) besitzen

> Sind $X_1, \ldots, X_n$ unabhängig $N(\mu, \sigma^2)$-verteilt mit bekannter Varianz σ^2, ist die Länge gegeben durch (vgl. Folie 672):

$$L_Z = 2z_{1-\alpha/2}\sqrt{\sigma^2/n}$$

Aufgelöst nach n ergibt dies:

$$n = \left(2z_{1-\alpha/2}\sigma/L_Z\right)^2.$$

Anwendungsbeispiel: Siehe Folie 686

> Was ist bei unbekannter Varianz? Was ist, falls keine Normalverteilung vorliegt?

● Zusammenfassung von Resultaten ●

10.2 Intervallschätzung

Satz 10.2.1: Konfidenzintervalle für Erwartungswerte

Seien $X_1, \ldots, X_n$ unabhängig und identisch verteilt mit Erwartungswert μ und Varianz $0 < \sigma^2 < \infty$. Dann ist ein $(1-\alpha)$-Konfidenzintervall für μ, falls

(i) $X_1, \ldots, X_n$ normalverteilt sind mit bekannter Varianz, gegeben durch

$$[\overline{X} - z_{1-\alpha/2}\sqrt{\sigma^2/n}, \overline{X} + z_{1-\alpha/2}\sqrt{\sigma^2/n}],$$

(ii) $X_1, \ldots, X_n$ normalverteilt sind mit unbekannter Varianz, gegeben durch

$$[\overline{X} - t_{n-1,1-\alpha/2}\sqrt{S^2/n}, \overline{X} + t_{n-1,1-\alpha/2}\sqrt{S^2/n}],$$

(iii) n groß ist und die Varianz bekannt ist, gegeben durch

$$[\overline{X} - z_{1-\alpha/2}\sqrt{\sigma^2/n}, \overline{X} + z_{1-\alpha/2}\sqrt{\sigma^2/n}],$$

(iv) n groß ist und die Varianz unbekannt ist, gegeben durch

$$[\overline{X} - z_{1-\alpha/2}\sqrt{\widetilde{S}^2/n}, \overline{X} + z_{1-\alpha/2}\sqrt{\widetilde{S}^2/n}].$$

Achtung: Quantile der N(0, 1)!

In den Fällen (iii) und (iv) handelt es sich um approximative Konfidenintervalle für $n \geq 30$. In (iv) kann sowohl die nichtkorrigierte als auch die korrigierte Stichprobenvarianz verwendet werden.

● Allgemeine Merkregel via Standardfehler ●

Schätzwert ± Quantil · (geschätzter) Standardfehler

Beispiel: Approximatives Konfidenzintervall bei unbekannter Varianz

... inklusive notationeller Besonderheiten

KI lautet: $\left[\hat{\mu} - z_{1-\alpha/2} \times \hat{\sigma}_{\hat{\mu}}\,, \hat{\mu} + z_{1-\alpha/2} \times \hat{\sigma}_{\hat{\mu}}\right]$

Dabei gilt:

$\hat{\mu} = \bar{X} \sim N(\mu, \sigma^2/n)$ ...Verteilung des Schätzers

$\sigma^2_{\hat{\mu}} = Var(\hat{\mu}) = \sigma^2/n$ ...Varianz des Schätzers („Genauigkeit“)

$\hat{\sigma}^2_{\hat{\mu}} = S^2/n$ ...geschätzte Varianz des Schätzers

$\hat{\sigma}_{\hat{\mu}} = \sqrt{\hat{\sigma}^2_{\hat{\mu}}} = \sqrt{S^2/n}$ ...geschätzte Standardabweichung des Schätzers ⇒ heißt auch **Standardfehler**

● Beispiel 10.2.2: Gepäckabfertigungszeiten ●

> Gegeben sei nochmals die Situation von Folie 672, wobei jedoch die Varianz σ^2 unbekannt ist. $X_1, \ldots, X_{10}$ sind also unabhängig $N(\mu, \sigma^2)$-verteilt. Wir betrachten wiederum die Beobachtungswerte (Realisationen):

7.8, 10.1, 9.0, 8.0, 11.6, 10.7, 8.1, 8.6, 9.4, 11.9.

> Wie lautet dann das 0.95-KI für μ? Es liegt Fall (ii) von Satz 10.2.1 vor.

$$\left[\bar{X} - t_{n-1,1-\alpha/2}\sqrt{S^2/n}\,, \bar{X} + t_{n-1,1-\alpha/2}\sqrt{S^2/n}\right]$$

9.52 2.24 10 2.26 ← Nachzuschlagendes Quantil:

$$0.95 = 1 - \alpha \quad \Rightarrow \quad \alpha = 0.05$$

$$\Rightarrow \quad t_{n-1,1-\alpha/2} = t_{9,0.975} = 2.26$$

$$= \left[9.52 - 2.26\sqrt{2.24/10}\,, 9.52 + 2.26\sqrt{2.24/10}\right]$$

$$= [8.45, 10.59]$$

> Angenommen die Populationsvarianz ist bekannt mit $\sigma^2 = 1.5$. Wie viele Abfertigungszeiten müssen erhoben werden, damit das 0.95-KI für μ (die mittlere Abfertigungszeit in der Grundgesamtheit) nicht länger ist als 1 (± 0.5 Minuten genaue Schätzung)?

Maßgebliche Formel (vgl. Folie 683):

$$n = \left(2 z_{1-\alpha/2} \sigma / L_Z\right)^2$$

Mit $1 - \alpha = 0.95$, d. h. $z_{0.975} = 1.96$, $\sigma = \sqrt{1.5}$ und $L_Z = 1$ folgt

$$n = \left(2 \times 1.96 \times \sqrt{1.5}/1\right)^2 \approx 23.04$$

⇒ Es sollten mindestens 23 Zeiten (besser 24) erhoben werden.

10.2 Intervallschätzung

Tab. A.1: Wertetabelle der Standardnormalverteilung

Tabelliert sind die Werte der Verteilungsfunktion der Normalverteilung, $\Phi(z)$.
Ablesebeispiel: $\Phi(1.23) = \Phi(1.2 + 0.03) \approx 0.8907$.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

10.2 Intervallschätzung

2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

10.2 Intervallschätzung

Tab. A.2: Quantilstabelle der t-Verteilung

Tabelliert sind die Quantile der t-Verteilung für n Freiheitsgrade.
Für $n > 30$ gilt: $t_{n,\alpha} \approx z_\alpha$, wobei z_α das α-Quantil der Standardnormalverteilung ist.
Ablesebeispiel: $t_{20,0.99} \approx 2.528$.

n	0.6	0.8	0.9	0.95	0.975	0.99	0.995	0.999	0.9995
1	0.3249	1.3764	3.0777	6.3138	12.706	31.821	63.657	318.309	636.619
2	0.2887	1.0607	1.8856	2.9200	4.3027	6.9646	9.9248	22.3271	31.5991
3	0.2767	0.9785	1.6377	2.3534	3.1824	4.5407	5.8409	10.2145	12.9240
4	0.2707	0.9410	1.5332	2.1318	2.7764	3.7469	4.6041	7.1732	8.6103
5	0.2672	0.9195	1.4759	2.0150	2.5706	3.3649	4.0321	5.8934	6.8688
6	0.2648	0.9057	1.4398	1.9432	2.4469	3.1427	3.7074	5.2076	5.9588
7	0.2632	0.8960	1.4149	1.8946	2.3646	2.9980	3.4995	4.7853	5.4079
8	0.2619	0.8889	1.3968	1.8595	2.3060	2.8965	3.3554	4.5008	5.0413
9	0.2610	0.8834	1.3830	1.8331	2.2622	2.8214	3.2498	4.2968	4.7809
10	0.2602	0.8791	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437	4.5869
11	0.2596	0.8755	1.3634	1.7959	2.2010	2.7181	3.1058	4.0247	4.4370
12	0.2590	0.8726	1.3562	1.7823	2.1788	2.6810	3.0545	3.9296	4.3178
13	0.2586	0.8702	1.3502	1.7709	2.1604	2.6503	3.0123	3.8520	4.2208
14	0.2582	0.8681	1.3450	1.7613	2.1448	2.6245	2.9768	3.7874	4.1405
15	0.2579	0.8662	1.3406	1.7531	2.1314	2.6025	2.9467	3.7328	4.0728

10.2 Intervallschätzung

16	0.2576	0.8647	1.3368	1.7459	2.1199	2.5835	2.9208	3.6862	4.0150
17	0.2573	0.8633	1.3334	1.7396	2.1098	2.5669	2.8982	3.6458	3.9651
18	0.2571	0.8620	1.3304	1.7341	2.1009	2.5524	2.8784	3.6105	3.9216
19	0.2569	0.8610	1.3277	1.7291	2.0930	2.5395	2.8609	3.5794	3.8834
20	0.2567	0.8600	1.3253	1.7247	2.0860	2.5280	2.8453	3.5518	3.8495
21	0.2566	0.8591	1.3232	1.7207	2.0796	2.5176	2.8314	3.5272	3.8193
22	0.2564	0.8583	1.3212	1.7171	2.0739	2.5083	2.8188	3.5050	3.7921
23	0.2563	0.8575	1.3195	1.7139	2.0687	2.4999	2.8073	3.4850	3.7676
24	0.2562	0.8569	1.3178	1.7109	2.0639	2.4922	2.7969	3.4668	3.7454
25	0.2561	0.8562	1.3163	1.7081	2.0595	2.4851	2.7874	3.4502	3.7251
26	0.2560	0.8557	1.3150	1.7056	2.0555	2.4786	2.7787	3.4350	3.7066
27	0.2559	0.8551	1.3137	1.7033	2.0518	2.4727	2.7707	3.4210	3.6896
28	0.2558	0.8546	1.3125	1.7011	2.0484	2.4671	2.7633	3.4082	3.6739
29	0.2557	0.8542	1.3114	1.6991	2.0452	2.4620	2.7564	3.3962	3.6594
30	0.2556	0.8538	1.3104	1.6973	2.0423	2.4573	2.7500	3.3852	3.6460

10.2.3 Konfidenzintervalle für Erwartungswertdifferenzen

● Hintergrund, Modellrahmen und Überblick ●

Im Folgenden betrachten wir stets $n_0 + n_1$ Stichprobenvariablen

$Y_{01}, Y_{02}, \dots, Y_{0n_0}$ mit $E(Y_{0i}) = \mu_0$ und $0 < Var(Y_{0i}) = \sigma_0^2 < \infty$

$Y_{11}, Y_{12}, \dots, Y_{1n_1}$ mit $E(Y_{1i}) = \mu_1$ und $0 < Var(Y_{1i}) = \sigma_1^2 < \infty$

Bei Gültigkeit weiterer Annahmen können dann spezifische KI‘s für die theoretische Differenz $\mu_1 - \mu_0$ konstruiert werden.

● Schätzung von $\mu_1 - \mu_0$ bei Normalverteilung und bekannten Varianzen ●

> Annahmen:

$Y_{01}, \dots, Y_{0n_0}$ unabhängig $N(\mu_0, \sigma_0^2)$-verteilt, μ_0 unbekannt, σ_0^2 bekannt

$Y_{11}, \dots, Y_{1n_1}$ unabhängig $N(\mu_1, \sigma_1^2)$-verteilt, μ_1 unbekannt, σ_1^2 bekannt

$Y_{01}, \dots, Y_{0n_0}, Y_{11}, \dots, Y_{1n_1}$ (vollständig) unabhängig

> Herleitung eines $(1-\alpha)$ −Konfidenzintervalls für $\mu_1 - \mu_0$:

$$\bar{Y}_0 \sim N(\mu_0, \sigma_0^2/n_0), \quad \bar{Y}_1 \sim N(\mu_1, \sigma_1^2/n_1)$$

$$\Rightarrow \quad \bar{Y}_1 - \bar{Y}_0 \sim N(\mu_1 - \mu_0, \sigma_0^2/n_0 + \sigma_1^2/n_1),$$

$$\Rightarrow \quad Z_D = \frac{\bar{Y}_1 - \bar{Y}_0 - (\mu_1 - \mu_0)}{\sqrt{\sigma_0^2/n_0 + \sigma_1^2/n_1}} \sim N(0,1),$$

$$\Rightarrow \quad 1 - \alpha = P\left(-z_{1-\alpha/2} \leq Z_D \leq z_{1-\alpha/2}\right)$$

$$= \cdots$$

$$= P\left(\begin{array}{l} \bar{Y}_1 - \bar{Y}_0 - z_{1-\alpha/2}\sqrt{\sigma_0^2/n_0 + \sigma_1^2/n_1} \leq \mu_1 - \mu_0 \\ \qquad \leq \bar{Y}_1 - \bar{Y}_0 + z_{1-\alpha/2}\sqrt{\sigma_0^2/n_0 + \sigma_1^2/n_1} \end{array} \right)$$

> Resultierendes KI:

$$\left[\bar{Y}_1 - \bar{Y}_0 - z_{1-\alpha/2}\sqrt{\sigma_0^2/n_0 + \sigma_1^2/n_1}\,,\, \bar{Y}_1 - \bar{Y}_0 + z_{1-\alpha/2}\sqrt{\sigma_0^2/n_0 + \sigma_1^2/n_1}\right]$$

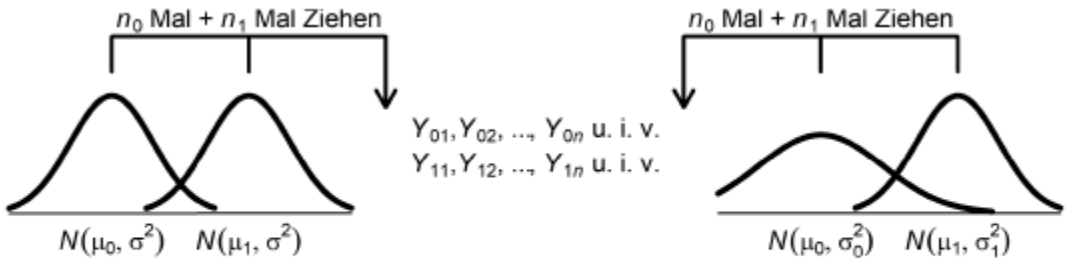


Abb. 10.2.7: Stichprobe bei homoskedastischer und heteroskedastischer Varianz

● Schätzung von $\mu_1 - \mu_0$ bei Normalverteilung und unbekannten Varianzen ●

> Annahmen:

$Y_{01}, \ldots, Y_{0n_0}$ unabhängig $N(\mu_0, \sigma_0^2)$-verteilt, μ_0 und σ_0^2 unbekannt

$Y_{11}, \ldots, Y_{1n_1}$ unabhängig $N(\mu_1, \sigma_1^2)$-verteilt, μ_1 und σ_1^2 unbekannt

$Y_{01}, \ldots, Y_{0n_0}, Y_{11}, \ldots, Y_{1n_1}$ (vollständig) unabhängig

> Problem dabei ist, dass die Statistik

$$T_D = \frac{\bar{Y}_1 - \bar{Y}_0 - (\mu_1 - \mu_0)}{\sqrt{\tilde{S}_0^2/n_0 + \tilde{S}_1^2/n_1}}$$

keiner einfach zu spezifizierenden Verteilung genügt. Auch die Verwendung von S_0^2 und S_1^2 als Varianzschätzer bringt hier keine Lösung. Einzig für großes n_0 bzw. n_1 erhalten wir eine asymptotische Normalverteilung.

10.2 Intervallschätzung

> Zusätzliche Annahme bei kleinen Stichproben, d. h. $n_0, n_1 < 30$:

$$\sigma_0^2 = \sigma_1^2$$ **Homoskedastizität**

Mit dieser zusätzlichen Annahme lässt sich dann zeigen, dass gilt:

$$\tilde{T}_D = \frac{\bar{Y}_1 - \bar{Y}_0 - (\mu_1 - \mu_0)}{\sqrt{S_p^2/n_0 + S_p^2/n_1}} \sim t(n_0 + n_1 - 2) \quad \text{mit}$$

$$S_p^2 = \frac{1}{n_0 + n_1 - 2}\left(n_0 \tilde{S}_0^2 + n_1 \tilde{S}_1^2\right)$$

> Herleitung eines $(1-\alpha)$ −Konfidenzintervalls für $\mu_1 - \mu_0$:

$$1 - \alpha = P\left(-t_{n_0+n_1-2,1-\alpha/2} \leq \tilde{T}_D \leq t_{n_0+n_1-2,1-\alpha/2}\right) \ldots$$

> Resultierendes KI:

$$\left[\bar{Y}_1 - \bar{Y}_0 - t_{n_0+n_1-2,1-\alpha/2}\sqrt{\frac{S_p^2}{n_0} + \frac{S_p^2}{n_1}}, \bar{Y}_1 - \bar{Y}_0 + t_{n_0+n_1-2,1-\alpha/2}\sqrt{\frac{S_p^2}{n_0} + \frac{S_p^2}{n_1}}\right]$$

● Schätzung von $\mu_1 - \mu_0$ bei beliebigen Ausgangsverteilungen ●

> Liegen keine Normalverteilungen vor, so können nur für großes n_0 und n_1 noch „auf einfache Weise“ Konfidenzintervalle konstruiert werden.

> Mit dem *ZGWS* für u. h. v.-Schema (vgl. Folie 577) folgt:

$$\bar{Y}_1 - \bar{Y}_0 \overset{approx.}{\sim} N\left(\mu_1 - \mu_0, \frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1}\right)$$

> Weiter kann (für den Fall unbekannter Varianzen) gezeigt werden, dass auch gilt:

$$T_D = \frac{\bar{Y}_1 - \bar{Y}_0 - (\mu_1 - \mu_0)}{\sqrt{\tilde{S}_0^2/n_0 + \tilde{S}_1^2/n_1}} \overset{a}{\sim} N(0,1)$$

auch mit S_0^2 und S_1^2 statt $\tilde{S}_0^2$ bzw. $\tilde{S}_1^2$

> Daraus resultieren dann approximative Konfidenzintervalle der Fälle (iii) und (iv) auf Folie 701.

● Schätzung von $\mu_1 - \mu_0$ bei Abhängigkeit in Form verbundener Werte ●

> Angenommen, eine weltweit operierende Schnellrestaurant-Kette betreibt eine Werbekampagne. Zur Überprüfung der Wirksamkeit der Kampagne werden stichprobenartig einige Filialen ausgewählt und deren wöchentliche Umsätze sowohl vor als auch nach der Werbekampagne ermittelt. Wie lässt sich nun eine Intervallschätzung für den Werbeeffekt konstruieren?
> In diesem Fall muss beachtet werden, dass die Umsätze vor bzw. nach der Werbekampagne für die jeweils gleichen Filialen betrachtet werden.

> Problem: $Y_{01}, \dots, Y_{0n}, Y_{11}, \dots, Y_{1n}$ nicht (vollständig) unabhängig

> Stattdessen: $(Y_{01}, Y_{11}), (Y_{02}, Y_{12}) \dots, (Y_{0n}, Y_{1n})$ sind u. i. v.

$\Rightarrow \quad Y_{11} - Y_{01}, Y_{12} - Y_{02}, \dots, Y_{1n} - Y_{0n}$ sind u. i. v.

(im Beispiel: Umsatzdifferenzen aller Filialen)

> Definiere $D_i = Y_{1i} - Y_{0i}$ für $i = 1, \dots, n$

$\Rightarrow D_1, \dots, D_n$ u. i. v. mit $E(D_i) = \delta$ und $Var(D_i) = \sigma_D^2$

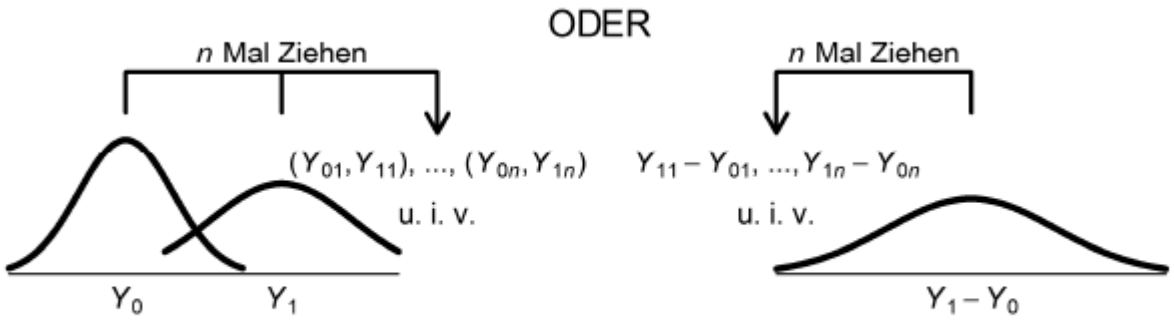


Abb. 10.2.8: Modellierung einer Stichprobe bei verbundenen Werten

> Lösung: Konstruiere nun Konfidenzintervalle für δ auf der Basis von $D_1, \ldots, D_n$ unter Zugrundelegung der „üblichen Annahmen“

> So können nun etwa $D_1, \ldots, D_n$ unabhängig $N(\delta, \sigma_D^2)$-verteilt sein mit bekannter oder unbekannter Varianz. Oder sie sind nur unabhängig, genügen einer beliebigen Verteilung und n ist groß.

> Im Falle unbekannter Varianz wird σ_D^2 durch die Stichprobenvarianz der Differenzenwerte S_D^2 geschätzt.

> Es gelten dann die Formeln von Satz 10.2.1 (Folie 684)

10.2 Intervallschätzung

Satz 10.2.2: Konfidenzintervalle für Erwartungswertdifferenzen

Gegeben seien $n = n_0 + n_1$ unabhängige Stichprobenvariablen

$$Y_{01}, Y_{02}, \ldots, Y_{0n_0} \quad \text{und} \quad Y_{11}, Y_{12}, \ldots, Y_{1n_1},$$

die innerhalb der beiden Gruppen jeweils identisch verteilt sind mit

$$E(Y_{0i}) = \mu_0, \; Var(Y_{0i}) = \sigma_0^2 \quad \text{mit } 0 < \sigma_0^2 < \infty \text{ für } i = 1, \ldots, n_0,$$
$$E(Y_{1i}) = \mu_1, \; Var(Y_{1i}) = \sigma_1^2 \quad \text{mit } 0 < \sigma_1^2 < \infty \text{ für } i = 1, \ldots, n_1.$$

Dann ist ein $(1 - \alpha)$-Konfidenzintervall für $\mu_1 - \mu_0$, falls

(i) alle Variablen normalverteilt sind mit bekannten Varianzen, gegeben durch

$$\left[\overline{Y}_1 - \overline{Y}_0 - z_{1-\alpha/2}\sqrt{\frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1}}, \overline{Y}_1 - \overline{Y}_0 + z_{1-\alpha/2}\sqrt{\frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1}}\right],$$

(ii) alle Variablen normalverteilt sind mit unbekannter gleicher Varianz ($\sigma_0^2 = \sigma_1^2$, homoskedastischer Fall), gegeben durch

$$\left[\overline{Y}_1 - \overline{Y}_0 - t_{n_0+n_1-2,1-\alpha/2}\sqrt{\frac{S_p^2}{n_0} + \frac{S_p^2}{n_1}}, \overline{Y}_1 - \overline{Y}_0 + t_{n_0+n_1-2,1-\alpha/2}\sqrt{\frac{S_p^2}{n_0} + \frac{S_p^2}{n_1}}\right]$$

$$\text{mit } S_p^2 = \frac{1}{n_0 + n_1 - 2}(n_0\tilde{S}_0^2 + n_1\tilde{S}_1^2),$$

(iii) n_0 und n_1 groß sind, Annahme (B) von Definition 8.2.1 und die Bedingung

$$\frac{n_1}{n} \xrightarrow{n\to\infty} \pi \in (0,1)$$

erfüllt und die Varianzen bekannt sind, gegeben durch

$$\left[\overline{Y}_1 - \overline{Y}_0 - z_{1-\alpha/2}\sqrt{\frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1}}, \overline{Y}_1 - \overline{Y}_0 + z_{1-\alpha/2}\sqrt{\frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1}}\right],$$

(iv) die Varianzen unbekannt und sonst alle Annahmen von (iii) erfüllt sind, gegeben durch

$$\left[\overline{Y}_1 - \overline{Y}_0 - z_{1-\alpha/2}\sqrt{\frac{\widetilde{S}_0^2}{n_0} + \frac{\widetilde{S}_1^2}{n_1}}, \overline{Y}_1 - \overline{Y}_0 + z_{1-\alpha/2}\sqrt{\frac{\widetilde{S}_0^2}{n_0} + \frac{\widetilde{S}_1^2}{n_1}}\right].$$

In den Fällen (iii) und (iv) handelt es sich um approximative Konfidenzintervalle für $n_0, n_1 \geq 30$. Hierzu beachte man, dass Annahme (B) erfüllt ist, sofern die Trägermengen der Verteilungen der beiden Gruppen beschränkt sind. Außerdem können im Fall (iv) sowohl nichtkorrigierte als auch korrigierte Stichprobenvarianzen verwendet werden.

● Allgemeine Merkregel via Standardfehler ●

In erweitertem Sinne analog wie auf Folie 685 (vgl. LB, S. 542-543)

● Beispiel 10.2.3: Pflanzenexperiment Nr. 1 ●

Gruppe 0 (ohne Düngung): 14, 25, 36 $\Rightarrow \; n_0 = 3, \bar{y}_0 = 25, \tilde{s}_0^2 = 80.67$

Gruppe 1 (mit Düngung): 64, 46, 55 $\Rightarrow \; n_1 = 3, \bar{y}_1 = 55, \tilde{s}_1^2 = 54.0$

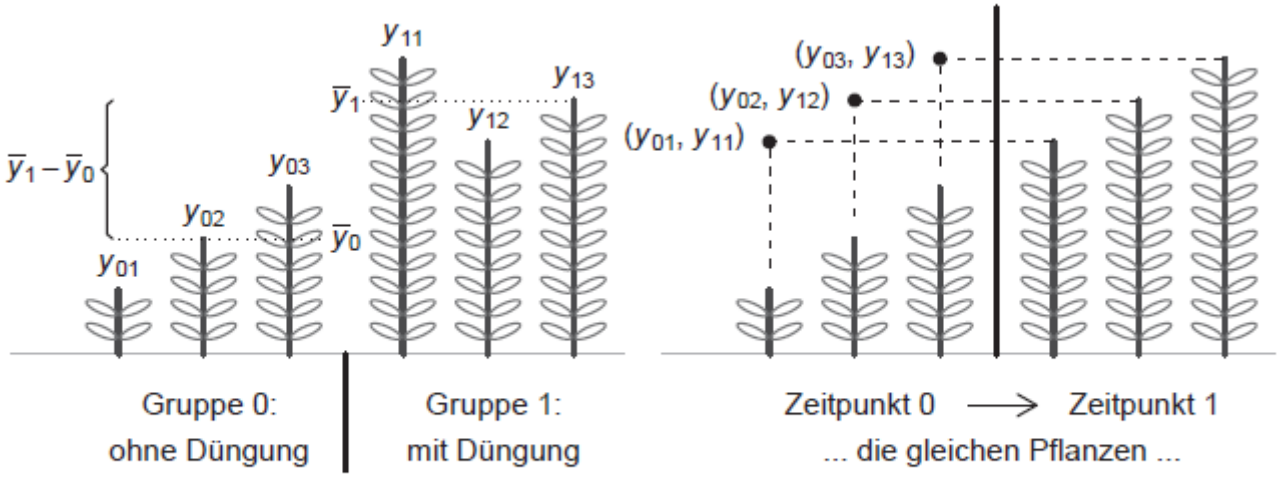


Abb. 10.2.9: Pflanzenexperimente der Beispiele 10.2.3 und 10.2.4

Annahmen: Situation von Abb. 10.2.9 links mit den Annahmen von Fall (ii) in Satz 10.2.2 (Folie 700), 0.95-KI für $\mu_1 - \mu_0$

$$\bar{y}_1 - \bar{y}_0 = 55 - 25 = 30$$

$$s_p^2 = \frac{1}{n_0 + n_1 - 2}(n_0 \tilde{s}_0^2 + n_1 \tilde{s}_1^2)$$

$$= \frac{1}{3 + 3 - 2}(3 \times 80.64 + 3 \times 54.0) \approx 101$$

$$t_{3+3-2,0.975} = t_{4,0.975} = 2.7764$$

$$\Rightarrow \quad \left[30 - 2.7764\sqrt{\frac{101}{3} + \frac{101}{3}}, 30 + 2.7764\sqrt{\frac{101}{3} + \frac{101}{3}}\right]$$

$$= [7.2, 52.8]$$

● Beispiel 10.2.4: Pflanzenexperiment Nr. 2 ●

	Pflanze 1	Pflanze 2	Pflanze 3
Zeitpunkt 0	14	25	36
Zeitpunkt 1	46	55	64

Annahmen: Situation von Abb. 10.2.9 rechts, verbundene Stichprobe
mit den Annahmen von Fall (ii) in Satz 10.2.1 (Folie 684),
0.95-KI für $\delta = \mu_1 - \mu_0$

$$d_1 = y_{11} - y_{01} = 46 - 14 = 32, \quad d_2 = y_{12} - y_{02} = 55 - 25 = 30,$$

$$d_3 = y_{13} - y_{03} = 64 - 36 = 28$$

$$\hat{\delta} = \bar{d} = \frac{1}{3}(32 + 30 - 28) = 30$$ gleiches Ergebnis wie zuvor (Zufall?)

$$\hat{\sigma}_D^2 = s_D^2 = \frac{1}{2}\left((32 - 30)^2 + (30 - 30)^2 + (28 - 30)^2\right) = 4.0$$

$$t_{3-1,0.975} = t_{2,0.975} = 4.3027$$

$$\Rightarrow \quad \left[30 - 4.3027\sqrt{\frac{4}{3}}, 30 + 4.3027\sqrt{\frac{4}{3}}\right]$$

$$= [25.0, 35.0]$$

• Bemerkung zum Zweistichproben-Gedanken •

> In manchen Lehrbüchern wird die Gruppierung

$$Y_{01}, \ldots, Y_{0n_0}, \qquad Y_{11}, \ldots, Y_{1n_1}$$

als Vorliegen von 2 Stichproben umschrieben...

> Wir ziehen den Ansatz vor, immer nur von einer einzigen Stichprobe auszugehen. Dabei sind die Stichprobenvariablen jedoch u. U. heterogen verteilt (wie etwa im Falle von Differenzenschätzungen relevant)

• Äquivalente Darstellungen im Regressionsmodell •

10.2.4 Weitere Konfidenzintervalle

• Konfidenzintervalle für Anteilswerte •

> Fall (iv) von Satz 10.2.1 (Folie 684) kann für den Fall einer $B(1,\pi)$-Verteilung spezialisiert werden. Dies führt zu einer Anteilsschätzung.

> Sind $X_1, \ldots, X_n$ unabhängig $B(1,\pi)$-verteilt, gilt:

$$E(X_i) = \pi, \qquad Var(X_i) = \pi(1-\pi)$$

$$E(\bar{X}) = \pi, \qquad Var(\bar{X}) = \frac{\pi(1-\pi)}{n}$$

> Bezogen auf Fall (iv) gilt

$$\mu = \pi, \qquad \sigma^2 = \pi(1-\pi)$$

> Da die X_i nur die Werte 0 und 1 annehmen können, gilt:

$$X_i^2 = X_i \quad \Rightarrow \quad \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

$$\Rightarrow \quad \tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \bar{X} - \bar{X}^2 = \bar{X}(1-\bar{X})$$

> Damit erhalten wir approximatives KI für π ...

Satz 10.2.3: Konfidenzintervalle für Anteilswerte

Seien $X_1, \ldots, X_n$ unabhängig $B(1, \pi)$-verteilt mit $0 < \pi < 1$ und $n \geq 30$. Dann ist ein approximatives $(1 - \alpha)$-Konfidenzintervall für π gegeben durch

$$\left[\hat{\pi} - z_{1-\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}, \hat{\pi} + z_{1-\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}\right].$$

Dabei ist $\hat{\pi} = \overline{X}$ die relative Häufigkeit der Einsen in der Stichprobe.

● Notwendiger Stichprobenumfang zur Erzielung bestimmter Längen ●

> Länge des approximativen KI‘s:

$$L_{\hat{\pi}} = 2z_{1-\alpha/2}\sqrt{\hat{\pi}(1-\hat{\pi})/n}$$

> Da $\hat{\pi} = \bar{X} \in [0,1]$ ist $\hat{\pi}(1-\hat{\pi})$ nach oben beschränkt. Es gilt:

$$\hat{\pi}(1-\hat{\pi}) \in [0,0.25], \text{ d. h. } \hat{\pi}(1-\hat{\pi}) \leq 0.25$$

> Damit ist die Länge des KI‘s zu vorgegebenen Niveau beschränkt:

$$L_{\hat{\pi}}^{max} = 2z_{1-\alpha/2}\sqrt{0.25/n} = z_{1-\alpha/2}\sqrt{1/n}$$

> Damit lässt sich ein Mindeststichprobenumfang bestimmen, bei dem eine vorgegebene theoretische Maximallänge nicht überschritten wird:

$$n = \left(z_{1-\alpha/2}/L_{\hat{\pi}}^{max}\right)^2$$

> Beispiel: $L_{\hat{\pi}}^{max} = 0.02$ ($\pm 1\%$ genaue Schätzung), $1-\alpha = 0.95$:

$$n = (z_{0.975}/0.02)^2 = (1.96/0.02)^2 = 9604$$

> Man beachten die vorgebene Maximallänge würde nur bei einem geschätzten Anteilswert von $\hat{\pi} = 0.5$ eintreten.

● **Beispiel 10.2.5: Mensabefragung** ●

Es votieren 40 von 50 Befragten für die Mensa. Wo liegt der Anteilswert? Das (approximative) 0.95-KI für π lautet dann:

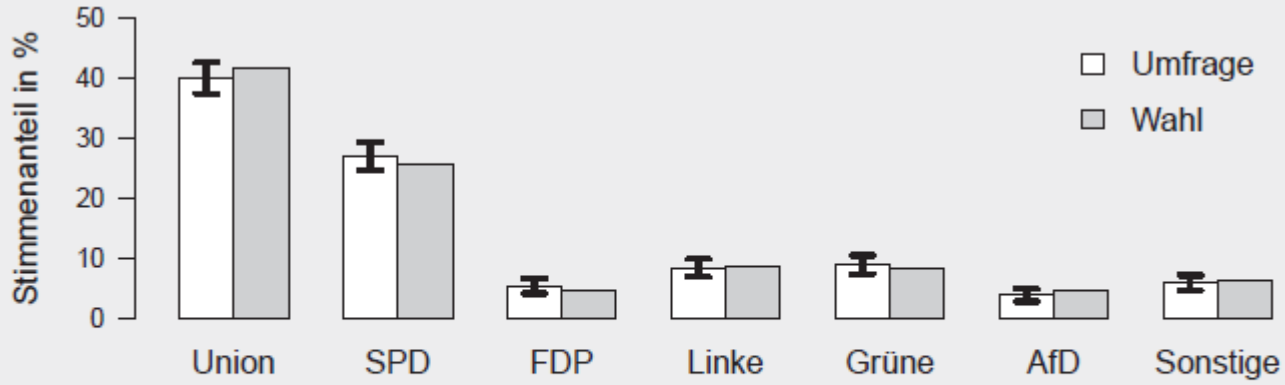
$$\left[\hat{\pi} - z_{0.975}\sqrt{\hat{\pi}(1-\hat{\pi})/n}\,, \hat{\pi} + z_{0.975}\sqrt{\hat{\pi}(1-\hat{\pi})/n}\right]$$

$$\hat{\pi} = 40/50 = 0.8, \quad \hat{\pi}(1-\hat{\pi}) = 0.8 \times 0.2 = 0.16$$

$$\Rightarrow \quad \left[0.8 - 1.96\sqrt{0.16/50}\,, 0.8 + 1.96\sqrt{0.16/50}\right]$$

$$= [0.69, 0.91]$$

10.2 Intervallschätzung



Daten: Der Bundeswahlleiter [2016], Forschungsgruppe Wahlen [2016]

Abb. 10.2.10: Politbarometer und Wahlergebnis der Bundestagswahl 2013

Tab. 10.2.1: Umfrageergebnis und Wahlergebnis (in %) der Bundestagswahl 2013

Partei	Union	SPD	FDP	Linke	Grüne	AfD	Sonst.
Umfrage und	40.0	27.0	5.5	8.5	9.0	4.0	6.0
Genauigkeit	±2.6%	±2.4%	±1.2%	±1.5%	±1.5%	±1.0%	±1.3%
Endergebnis	41.5	25.7	4.8	8.6	8.4	4.7	6.3

Daten: Der Bundeswahlleiter [2016], Forschungsgruppe Wahlen [2016]

$n = 1369$

710

● Konfidenzintervalle für Anteilswertdifferenzen ●

> Fall (iv) von Satz 10.2.2 (Folie 701) kann für den Fall einer Differenzschätzung zweier Anteilswerte $(\pi_1 - \pi_0)$ spezialisiert werden...

> Nach einigen Umformungen gelangt man dann zur Formel:

$$\left[\hat{\pi}_1 - \hat{\pi}_0 - z_{1-\alpha/2}\sqrt{\frac{\hat{\pi}_0(1-\hat{\pi}_0)}{n_0} + \frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1}},\right.$$

$$\left.\hat{\pi}_1 - \hat{\pi}_0 + z_{1-\alpha/2}\sqrt{\frac{\hat{\pi}_0(1-\hat{\pi}_0)}{n_0} + \frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1}}\right]$$

● Konfidenzintervalle für σ^2 ●

● Konfidenzintervalle für sonstige Größen ●

10.2.5 Adäquatheit bestimmter Modellannahmen

- Hintergrund und Überblick
- Zuallererst: Repräsentativität
- Identisch vs. heterogen verteilt
- Normal- vs. nicht normalverteilt
- Verwendung der Approximationsregeln
- Unabhängig vs. abhängig
- Homoskedastizität vs. Heteroskedastizität
- Nichtstochastische vs. stochastische Gruppenumfänge

10.2 Intervallschätzung

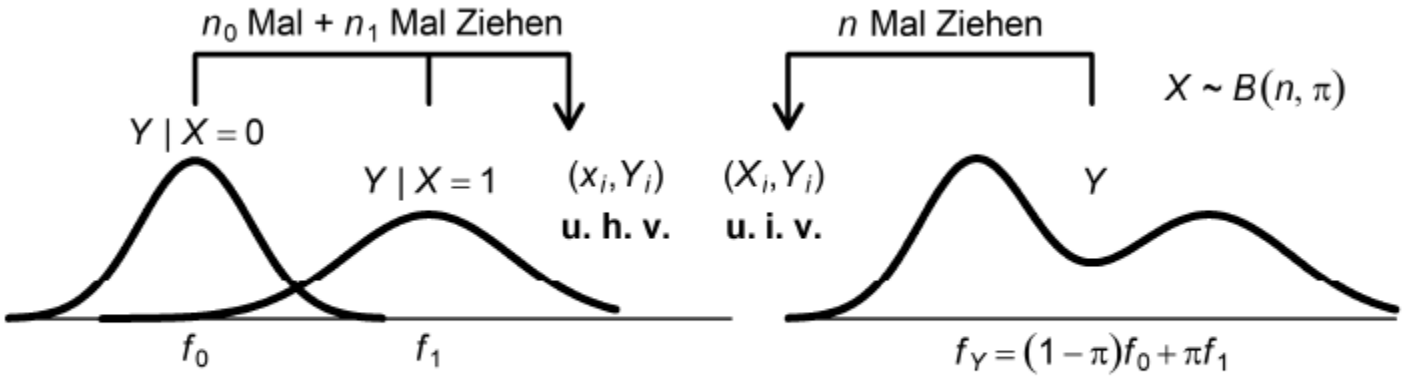


Abb. 10.2.11: Nichtstochastische vs. stochastische Gruppenumfänge

10.3 Schätzmethoden

10.3.1 Momentenmethode

● Allgemeiner Ansatz ●

Die Momentenmethode basiert darauf, dass **Stichprobenmomente** stochastisch gegen die korrespondierenden **theoretischen Momente** konvergieren:

$$\frac{1}{n}\sum_{i=1}^{n} X_i^m \xrightarrow{p} E(X_i^m) \qquad \text{(GGZ für } k\text{-te Momente, Folie 581)}$$

> Stelle, sofern möglich, zunächst die zu schätzenden Parameter als Funktionen theoretischer Momente dar. Ersetze dann letztere durch die korrespondierenden Stichprobenmomente.

● Beispiel 10.3.1: Exponentialverteilung ●

Angenommen, $X_1, X_2, \ldots, X_n$ sind unabhängig $Exp(\lambda)$-verteilt, wobei der Parameter λ geschätzt werden soll. Dann gilt zunächst einmal

$$E(X_i) = 1/\lambda. \qquad \text{(Momentengleichung)}$$

Durch Umstellen erhalten wir dann

$$\lambda = 1/E(X_i).$$

In dieser Darstellung ist der zu schätzende Parameter λ nun als Funktion des ersten theoretischen Moments dargestellt. Das theoretische Moment wird nun durch das entsprechende Stichprobenmoment ersetzt. Daraus erhält man dann

$$\hat{\lambda}_{MM} = 1/\bar{X}$$

als **Momentenschätzer** für λ.

● Beispiel 10.3.2: Schätzung von μ und σ^2 ●

Angenommen, $X_1, X_2, \ldots, X_n$ sind u. i. v. mit $E(X_i) = \mu$ und $Var(X_i) = \sigma^2$, wobei beide Parameter unbekannt sind und geschätzt werden sollen.

(i) $E(X_i) = \mu$,

(ii) $E(X_i^2) = \sigma^2 + \mu^2$

(2 Momentengleichungen)

Darstellung der Parameter als Funktion der theoretischen Momente:

(i) $\mu = E(X_i)$,

(ii) $\sigma^2 = E\left(X_i^2\right) - \mu^2$

Einsetzen von (i) in (ii):

(ii) $\sigma^2 = E\left(X_i^2\right) - E(X_i)^2$

Ersetzen der theoretischen Momente durch Stichprobenmomente liefert als Momentenschätzer:

(i) $\hat{\mu}_{MM} = \bar{X}$,

(ii) $\hat{\sigma}^2_{MM} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \tilde{S}^2$

● Beispiel 10.3.3: Approximatives Konfidenzintervall für die Varianz ●

Siehe LB, S. 558-559 (weniger triviales Beispiel)

10.3.2 Maximum-Likelihood-Methode

- **Vorbemerkung**
- **Allgemeiner Ansatz**

Unter welchem Parameterwert sind die Beobachtungen am wahrscheinlichsten?

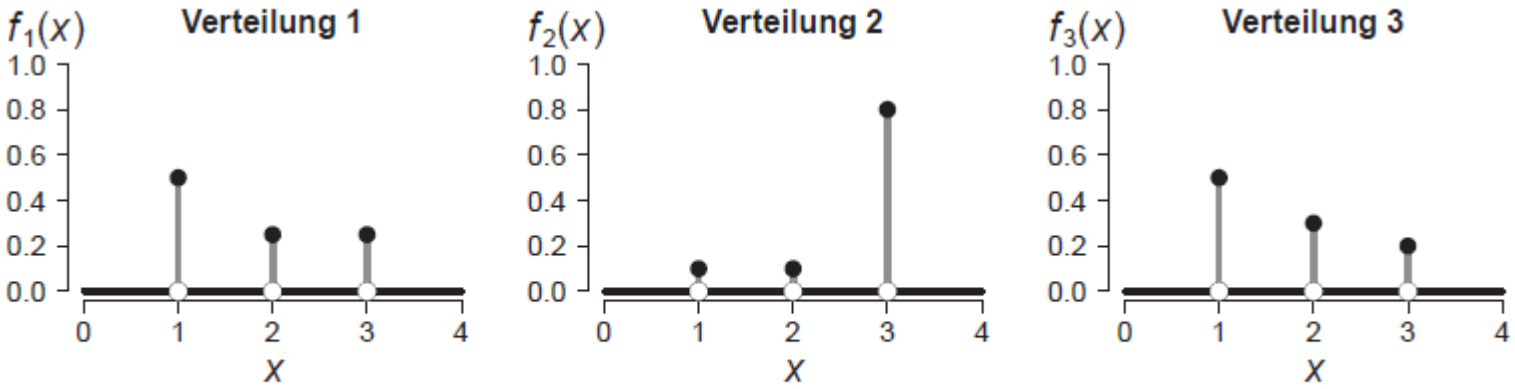


Abb. 10.3.1: Drei zur Wahl stehende Verteilungen

● **Beispiel 10.3.4: Schätzung von λ bei einer Poisson-Verteilung für $n = 1$** ●

$$X \sim Po(\lambda)$$

$$\Rightarrow \quad f_X(x) = P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}, \qquad x = 0, 1, 2, \ldots$$

Likelihood-Funktion:

$$L(\lambda|x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

Ziel: Maximiere L bezüglich λ.

Problem: L analytisch schwer handhabbar

Lösung: **Log-Likelihood-Funktion** einfacher handhabbar

$$l(\lambda|x) = ln\big(L(\lambda|x)\big) = x ln(\lambda) - ln(x!) - \lambda$$

Die Maximierung von l führt aufgrund der Monotonie des Logarithmus zur identischen Lösung wie die Maximierung von L.

Ableitung der Log-Likelihood:

$$\frac{\partial l}{\partial \lambda} = \frac{x}{\lambda} - 1 \qquad \text{(Likelihood-Gleichung)}$$

Nullsetzen und Auflösen nach λ:

$$\frac{x}{\lambda} - 1 = 0 \quad \Leftrightarrow \quad \lambda = x$$

Resultierender Maximum-Likelihood-Schätzer (ML-Schätzer):

$$\hat{\lambda}_{ML} = X$$

● Beispiel 10.3.5: Schätzung von μ bei einer Normalverteilung für $n = 1$ ●

$$X \sim N(\mu, 1)$$

$$\Rightarrow \quad f_X(x) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{1}{2}(x-\mu)^2\right)$$

Likelihood-Funktion:

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

Ziel: Maximiere L bezüglich μ.

Problem: L analytisch schwer handhabbar

Lösung: **Log-Likelihood-Funktion** einfacher handhabbar

$$l(\mu|x) = -0.5ln(2\pi) - 0.5(x-\mu)^2$$

Ableitung der Log-Likelihood:

$$\frac{\partial l}{\partial \mu} = x - \mu$$

Nullsetzen und Auflösen nach μ:

$$x - \mu = 0 \quad \Leftrightarrow \quad \mu = x$$

Resultierender ML-Schätzer:

$$\hat{\mu}_{ML} = X$$

- **Beispiel 10.3.6: Schätzung von λ bei einer Exponentialverteilung für $n = 1$** ●

$$X \sim Exp(\lambda)$$

$$\Rightarrow \quad f_X(x) = \lambda e^{-\lambda x} I_{[0,\infty)}(x)$$

Likelihood-Funktion:

$$L(\lambda|x) = \lambda e^{-\lambda x}$$

Ziel: Maximiere L bezüglich λ.

Problem: L analytisch schwer handhabbar

Lösung: **Log-Likelihood-Funktion** einfacher handhabbar

$$l(\lambda|x) = ln\big(L(\lambda|x)\big) = ln(\lambda) - \lambda x$$

Ableitung der Log-Likelihood:

$$\frac{\partial l}{\partial \lambda} = \frac{1}{\lambda} - x$$

Nullsetzen und Auflösen nach λ:

$$\frac{1}{\lambda} - x = 0 \quad \Leftrightarrow \quad \lambda = \frac{1}{x}$$

Resultierender ML-Schätzer:

$$\hat{\lambda}_{ML} = \frac{1}{X}$$

● **Beispiel 10.3.7: Schätzung von λ bei einer Poisson-Verteilung für $n \geq 1$** ●

$X_1, \dots, X_n$ sind unabhängig $Po(\lambda)$-verteilt

Damit klar: $P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \times \cdots \times P(X_n = x_n)$

$$\Rightarrow \quad f_{X_1,\dots,X_n}(x_1, \dots, x_n) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}$$

Likelihood-Funktion:

$$L(\lambda | x_1, \dots, x_n) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \frac{\lambda^{x_1}}{x_1!} e^{-\lambda} \times \frac{\lambda^{x_2}}{x_2!} e^{-\lambda} \times \cdots \times \frac{\lambda^{x_n}}{x_n!} e^{-\lambda}$$

Ziel: Maximiere L bezüglich λ.

Problem: L analytisch schwer handhabbar

Lösung: **Log-Likelihood-Funktion** einfacher handhabbar

$$l(\lambda | x_1, \dots, x_n) = ln(\lambda) \sum_{i=1}^{n} x_i - n\lambda - \sum_{i=1}^{n} ln(x_i!)$$

Ableitung der Log-Likelihood:

$$\frac{\partial l}{\partial \lambda} = \sum_{i=1}^{n} x_i/\lambda - n$$

Nullsetzen und Auflösen nach λ:

$$\sum_{i=1}^{n} x_i/\lambda - n = 0 \quad \Leftrightarrow \quad \lambda = \bar{x}$$

(Likelihood-Gleichung)

Resultierender ML-Schätzer:

$$\hat{\lambda}_{ML} = \bar{X}$$

- **Beispiel 10.3.8: Schätzung von μ und σ^2 bei einer Normalverteilung für $n \geq 1$** -

$X_1, \dots, X_n$ sind unabhängig $N(\mu, \sigma^2)$-verteilt, wobei beide Parameter, d. h. μ und σ^2 unbekannt sind.

$$\Rightarrow \quad f_{X_1,\dots,X_n}(x_1, \dots, x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right)$$

$$= (2\pi\sigma^2)^{-n/2} exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

Likelihood-Funktion:

$$L(\mu, \sigma^2 | x_1, \dots, x_n) = (2\pi\sigma^2)^{-n/2} exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

Ziel: Maximiere L bezüglich μ und σ^2

Problem: L analytisch schwer handhabbar

Lösung: **Log-Likelihood-Funktion** einfacher handhabbar

$$l(\mu, \sigma^2 | x_1, \dots, x_n) = -\frac{n}{2} ln(2\pi) - \frac{n}{2} ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

Partielle Ableitungen der Log-Likelihood:

(i) $\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu)$

(ii) $\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2$

(2 Likelihood-Gleichungen)

Nullsetzen und Lösen nach den Parametern μ und σ^2:

(i) $\frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0$

(ii) $-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2 = 0$

(i) $\mu = \bar{x}$

Einsetzen von (i) in (ii) und Lösen nach σ^2:

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

Resultierende ML-Schätzer:

$$\hat{\mu}_{ML} = \bar{X} \quad \text{und} \quad \hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \tilde{S}^2$$

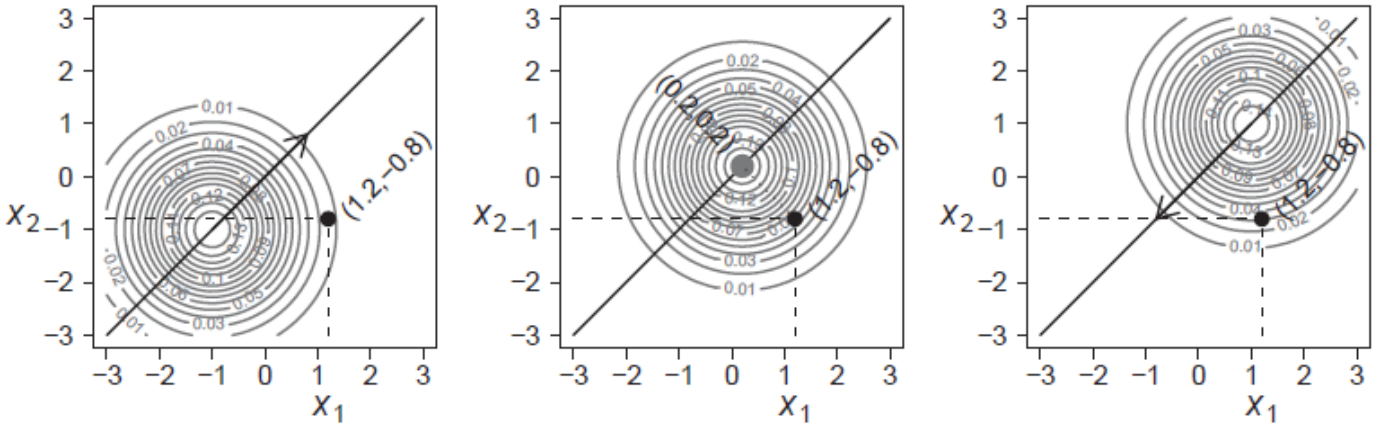


Abb. 10.3.2: ML-Schätzung von μ bei Vorliegen von zwei Beobachtungen

- **Beispiel 10.3.9: Schätzung von λ bei einer Exponentialverteilung für $n \geq 1$** -

$X_1, \dots, X_n$ sind unabhängig $Exp(\lambda)$-verteilt

$$\Rightarrow \quad f_{X_1,\dots,X_n}(x_1, \dots, x_n) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n exp\left(-\lambda \sum_{i=1}^{n} x_i\right)$$

Likelihood-Funktion:

$$L(\lambda | x_1, \dots, x_n) = \lambda^n exp\left(-\lambda \sum_{i=1}^{n} x_i\right)$$

Ziel: Maximiere L bezüglich λ.

Problem: L analytisch schwer handhabbar

Lösung: **Log-Likelihood-Funktion** einfacher handhabbar

$$l(\lambda | x_1, \dots, x_n) = n\, ln(\lambda) - \lambda \sum_{i=1}^{n} x_i$$

Ableitung der Log-Likelihood:

$$\frac{\partial l}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i$$

Nullsetzen und Auflösen nach λ:

$$\frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0 \quad \Leftrightarrow \quad \lambda = \frac{1}{\bar{x}}$$

Resultierender ML-Schätzer:

$$\hat{\lambda}_{ML} = \frac{1}{\bar{X}}$$

● Abschließende Bemerkungen ●

> Im Falle mehrerer Parameter sind Likelihood-Gleichungen analytisch nicht immer lösbar. Ebenso müssen Lösungen nicht eindeutig sein

> Unter bestimmten Bedingungen (Regularitätsbedingungen) ist ML-Schätzung asymptotisch effizient

10.3.3 Weitere Schätzmethoden

● Bayes-Methode ●

> Zur Erinnerung: Schlangenbeispiel (Beispiel 6.2.6, Folie 327)

> Jetzt: Statt Frage nach der richtigen Schlange, Frage nach dem richtigen Parameterwert (Schätzung)

> Bayes-Kalkül:
+ Gegeben ein bestimmtes Symptom, welche Schlange erscheint am wahrscheinlichsten?
+ Jetzt: ...

Welcher Parameterwert ist unter den Beobachtungen am wahrscheinlichsten?

● Kleinste-Quadrate-Methode ●

siehe Kapitel 12

11 Statistisches Testen

11 Statistisches Testen

11.1 Was versteht man unter einem Test?	735
11.1.1 Einführende Beispiele	735
11.1.2 Grundstruktur und Durchführung	749
11.1.3 Zusammenhang zur Intervallschätzung	754
11.2 Wichtige Aspekte beim Testen	756
11.2.1 Hypothesenwahl und Fehlerarten	756
11.2.2 Irrtumswahrscheinlichkeiten und Güte	759
11.2.3 *p*-Werte	774
11.2.4 Signifikanz vs. Relevanz	780
11.3 Ausgewählte Testverfahren	783
11.3.1 Tests über Erwartungswerte	783
11.3.2 Tests über Erwartungswertdifferenzen	790

11 Statistisches Testen

11.3.3 Nichtparametrische χ^2-Tests	795
Allgemeiner Überblick	795
χ^2-Anpassungstest	795
χ^2-Unabhängigkeitstest	810
11.3.4 Weitere Tests	823
Tests über Anteilswerte	823
Tests auf Unkorreliertheit und Unabhängigkeit	836
Sonstige Tests und Testmethoden	844
11.3.5 Allgemein zu beachtende Punkte	845

11.1 Was versteht man unter einem Test?

11.1.1 Einführende Beispiele

● Vorbemerkung ●

● Beispiel 11.1.1: Raten vs. Wissen ●

Angenommen eine Multiple-Choice-Klausur besteht aus 30 Aussagen, die entweder richtig oder falsch sind. Falls ein Student 19 dieser Fragen korrekt und 11 Fragen falsch beantwortet, wie ist dann eine solche Leistung einzustufen?

Statistisches Modell: $X_1, \ldots, X_{30}$ unabhängig $B(1, \pi)$-verteilt

mit Kodierung 0 = falsch, 1 = richtig

Entscheidungsproblem: Raten vs. Wissen

Testproblem: $H_0: \pi = 0.5$ vs. $H_1: \pi > 0.5$

Nullhypothese **Alternativhypothese**

Teststatistik: (Prüfgröße)

$$S_{30} = \sum_{i=1}^{30} X_i \overset{\pi = 0.5}{\sim} B(30, 0.5)$$

Testverteilung

11.1 Was versteht man unter einem Test?

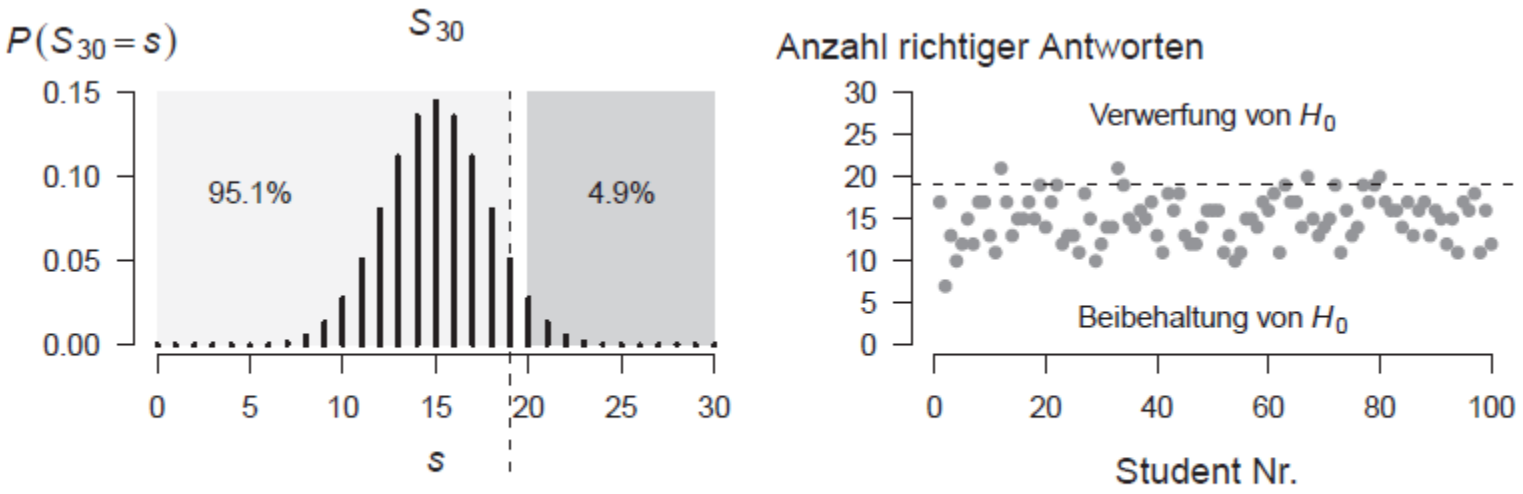


Abb. 11.1.1: Einseitiger oberer Binomialtest – Testverteilung und Simulation

Testniveau (Signifikanzniveau): $\alpha = 0.049$

Kritischer Wert: $c = 19$

Testprozedur: Falls $S_{30} > 19$, Entscheidung für H_1 (Verwerfung von H_0), sonst Beibehaltung von H_0

Testergebnis hier:

Mit $s_{30} = 19$ wird hier H_0 beibehalten. Somit sprechen die Daten bei einem Testniveau von 4.9% nicht signifikant gegen Raten.

● Beispiel 11.1.2: Kein Rückgang vs. Rückgang ●

Studierende werden zu ihrer Zufriedenheit mit der Mensa befragt. Angenommen die Zufriedenheitsquote lag in einem vorhergehenden Jahr bei 80%. In einer Blitzumfrage äußerten sich aktuell nun 15 von 20 Befragten zufrieden. Dies entspricht einem aktuellen Anteil von nur noch 75%. Ist die Zufriedenheit damit nun tatsächlich zurückgegangen im Vergleich zum Vorjahr?

Statistisches Modell: $X_1, \ldots, X_{20}$ unabhängig $B(1,\pi)$-verteilt

mit Kodierung 0 = unzufrieden, 1 = zufrieden

Entscheidungsproblem: Kein Rückgang vs. Rückgang

Testproblem: $H_0\colon\ \pi \geq 0.8$ vs. $H_1\colon \pi < 0.8$

Teststatistik: $S_{20} = \sum_{i=1}^{20} X_i \overset{\pi = 0.8}{\sim} B(20,0.8)$

Testniveau (Signifikanzniveau): $\alpha = 0.032$

11.1 Was versteht man unter einem Test?

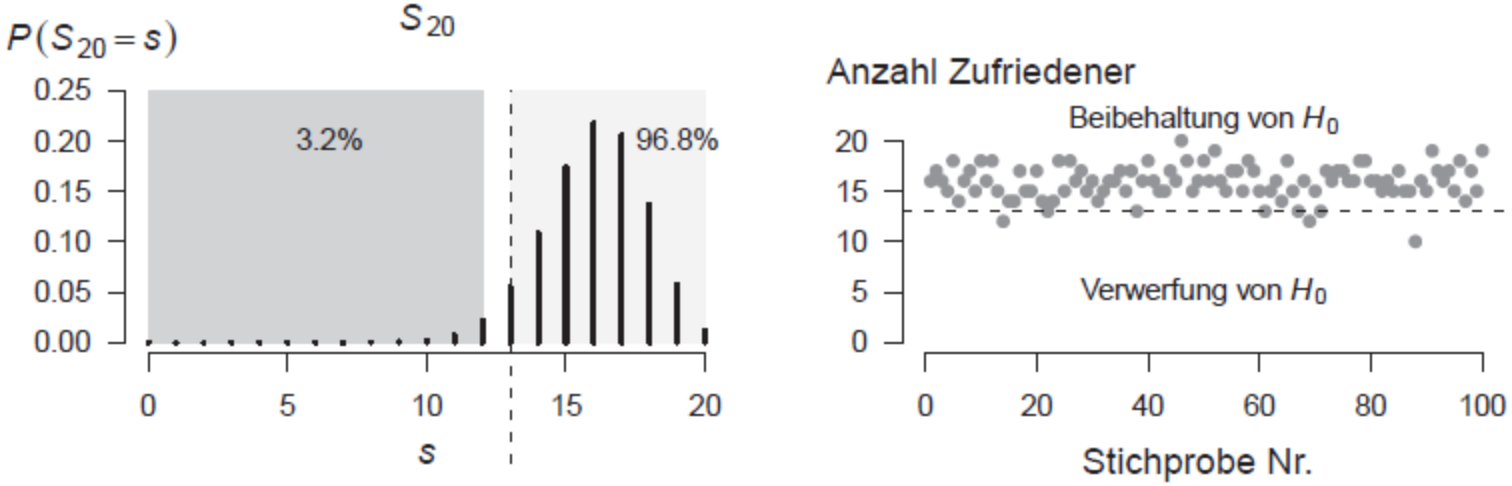


Abb. 11.1.2: Einseitiger unterer Binomialtest – Testverteilung und Simulation

Kritischer Wert: $c = 13$

Testprozedur: Falls $S_{20} < 13$, Entscheidung für H_1 (Verwerfung von H_0), sonst Beibehaltung von H_0

Testergebnis hier:

Mit $s_{20} = 15$ wird hier H_0 beibehalten. Somit sprechen die Daten bei einem Testniveau von 3.2% für keinen signifikanten Rückgang der Zufriedenheitsquote.

• Beispiel 11.1.3: Norm vs. Abweichung •

In der Kunststoffverarbeitung müssen bei der Herstellung eines bestimmten Produktes ein weißes und ein schwarzes Kunststoffgranulat im Verhältnis von 2:3 miteinander vermischt werden. Das korrekte Mischungsverhältnis wird dabei in regelmäßigen Abständen kontrolliert. Dazu wird mithilfe eines Messbechers der laufenden Produktion eine Stichprobe entnommen, von der dann 500 Körner maschinell nach Farben sortiert und ausgezählt werden. Der Anteil schwarzer Körner darf hierbei vom Sollanteil 0.6 nicht signifikant abweichen. Doch wo sind hier die Grenzen zu setzen?

Statistisches Modell: $X_1, \dots, X_{500}$ unabhängig $B(1, \pi)$-verteilt
mit Kodierung 0 = weiß, 1 = schwarz

Entscheidungsproblem: Norm vs. Abweichung

Testproblem: $H_0: \pi = 0.6$ vs. $H_1: \pi \neq 0.6$

Teststatistik: $S_{500} = \sum_{i=1}^{500} X_i \overset{\pi = 0.6}{\sim} B(500, 0.6)$

11.1 Was versteht man unter einem Test?

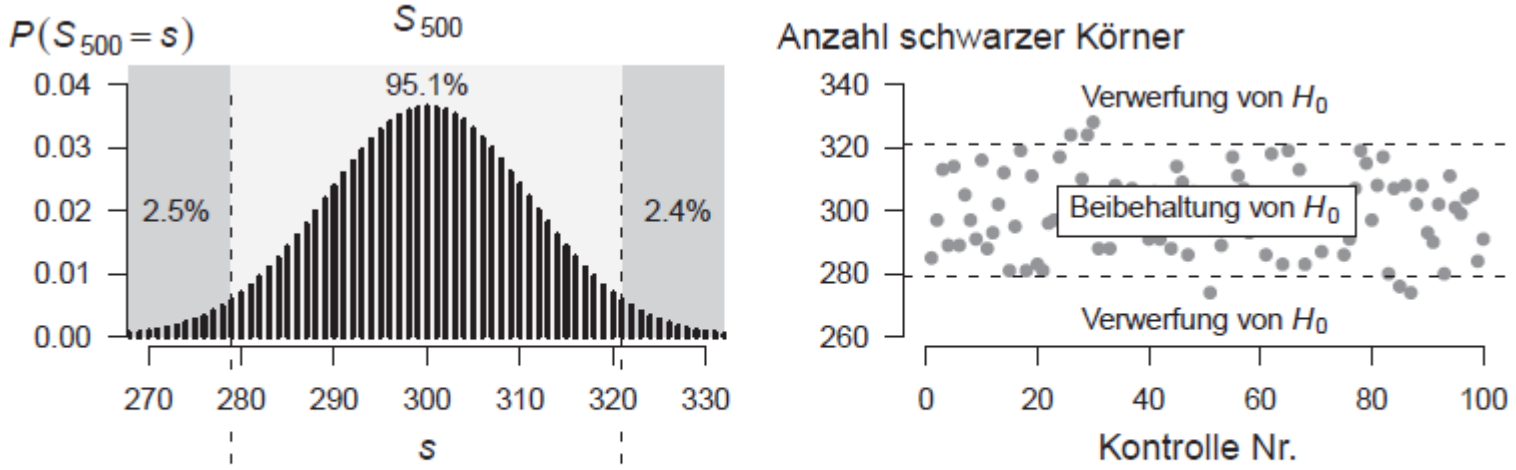


Abb. 11.1.3: Zweiseitiger Binomialtest – Testverteilung und Simulation

Testniveau (Signifikanzniveau): $\alpha = 0.025 + 0.024 = 0.049$

Kritische Werte: $c_1 = 279, c_2 = 321$

Testprozedur: Falls $S_{500} < 279$ oder $S_{500} > 321$, Entscheidung für H_1 (Verwerfung von H_0), sonst Beibehaltung von H_0

Man beachte, dass die kritischen Werte 279 und 321 zu den Anteilswerten 55.8% und 64.2% korrespondieren. Außerhalb dieser Schranken würde man ein Abweichung vom Sollwert 60% als signifikant erachten.

● Approximative Binomialtests - nichtstandardisierte Varianten ●

Wir betrachten erneut das vorhergehende Beispiel (Norm vs. Abweichung)

Jetzt: Approximation der Testverteilung durch Normalverteilung

Statistisches Modell: $X_1, \ldots, X_{500}$ unabhängig $B(1, \pi)$-verteilt

mit Kodierung 0 = weiß, 1 = schwarz

Testproblem: $H_0: \pi = 0.6$ vs. $H_1: \pi \neq 0.6$

Teststatistik: $$S_{500} = \sum_{i=1}^{500} X_i \overset{approx.}{\underset{\pi = 0.6}{\sim}} N(300, 120)$$

Testniveau (Signifikanzniveau): $\alpha = 0.05$

Kritische Werte: $$c_1 = q_{0.025} = 300 + \sqrt{120} \times z_{0.025}$$
$$= 300 - \sqrt{120} \times 1.96 = 278.5$$
$$c_2 = q_{0.975} = 300 + \sqrt{120} \times z_{0.975}$$
$$= 300 + \sqrt{120} \times 1.96 = 321.5$$

11.1 Was versteht man unter einem Test?

Testprozedur: Unter Berücksichtigung, dass S_{500} nur ganze Zahlen annehmen kann, resultiert die gleiche Testprozedur wie zuvor, d. h.

Falls $S_{500} < 279$ oder $S_{500} > 321$, Entscheidung für H_1 (Verwerfung von H_0), sonst Beibehaltung von H_0

Fazit:

Somit bestünde in diesem Beispiel kein Unterschied zwischen dem sog. **exakten Binomialtest** und dem **approximativen Binomialtest**. Die Testentscheidungen wären jeweils immer identisch.

In analoger Weise lassen sich approximative Tests für die Beispiele 11.1.1 (Raten vs. Wissen) und 11.1.2 (kein Rückgang vs. Rückgang) konstruieren, welche ebenfalls zu identischen oder fast identischen Testprozeduren führen. Nähere Details dazu im LB, S. 574-576.

● Approximative Binomialtests - standardisierte Varianten ●

Wir betrachten weiterhin das Beispiel 11.1.3 („Norm vs. Abweichung“)

Jetzt: Approximation durch Normalverteilung und Standardisierung

Statistisches Modell: $X_1, \ldots, X_{500}$ unabhängig $B(1, \pi)$-verteilt
mit Kodierung 0 = weiß, 1 = schwarz

Testproblem: $H_0: \pi = 0.6$ vs. $H_1: \pi \neq 0.6$

Teststatistik: $$Z_{500} = \frac{\sum_{i=1}^{500} X_i - 300}{\sqrt{120}} \overset{a}{\underset{\pi = 0.6}{\sim}} N(0,1)$$

Testniveau (Signifikanzniveau): $\alpha = 0.05$

Kritische Werte: $c_1 = z_{0.025} = -1.96$ und $c_2 = z_{0.975} = 1.96$

Testprozedur: Falls $Z_{500} < -1.96$ oder $Z_{500} > 1.96$, Entscheidung für H_1 (Verwerfung von H_0), sonst Beibehaltung von H_0

Bemerkungen:

> Analog funktionieren die Beispiele 11.1.1 und 11.1.2.

> Die Testprozeduren von standardisierter und nichtstandardisierter Variante sind völlig äquivalent (führen stets zu identischen Entscheidungen).

> Mit $S_n \overset{a}{\sim} N(\mu, \sigma^2)$ und $Z_n \overset{a}{\sim} N(0,1)$ gilt dann beispielsweise:

$$S_n < q_{\alpha/2} \quad \Leftrightarrow \quad Z_n = \frac{S_n - \mu}{\sigma} < \frac{q_{\alpha/2} - \mu}{\sigma} = z_{\alpha/2}$$

Beachte, dass gilt: $q_{\alpha/2} = \mu + \sigma z_{\alpha/2}$

> Allgemeine Form der Teststatistik unter H_0: $\pi = \pi_0$:

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}} = \frac{\bar{X} - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}} \overset{a}{\underset{\pi = \pi_0}{\sim}} N(0,1)$$

Standardisierte Summe

Standardisiertes Stichprobenmittel

> Die standardisierte Variante ist in der Praxis üblicher.

11.1 Was versteht man unter einem Test?

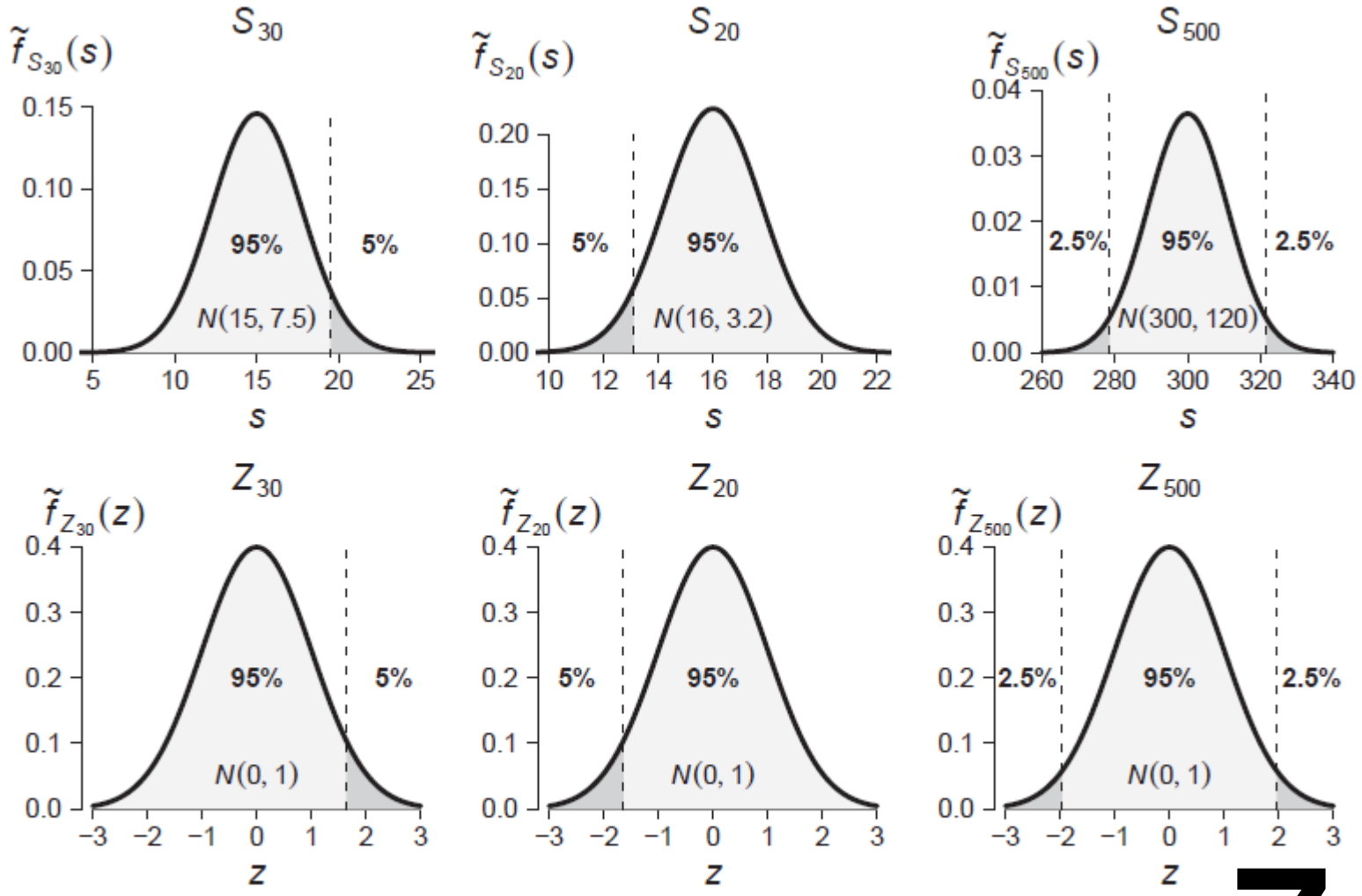


Abb. 11.1.4: Approximative Binomialtests – Testverteilungen

● Beispiel 11.1.4: Wirkung vs. keine Wirkung ●

Gegeben sei das Pflanzenexperiment Nr. 1 aus Beispiel 10.2.3 von Folie 702. Wie lässt sich hier statistisch prüfen und entscheiden, ob ein gemessener Unterschied zwischen den beiden Behandlungsgruppen signifikant oder möglicherweise nur Zufall ist? Ab wann kann man von einer systematischen Wirkung der Düngung ausgehen?

Statistisches Modell:

Y_{01}, Y_{02}, Y_{03} unabhängig $N(\mu_0, \sigma^2)$-verteilt, μ_0 und σ^2 unbekannt

Y_{11}, Y_{12}, Y_{13} unabhängig $N(\mu_1, \sigma^2)$-verteilt, μ_1 und σ^2 unbekannt

$Y_{01}, Y_{02}, Y_{03}, Y_{11}, Y_{12}, Y_{13}$ (vollständig) unabhängig

Entscheidungsproblem: Wirkung vs. keine Wirkung

Testproblem: $H_0\colon\ \mu_1 - \mu_0 = 0$ vs. $H_1\colon \mu_1 - \mu_0 \neq 0$

bzw. $H_0\colon\ \mu_0 = \mu_1$ vs. $H_1\colon \mu_0 \neq \mu_1$

11.1 Was versteht man unter einem Test?

Teststatistik: $\tilde{T}_D = \dfrac{\bar{Y}_1 - \bar{Y}_0}{\sqrt{S_p^2/3 + S_p^2/3}} \overset{\mu_0 = \mu_1}{\sim} t(4)$

Testniveau (Signifikanzniveau): $\alpha = 0.05$

Kritische Werte: $c_1 = t_{4,0.025} = -2.78$ und $c_2 = t_{4,0.975} = 2.78$

Testprozedur: Falls $\tilde{T}_D < -2.78$ oder $\tilde{T}_D > 2.78$, Entscheidung für H_1 (Verwerfung von H_0), sonst Beibehaltung von H_0

Testergebnis hier:

Mit $n_0 = 3, n_1 = 3, \bar{y}_1 - \bar{y}_0 = 30$ und $s_P^2 = 101$ erhalten wir

$$\tilde{t}_D = \frac{30}{\sqrt{101/3 + 101/3}} = 3.66 > 2.78$$

Damit kann H_0 verworfen werden. Bei einem Testniveau von 5% ist der Unterschied signifikant.

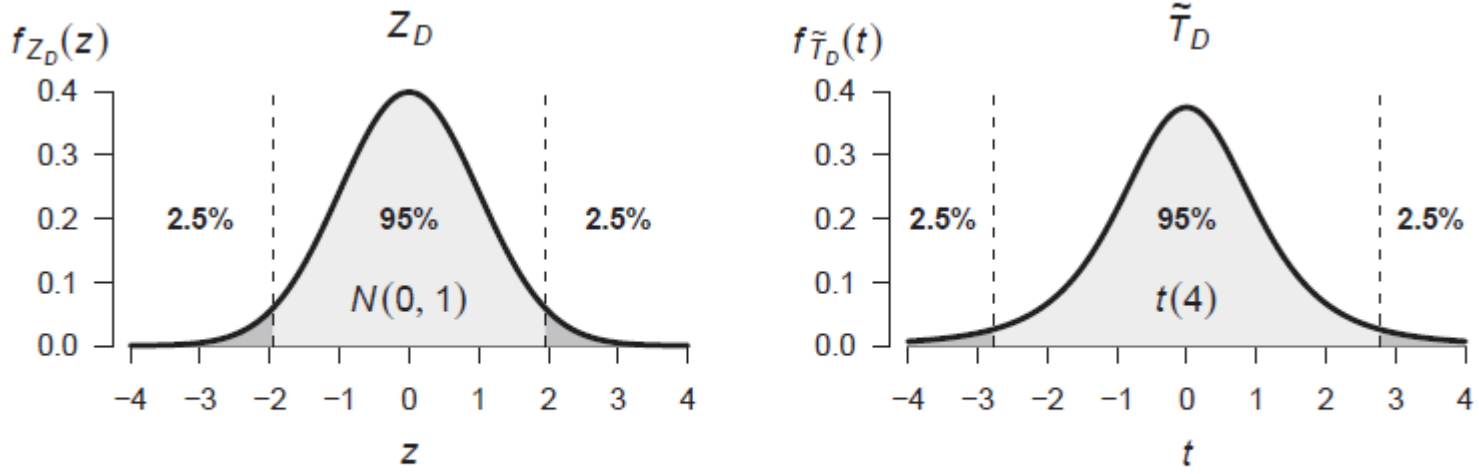


Abb. 11.1.5: Zweiseitiger Gauß-Test und zweiseitiger t-Test – Testverteilungen

Bemerkung:
Im Falle einer bekannten (homoskedastischen) Varianz σ^2
nutzt man als Statistik das Verteilungsresultat

$$Z_D = \frac{\bar{Y}_1 - \bar{Y}_0 - (\mu_1 - \mu_0)}{\sqrt{\sigma^2/3 + \sigma^2/3}} \quad \overset{\mu_0 = \mu_1}{\sim} \quad N(0,1)$$

und verwendet die Standardnormalverteilung als Testverteilung.

11.1.2 Grundstruktur und Durchführung

● Überblick ●

Die Grundstruktur eines statistischen Tests ist im Prinzip immer gleich. Im Folgenden soll dies anhand der Beispiele des vorhergehenden Abschnitts, hier als (B1)-(B4) bezeichnet, nochmals in kompakter Form verdeutlicht werden. Die allgemeine Grundstruktur kann in folgende 5 Bestandteile zerlegt werden:

- das statistische Entscheidungsproblem,
- das statistische Modell,
- die Teststatistik und die Testverteilung,
- das Testniveau und die Entscheidungsregel,
- die Testdurchführung und die Testentscheidung.

11.1 Was versteht man unter einem Test?

• Statistisches Entscheidungsproblem •

B1: Multiple Choice	Raten vs. Wissen
B2: Mensa-Befragung	Kein Rückgang vs. Rückgang
B3: Qualitätskontrolle	Norm vs. Abweichung
B4: Pflanzenexperiment	Keine Wirkung vs. Wirkung

• Statistisches Modell •

B1	Modellrahmen: $X_1, \ldots, X_{30}$ unabhängig $B(1, \pi)$-verteilt Testproblem: $H_0 : \pi = 0.5$ vs. $H_1 : \pi > 0.5$.
B2	Modellrahmen: $X_1, \ldots, X_{20}$ unabhängig $B(1, \pi)$-verteilt Testproblem: $H_0 : \pi \geq 0.8$ vs. $H_1 : \pi < 0.8$.
B3	Modellrahmen: $X_1, \ldots, X_{500}$ unabhängig $B(1, \pi)$-verteilt Testproblem: $H_0 : \pi = 0.6$ vs. $H_1 : \pi \neq 0.6$.
B4	Modellrahmen: Y_{01}, Y_{02}, Y_{03} unabhängig $N(\mu_0, \sigma^2)$-verteilt und Y_{11}, Y_{12}, Y_{13} unabhängig $N(\mu_1, \sigma^2)$-verteilt Testproblem: $H_0 : \mu_0 = \mu_1$ vs. $H_1 : \mu_0 \neq \mu_1$.

11.1 Was versteht man unter einem Test?

- **Teststatistik und Testverteilung**

B1:	$S_{30} = \sum_{i=1}^{30} X_i \sim B(30, 0.5)$ für $\pi = 0.5$.
B2:	$S_{20} = \sum_{i=1}^{20} X_i \sim B(20, 0.8)$ für $\pi = 0.8$.
B3:	$S_{500} = \sum_{i=1}^{500} X_i \sim B(500, 0.6)$ für $\pi = 0.6$.
B4:	$\tilde{T}_D = \frac{\overline{Y}_1 - \overline{Y}_0}{\sqrt{2S_p^2/3}} \sim t(4)$ für $\mu_0 = \mu_1$.

11.1 Was versteht man unter einem Test?

• Testniveau und Entscheidungsregel •

B1	Testniveau:	$\alpha = 0.049$.
	Entscheidungsregel:	Falls $S_{30} > 19$, verwerfe H_0,
		falls $S_{30} \leq 19$, behalte H_0 bei.
B2	Testniveau:	$\alpha = 0.032$.
	Entscheidungsregel:	Falls $S_{20} < 13$, verwerfe H_0,
		falls $S_{20} \geq 13$, behalte H_0 bei.
B3	Testniveau:	$\alpha = 0.049$.
	Entscheidungsregel:	Falls $S_{500} < 279$ oder $S_{500} > 321$, verwerfe H_0,
		falls $279 \leq S_{500} \leq 321$, behalte H_0 bei.
B4	Testniveau:	$\alpha = 0.05$.
	Entscheidungsregel:	Falls $\tilde{T}_D < -2.78$ oder $\tilde{T}_D > 2.78$, verwerfe H_0,
		falls $-2.78 \leq \tilde{T}_D \leq 2.78$, behalte H_0 bei.

11.1 Was versteht man unter einem Test?

- **Testdurchführung und Testentscheidung**

B1	Sofern von 30 Fragen mindestens 20 korrekt beantwortet werden, liegt bei einem Niveau von 4.9% signifikantes Wissen vor.
B2	Sofern weniger als 13 von 20 Befragten die Mensa präferieren, ist ihr Anteil bei einem Niveau von 3.2% signifikant geringer als 80%.
B3	Sofern mehr als 321 oder weniger als 279 von 500 Körnern schwarz sind, weicht dies bei einem Niveau von 4.9% signifikant vom Normanteil 0.6 ab.
B4	Sofern $\tilde{T}_D$ den Wert 2.78 betragsmäßig überschreitet, wirkt sich die Düngung bei einem Niveau von 5% signifikant auf das Wachstum aus.

Definition 11.1.1: Statistischer Test

Ein **statistischer Test** ist eine auf einer Stichprobe basierende Entscheidungsregel, die zugunsten einer der beiden Hypothesen eines Testproblems eine Entscheidung herbeiführt.

- **Deskriptiv vs. konfirmatorisch**

11.1.3 Zusammenhang zur Intervallschätzung

• Hintergrund •

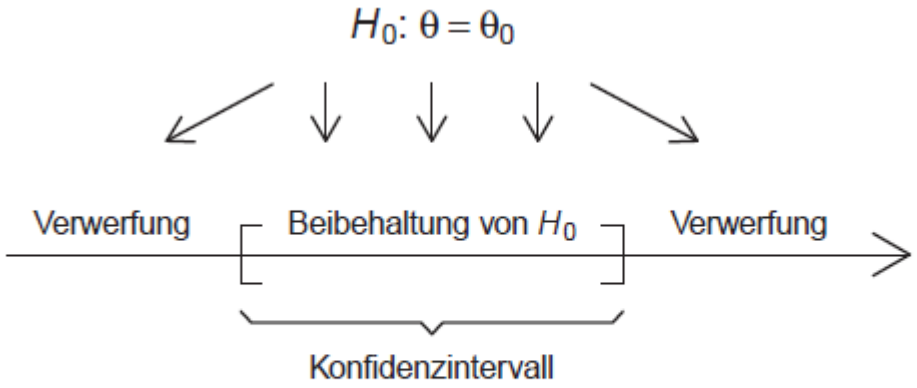


Abb. 11.1.6: Zusammenhang von Konfidenzintervallen und Testentscheidungen

• Beispiel 11.1.5 •

Betrachten wir nochmals das Pflanzenexperiment Nr. 1 aus Beispiel 10.2.3 und Beispiel 11.1.4 von Folie 702 bzw. 745.

0.95-KI für $\mu_1 - \mu_0$: $[7.2, 52.8]$

$\Rightarrow$ Bei einem Test zum Niveau 5% über das Testproblem

$$H_0\colon\ \mu_1 - \mu_0 = \delta_0 \quad \text{vs. } H_1\colon\ \mu_1 - \mu_0 \neq \delta_0$$

wird jedes $\delta_0 \notin [7.2, 52.8]$ verworfen und jedes $\delta_0 \in [7.2, 52.8]$ beibehalten. Man beachte, dass im vorliegenden Fall insbesondere $\delta_0 = 0$, d. h. $H_0\colon\ \mu_0 = \mu_1$ verworfen wird.

Satz 11.1.1: Konfidenzintervalle und Tests

Es sei $KI = [G_u, G_o]$ ein $(1 - \alpha)$-Konfidenzintervall für einen Parameter θ. Wir betrachten das zweiseitige Testproblem

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta \neq \theta_0. \qquad (*)$$

Einen α-Niveau-Test für (*) erhalten wir dann durch die Entscheidungsregel:

Falls $\theta_0 \notin KI$, wird H_0 verworfen,
falls $\theta_0 \in KI$, wird H_0 beibehalten.

11.2 Wichtige Aspekte beim Testen

11.2.1 Hypothesenwahl und Fehlerarten

• Arten von Testproblemen •

Definition 11.2.1: Typische parametrische Testprobleme

Sei θ ein Parameter. Dann bezeichnet man ein Testproblem der Form

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta \neq \theta_0$$

als **zweiseitiges Testproblem** und den dazu korrespondierenden Test als **zweiseitigen Test** („Test nach unten und oben"). Die Testprobleme

$$H_0 : \theta \geq \theta_0 \quad \text{vs.} \quad H_1 : \theta < \theta_0 \text{ und}$$
$$H_0 : \theta \leq \theta_0 \quad \text{vs.} \quad H_1 : \theta > \theta_0$$

werden als **einseitige Testprobleme** bezeichnet und die korrespondierenden Tests als **einseitiger unterer Test** („Test nach unten") und **einseitiger oberer Test** („Test nach oben").

Besteht H_0 (H_1) nur aus einem einzelnen Wert (wie etwa bei einem zweiseitigen Test), so liegt eine **einfache Nullhypothese** (Alternative) vor. Umfasst H_0 (H_1) hingegen mehrere Werte, so spricht man von einer **zusammengesetzten Nullypothese** (Alternative).

11.2 Wichtige Aspekte beim Testen

• Asymmetrische Auslegung von Testentscheidungen •

Bei Testentscheidungen wird H_0 verworfen (abgelehnt) und H_1 angenommen, falls H_0 sehr unplausibel und H_1 sehr plausibel erscheint. Ist dies nicht der Fall, bleibt man bei H_0.

• Wahl der Hypothesen •

Nachzuweisendes wird in der Alternativhypothese formuliert, zu Widerlegendes hingegen in der Nullhypothese.

• Grenzen bei der Hypothesenwahl •

Sinnvolle Testverfahren für Problemvarianten wie etwa

$$H_0\colon\ \pi > 0.5 \ \text{ vs. } H_1\colon\ \pi = 0.5 \text{ oder}$$

$$H_0\colon\ \theta \neq \theta_0 \ \text{ vs. } H_1\colon \theta = \theta_0$$

nicht konstruierbar.

• Verhältnis zum Signifikanzbegriff •

Nur falls die Nullhypothese verworfen wird, gilt das in der Alternativ hypothese repräsentierte Ergebnis als signifikant. Bei Beibehaltung der Nullhypothese spricht man nicht von einem signifikanten Ergebnis. Dieser Umstand ist auf die zuvor beschriebene Asymmetrie zurückzuführen. Auf die dem Signifikanzbegriff innewohnende Deutungsproblematik gehen wir gesondert in Abschnitt 11.2.4 ein.

• Fehlerarten •

Tab. 11.2.1: Entscheidungsmatrix beim Testen

Wahrer Zustand \ Testentscheidung	H_0	H_1
H_0	richtig	Fehler 1.Art
H_1	Fehler 2.Art	richtig

11.2.2 Irrtumswahrscheinlichkeiten und Güte

- **Allgemeines Optimierungsbestreben**
- **Ein nichtstatistisches Beispiel zur Begriffserläuterung**

Tab. 11.2.2: Entscheidungsmatrix bei einem Strafprozess

Urteil / Angeklagter	unschuldig	schuldig
unschuldig	richtig	Fehler 1.Art
schuldig	Fehler 2.Art	richtig

Tab. 11.2.3: Fiktive Historie von 10 Strafprozessen mit jeweiligen Entscheidungen

Prozess-Nr.	1	2	3	4	5	6	7	8	9	10
Angeklagter	S	U	U	U	S	S	S	U	S	U
Endurteil	U	U	S	S	U	U	S	U	U	U

Fehler 1. Art: 2

Fehler 2. Art: 4

Richtige Entscheidungen: 4

11.2 Wichtige Aspekte beim Testen

● Globale Irrtumswahrscheinlichkeit ●

> Empirische Fehlerrate bei Strafprozesse: 60%

> Angabe einer globalen Fehlerwahrscheinlichkeit beim Testen möglich?

Bei einem klassischen statistischen Test kann keine globale Wahrscheinlichkeit für eine falsche (oder eine richtige) Entscheidung angegeben werden.

● Irrtumswahrscheinlichkeit 1. Art ●

Definition 11.2.2: Irrtumswahrscheinlichkeit 1. Art

Unter einer **Irrtumswahrscheinlichkeit 1. Art** (auch **Fehlerwahrscheinlichkeit 1. Art**) eines Tests versteht man eine Wahrscheinlichkeit, sich für H_1 zu entscheiden, falls ein bestimmter Wert aus H_0 zutrifft.

Im Beispiel wäre die (geschätzte) **Irrtumswahrscheinlichkeit 1. Art**

$$P(Urteil = S | Angeklagter = U) = 2/5 = 0.4$$

und nicht wie man etwa meinen könnte

$$P(Urteil = S, Angeklagter = U) = 2/10 = 0.2.$$

Letztere könnte man als **globale Irrtumswahrscheinlichkeit 1. Art** bezeichnen.

- **Irrtumswahrscheinlichkeit 2. Art** -

Definition 11.2.3: Irrtumswahrscheinlichkeit 2. Art

Unter einer **Irrtumswahrscheinlichkeit 2. Art** (auch **Fehlerwahrscheinlichkeit 2. Art**) eines Tests versteht man eine Wahrscheinlichkeit, H_0 beizubehalten, falls ein bestimmter Wert aus H_1 zutrifft.

Im Beispiel wäre die (geschätzte) **Irrtumswahrscheinlichkeit 2. Art**

$$P(Urteil = U | Angeklagter = S) = 4/5 = 0.8$$

und nicht wie man etwa meinen könnte

$$P(Urteil = U, Angeklagter = S) = 4/10 = 0.4.$$

Letztere könnte man als **globale Irrtumswahrscheinlichkeit 2. Art** bezeichnen.

11.2 Wichtige Aspekte beim Testen

● Quantifizierung von Irrtumswahrscheinlichkeiten ●

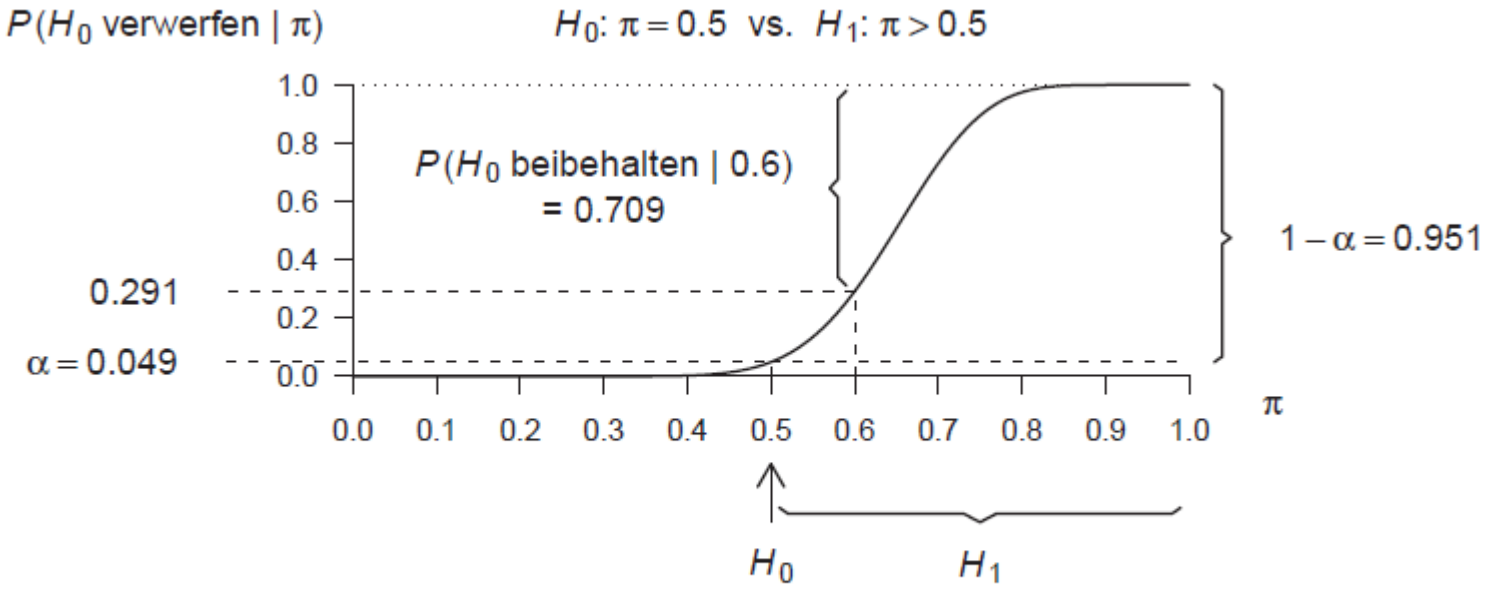


Abb. 11.2.1: Verwerfungswahrscheinlichkeiten der Nullhypothese in Beispiel 11.1.1

> Ablesebeispiele sofern $X \sim B(30, \pi)$:

- $P(Verwerfung\ von\ H_0 | \pi = 0.5) = P_{0.5}(X > 19) \approx 0.049$
  Dies ist eine Fehlerwahrscheinlichkeit 1. Art

11.2 Wichtige Aspekte beim Testen

- $P(Verwerfung\ von\ H_0|\pi = 0.6) = P_{0.6}(X > 19) \approx 0.291$

  Die Fehlerwahrscheinlichkeit 2. Art für $\pi = 0.6$: $1 - 0.291 = 0.709$
- $P(Verwerfung\ von\ H_0|\pi = 0.8) = P_{0.8}(X > 19) \approx 0.974$

  Die Fehlerwahrscheinlichkeit 2. Art für $\pi = 0.8$: $1 - 0.974 = 0.026$
- $P(Verwerfung\ von\ H_0|\pi) = P_\pi(X > 19) < 0.049$ für $\pi < 0.5$

  Dies wäre eine Fehlerwahrscheinlichkeit 1. Art beim Testproblem

  $H_0: \pi \leq 0.5$ vs. $H_1: \pi > 0.5$

> Das Testniveau α bildet eine Obergrenze für die Fehlerwahrscheinlichkeit 1. Art für alle Werte aus H_0 und wird an der Schwelle zwischen H_0 und H_1 exakt eingehalten

> Für Werte π aus H_1 ergeben sich Fehlerwahrscheinlichkeiten 2. Art als

$$P(Beibehaltung\ von\ H_0|\pi) = 1 - P(Verwerfung\ von\ H_0|\pi)$$

11.2 Wichtige Aspekte beim Testen

Definition 11.2.4: Niveau-α-Test

Es sei $\alpha \in (0, 1)$. Ein statistischer Test zu einem Testproblem über einen Parameter θ heißt **Niveau-α-Test** oder **Test zum Signifikanzniveau** α, wenn die Irrtumswahrscheinlichkeit 1. Art stets kleiner oder gleich α ist:

$$P(\text{Verwerfung von } H_0 | \theta) \leq \alpha \quad \text{für alle Werte } \theta \text{ aus } H_0.$$

> Ist die Verwerfungswahrscheinlichkeit bei einem **Niveau-α-Test** für Werte aus H_1 niemals geringer als für Werte aus H_0 (was ja sinnvoll ist), spricht man von einem **unverfälschten Test**

> Eine Obergrenze für die Fehlerwahrscheinlichkeit 2. Art bildet bei einem unverfälschten Test die Gegenwahrscheinlichkeit des Testniveaus:

$$P(Beibehaltung\ von\ H_0 | \pi) \leq 1 - \alpha$$

• Güte und Gütefunktion •

Definition 11.2.5: Gütefunktion und Güte

Gegeben sei ein parametrischer Test bezüglich eines Parameters θ. Dann bezeichnet man die Funktion g mit

$$g(\theta) = P(\text{Verwerfung von } H_0 | \theta)$$

als **Gütefunktion** des Tests. Für jeden Wert θ aus H_1 wird $g(\theta)$ als **Güte** oder **Macht** des Tests an der Stelle θ bezeichnet.

Irrtumswahrscheinlichkeit 2. Art = 1 – Güte

Beispiel: Abb. 11.2.1 (Folie 761)

$g(0.5) = 0.049$ oder $g(0.6) = 0.291$

• Gütefunktionen der Gauß-Test-Varianten •

Satz 11.2.1: Gütefunktionen des Gauß-Tests

Die Gütefunktion des Gauß-Tests ist im Falle des Testproblems

(a) $H_0 : \mu \geq \mu_0$ vs. $H_1 : \mu < \mu_0$ gegeben durch

$$g(\mu) = \Phi\Big(-z_{1-\alpha} - \frac{\mu - \mu_0}{\sqrt{\sigma^2/n}}\Big),$$

(b) $H_0 : \mu \leq \mu_0$ vs. $H_1 : \mu > \mu_0$ gegeben durch

$$g(\mu) = \Phi\Big(-z_{1-\alpha} + \frac{\mu - \mu_0}{\sqrt{\sigma^2/n}}\Big),$$

(c) $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$ gegeben durch

$$g(\mu) = \Phi\Big(-z_{1-\alpha/2} - \frac{\mu - \mu_0}{\sqrt{\sigma^2/n}}\Big) + \Phi\Big(-z_{1-\alpha/2} + \frac{\mu - \mu_0}{\sqrt{\sigma^2/n}}\Big).$$

Herleitung: Betrachte dazu exemplarisch nachfolgendes Beispiel 11.2.1 für Fall a)

• Beispiel 11.2.1: Gütefunktionen eines einseitigen Gauß-Tests •

Angenommen eine Maschine produziere Metallstifte der Länge 40 mm. Laut Herstellerangaben arbeite die Maschine mit einer Standardabweichung von $\sigma = 1\,mm$. Die Länge der Metallstifte kann dabei als normalverteilt angenommen werden. Im Rahmen der Qualitätskontrolle werden in regelmäßigen Abständen n Stifte entnommen und deren mittlere Länge festgestellt. Bei einer signifikanten Abweichung von der Solllänge, muss die Produktion auf jeden Fall kurzeitig unterbrochen werden, wobei irrtümliche Produktionsstopps möglichst zu vermeiden sind.

Die Situation lässt sich im Rahmen eines **Gauß-Tests** statistisch lösen, wobei die Wahl des konkreten Testproblems davon abhängt, in welche Richtung primär keine Abweichung auftreten darf. Wir nehmen im Folgenden an, dass primär zu kurze Stifte vermieden werden sollen. Dies führt zu einem „unteren Gauß-Test“ für das Testproblem:

$$H_0\colon \mu \geq 40 \quad \text{vs.}\ H_1\colon \mu < 40$$

> Unterer Gauß-Test

Statistisches Modell: $X_1, \ldots, X_n$ unabhängig $N(\mu, 1)$-verteilt

Testproblem: $H_0\colon\ \mu \geq 40$ vs. $H_1\colon \mu < 40$

Teststatistik: entweder $\bar{X} \overset{\mu = 40}{\sim} N(40, 1/n)$ (nichtstandardisiert)

oder $Z = \dfrac{\bar{X} - 40}{\sqrt{1/n}} \overset{\mu = 40}{\sim} N(0,1)$ (standardisiert)

Testniveau (Signifikanzniveau): α

Kritischer Wert: $c = q_\alpha = 40 + \sqrt{1/n} \times z_\alpha$

bzw. $c = z_\alpha$

Testprozedur: Falls $\bar{X}$ bzw. $Z < c$ Entscheidung für H_1 (Verwerfung von H_0, sonst Beibehaltung von H_0

> Herleitung der Gütefunktion des unteren Gauß-Tests für das Beispiel:

$$g(\mu) = P(H_0\ verwerfen \mid \mu)$$

$$= P(\bar{X} < q_\alpha \mid \mu) = P\left(\bar{X} < 40 + z_\alpha\sqrt{1/n} \mid \mu\right)$$

$$= P\left(\underbrace{\frac{\bar{X} - \mu}{\sqrt{1/n}}}_{N(0,1)} < \frac{40 - \mu + z_\alpha\sqrt{1/n}}{\sqrt{1/n}} \mid \mu\right)$$

$$= \Phi\left(z_\alpha + \frac{40 - \mu}{\sqrt{1/n}} \mid \mu\right) = \Phi\left(-z_{1-\alpha} - \frac{\mu - 40}{\sqrt{1/n}} \mid \mu\right)$$

> Verallgemeinerung davon ergibt Fall a) von Satz 11.2.1, Folie 765):

$$g(\mu) = \Phi\left(-z_{1-\alpha} - \frac{\mu - \mu_0}{\sqrt{\sigma^2/n}} \mid \mu\right)$$

11.2 Wichtige Aspekte beim Testen

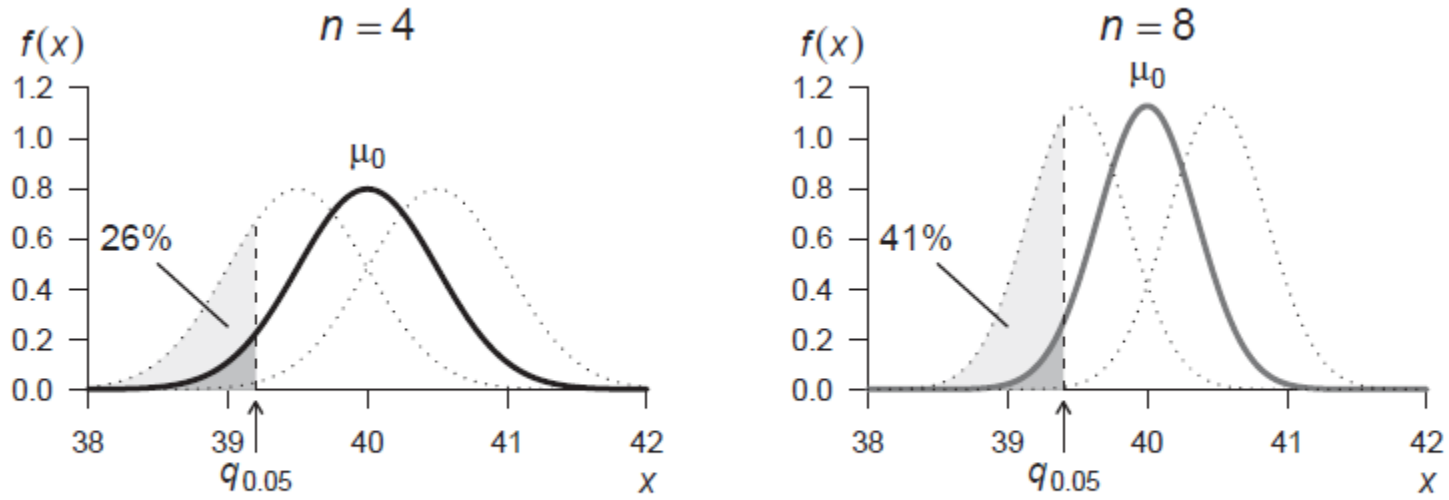


Abb. 11.2.2: Verteilung des Stichprobenmittels für verschiedene Konstellationen

> Beispielrechnungen: $\mu_0 = 40, \sigma^2 = 1, \alpha = 0.05$ und ...

- $n = 4$: $g(39.5) = \Phi\left(\underbrace{-1.64}_{-z_{0.95}} - \frac{39.5 - 40}{\sqrt{1/4}}\right) = \Phi(-0.64) = 0.26$

- $n = 8$: $g(39.5) = \Phi\left(-1.64 - \frac{39.5 - 40}{\sqrt{1/8}}\right) = \Phi(-0.23) = 0.41$

11.2 Wichtige Aspekte beim Testen

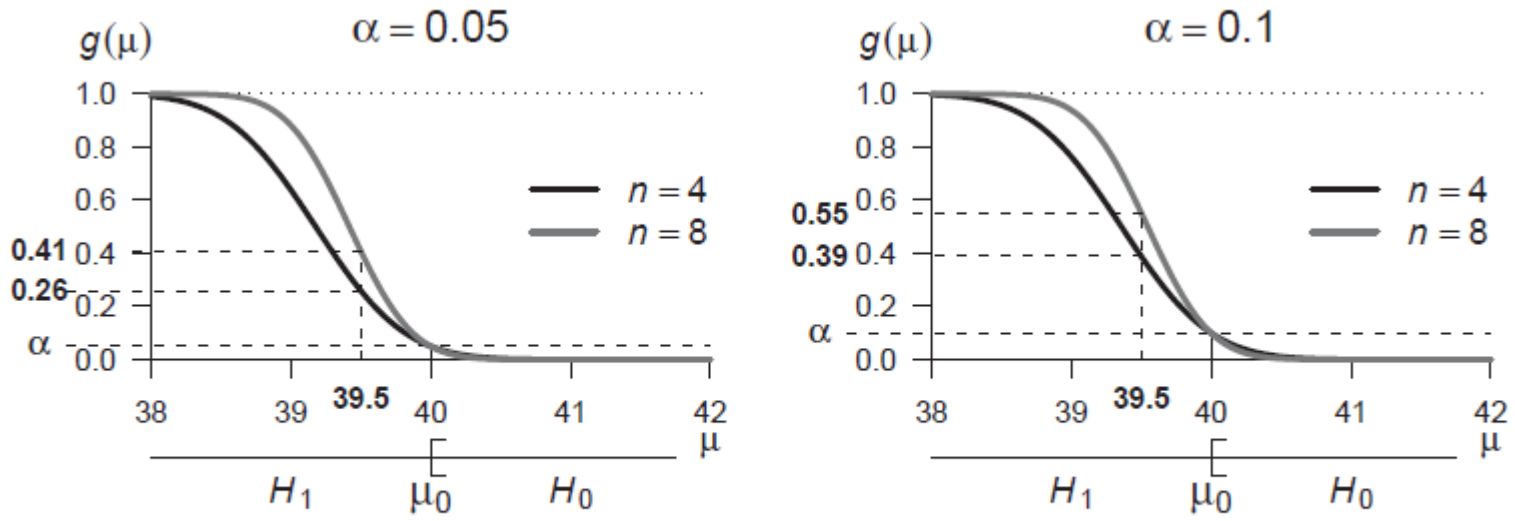


Abb. 11.2.3: Gütefunktionen eines einseitigen unteren Gauß-Tests

> Beispielrechnungen: $\mu_0 = 40, \sigma^2 = 1, \alpha = 0.1$ und ...

- $n = 4$: $g(39.5) = \Phi\left(\underbrace{-1.28}_{-z_{0.9}} - \frac{39.5 - 40}{\sqrt{1/4}}\right) = \Phi(-0.28) = 0.39$
- $n = 8$: $g(39.5) = \Phi\left(-1.28 - \frac{39.5 - 40}{\sqrt{1/8}}\right) = \Phi(0.13) = 0.55$

● Beispiel 11.2.2: Gütefunktionen eines zweiseitigen Gauß-Tests ●

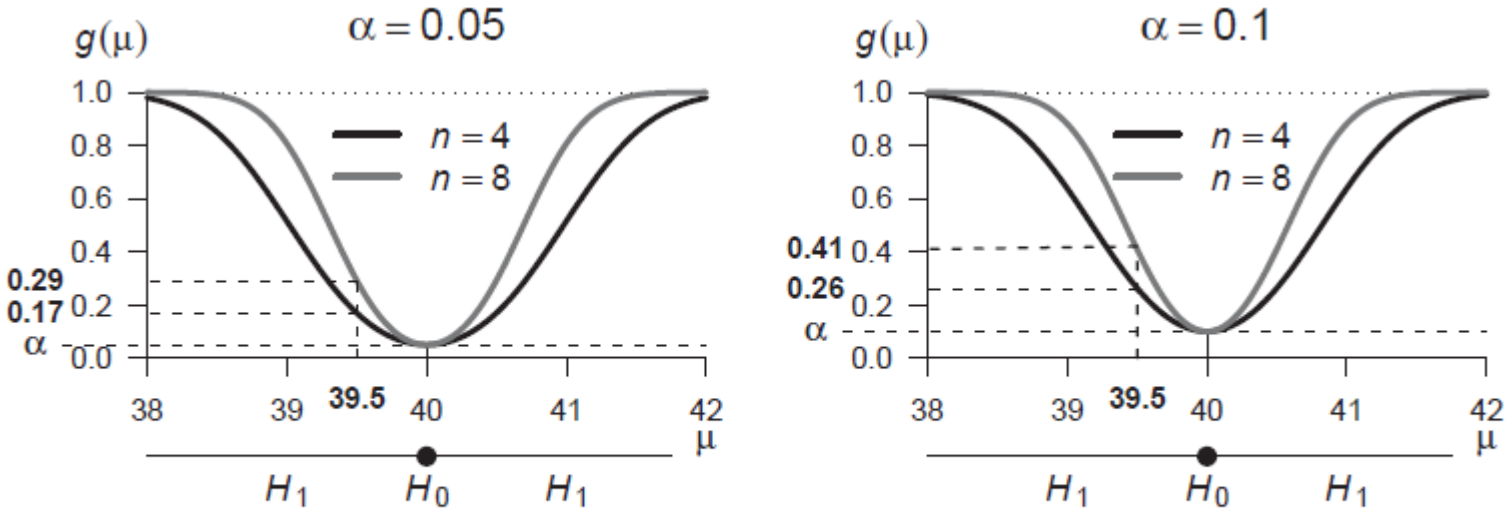


Abb. 11.2.4: Gütefunktionen eines zweiseitigen Gauß-Tests

> Gleicher Modellrahmen wie auf Folie 766, allerdings für das Testproblem

$$H_0\colon\ \mu = 40 \quad \text{vs.} \quad H_1\colon \mu \neq 40$$

> Gütefunktion des zweiseitigen Gauß-Tests ergibt Fall c) von Satz 11.2.1

> Angaben zur Güte in Abb. 11.2.4 basieren auf Formel für Fall c).

11.2 Wichtige Aspekte beim Testen

- **Eigenschaften von Gütefunktionen**

Eigenschaften von Gütefunktionen

Im Rahmen eines Testproblems bezüglich eines Parameters θ gelten für die Gütefunktion $g(\theta)$ im Allgemeinen folgende Eigenschaften:

(i) $g(\theta) \leq \alpha$ für alle Werte θ aus H_0.
(ii) Die Güte nimmt zu, „je weiter man in die Alternative vordringt".
(iii) Die Güte nimmt für wachsendes n zu, d.h. die Fehlerwahrscheinlichkeit 2. Art nimmt für wachsendes n ab.
(iv) Die Güte nimmt für wachsendes α zu, d.h. die Fehlerwahrscheinlichkeit 2. Art nimmt für wachsendes α ab.

- **Praktische Implikation zur Fehlervermeidung**

Wähle α möglichst klein und n möglichst groß!

- **Optimale Tests**

11.2.3 *p*-Werte

● Hintergrund und Definition ●

Definition 11.2.6: *p*-Wert

Der **p-Wert** ist das ex post ermittelte Testniveau, zu dem H_0 für die vorliegenden Daten gerade noch verworfen werden könnte.

Bei stetigen Testverteilungen: „gerade noch/nicht mehr (vgl. Beispiel 11.2.4)

● *p*-Werte als Ausgabewerte statistischer Software ●

● Beispiel 11.2.3: *p*-Wert bei diskreter Testverteilung ●

> Betrachten wir nochmals die Situation von Beispiel 11.1.1 (Raten vs. Wissen) von Folie 734 mit dem Testproblem

$$H_0\colon\ \pi = 0.5 \quad \text{vs.} \quad H_1\colon \pi > 0.5$$

Angenommen, der Wert der Teststatistik lautet $s_{30} = 19$. Um gerade noch verwerfen zu können, müsste der kritische Wert $c = 18$ lauten, da 19 gerade noch größer als 18 ist.

Da für $X \sim B(30,0.5)$ gilt $P(X > 18) = 1 - 0.9 = 0.1$ (Tab. 11.2.4) entspräche dies einem Test zum Niveau 0.1. Somit ist der p-Wert 0.1.
Analog:

$$s_{30} = 16 \quad \Rightarrow \quad p\text{-Wert:} \quad 1 - 0.572 = 0.428$$

$$s_{30} = 20 \quad \Rightarrow \quad p\text{-Wert:} \quad 1 - 0.951 = 0.049$$

> Wir betrachten (ungeachtet der Sinnhaftigkeit) nun das Testproblem

$$H_0\colon\ \pi = 0.5 \ \text{ vs. } \ H_1\colon \pi < 0.5$$

In diesem Fall würde man nun „bei kleinen Werten“ verwerfen, d. h. wenn bestimmte kritische Werte unterschritten werden.
Angenommen, der Wert der Teststatistik lautet $s_{30} = 14$. Um gerade noch verwerfen zu können, müsste der kritische Wert $c = 15$ lauten, da 14 gerade noch kleiner als 15 ist.
Da für $X \sim B(30,0.5)$ gilt $P(X < 15) = P(X \leq 14) = 0.428$ (Tab. 11.2.4) entspräche dies einem Test zum Niveau 0.428. Somit ist der p-Wert 0.428.

11.2 Wichtige Aspekte beim Testen

Analog:

$s_{30} = 16 \quad \Rightarrow \quad p$-Wert: 0.708

$s_{30} = 20 \quad \Rightarrow \quad p$-Wert: 0.979

Tab. 11.2.4: Verteilungsfunktion der $B(30, 0.5)$-Verteilung – auszugsweise

s	14	15	16	17	18	19	20
$P(S_{30} \leq s)$	0.428	0.572	0.708	0.819	0.900	0.951	0.979

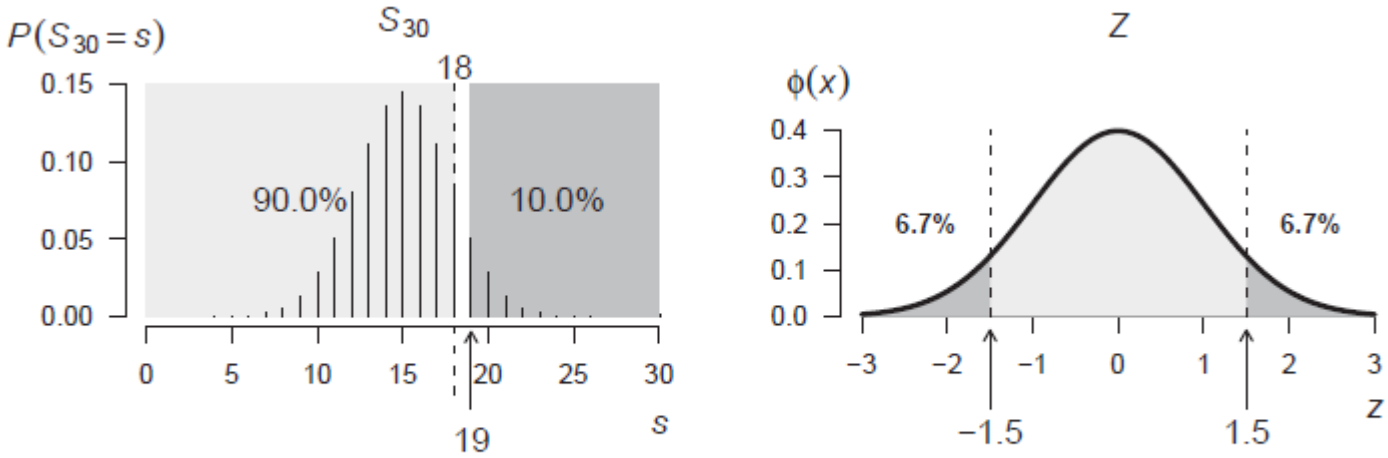


Abb. 11.2.5: Ermittlung von p-Werten in den Beispielen 11.2.3 und 11.2.4

● Beispiel 11.2.4: *p*-Wert bei stetiger Testverteilung ●

> Gegeben sei ein zweiseitiger Gauß-Test für das Testproblem

$$H_0\colon\ \mu = \mu_0 \ \text{ vs. } H_1\colon\ \mu \neq \mu_0$$

Beispielsweise könnte eine Situation wie in Beispiel 11.2.1 (Folie 766), wobei Abweichungen weder nach oben noch nach unten hin vermieden werden sollen.

Die Teststatistik der standardisierten Variante lautet dann

$$Z = \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}} \overset{\mu = \mu_0}{\sim} N(0,1)$$

Die kritischen Werte beispielsweise zum Niveau $\alpha = 0.05$ wären dann

$$c_1 = -z_{0.975} = -1.96 \ \text{ und } \ c_2 = z_{0.975} = 1.96$$

Angenommen, aus den Daten heraus ergibt sich ein Wert von

$$z = 1.5 \qquad \text{(keine Verwerfung)}$$

Dann ist der *p*-Wert $= 0.134$, da für $Z \sim N(0,1)$ gilt:

$$P(|Z| > 1.5) = P(Z < -1.5) + P(Z > 1.5)$$
$$= 2 \times 0.067 = 0.134$$

> Betrachtet man den „unteren Gauß-Test“ mit dem Testproblem

$$H_0\colon\ \mu \geq \mu_0 \ \text{ vs. } H_1\colon\ \mu < \mu_0$$

so würde man nur „bei kleinen Werten“ verwerfen, d. h. wenn bestimmte kritische Werte unterschritten werden.

Mit $z = 1.5$ ergibt sich dann ein p-Wert von 0.933, da

$$P(Z < 1.5) = 0.933.$$

> Betrachtet man den „oberen Gauß-Test“ mit dem Testproblem

$$H_0\colon\ \mu \leq \mu_0 \ \text{ vs. } H_1\colon\ \mu > \mu_0$$

so würde man nur „bei großen Werten“ verwerfen, d. h. wenn bestimmte kritische Werte überrschritten werden.

Mit $z = 1.5$ ergibt sich dann ein p-Wert von 0.067, da

$$P(Z > 1.5) = 0.067.$$

Berechnung von p-Werten

Sei t die Realisierung der Teststatistik eines Tests und die Zufallsvariable T entsprechend der Testverteilung des Tests verteilt. Für die in diesem Lehrbuch besprochenen Tests gelten die folgenden Berechnungsformeln: Ist der Test ein

a) unterer einseitiger Test, der für kleine Werte der Teststatistik ablehnt, dann berechnet sich der p-Wert als $P(T \leq t)$.
b) oberer einseitiger Test, der für große Werte der Teststatistik ablehnt, dann berechnet sich der p-Wert als $P(T \geq t)$.
c) zweiseitiger Test, der für betragsmäßig große Werte der Teststatistik ablehnt, und die Testverteilung symmetrisch um 0, dann berechnet sich der p-Wert als $P(|T| \geq |t|)$.

• Sachgerechte Verwendung von *p*-Werten •

p-Werte dürfen niemals zu einer nachträglichen Anpassung eines Testniveaus dienen (etwa um signifikante Resultate zu erzielen).

11.2.4 Signifikanz vs. Relevanz

● Deutung von Signifikanz ●

Aus Ausgangsfragen wie...

„*Ist das tatsächlich oder nur zufällig größer als...* ?“

„*Ist das tatsächlich oder nur zufällig kleiner als...* ?“

„*Ist das tatsächlich oder nur zufällig verschieden*?“

werden bei Verwerfungen von H_0 Aussagen wie...

„*Der Wert ist signifikant größer* als...“

„*Der Wert ist signifikant kleiner als...*“

„*Die beiden Werte sind signifikant verschieden.*“

● Übliche Signifikanzniveaus ●

> Verbreiteter Standard: 5%

> Je nach Kontext auch deutlich kleinere Niveaus...

11.2 Wichtige Aspekte beim Testen

- **Güte und Signifikanz für großes *n***

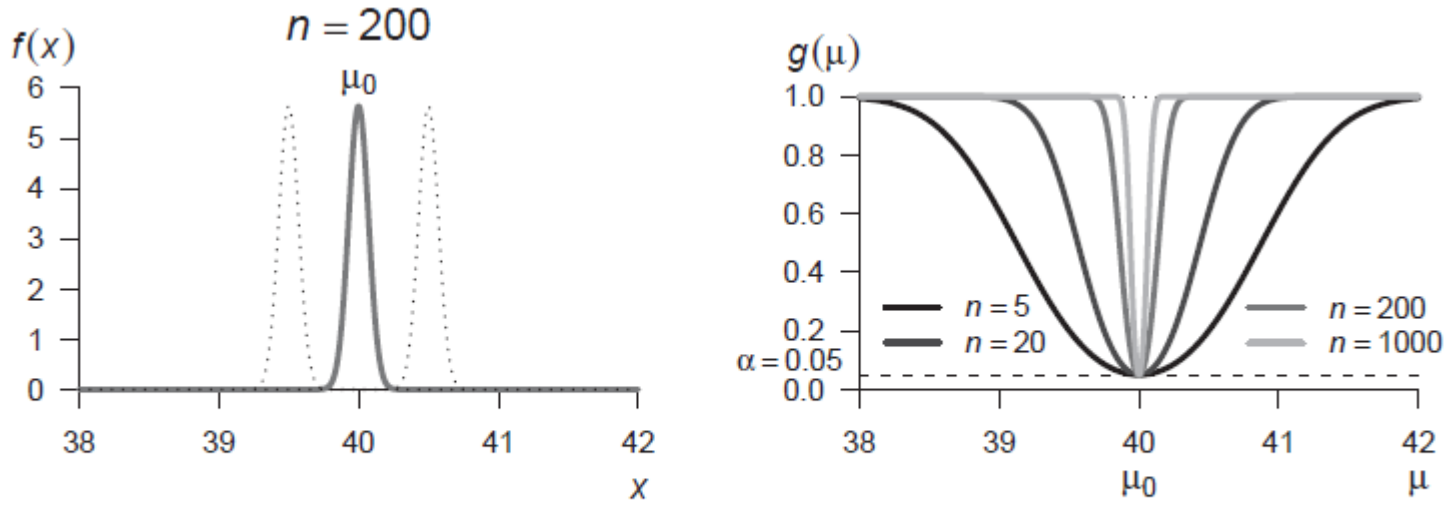


Abb. 11.2.6: Güte und Signifikanz für wachsendes *n*

● Signifikanz ≠ Relevanz ●

> Signifikanz ist nicht gleich Relevanz! Ein signifikanter Befund ist lediglich ein als nicht zufällig erachteter.

Jede noch so kleine Abweichung von der Nullhypothese kann zu einer signifikanten Abweichung werden, sofern die Stichprobe nur groß genug gewählt wird. Zumindest ist dann die Wahrscheinlichkeit einer Verwerfung von H_0 sehr groß. In dem im vorhergehenden Punkt verwendeten Beispiel etwa beträgt die Güte an der Stelle 40.1 für $n = 1000$ bereits 88.5%. Somit wird eine Abweichung von nur 0.1 mm vom Sollwert bereits mit einer Wahrscheinlichkeit von fast 90% entdeckt und als signifikant eingestuft. Es sollte damit klar werden, dass das Adjektiv „signifikant“ eine völlig andere Bedeutung hat als „groß“, „bedeutsam“ oder „relevant“.

11.3.1 Tests über Erwartungswerte

- **Modellrahmen und Überblick**
- **Tests über μ bei Normalverteilung und bekannter Varianz**

... führt zur Klasse der Gauß-Tests: Satz 11.3.1 Fall (i) (Folie 787)

- **Beispiel 11.3.1: Gauß-Tests über μ**

Man beachte hierzu Beispiel 11.2.1 auf Folie 766-767

- **Tests über μ bei Normalverteilung und unbekannter Varianz**

... führt zur Klasse der *t*-Tests: Satz 11.3.1 Fall (ii) (Folie 787)

- **Beispiel 11.3.2: *t*-Tests über μ**

> Gegeben seien die folgenden Beobachtungswerte:

7.8, 10.1, 9.0, 8.0, 11.6, 10.7, 8.1, 8.6, 9.4, 11.9

mit $\bar{x} = 9.52$ und $s^2 = 2.24$.

> Frage 1: Ist der Mittelwert signifikant kleiner als 10?

Statistisches Modell: $X_1, \ldots, X_{10}$ unabhängig $N(\mu, \sigma^2)$-verteilt

Testproblem: $H_0: \mu \geq 10$ vs. $H_1: \mu < 10$

Teststatistik: $T = \dfrac{\bar{X} - 10}{\sqrt{S^2/10}} \overset{\mu = 10}{\sim} t(9)$

Testniveau (Signifikanzniveau): $\alpha = 0.05$

Kritischer Wert: $c = t_{9,0.05} = -t_{9,0.95} = -1.83$

Testprozedur: Falls $t < -1.83$, Entscheidung für H_1, sonst Beibehaltung von H_0

Testergebnis hier:

$$t = \frac{9.52 - 10}{\sqrt{2.24/10}} = -1.01 > -1.83 \quad \Rightarrow \quad H_0$$

$\Rightarrow$ Mittelwert ist nicht signifikant kleiner als 10.

p-Wert: Abschätzung mittels Quantilstabelle (Folie 690-691)

$$0.1 < p - Wert < 0.2$$

Exakter Wert: 0.169 (Berechnung etwa mit R)

> Frage 2: Ist der Mittelwert signifikant von 10 verschieden?

Statistisches Modell: $X_1, \dots, X_{10}$ unabhängig $N(\mu, \sigma^2)$-verteilt

Testproblem: $H_0: \mu = 10$ vs. $H_1: \mu \neq 10$

Teststatistik: $$T = \frac{\bar{X} - 10}{\sqrt{S^2/10}} \overset{\mu = 10}{\sim} t(9)$$

Testniveau (Signifikanzniveau): $\alpha = 0.05$

Kritische Werte: $c_1 = t_{9,0.025} = -2.26, c_2 = t_{9,0.975} = 2.26$

Testprozedur: Falls $t < -2.26$ oder $t > 2.26$, Entscheidung für H_1, sonst Beibehaltung von H_0

Testergebnis hier:

$$t = \frac{9.52 - 10}{\sqrt{2.24/10}} = -1.01 \quad \Rightarrow \quad H_0$$

$\Rightarrow$ Mittelwert ist nicht signifikant von 10 verschieden.

p-Wert: Abschätzung mittels Quantilstabelle (Folie 690-691)

$$0.2 < p - Wert < 0.4$$

Exakter Wert: 0.339 (das Doppelte wie zuvor)

> Bemerkung: Das 0.95-KI für μ lautete (vgl. Folie 686):

$$[8.45, 10.59]$$

Dieses enthält den Wert 10.

11.3 Ausgewählte Testverfahren

• Tests über μ bei beliebiger Ausgangsverteilung •

> Statistisches Modell: $X_1, \ldots, X_n$ sind u. i. v. mit großem n

> Testprobleme und Durchführung wie beim Gauß-Test wg.

$$Z = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \overset{a}{\sim} N(0,1) \quad \text{bzw.} \quad T = \frac{\bar{X} - \mu}{\sqrt{S^2/n}} \overset{a}{\sim} N(0,1)$$

• Zusammenfassung •

Satz 11.3.1: Tests über Erwartungswerte

Seien $X_1, \ldots, X_n$ unabhängig und identisch verteilt mit Erwartungswert μ und Varianz $0 < \sigma^2 < \infty$. Gegeben seien folgende Testprobleme über μ:

a) $H_0 : \mu \geq \mu_0$ vs. $H_1 : \mu < \mu_0$,
b) $H_0 : \mu \leq \mu_0$ vs. $H_1 : \mu > \mu_0$,
c) $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$.

Dann basieren die Entscheidungsregeln eines Tests zum Niveau α, falls

(i) $X_1, \ldots, X_n$ normalverteilt sind mit bekannter Varianz, auf dem Resultat

$$Z = \frac{\overline{X} - \mu_0}{\sqrt{\sigma^2/n}} \sim N(0,1) \quad \text{für } \mu = \mu_0$$

und lauten: Verwerfe H_0 in

a), falls $z < -z_{1-\alpha}$,

b), falls $z > z_{1-\alpha}$,

c), falls $|z| > z_{1-\alpha/2}$, d.h. $z < -z_{1-\alpha/2}$ oder $z > z_{1-\alpha/2}$.

(ii) $X_1, \ldots, X_n$ normalverteilt sind mit unbekannter Varianz, auf dem Resultat

$$T = \frac{\overline{X} - \mu_0}{\sqrt{S^2/n}} \sim t(n-1) \quad \text{für } \mu = \mu_0$$

und lauten: Verwerfe H_0 in

a), falls $t < -t_{n-1,1-\alpha}$,

b), falls $t > t_{n-1,1-\alpha}$,

c), falls $|t| > t_{n-1,1-\alpha/2}$, d.h. $t < -t_{n-1,1-\alpha/2}$ oder $t > t_{n-1,1-\alpha/2}$.

(iii) n groß und die Varianz bekannt ist, auf dem Resultat

$$Z \overset{a}{\sim} N(0,1) \quad \text{für } \mu = \mu_0$$

und sind identisch zu denen in (i),

(iv) n groß und die Varianz unbekannt ist, auf dem Resultat

$$T \overset{a}{\sim} N(0,1) \quad \text{für } \mu = \mu_0$$

und sind identisch zu denen in (i), wobei mit z durch t zu ersetzen ist.

Die Tests in (i) werden als **Gauß-Tests** und die Tests in (ii) als **t-Tests** bezeichnet. In den Fällen (iii) und (iv) handelt es sich um **approximative Gauß-Tests**, die man für $n \geq 30$ anwenden kann. Anstelle der korrigierten Stichprobenvarianz kann in Fall (iv) auch die nichtkorrigierte verwendet werden.

- **Allgemeine Regel via Standardfehler** -

(Schätzwert – Hypothetischer Wert) / (geschätzter) Standardfehler.

11.3.2 Tests über Erwartungswertdifferenzen

- **Modellrahmen und Überblick**

  Unter den gleichen Rahmenbedingungen wie in Abschnitt 10.2.3 (Konfidenzintervalle für Erwartungswertdifferenzen lassen sich Testverfahren für solche konstruieren.

- **Tests über $\mu_1 - \mu_0$ bei Normalverteilung und bekannten Varianzen**

- **Tests über $\mu_1 - \mu_0$ bei Normalverteilung und unbekannten Varianzen**

- **Beispiele**

  Beachte hierzu Beispiel 11.1.4 (Pflanzenexperiment) auf Folien 745-747

- **Tests über $\mu_1 - \mu_0$ bei beliebiger Ausgangsverteilung**

- **Tests über $\mu_1 - \mu_0$ bei Abhängigkeit in Form verbundener Werte**

  Führe Tests in Bezug auf den Erwartungswert der Differenzen durch. Beispiel im LB, S. 615.

• Zusammenfassung •

Satz 11.3.2: Tests über Erwartungswertdifferenzen

Gegeben seien $n = n_0 + n_1$ unabhängige Stichprobenvariablen

$$Y_{01}, Y_{02}, \ldots, Y_{0n_0} \quad \text{und} \quad Y_{11}, Y_{12}, \ldots, Y_{1n_1},$$

die innerhalb der beiden Gruppen jeweils identisch verteilt sind mit

$$E(Y_{0i}) = \mu_0, \; Var(Y_{0i}) = \sigma_0^2 \quad \text{mit } 0 < \sigma_0^2 < \infty \text{ für } i = 1, \ldots, n_0,$$
$$E(Y_{1i}) = \mu_1, \; Var(Y_{1i}) = \sigma_1^2 \quad \text{mit } 0 < \sigma_1^2 < \infty \text{ für } i = 1, \ldots, n_1.$$

Gegeben seien folgende Testprobleme über $\mu_1 - \mu_0$:

a) $H_0 : \mu_1 - \mu_0 \geq \delta_0$ vs. $H_1 : \mu_1 - \mu_0 < \delta_0$,
b) $H_0 : \mu_1 - \mu_0 \leq \delta_0$ vs. $H_1 : \mu_1 - \mu_0 > \delta_0$,
c) $H_0 : \mu_1 - \mu_0 = \delta_0$ vs. $H_1 : \mu_1 - \mu_0 \neq \delta_0$.

Dann basieren die Entscheidungsregeln eines Tests zum Niveau α, falls

11.3 Ausgewählte Testverfahren

(i) alle Variablen normalverteilt sind mit bekannten Varianzen, auf dem Resultat

$$Z_D = \frac{\overline{Y}_1 - \overline{Y}_0 - \delta_0}{\sqrt{\sigma_0^2/n_0 + \sigma_1^2/n_1}} \sim N(0,1) \quad \text{für } \mu_1 - \mu_0 = \delta_0$$

und lauten: Verwerfe H_0 in

a), falls $z_D < -z_{1-\alpha}$,

b), falls $z_D > z_{1-\alpha}$,

c), falls $|z_D| > z_{1-\alpha/2}$, d.h. $z_D < -z_{1-\alpha/2}$ oder $z_D > z_{1-\alpha/2}$.

(ii) alle Variablen normalverteilt sind mit unbekannter, homoskedastischer Varianz ($\sigma_0^2 = \sigma_1^2$), auf dem Resultat

$$\tilde{T}_D = \frac{\overline{Y}_1 - \overline{Y}_0 - \delta_0}{\sqrt{S_p^2/n_0 + S_p^2/n_1}} \sim t(n_0 + n_1 - 2) \quad \text{für } \mu_1 - \mu_0 = \delta_0,$$

$$\text{mit } S_p^2 = \frac{1}{n_0 + n_1 - 2}(n_0\tilde{S}_0^2 + n_1\tilde{S}_1^2) = \frac{1}{n_0 + n_1 - 2}((n_0 - 1)S_0^2 + (n_1 - 1)S_1^2),$$

und lauten: Verwerfe H_0 in

a), falls $\tilde{t}_D < -t_{n_0+n_1-2,1-\alpha}$,

b), falls $\tilde{t}_D > t_{n_0+n_1-2,1-\alpha}$,

c), falls $|\tilde{t}_D| > t_{n_0+n_1-2,1-\alpha/2}$.

(iii) n_0 und n_1 groß sind, Annahme (B) von Definition 8.2.1 und die Bedingung

$$n_1/n \xrightarrow{n\to\infty} \pi \in (0,1)$$

gelten und die Varianzen bekannt sind, auf dem Resultat

$$Z_D \overset{a}{\sim} N(0,1) \quad \text{für } \mu_1 - \mu_0 = \delta_0$$

und sind identisch, wie die in (i).

(iv) die Varianzen unbekannt und sonst alle Annahmen von (iii) erfüllt sind, auf dem Resultat

$$\widehat{T}_D = \frac{\overline{Y}_1 - \overline{Y}_0 - \delta_0}{\sqrt{\widetilde{S}_0^2/n_0 + \widetilde{S}_1^2/n_1}} \overset{a}{\sim} N(0,1) \quad \text{für } \mu_1 - \mu_0 = \delta_0$$

und lauten wie in (i), wobei z_D durch $\widehat{t}_D$ zu ersetzen ist.

Die Tests in (i) werden hier als **Gauß-Tests über Erwartungswertdifferenzen**, die Tests in (ii) als t-**Tests über Erwartungswertdifferenzen** bezeichnet. In den Fällen (iii) und (iv) handelt es sich um **approximative Gauß-Tests über Erwartungswertdifferenzen**, die für $n_0, n_1 \geq 30$ anwendbar sind. Hierzu beachte man, dass Annahme (B) erfüllt ist, sofern die Trägermengen der Verteilungen der beiden Gruppen beschränkt sind. Außerdem können in Fall (iv) sowohl nichtkorrigierte als auch korrigierte Stichprobenvarianzen verwendet werden.

11.3 Ausgewählte Testverfahren

- **Bemerkung zu anderen Fällen**
- **Bemerkung zum Zweistichproben-Gedanken**
- **Äquivalente Darstellungen im Regressionsmodell**

11.3.3 Nichtparametrische χ^2-Tests

Allgemeiner Überblick

χ^2-Anpassungstest

● **Verteilungstheoretische Grundlagen** ●

> X diskret verteilt mit

$$P(X = i) = \pi_i \quad \text{für } i = 1, \dots, k.$$

> Stichprobe: $X_1, \dots, X_n$ unabhängig und identisch verteilt wie X

> N_i ... absolute Häufigkeit (stochastisch) der Ausprägung i in Stichprobe

$$N_i = \sum_{j=1}^{n} I_{\{i\}}(X_j) \quad \text{mit} \quad I_{\{i\}}(X_j) = \begin{cases} 1, & falls\ X_j = i \\ 0, & sonst. \end{cases}$$

> Es lässt sich dann (mit einigem Aufwand) zeigen, dass für großes n gilt:

$$\chi_A^2 = \sum_{i=1}^{k} \frac{(N_i - n\pi_i)^2}{n\pi_i} \overset{a}{\sim} \chi^2(k-1)$$

- **Approximationsregel** -

Unsere Konvention: $n\pi_i \geq 5$ für $i = 1, \ldots, k$

- **Beispiel 11.3.3: Unterschied von Umfragewerten und Wahlergebnissen** -

Tab. 10.2.1: Umfrageergebnis und Wahlergebnis (in %) der Bundestagswahl 2013

Partei	Union	SPD	FDP	Linke	Grüne	AfD	Sonst.
Umfrage und	40.0	27.0	5.5	8.5	9.0	4.0	6.0
Genauigkeit	±2.6%	±2.4%	±1.2%	±1.5%	±1.5%	±1.0%	±1.3%
Endergebnis	41.5	25.7	4.8	8.6	8.4	4.7	6.3

Daten: Der Bundeswahlleiter [2016], Forschungsgruppe Wahlen [2016]

11.3 Ausgewählte Testverfahren

> X ... Partei mit $X = 1$ für Union, $X = 2$ für SPD, ..., $X = 7$ für Sonstige

> f_i ... relativer Stimmenanteil von Partei i in der Umfrage ($i = 1, \ldots, 7$)

> n_i ... absolute Stimmenzahl von Partei i in der Umfrage

> π_i ... relativer Stimmenanteil von Partei i bei der Wahl

> $n\pi_i$ ... zu erwartende absolute Stimmenzahl für Partei i, falls die Wahlpräferenz zum Zeitpunkt der Umfrage genau gleich gewesen wäre wie am eigentlichen Wahltag

Tab. 11.3.1: Beobachtete vs. erwartete Häufigkeiten in Beispiel 11.3.3

Ausprägung i	1	2	3	4	5	6	7
f_i	0.400	0.270	0.055	0.085	0.09	0.040	0.060
$n_i = n \cdot f_i$	548	370	75	116	123	55	82
π_i	0.415	0.257	0.048	0.086	0.084	0.047	0.063
$n\pi_i$	568.1	351.8	65.7	117.7	115.0	64.3	86.2

> Statistisches Entscheidungsproblem:
Unterscheiden sich Umfrage- und Wahlergebnis signifikant?

Testproblem:

$$H_0\colon\ P(X = i) = \pi_i \ \text{ für } \ i = 1, \ldots, 7 \ \text{ vs.}$$

$$H_1\colon\ P(X = i) \neq \pi_i \ \text{ für mindestens ein } i$$

Teststatistik: $$\chi_A^2 = \sum_{i=1}^{7} \frac{(N_i - n\pi_i)^2}{n\pi_i} \overset{a}{\sim} \chi^2(6)$$

Testniveau (Signifikanzniveau): $\alpha = 0.05$

Kritischer Wert: $c = \chi^2_{6,0.95} = 12.592$

Testprozedur: Falls $\chi_A^2 > c$, Entscheidung für H_1, sonst Beibehaltung von H_0

Testergebnis hier:

$$\chi_A^2 = \sum_{i=1}^{7} \frac{(n_i - n\pi_i)^2}{n\pi_i}$$

$$= \frac{(548 - 568.1)^2}{568.1} + \frac{(370 - 351.8)^2}{351.8} + \cdots + \frac{(82 - 86.2)^2}{86.2}$$

$$= 5.10 < c$$

⇒ Bleibe bei H_0, d. h. Umfrage- und Wahlergebnis unterscheiden sich nicht signifikant.

• Beispiel 11.3.4: Überprüfung einer hypothetischen Normalverteilung •

Gegeben sei die Situation von Beispiel 11.2.1 (Folie 766) mit folgenden 50 Realisationen einer Stichprobe:

37.4, 38.1, 38.2, 38.3, 38.4, 38.5, 38.5, 38.5, 38.6, 38.6,
38.7, 38.8, 38.8, 39.0, 39.1, 39.2, 39.4, 39.4, 39.5, 39.5,
39.5, 39.6, 39.6, 39.8, 39.8, 39.8, 39.8, 39.9, 39.9, 40.0,
40.4, 40.5, 40.5, 40.6, 40.8, 40.8, 41.0, 41.0, 41.2, 41.2,
41.3, 41.4, 41.8, 41.9, 41.9, 42.0, 42.4, 42.6, 43.7, 44.0

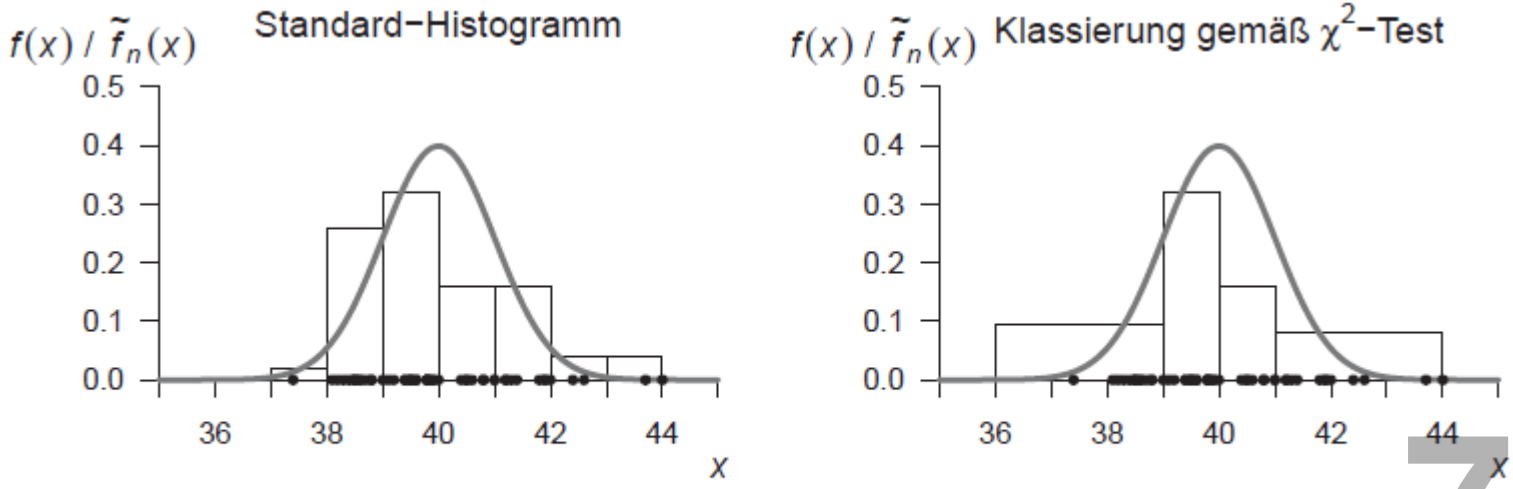


Abb. 11.3.1: Empirische vs. hypothetische Verteilung in Beispiel 11.3.4

11.3 Ausgewählte Testverfahren

> Y ... Länge eines Metallstifts

> Für Y gelten dann folgende Klassen-Einfallswahrscheinlichkeiten:

$$\pi_1 = P(Y \in [36{,}39]) \approx 0.16, \quad \pi_2 = P(Y \in (39, 40]) \approx 0.34$$

$$\pi_3 = P(Y \in (40, 41]) \approx 0.34, \pi_4 = P(Y \in (41, 44]) \approx 0.16$$

> Definiere X mit

$$X = \begin{cases} 1, & \text{für } Y \in [36{,}39], \\ 2, & \text{für } Y \in (39, 40], \\ 3, & \text{für } Y \in (40, 41], \\ 4, & \text{für } Y \in (41, 44], \end{cases}$$

> Statistisches Entscheidungsproblem:
Ist die Länge der Metallstifte $N(40{,}1)$-verteilt?

> Anpassungstest bezieht sich jedoch auf (diskretisierte) Variable X:
Genügt X einer diskreten Verteilung mit den hypothetischen Klassen-Einfallswahrscheinlichkeiten π_1, π_2, π_3 und π_4?

Testproblem:

H_0: $P(X = i) = \pi_i$ für $i = 1, \ldots, 4$ vs.

H_1: $P(X = i) \neq \pi_i$ für mindestens ein i

$\Rightarrow$ Immanentes Testproblem:

H_0: $Y \sim N(40,1)$ vs. H_1: $Y \nsim N(40,1)$

Teststatistik: $\chi_A^2 = \sum_{i=1}^{4} \frac{(N_i - n\pi_i)^2}{n\pi_i} \overset{a}{\sim} \chi^2(3)$

Testniveau (Signifikanzniveau): $\alpha = 0.05$

Kritischer Wert: $c = \chi^2_{3,0.95} = 7.81$

Testprozedur: Falls $\chi_A^2 > c$, Entscheidung für H_1, sonst Beibehaltung von H_0

Tab. 11.3.2: Beobachtete vs. erwartete Häufigkeiten in Beispiel 11.3.4 (gerundet)

Ausprägung i	1	2	3	4
f_i	0.280	0.320	0.160	0.240
$n_i = n \cdot f_i$	14	16	8	12
π_i	0.16	0.34	0.34	0.16
$n\pi_i$	8	17	17	8

Testergebnis hier:

$$\chi_A^2 = \sum_{i=1}^{4} \frac{(n_i - n\pi_i)^2}{n\pi_i}$$

$$= \frac{(14-8)^2}{8} + \frac{(16-17)^2}{17} + \frac{(8-17)^2}{17} + \frac{(12-8)^2}{8}$$

$$= 11.32 > c$$

$\Rightarrow$ Verwerfung von H_0, d. h. es liegt eine signifikante Abweichung von einer $N(40,1)$-Verteilung vor.

• Zusammenfassung •

Satz 11.3.3: χ^2-Anpassungstest

Seien $X_1, \ldots, X_n$ u.i.v. wie X verteilt, wobei X diskret verteilt sei mit Trägermenge $T_X = \{1, \ldots, k\}$ und $k \geq 2$. Gegeben sei folgendes Testproblem:

$$H_0 : P(X = i) = \pi_i \quad \text{für jedes } i \in T_X \text{ vs.}$$
$$H_1 : P(X = i) \neq \pi_i \quad \text{für mindestens ein } i \in T_X.$$

Dann basiert die Entscheidungsregel eines Tests zum Niveau α auf dem Resultat

$$\chi^2_A = \sum_{i=1}^{k} \frac{(N_i - n\pi_i)^2}{n\pi_i} \overset{a}{\sim} \chi^2(k-1) \quad \text{unter } H_0,$$

wobei N_i die absolute Häufigkeit der Ausprägung i ist, und lautet:

$$\text{Verwerfe } H_0, \text{ falls } \chi^2_A > \chi^2_{k-1,1-\alpha}, \text{ und sonst nicht.}$$

Der Test ist anwendbar, falls $n\pi_i \geq 5$ für $i = 1, \ldots, k$. Ferner kann X auch eine kategorisierte (diskretisierte) Zufallsvariable sein.

11.3 Ausgewählte Testverfahren

- **Die Frage der Klassenwahl bei Kategorisierung**

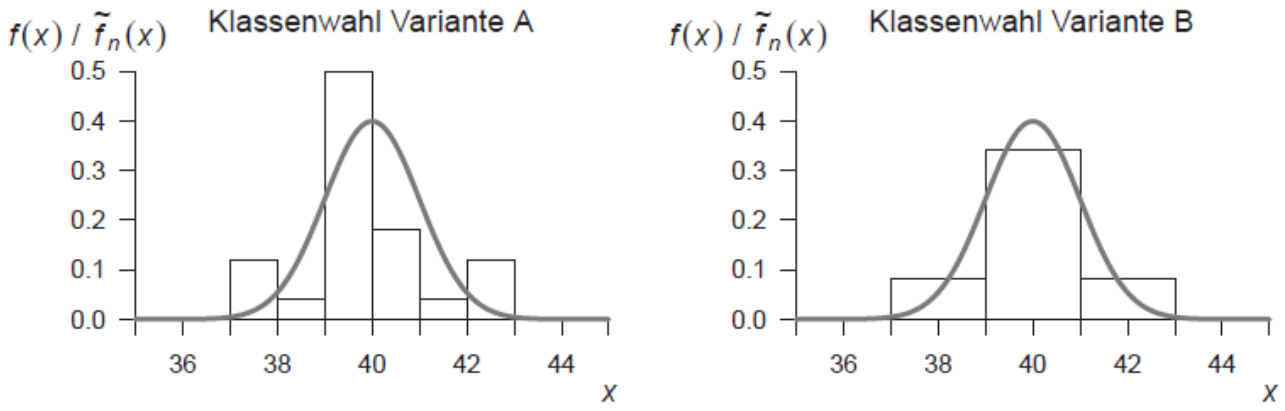


Abb. 11.3.2: Einfluss der Klassenwahl auf die Testentscheidung – fiktives Beispiel

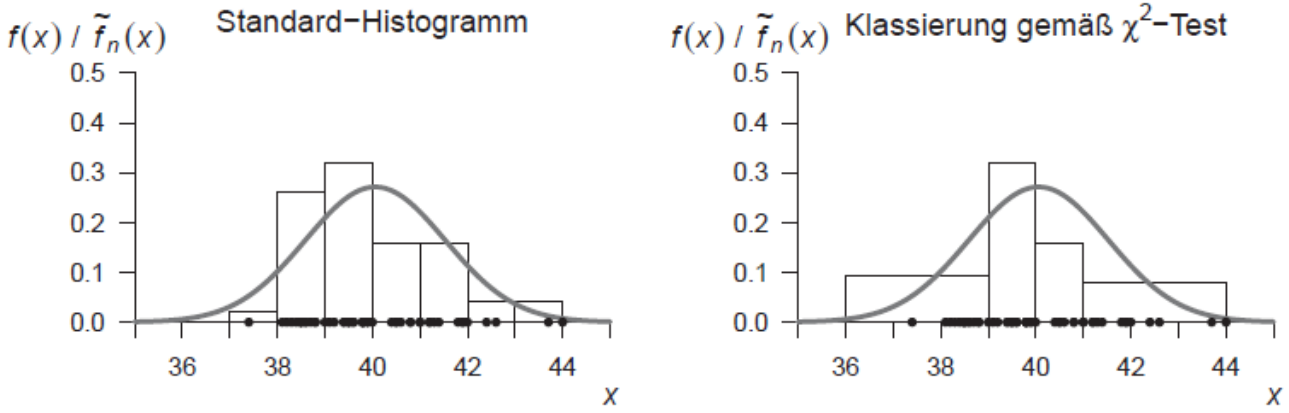


Abb. 11.3.3: Empirische vs. hypothetische Verteilung nach Parameterschätzung

● Abzug von Freiheitsgraden bei Parameterschätzungen ●

> Schätze zunächst die Verteilungsparameter

$$\hat{\mu} = \bar{x} = 40.06 \text{ und } \hat{\sigma}^2 = s^2 = 1.47^2$$

> Modifiziere das Testproblem dann zu

$$H_0\colon\ Y \sim N(40.06, 1.47^2) \text{ vs. } H_1\colon\ Y \nsim N(40.06, 1.47^2)$$

bzw. zu

$$H_0\colon\ P(X = i) = \pi_i \text{ für } i = 1, \ldots, 4 \text{ vs.}$$

$$H_1\colon\ P(X = i) \neq \pi_i \text{ für mindestens ein } i$$

mit den neuen implizierten Klassen-Einfallswahrscheinlichkeiten

$$\pi_1 = P(Y \in [36{,}39]) \approx 0.23, \quad \pi_2 = P(Y \in (39, 40]) \approx 0.25$$

$$\pi_3 = P(Y \in (40, 41]) \approx 0.26, \pi_4 = P(Y \in (41, 44]) \approx 0.26$$

für $Y \sim N(40.06, 1.47^2)$

Tab. 11.3.3: Beobachtete vs. erwartete Häufigkeiten nach Parameterschätzung (gerundet)

Ausprägung i	1	2	3	4
f_i	0.28	0.32	0.16	0.24
$n_i = n \cdot f_i$	14	16	8	12
π_i	0.23	0.25	0.26	0.26
$n\pi_i$	11.5	12.5	13	13

> Testergebnis wäre dann:

$$\chi_A^2 = \sum_{i=1}^{4} \frac{(n_i - n\pi_i)^2}{n\pi_i}$$

$$= \frac{(14 - 11.5)^2}{11.5} + \frac{(16 - 12.5)^2}{12.5} + \frac{(8 - 13)^2}{13} + \frac{(12 - 13)^2}{13}$$

$$= 3.52 < c$$

$\Rightarrow$ Beibehaltung von H_0

> **ABER!** Die verteilungstheoretischen Grundlagen sind bei dieser Vorgehensweise (vorhergehende Schätzung von Parametern) nicht mehr gegeben.

> Man kann nicht mit den gleichen Daten zunächst schätzen und dann testen. Die Schätzung der Parameter bewirkt eine Anpassung der Null-hypothese an die Daten (Verwerfung wird damit erschwert).

> Übliche (häufige) Vorgehensweise: Abzug von Freiheitsgraden; so viele wie Parameter geschätzt wurden. Dadurch werden kritische Werte kleiner, Verwerfung wird erleichtert (vgl. Abb. 11.3.4, Folie 808)

> Im vorliegenden Fall: 2 Freiheitsgrade abziehen!

⇒ neuer kritischer Wert: $c = \chi^2_{1,0.95} = 3.84$ (kleiner als zuvor)

⇒ Jedoch weiterhin keine Verwerfung von H_0

> **Allerdings weiterer Einwand!** Siehe nächster Punkt

11.3 Ausgewählte Testverfahren

- **Abzug von Freiheitsgraden setzt bestimmte Schätzmethoden voraus** -

Die Parameter müssten eigentlich mit der sog. χ^2-Minimum-Methode geschätzt werden...

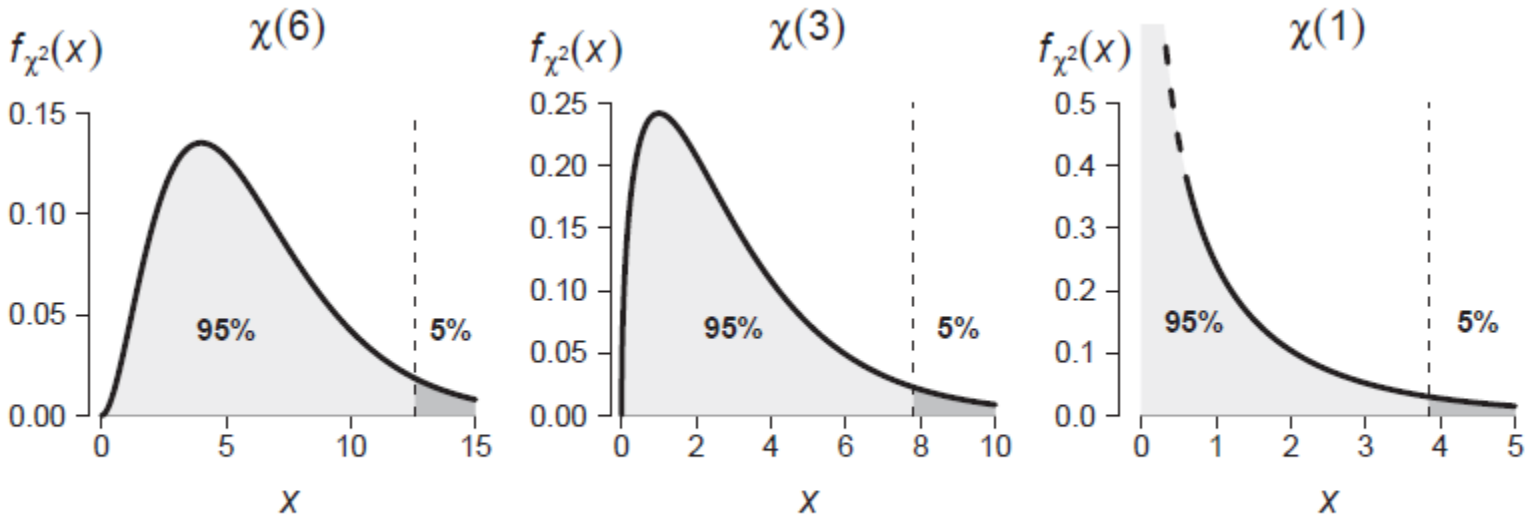


Abb. 11.3.4: Testverteilungen von χ^2-Tests

χ^2-Unabhängigkeitstest

● Verteilungstheoretische Grundlagen ●

> (X, Y) zweidimensional diskret verteilt mit

$$P(X = i, Y = j) = \pi_{ij} \text{ für } i = 1, \dots, k \text{ und } j = 1, \dots, l$$

$$P(X = i) = \pi_{i\bullet},\ P(Y = j) = \pi_{\bullet j}$$

> Stichprobe: $X_1, \dots, X_n$ unabhängig und identisch verteilt wie X

$(X_1, Y_1), \dots, (X_n, Y_n)$ unabhängig und identisch verteilt wie (X, Y)

> N_{ij} ... absolute Häufigkeit (stochastisch) der Ausprägung (i, j) in der Stichprobe

> $N_{i\bullet}$ ... absolute Häufigkeit (stochastisch) der Ausprägung i (von X)

> $N_{i\bullet}$ ... absolute Häufigkeit (stochastisch) der Ausprägung j (von Y)

> Es lässt sich dann (mit einigem Aufwand) zeigen, dass für großes n gilt:

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{\left(N_{ij} - \frac{N_{i\bullet} N_{\bullet j}}{n}\right)^2}{\frac{N_{i\bullet} N_{\bullet j}}{n}} \overset{a}{\sim} \chi^2\big((k-1)(l-1)\big)$$

falls X und Y stochastisch unabhängig sind.

● Verbindung zum Chi-Quadrat- und Kontingenzkoeffizienten ●

> Der **χ^2-Koeffizient** wurde schon einmal in Statistik 1 besprochen und inhaltlich gedeutet. Er ist ein nichtnormiertes Zusammenhangsmaß für kategoriale Daten (Kontingenztabellen)

> Der Kontingenzkoeffizient basiert hauptsächlich auf dem χ^2-Koeffizient

● Approximationsregel ●

Unsere Konvention: $\frac{n_{i\bullet} n_{\bullet j}}{n} \geq 5$ für alle i, j

11.3 Ausgewählte Testverfahren

- **Beispiel 11.3.5: Geschlecht und Raucherstatus**

Tab. 5.1.1: Gemeinsame Verteilung von Geschlecht und Raucherstatus

	Raucher	Gelegenheits-raucher	Nichtraucher	Summe	
weiblich	n_{11} 4	n_{12} 8	n_{13} 28	40	$n_{1\bullet}$
männlich	n_{21} 12	n_{22} 12	n_{23} 36	60	$n_{2\bullet}$
Summe	$n_{\bullet 1}$ 16	$n_{\bullet 2}$ 20	$n_{\bullet 3}$ 64	100	n

> Zur Erinnerung:

$f_{ij} = n_{ij}/n$, $f_{\bullet j} = n_{\bullet j}/n$ und $f_{i\bullet} = n_{i\bullet}/n$

Empirisches Unabhängigkeit von Merkmalen (Multiplikationskriterium):

$f_{ij} = f_{i\bullet} f_{\bullet j}$ für alle i und j

Das Produkt der Randverteilungen ergibt die gemeinsame Verteilung

Bei Unabhängigkeit **erwartet** man für Zelle (i, j) genau

$n \times f_{ij} = n \times f_{i\bullet} f_{\bullet j} = \frac{n_{i\bullet} n_{\bullet j}}{n}$ Beobachtungswerte.

11.3 Ausgewählte Testverfahren

> Jetzt als statistisches Entscheidungsproblem:
Sind die Zufallsvariablen X und Y stochastisch unabhängig oder nicht?

> Kodierung von Geschlecht und Raucherstatus mit 1, 2 bzw. 1, 2 und 3

> Testproblem

H_0: $P(X = i, Y = j) = P(X = i)P(Y = j)$ für $i = 1,2, j = 1,2,3$

H_1: $P(X = i, Y = j) \neq P(X = i)P(Y = j)$ für mindestens ein Paar (i, j)

> Teststatistik: $$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{3} \frac{\left(N_{ij} - \frac{N_{i\bullet} N_{\bullet j}}{n}\right)^2}{\frac{N_{i\bullet} N_{\bullet j}}{n}} \overset{a}{\sim} \chi^2(2)$$

> Testniveau (Signifikanzniveau): $\alpha = 0.05$

> Kritischer Wert: $c = \chi^2_{2,0.95} = 5.99$

> Testprozedur: Falls $\chi^2 > c$, Entscheidung für H_1, sonst Beibehaltung von H_0

11.3 Ausgewählte Testverfahren

Tab. 5.1.9: Tatsächliche und zu erwartende Häufigkeiten mit jeweiligen Abweichungen

	Raucher	Gelegenheits-raucher	Nichtraucher	Summe
weiblich	4 (6.4)	8 (8)	28 (25.6)	40
	−2.4	**0**	**2.4**	
männlich	12 (9.6)	12 (12)	36 (38.4)	60
	2.4	**0**	**−2.4**	
Summe	16	20	64	100

> Beispielrechnungen für die Terme $n_{i\bullet}n_{\bullet j}/n$:

$$\frac{n_{1\bullet}n_{\bullet 1}}{n} = \frac{40\times 16}{100} = 6.4, \quad \frac{n_{1\bullet}n_{\bullet 2}}{n} = \frac{40\times 20}{100} = 8, \quad \text{etc.}$$

> Testergebnis wäre dann:

$$\chi^2 = \frac{(4-6.4)^2}{6.4} + \frac{(8-8)^2}{8} + \cdots + \frac{(36-38.4)^2}{38.4} = 1.875 < c$$

$\Rightarrow$ Beibehaltung von H_0 (keine signifikante Abhängigkeit)

11.3 Ausgewählte Testverfahren

- **Beispiel 11.3.6: Weißer und schwarzer Würfel**

Tab. 11.3.4: Tatsächliche und zu erwartende Häufigkeiten mit jeweiligen Abweichungen

Zahl / Würfel	1	2	3	4	5	6	$n_{i\bullet}$
Weiß	9 (9)	6 (11)	14 (12)	10 (8)	8 (7.5)	13 (12.5)	60
	0	-5	2	2	0.5	0.5	
Schwarz	9 (9)	16 (11)	10 (12)	6 (8)	7 (7.5)	12 (12.5)	60
	0	5	−2	−2	−0.5	−0.5	
$n_{\bullet j}$	18	22	24	16	15	25	120

11.3 Ausgewählte Testverfahren

> Statistisches Entscheidungsproblem:
Gibt es einen Zusammenhang zwischen Würfelfarbe und Ergebnis?

> Kodierung von Farbe (X) mit 1 und 2 und Zahl (Y) mit 1, 2, ..., 6

> Testproblem

H_0: $P(X = i, Y = j) = P(X = i)P(Y = j)$ für $i = 1,2, j = 1, \dots, 6$

H_1: $P(X = i, Y = j) \neq P(X = i)P(Y = j)$ für mindestens ein Paar (i, j)

> Teststatistik: $$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{6} \frac{\left(N_{ij} - \frac{N_{i\bullet} N_{\bullet j}}{n}\right)^2}{\frac{N_{i\bullet} N_{\bullet j}}{n}} \overset{a}{\sim} \chi^2(5)$$

> Testniveau (Signifikanzniveau): $\alpha = 0.05$

> Kritischer Wert: $c = \chi^2_{5,0.95} = 11.07$

> Testprozedur: Falls $\chi^2 > c$, Entscheidung für H_1, sonst Beibehaltung von H_0

> Beispielrechnungen für die Terme $n_{i\bullet}n_{\bullet j}/n$:

$$\frac{n_{1\bullet}n_{\bullet 1}}{n} = \frac{60\times 18}{120} = 9, \quad \frac{n_{1\bullet}n_{\bullet 2}}{n} = \frac{60\times 22}{120} = 11, \quad \text{etc.}$$

> Testergebnis wäre dann:

$$\chi^2 = \frac{(9-9)^2}{9} + \frac{(6-11)^2}{11} + \cdots + \frac{(12-12.5)^2}{12.5} = 6.32 < c$$

$\Rightarrow$ Beibehaltung von H_0 (kein signifikanter Zusammenhang)

> Bemerkung:

Da jeder Würfel a priori (geplant) jeweils 60 Mal geworfen wird, sind die Randsummen ($n_{1\bullet} = n_{2\bullet} = 60$) nichtstochastisch. Tatsächlich führt man hier einen **χ^2-Homogenitätstest** durch, der verteilungstheoretisch etwa anders fundiert ist, numerisch jedoch auf genau die gleiche Weise berechnet wird.

- **Beispiel 11.3.7: Placebo vs. Medikament**

Tab. 11.3.5: Gleichstarker Zusammenhang bei unterschiedlicher Signifikanz

	krank	gesund
Placebo	40	45
Medikament	48	62

	krank	gesund
Placebo	4000	4500
Medikament	4800	6200

> Empirische Analyse:
Gesundanteil bei Medikament nur um ca. 3% höher. Kontingenzkoeffizient links und rechts jeweils $C_K = 0.05 \Rightarrow$ schwache empirische Abhängigkeit

> Statistisches Entscheidungsproblem:
Ist die gemessene Abhängigkeit überhaupt signifikant („nichtzufällig“)?

> Kodierung von Medikation und Gesundheitszustand jeweils mit 1 und 2

> Testproblem

H_0: X und Y sind stochastisch unabhängig

H_1: X und Y sind stochastisch abhängig

> Teststatistik: $\chi^2 = \frac{n(N_{11}N_{22} - N_{12}N_{21})^2}{N_{\bullet 1}N_{\bullet 2}N_{1\bullet}N_{2\bullet}} \overset{a}{\sim} \chi^2(1)$

> Kritischer Wert: $c = \chi^2_{1,0.95} = 3.84$

> Testprozedur: Falls $\chi^2 > c$, Entscheidung für H_1, sonst Beibehaltung von H_0

> Testergebnisse wäre dann:

$$\chi^2_{links} = \frac{195 \times (40 \times 62 - 48 \times 45)^2}{95 \times 110 \times 88 \times 107} = 0.2268$$

$\Rightarrow$ Beibehaltung von H_0 (kein signifikanter Zusammenhang)

$$\chi^2_{rechts} = \frac{19500 \times (4000 \times 6200 - 4800 \times 4500)^2}{9500 \times 11000 \times 8800 \times 10700} = 22.68$$

$\Rightarrow$ Verwerfung von H_0.
Obwohl der Zusammenhang nur schwach ist, ist er für großes n dennoch statistisch signifikant

• Zusammenfassung •

Satz 11.3.4: χ^2-Unabhängigkeitstest

Seien $(X_1, Y_1), \ldots, (X_n, Y_n)$ u.i.v. wie (X, Y), wobei (X, Y) diskret verteilt ist mit Trägermenge $T_{XY} = \{1, \ldots, k\} \times \{1, \ldots, l\}$, wobei $k, l \geq 2$. Gegeben sei folgendes Testproblem:

$$H_0 : P(X = i, Y = j) = P(X = i)P(Y = j) \quad \text{für alle } (i, j) \in T_{XY} \text{ vs.}$$
$$H_1 : P(X = i, Y = j) \neq P(X = i)P(Y = j) \quad \text{für mindestens ein } (i, j) \in T_{XY}.$$

Dann basiert die Entscheidungsregel eines Tests zum Niveau α auf dem Resultat

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(N_{ij} - \frac{N_{i\bullet} N_{\bullet j}}{n})^2}{\frac{N_{i\bullet} N_{\bullet j}}{n}} \overset{a}{\sim} \chi^2((k-1)(l-1)) \quad \text{unter } H_0.$$

Dabei lehnt sich die Notation an diejenige für Kontingenztabellen an. Die Entscheidungsregel lautet dann:

$$\text{Verwerfe } H_0, \text{ falls } \chi^2 > \chi^2_{(k-1)(l-1),1-\alpha} \text{ und sonst nicht.}$$

Der Test ist anwendbar, falls

$$\frac{n_{i\bullet} n_{\bullet j}}{n} \geq 5 \quad \text{für alle } (i, j) \in T_{XY}.$$

Ferner können X und Y auch kategorisierte (diskretisierte) Zufallsvariablen sein.

11.3 Ausgewählte Testverfahren

Tab. A.3: Quantilstabelle der χ^2-Verteilung

Tabelliert sind die Quantile der χ^2-Verteilung für n Freiheitsgrade.
Ablesebeispiel: $\chi^2_{10,0.95} = 18.307$.

n	0.01	0.025	0.05	0.1	0.5	0.9	0.95	0.975	0.99
1	0.000	0.001	0.004	0.016	0.455	2.705	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	1.386	4.605	5.992	7.378	9.210
3	0.115	0.216	0.352	0.584	2.366	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	3.357	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	4.351	9.236	11.070	12.832	15.086
6	0.872	1.237	1.635	2.204	5.348	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	6.346	12.017	14.067	16.013	18.475
8	1.647	2.180	2.733	3.490	7.344	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	8.343	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	9.342	15.987	18.307	20.483	23.209
11	3.054	3.816	4.575	5.578	10.341	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	11.340	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	12.340	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	13.339	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	14.339	22.307	24.996	27.488	30.578

16	5.812	6.908	7.962	9.312	15.338	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	10.085	16.338	24.769	27.587	30.191	33.409
18	7.015	8.231	9.390	10.865	17.338	25.989	28.869	31.526	34.805
19	7.633	8.906	10.117	11.651	18.338	27.204	30.143	32.852	36.191
20	8.260	9.591	10.851	12.443	19.337	28.412	31.410	34.170	37.566
21	8.897	10.283	11.591	13.240	20.337	29.615	32.671	35.479	38.932
22	9.543	10.982	12.338	14.041	21.337	30.813	33.924	36.781	40.289
23	10.196	11.689	13.091	14.848	22.337	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	15.659	23.337	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	16.473	24.337	34.382	37.653	40.647	44.314
26	12.198	13.844	15.379	17.292	25.337	35.563	38.885	41.923	45.642
27	12.879	14.573	16.151	18.114	26.336	36.741	40.113	43.194	46.963
28	13.565	15.308	16.928	18.939	27.336	37.916	41.337	44.461	48.278
29	14.257	16.047	17.708	19.768	28.336	39.087	42.557	45.722	49.588
30	14.954	16.791	18.493	20.599	29.336	40.256	43.773	46.979	50.892

11.3.4 Weitere Tests

Tests über Anteilswerte

● Exakter Binomialtest ●

> Ausgangspunkt: $X_1, \ldots, X_n$ sind unabhängig $B(1, \pi)$-verteilt

> Testprobleme bezogen auf π.

> Teststatistik:

$$S_n = \sum_{i=1}^{n} X_i \overset{\pi = \pi_0}{\sim} B(n, \pi_0)$$

> Beispiele: Vgl. Beispiele 11.1.1 bis 11.1.3 (Folien 734-738)

> Bei der Festlegung von kritischen Werten einseitiger Tests gilt die Konvention der „nächstmöglichen Unterbietung“ des Testniveaus (vgl. Übungsblatt 18).

> Die Festlegung von kritischen Werten beim zweiseitigen Testen sowie ggf. *p*-Wert-Bestimmung sind nicht prüfungsrelevant

• Approximativer Binomialtest •

Satz 11.3.5: Approximativer Binomialtest

Seien $X_1, \ldots, X_n$ unabhängig $B(1, \pi)$-verteilt mit $0 < \pi < 1$. Gegeben seien folgende Testprobleme:

a) $H_0 : \pi \geq \pi_0$ vs. $H_1 : \pi < \pi_0$,
b) $H_0 : \pi \leq \pi_0$ vs. $H_1 : \pi > \pi_0$,
c) $H_0 : \pi = \pi_0$ vs. $H_1 : \pi \neq \pi_0$,

Dann basieren die Entscheidungsregeln eines Tests zum Niveau α, falls n groß ist, auf dem Resultat

$$Z = \frac{\hat{\pi} - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}} \overset{a}{\sim} N(0,1) \quad \text{für } \pi = \pi_0$$

und lauten: Verwerfe H_0 in

a), falls $z < -z_{1-\alpha}$,
b), falls $z > z_{1-\alpha}$,
c), falls $|z| > z_{1-\alpha/2}$, d.h. $z < -z_{1-\alpha/2}$ oder $z > z_{1-\alpha/2}$.

Dabei ist $\hat{\pi}$ der Anteil von Einsen in der Stichprobe. Der Test wird als **approximativer Binomialtest** bezeichnet und ist für $n \geq 30$ anwendbar.

● Binomialtests als Tests über den Median ●

> Ausgangspunkt: $X_1, \ldots, X_n$ sind unabhängig stetig verteilt mit theoretischem Median $q_{0.5}$

> Für $q_{0.5} = m_0$ gilt dann:

$$P(X_i = m_0) = 0, P(X_i < m_0) = 0.5, P(X_i > m_0) = 0.5$$

> Testprobleme bezogen auf den theoretische Median $q_{0.5}$:

$$H_0: q_{0.5} = m_0 \text{ vs. } H_1: q_{0.5} \neq m_0$$

$$H_0: q_{0.5} \geq m_0 \text{ vs. } H_1: q_{0.5} < m_0$$

$$H_0: q_{0.5} \leq m_0 \text{ vs. } H_1: q_{0.5} > m_0$$

> Teststatistik:

$$S_n = \sum_{i=1}^{n} Y_i \overset{q_{0.5} = m_0}{\sim} B(n, 0.5) \quad \text{mit} \quad Y_i = I_{(m_0, \infty)}(X_i)$$

> Wir erhalten dann exakte Binomialtests über π mit $\pi_0 = 0.5$

> Bei großem n: Approximative Binomialtests möglich

> Beispiel: Durchführung eines approximativen Binomialtests über den Median

Gegeben sei die Situation aus Beispiel 11.3.4 (Folie 799), wobei wir annehmen, dass die Länge eines Metallstifts einer (unbekannten) stetigen Verteilung genügt. Die Realisationen der Stichprobe ($n = 50$) lauten:

37.4, 38.1, 38.2, 38.3, 38.4, 38.5, 38.5, 38.5, 38.6, 38.6,
38.7, 38.8, 38.8, 39.0, 39.1, 39.2, 39.4, 39.4, 39.5, 39.5,
39.5, 39.6, 39.6, 39.8, 39.8, 39.8, 39.8, 39.9, 39.9, 40.0,
40.4, 40.5, 40.5, 40.6, 40.8, 40.8, 41.0, 41.0, 41.2, 41.2,
41.3, 41.4, 41.8, 41.9, 41.9, 42.0, 42.4, 42.6, 43.7, 44.0

Entscheidungsproblem:
Ist signifikant mehr als die Hälfte der Stifte kürzer als 40 mm?
Alternativ (und äquivalent):
Ist der Median der Stiftlängen signifikant kleiner als 40 mm?
Alternativ (und äquivalent):
Ist der Anteil von Stiften, die länger sind als 40mm signifikant kleiner als 50%?

Statistisches Modell: $X_1, \ldots, X_{50}$ unabhängig stetig verteilt

Testproblem: $H_0: q_{0.5} \geq 40$ vs. $H_1: q_{0.5} < 40$

Teststatistik: $$Z = \frac{\hat{\pi} - 0.5}{\sqrt{0.5 \times (1 - 0.5)/50}} \overset{a}{\underset{\pi = 0.5}{\sim}} N(0,1)$$

wobei $\hat{\pi}$ der relative Anteil der Werte ist, die größer sind als 40

Man beachte: $$Y_i = I_{(40,\infty)}(X_i) \overset{q_{0.5} = 40}{\sim} B(1,0.5)$$

$$\Rightarrow \sum_{i=1}^{50} Y_i \sim B(50,0.5) \approx N(25,12.5)$$

$$\Rightarrow \hat{\pi} = \bar{Y} \sim N(0.5,0.005)$$

Testniveau (Signifikanzniveau): $\alpha = 0.05$

Kritischer Wert: $c = z_{0.05} = -1.64$

Testprozedur: Falls $Z < -1.64$, Entscheidung für H_1,
sonst Beibehaltung von H_0

Testergebnis hier:

$$z = \frac{0.4 - 0.5}{\sqrt{0.5 \times (1 - 0.5)/50}} = -1.41 > -1.64$$

$\Rightarrow$ Beibehaltung von H_0.
Nicht signifikant mehr als die Hälfte der Stifte sind kürzer als 40 mm bzw. der Median der Stiftlänge ist nicht signifikant kleiner als 40 mm.

• Approximativer Binomialtest über Anteilswertdifferenz •

Satz 11.3.6: Approximativer Binomialtest über eine Anteilswertdifferenz

Gegeben seien $n = n_0 + n_1$ unabhängige Stichprobenvariablen

$$Y_{01}, Y_{02}, \ldots, Y_{0n_0} \text{ und } Y_{11}, Y_{12}, \ldots, Y_{1n_1},$$

die innerhalb der beiden Gruppen jeweils identisch verteilt sind mit

$$Y_{0i} \sim B(1, \pi_0) \quad \text{mit } 0 < \pi_0 < 1 \text{ für } i = 1, \ldots, n_0,$$
$$Y_{1i} \sim B(1, \pi_1) \quad \text{mit } 0 < \pi_1 < 1 \text{ für } i = 1, \ldots, n_1.$$

Gegeben seien folgende Testprobleme über $\pi_1 - \pi_0$:

a) $H_0 : \pi_1 - \pi_0 \geq \delta_0$ vs. $H_1 : \pi_1 - \pi_0 < \delta_0$,
b) $H_0 : \pi_1 - \pi_0 \leq \delta_0$ vs. $H_1 : \pi_1 - \pi_0 > \delta_0$,
c) $H_0 : \pi_1 - \pi_0 = \delta_0$ vs. $H_1 : \pi_1 - \pi_0 \neq \delta_0$.

Dann basieren die Entscheidungsregeln eines Tests zum Niveau α, falls n_0 und n_1 groß und die Bedingung

$$n_1/n \xrightarrow{n\to\infty} \pi \in (0,1)$$

erfüllt sind, auf dem Resultat

$$\hat{T}_D = \frac{\hat{\pi}_1 - \hat{\pi}_0 - \delta_0}{\sqrt{\frac{\hat{\pi}_0(1-\hat{\pi}_0)}{n_0} + \frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1}}} \overset{a}{\sim} N(0,1) \quad \text{für } \pi_1 - \pi_0 = \delta_0$$

und lauten: Verwerfe H_0 in

a), falls $\hat{t}_D < -z_{1-\alpha}$,

b), falls $\hat{t}_D > z_{1-\alpha}$,

c), falls $|\hat{t}_D| > z_{1-\alpha/2}$, d.h. $\hat{t}_D < -z_{1-\alpha/2}$ oder $\hat{t}_D > z_{1-\alpha/2}$,

Dabei sind

$$\hat{\pi}_0 = \overline{Y}_0 \quad \text{und} \quad \hat{\pi}_1 = \overline{Y}_1$$

die Anteile von Einsen in den jeweiligen Gruppen. Der Test wird hier als **approximativer Binomialtest über eine Anteilswertdifferenz** bezeichnet und ist für $n_0, n_1 \geq 30$ anwendbar.

11.3 Ausgewählte Testverfahren

Satz 11.3.7: Test über eine Anteilswertdifferenz $\delta_0 = 0$

Gegeben seien die Voraussetzungen und Testprobleme von Satz 11.3.6. Dann basieren die Entscheidungsregeln für $\delta_0 = 0$ auf dem Resultat

$$\tilde{\tilde{T}}_D = \frac{\hat{\pi}_1 - \hat{\pi}_0}{\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n_0} + \frac{\hat{\pi}(1-\hat{\pi})}{n_1}}} \overset{a}{\sim} N(0,1) \quad \text{mit } \hat{\pi} = \frac{1}{n}(n_0\hat{\pi}_0 + n_1\hat{\pi}_1).$$

Man beachte, dass unter $\pi_0 = \pi_1 (= \pi)$ gilt:

$Y_{01}, \ldots, Y_{0n_0}, Y_{11}, \ldots, Y_{1n_1}$ sind unabhängig $B(1, \pi)$-verteilt

Deshalb erfolgt die Schätzung des einheitlichen π-Wertes durch eine „gepoolte Schätzung“ (rel. Häufigkeit über beide Gruppen hinweg)

Der approximative Binomialtest über „$H_0 : \pi_1 - \pi_0 = 0$“ ist äquivalent zu einem entsprechenden χ^2-Homogenitätstest.

Es gilt dabei: $\tilde{\tilde{T}}_D^2 = \chi^2 \overset{a}{\sim} \chi^2(1)$

• Beispiel 11.3.7 fortgesetzt (Placebo vs. Medikament) •

> Gegeben sei die Situation von Beispiel 11.3.7 (Folie 817) für $n = 195$:

	krank	gesund
Placebo	40	45
Medikament	48	62

π_0 ... Gesundungswahrscheinlichkeit mit Placebo

π_1 ... Gesundungswahrscheinlichkeit mit Medikament

> Statistisches Modell:

$Y_{01}, \ldots, Y_{0n_0}$ unabhängig $B(1, \pi_0)$-verteilt

$Y_{11}, \ldots, Y_{1n_1}$ unabhängig $B(1, \pi_1)$-verteilt

$Y_{01}, \ldots, Y_{0n_0}, Y_{11}, \ldots, Y_{1n_1}$ (vollständig) unabhängig

> Entscheidungsproblem: Sind die Wirkungen verschieden?

> Testproblem: $H_0: \pi_1 - \pi_0 = 0$ vs. $H_1: \pi_1 - \pi_0 \neq 0$

bzw. $H_0: \pi_0 = \pi_1$ vs. $H_1: \pi_0 \neq \pi_1$

11.3 Ausgewählte Testverfahren

> Teststatistik: $\tilde{\tilde{T}}_D = \dfrac{\hat{\pi}_1 - \hat{\pi}_0}{\sqrt{\dfrac{\hat{\pi}(1-\hat{\pi})}{n_0} + \dfrac{\hat{\pi}(1-\hat{\pi})}{n_1}}} \overset{a}{\underset{\pi_0 = \pi_1}{\sim}} N(0,1)$

mit $\hat{\pi} = \dfrac{1}{n}(n_0\hat{\pi}_0 + n_1\hat{\pi}_1)$

> Testniveau (Signifikanzniveau): $\alpha = 0.05$

> Kritische Werte: $c_1 = z_{0.025} = -1.96$ und $c_2 = z_{0.975} = 1.96$

> Testprozedur: Falls $\tilde{\tilde{T}}_D < -1.96$ oder $\tilde{\tilde{T}}_D > 1.96$, Entscheidung für H_1 (Verwerfung von H_0), sonst Beibehaltung von H_0

> Testergebnis hier:

Mit $n_0 = 85, n_1 = 110, \hat{\pi}_0 = 45/85 \approx 0.5294,\ \hat{\pi}_1 = 62/110 \approx 0.5636$ erhalten wir $\hat{\pi} = (45 + 62)/195 \approx 0.5487$ und damit

$$\tilde{\tilde{t}}_D = 0.4762$$

$\Rightarrow$ Beibehaltung von H_0 (kein signifikanter Unterschied)

> Bemerkung (vgl. Folie 818)

$$\tilde{\tilde{t}}_D^2 = (0.4762)^2 = 0.2268 = \chi^2$$

$$c_1^2 = (1.96)^2 = 3.84 = \chi^2_{1,0.95}$$

● Beispiel 11.3.8: Test auf klinische Relevanz ●

> Gegeben sei die Situation von Beispiel 11.3.7 (Folie 817) für $n = 19500$:

	krank	gesund
Placebo	4000	4500
Medikament	4800	6200

> Entscheidungsproblem:
Ist Medikament um mindestens 2 Prozentpunkte wirksamer?
(Annahme, dass dieser Wert „klinisch relevant“ ist

> Testproblem: $H_0: \pi_1 - \pi_0 \leq 0.02$ vs. $H_1: \pi_1 - \pi_0 > 0.02$

11.3 Ausgewählte Testverfahren

> Teststatistik: $\hat{T}_D = \dfrac{\hat{\pi}_1 - \hat{\pi}_0 - 0.02}{\sqrt{\dfrac{\hat{\pi}_0(1-\hat{\pi}_0)}{n_0} + \dfrac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1}}} \overset{a}{\underset{\pi_1 - \pi_0 = 0.02}{\sim}} N(0,1)$

> Testniveau (Signifikanzniveau): $\alpha = 0.05$

> Kritischer Wert: $c = z_{0.95} = 1.64$

> Testprozedur: Falls $\hat{T}_D > 1.64$, Entscheidung für H_1,
sonst Beibehaltung von H_0

> Testergebnis hier:

Mit $n_0 = 8500,\ n_1 = 11000,\ \hat{\pi}_0 \approx 0.5294,\ \hat{\pi}_1 \approx 0.5636$ erhalten wir

$$\hat{t}_D = 1.9755 > c$$

$\Rightarrow$ Verwerfung von H_0, klinische Relevanz vorhanden

Tests auf Unkorreliertheit und Unabhängigkeit

• Korrelationstest •

Satz 11.3.8: Korrelationstest

Seien $(X_1, Y_1), \ldots, (X_n, Y_n)$ unabhängig gemeinsam normalverteilt und ϱ_{XY} die theoretische Korrelation zwischen X_i und Y_i, d.h.

$$\varrho_{XY} = Corr(X_i, Y_i) \quad \text{für } i = 1, \ldots, n.$$

Gegeben seien folgende Testprobleme über ϱ_{XY}:

a) $H_0 : \varrho_{XY} \geq 0$ vs. $H_1 : \varrho_{XY} < 0$,
b) $H_0 : \varrho_{XY} \leq 0$ vs. $H_1 : \varrho_{XY} > 0$,
c) $H_0 : \varrho_{XY} = 0$ vs. $H_1 : \varrho_{XY} \neq 0$.

Dann basieren die Entscheidungsregeln eines Tests zum Niveau α auf dem Resultat

$$T_R = \frac{R_{XY}}{\sqrt{1 - R_{XY}^2}} \sqrt{n-2} \sim t(n-2) \quad \text{für } \varrho_{XY} = 0$$

und lauten: Verwerfe H_0 in

a), falls $t_R < -t_{n-2,1-\alpha}$,

b), falls $t_R > t_{n-2,1-\alpha}$,

c), falls $|t_R| > t_{n-2,1-\alpha/2}$, d.h. $t_R < -t_{n-2,1-\alpha/2}$ oder $t_R > t_{n-2,1-\alpha/2}$,

Dabei bezeichnet R_{XY} die Stichprobenkorrelation. Der Test wird hier als t-**Test auf Unkorreliertheit** oder kurz als **Korrelationstest** bezeichnet.

• Approximativer Gauß-Test auf Unabhängigkeit •

Satz 11.3.9: Approximativer Gauß-Test auf Unabhängigkeit

Seien $(X_1, Y_1), \ldots, (X_n, Y_n)$ unabhängig identisch verteilt wie (X, Y), wobei die Momente aller Stichprobenvariablen endlich sind, d.h. Annahme (M) von Definition 8.2.1 erfüllt ist. Gegeben sei folgendes Testproblem:

$$H_0\text{: } X \text{ und } Y \text{ sind unabhängig} \quad \text{vs.} \quad H_1\text{: } X \text{ und } Y \text{ sind abhängig.}$$

Dann basiert die Entscheidungsregel eines Tests zum Niveau α auf dem Resultat

$$T_R \overset{a}{\sim} N(0,1) \quad \text{unter } H_0$$

und lautet: Verwerfe H_0,

$$\text{falls } |t_R| > z_{1-\alpha/2}, \text{ d.h. } t_R < -z_{1-\alpha/2} \text{ oder } t_R > z_{1-\alpha/2}.$$

Der Test wird hier als **approximativer Gauß-Test auf Unabhängigkeit** bezeichnet und ist für $n \geq 30$ anwendbar.

● Welche Korrelation genügt zur Signifikanz ●

> Ab welcher Korrelation wird H_0 bei einem Testniveau von $\alpha \in (0,0.5)$ verworfen? Die Frage führt zum Kriterium

$$|T_R| > t_{n-2,1-\alpha/2}$$

> Mit entsprechenden Umformungen erhält man daraus

$$|r_{XY}| > \sqrt{\frac{t^2_{n-2,1-\alpha/2}}{n-2+t^2_{n-2,1-\alpha/2}}}$$

> Beispiel:

Für $n = 10$ und $\alpha = 0.05$ folgt

$$|r_{XY}| > \sqrt{\frac{t^2_{8,0.975}}{8+t^2_{8,0.975}}} = \sqrt{\frac{2.3060^2}{8+2.3060^2}} \approx 0.63$$

11.3 Ausgewählte Testverfahren

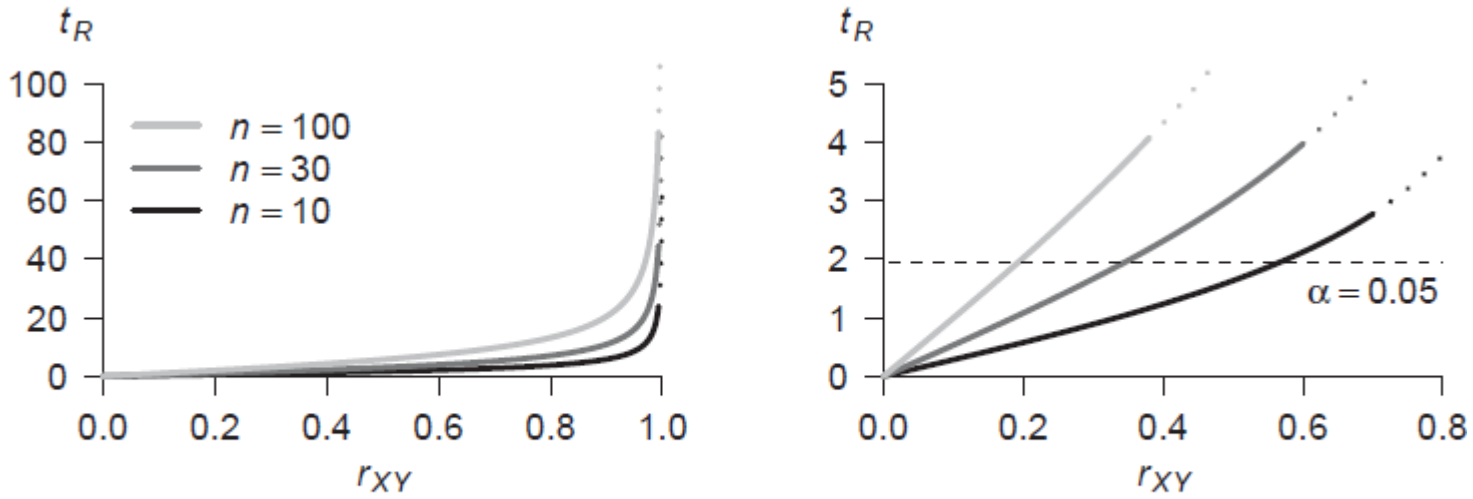


Abb. 11.3.5: Zusammenhang von Korrelation und Teststatistik

Jede Korrelation, die betragsmäßig kleiner als 0.20 ist, ist für $n \leq 100$ nicht signifikant.

- **Beispiel 11.3.9: Einige Beispiele für Tests auf Unkorreliertheit und Unabhängigkeit** ●

> Gegeben: (2, 1), (4, 2), (4, 3), (5, 2), (6, 4), (8, 5), (9, 6), (10, 4), (4, 5), (7, 3)

> Entscheidungsproblem: Sind X und Y korreliert oder nicht?

> Testproblem: $H_0\colon\ \rho_{XY} = 0$ vs. $H_1\colon \rho_{XY} \neq 0$

> Teststatistik: $T_R = \dfrac{R_{XY}}{\sqrt{1 - R_{XY}^2}}\sqrt{n-2} \overset{\rho_{XY}=0}{\sim} t(n-2)$

> Testniveau (Signifikanzniveau): $\alpha = 0.05$

> Kritischer Wert: $c_1 = t_{8,0.025} = -2.31,\ c_2 = t_{8,0.975} = 2.31$

> Testprozedur: Falls $t_R < -2.31$ oder $t_R > 2.31$, Entscheidung für H_1 (Verwerfung von H_0), sonst Beibehaltung von H_0

11.3 Ausgewählte Testverfahren

> Testergebnis hier:

Tab. 5.2.2: Arbeitstabelle zur Berechnung der Korrelation nach Pearson

i	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	2	1	4	1	2
2	4	2	16	4	8
3	4	3	16	9	12
4	5	2	25	4	10
5	6	4	36	16	24
6	8	5	64	25	40
7	9	6	81	36	54
8	10	4	100	16	40
9	4	5	16	25	20
10	7	3	49	9	21
Summe	59	35	407	145	231

$$\bar{x} = \frac{59}{10} = 5.9 \qquad \bar{y} = \frac{35}{10} = 3.5$$

$$r_{XY} = \frac{\tilde{s}_{XY}}{\tilde{s}_X \tilde{s}_Y} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} {x_i}^2 - \bar{x}^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n} y^2 - \bar{y}^2}}$$

$$= \frac{\frac{231}{10} - 5.9 \times 3.5}{\sqrt{\frac{407}{10} - 5.9^2}\sqrt{\frac{145}{10} - 3.5^2}}$$

$$\approx 0.6730$$

$$\Rightarrow \quad t_R = \frac{0.6730}{\sqrt{1 - 0.6730^2}}\sqrt{8} \approx 2.5736 > c_2$$

$\Rightarrow$ Verwerfung von H_0, signifikante Korrelation (lineare Abhängigkeit) vorhanden

Sonstige Tests und Testmethoden

- Vorbemerkung
- Binomialtests über Quantile
- *F*-Tests über Varianzen
- *F*-Tests auf Gleichheit von Erwartungswerten und ANOVA
- Tests auf Normalverteilung
- Korrelationstests nach Spearman
- Monte-Carlo-Tests
- Bayes-Tests

11.3.5 Allgemein zu beachtende Punkte

- **Adäquatheit von Modellannahmen**

> Modellannahmen sind in der Praxis niemals perfekt erfüllt.

- **Stochastische Gruppenumfänge**
- **Abhängigkeit und Kausalität**

> Korrelation impliziert keine Kausalität.

12 Das lineare Regressionsmodell

12 Das lineare Regressionsmodell

12.1 Einfaches lineares Regressionsmodell	849
12.1.1 Grundmodell und KQ-Methode	849
Statistisches Grundmodell	849
Herleitung der KQ-Schätzer	851
Eigenschaften der empirischen KQ-Regression	855
Bestimmtheitsmaß und Standardfehler der Regression	862
12.1.2 Modellannahmen und theoretische KQ-Regression	869
Ensembles von Modellannahmen	869
Theoretische KQ-Regression	880
12.1.3 Verteilungstheoretische Grundlagen	887
Verteilungen der KQ-Schätzer	887
Konsistenz und Effizienz der KQ-Schätzer	906

12 Das lineare Regressionsmodell

Schätzung der Varianzen der KQ-Schätzer	907
Verteilungen der Inferenzstatistiken	910
12.1.4 Schätzen und Testen	911
Konfidenzintervalle und Tests	911
Spezialfall: Binärer Regressor	916
Adäquatheit bestimmter Modellannahmen	922
Fallbeispiel 1: Bewässerung und Wachstum	930
Fallbeispiel 2: Klassengröße und Lernerfolg	937
12.2 Einführung in das multiple lineare Regressionsmodell	947
12.2.1 Partielle lineare KQ-Regression	947
Empirische partielle Regression	947
Theoretische partielle Regression	955

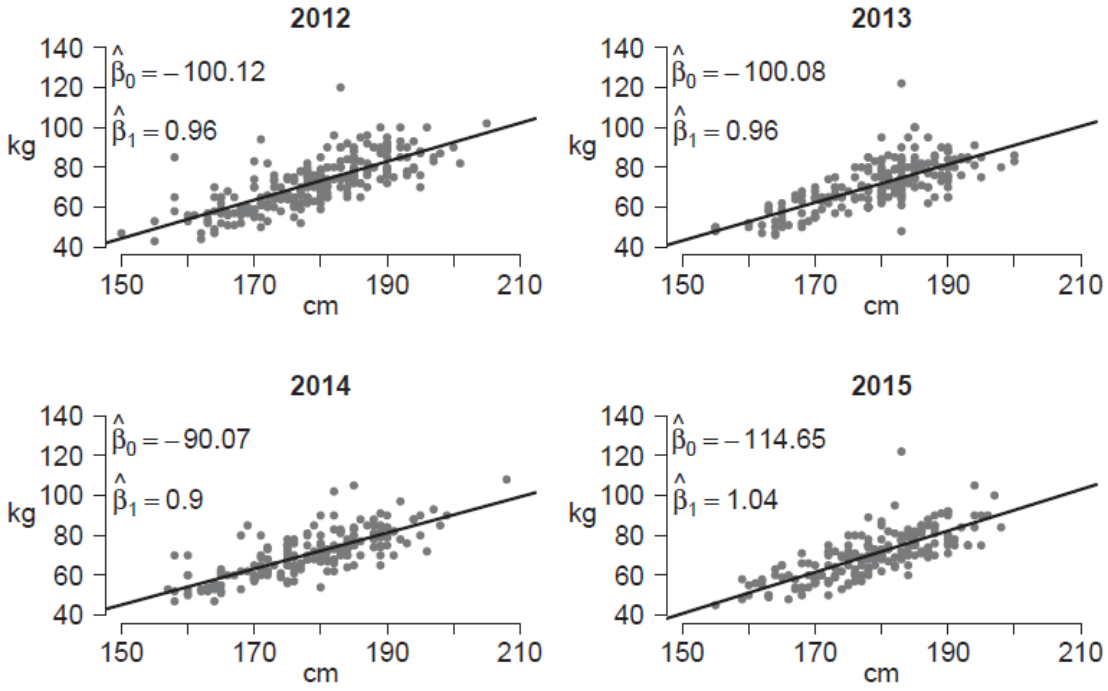
12 Das lineare Regressionsmodell

	Verbindung von Empirie und Theorie	958
12.2.2	Multiple lineare KQ-Regression	960
	Empirische multiple Regression	960
	Theoretische multiple Regression	966
	Verbindung von Empirie und Theorie	970
	Statistische Modelle und Inferenz	971
12.2.3	Fallbeispiele	981
	Fallbeispiel 2 fortgesetzt: Determinanten des Lernerfolgs	981
	Fallbeispiel 3: Gewicht und Geschlecht	984
	Fallbeispiel 4: Binäre Regressoren und ANOVA-Modelle	986

12.1.1 Grundmodell und KQ-Methode

Statistisches Grundmodell

- **Hintergrund**



Daten: Eigene Erhebungen

Abb. 12.1.1: Zusammenhang von Größe und Gewicht von Studenten

12.1 Einfaches lineares Regressionsmodell

• Definition •

Definition 12.1.1: Einfaches lineares Regressionsmodell

Gegeben sei eine Stichprobe aus n beobachtbaren Zufallsvektoren $(X_1, Y_1), \ldots, (X_n, Y_n)$ und n nicht beobachtbaren Zufallsvariablen $U_1, \ldots, U_n$. Gilt dann

(A0) $Y_i = \beta_0 + \beta_1 X_i + U_i$ für $i = 1, \ldots, n$,

so bezeichnet man (A0) als **einfaches lineares Regressionsmodell.** Die Parameter β_0 und β_1 werden als **theoretische Regressionskoeffizienten** und die Zufallsvariablen $U_1, \ldots, U_n$ als **stochastische Fehler** oder **stochastische Residuen** bezeichnet. Die Funktion

$$y(x) = \beta_0 + \beta_1 x$$

definiert die **theoretische Regressionsgerade,** wobei der Definitionsbereich von x fallabhängig einzuschränken ist.

• Interpretation •

Herleitung der KQ-Schätzer

- **Lösung des empirischen Kleinste-Quadrate-Problems**

$$\min_{\beta_0,\beta_1} \hat{Q}(\beta_0,\beta_1) \quad \text{mit} \quad \hat{Q}(\beta_0,\beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

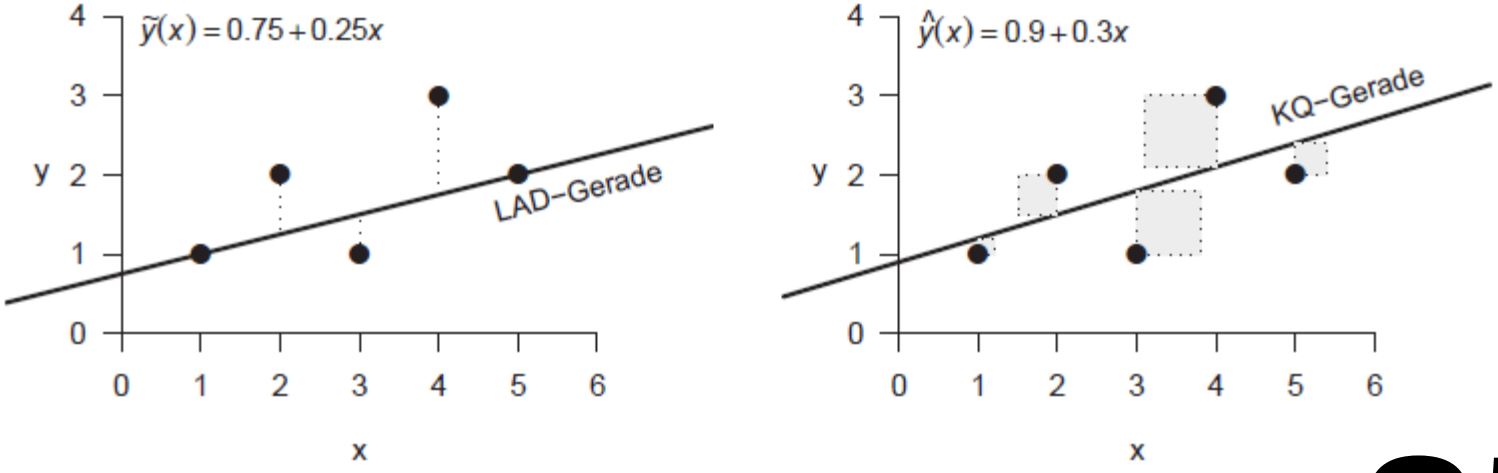


Abb. 5.2.15: Minimierung absoluter oder quadratischer Abweichungen

(i) $$\frac{\partial \hat{Q}}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)$$

$$= -2 \sum_{i=1}^{n} y_i + 2n\beta_0 + 2\beta_1 \sum_{i=1}^{n} x_i \overset{!}{=} 0$$

(ii) $$\frac{\partial \hat{Q}}{\partial \beta_1} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)\, x_i$$

$$= -2 \sum_{i=1}^{n} x_i y_i + 2\beta_0 \sum_{i=1}^{n} x_i + 2\beta_1 \sum_{i=1}^{n} x_i^2 \overset{!}{=} 0$$

(i) $$n\beta_0 = \sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i$$

$$\Leftrightarrow \quad \beta_0 = \bar{y} - \beta_1 \bar{x}$$

(ii) $$\sum_{i=1}^{n} x_i y_i = \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2$$

Setze (i) in (ii) ein:

$$\sum_{i=1}^{n} x_i y_i = (\bar{y} - \beta_1 \bar{x}) \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2$$

$$\Leftrightarrow \quad \beta_1 = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2} = \frac{\tilde{s}_{XY}}{\tilde{s}_X^2} \qquad \text{vorausgesetzt: } \tilde{s}_X^2 > 0$$

Die KQ-Koeffizienten (Lösungen):

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \quad \hat{\beta}_1 = \frac{\tilde{s}_{XY}}{\tilde{s}_X^2}$$

12.1 Einfaches lineares Regressionsmodell

● Der Fall einer nicht eindeutigen Lösung ●

Falls alle x-Werte gleich sind, ist die KQ-Gerade nicht eindeutig festgelegt.

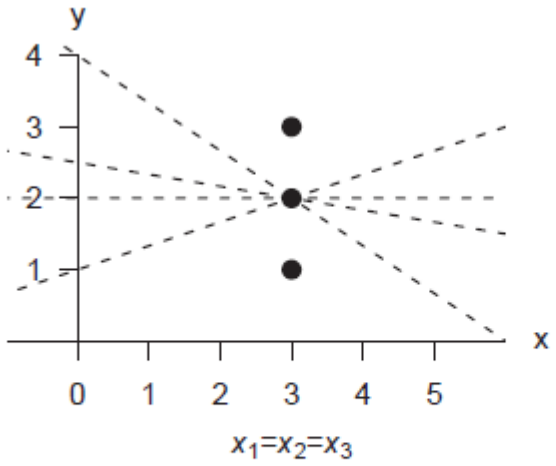


Abb. 5.2.17: L_2-Regression: Beispiel einer nicht eindeutigen Lösung

● Übersetzung in eine Schätzmethode ●

KQ-Schätzer: $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}, \quad \hat{\beta}_1 = \frac{\tilde{S}_{XY}}{\tilde{S}_X^2}$

Eigenschaften der empirischen KQ-Regression

• Übersicht •

Satz 12.1.1: Eigenschaften der empirischen KQ-Regression

1. Die KQ-Gerade geht durch den Schwerpunkt $(\overline{x}, \overline{y})$.
2. Die Summe der gefitteten Werte ist gleich der Summe der y-Werte.
3. Die Summe der KQ-Residuen ist gleich 0.
4. KQ-Residuen und x-Werte sind unkorreliert.
5. Gefittete Werte und KQ-Residuen sind unkorreliert.
6. Es gilt die **Streuungszerlegungsformel der KQ-Regression.**

Achtung! Eigenschaften nur gültig bei KQ-Regression mit β_0 und β_1 (inkl. Achsenabschnitt)

● Die KQ-Gerade geht durch den Schwerpunkt ●

> KQ-Gerade als Funktion: $\hat{y}(x) = \hat{\beta}_0 + \hat{\beta}_1 x$, wobei $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

> Daraus ergibt sich

$$\hat{y}(\bar{x}) = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \left(\bar{y} - \hat{\beta}_1 \bar{x}\right) + \hat{\beta}_1 \bar{x} = \bar{y}$$

● Die Summe der gefitteten Werte ist gleich der Summe der *y*-Werte ●

> Gefitteter Wert: $\hat{y}_i = \hat{y}(x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i$

> Mittelwert der gefitteten Werte: $\bar{\hat{y}} = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i$

> Daraus ergibt sich

$$\bar{\hat{y}} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right) = \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_0 + \hat{\beta}_1 \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$= \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y}$$

$$\Rightarrow \quad \bar{\hat{y}} = \bar{y}$$

$$\Rightarrow \quad \sum_{i=1}^{n} \hat{y}_i = \sum_{i=1}^{n} y_i$$

● Die Summe der KQ-Residuen ist gleich 0 ●

> KQ-Residuum: $\hat{u}_i = y_i - \hat{y}_i$

> Mittelwert der KQ-Residuen: $\bar{\hat{u}} = \frac{1}{n} \sum_{i=1}^{n} \hat{u}_i$

> Daraus ergibt sich

$$\bar{\hat{u}} = \frac{1}{n} \sum_{i=1}^{n} \hat{u}_i = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)$$

$$= \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i = \bar{y} - \bar{\hat{y}} = 0$$

$$\Rightarrow \quad \bar{\hat{u}} = 0$$

$$\Rightarrow \quad \sum_{i=1}^{n} \hat{u}_i = 0$$

● KQ-Residuen und *x*-Werte sind unkorreliert ●

> Kovarianz zwischen Residuen und *x*-Werten:

$$\tilde{s}_{X\hat{U}} = \frac{1}{n}\sum_{i=1}^{n} x_i \hat{u}_i - \bar{x}\bar{\hat{u}} = \frac{1}{n}\sum_{i=1}^{n} x_i \hat{u}_i$$

> Weiter gilt dann (Beweis, siehe LB S. 656):

$$\frac{1}{n}\sum_{i=1}^{n} x_i \hat{u}_i = 0$$

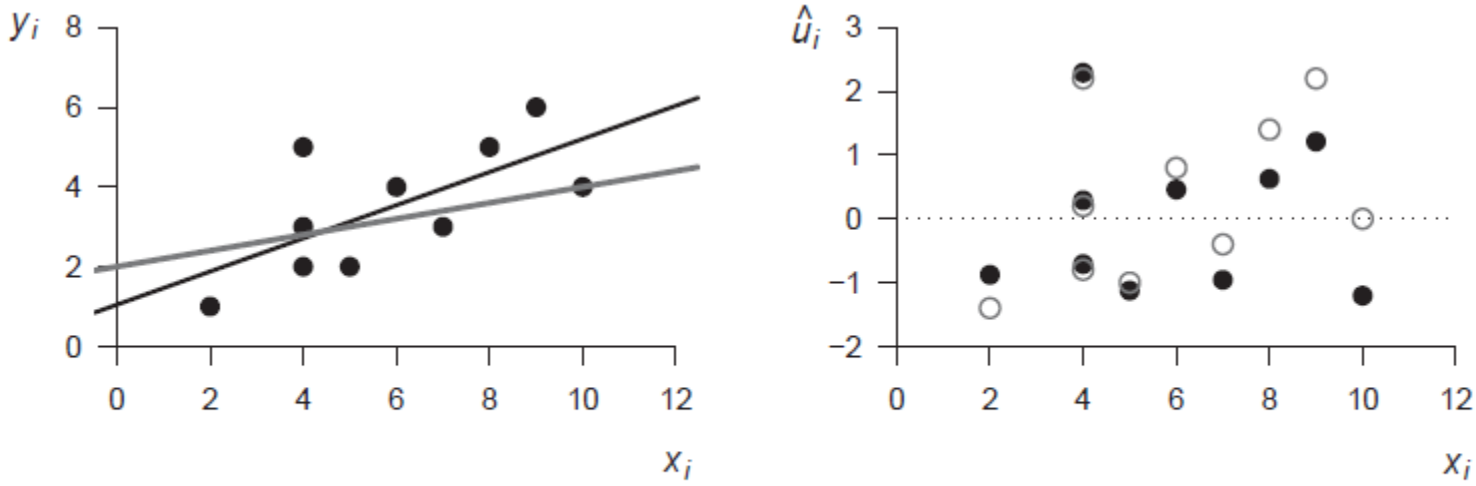


Abb. 12.1.2: Illustration zu Eigenschaft 4

- **Gefittete Werte und KQ-Residuen sind unkorreliert** -

> Kovarianz zwischen gefitteten Werte und KQ-Residuen:

$$\tilde{s}_{\hat{Y}\hat{U}} = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i \hat{u}_i - \bar{\hat{y}}\bar{\hat{u}} = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i \hat{u}_i$$

> Weiter gilt dann:

$$\frac{1}{n}\sum_{i=1}^{n} \hat{y}_i \hat{u}_i = 0$$

Dies ergibt sich aus

$$\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i\hat{u}_i = \frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}_0 + \hat{\beta}_1 x_i)\hat{u}_i$$

$$= \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_0\hat{u}_i + \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_1 x_i\hat{u}_i$$

$$= \hat{\beta}_0 \times \frac{1}{n}\sum_{i=1}^{n}\hat{u}_i + \hat{\beta}_1 \times \frac{1}{n}\sum_{i=1}^{n}x_i\hat{u}_i$$

$$= \hat{\beta}_0\underbrace{\bar{\hat{u}}}_{=0} + \hat{\beta}_1 \times \underbrace{\frac{1}{n}\sum_{i=1}^{n}x_i\hat{u}_i}_{=0}$$

$$= 0$$

- **Es gilt die Streuungszerlegungsformel** -

> Die **Streuungszerlegungsformel der KQ-Regression** lautet in der der Darstellung über Quadratsummen

$$\underbrace{\sum_{i=1}^{n}(y_i-\bar{y})^2}_{\text{Gesamtquadratsumme}} = \underbrace{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}_{\text{Erklärte Quadratsumme}} + \underbrace{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}_{\text{Residuenquadratsumme}}$$

$$\uparrow = \bar{\hat{y}} \qquad = \hat{u}_i^2$$

Gesamtquadratsumme | **Erklärte Quadratsumme** | **Residuenquadratsumme**

> Dividiert man auf beiden Seiten durch n und beachtet $\bar{\hat{u}} = 0$, erhält man

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2}_{\tilde{s}_Y^2} = \underbrace{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i-\bar{\hat{y}}\right)^2}_{\tilde{s}_{\hat{Y}}^2} + \underbrace{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{u}_i-\bar{\hat{u}}\right)^2}_{\tilde{s}_{\hat{U}}^2}$$

Gesamtstreuung = **Erklärte Streuung** + **Residualstreuung**

Bestimmtheitsmaß und Standardfehler der Regression

Bestimmtheitsmaß der KQ-Regression

Gegeben seien metrisch skalierte Beobachtungswerte $(x_1, y_1), \ldots, (x_n, y_n)$ mit $\tilde{s}_X^2 > 0$ und $\tilde{s}_Y^2 > 0$. Im Zusammenhang der KQ-Regression gilt dann:

$$\sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{12.1.14}$$

oder äquivalent

$$\tilde{s}_Y^2 = \tilde{s}_{\hat{Y}}^2 + \tilde{s}_{\hat{U}}^2. \tag{12.1.15}$$

Dabei wird $\tilde{s}_Y^2$ als **Gesamtstreuung**, $\tilde{s}_{\hat{Y}}^2$ als **erklärte Streuung** und $\tilde{s}_{\hat{U}}^2$ als **Residualstreuung** bezeichnet. Die Terme in (12.1.14) heißen **Gesamtquadratsumme**, **erklärte Quadratsumme** und **Residuenquadratsumme**. Darauf basierend ist das Bestimmtheitsmaß definiert als

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{12.1.16}$$

bzw.

$$R^2 = \frac{\tilde{s}^2_{\hat{Y}}}{\tilde{s}^2_Y} = 1 - \frac{\tilde{s}^2_{\hat{U}}}{\tilde{s}^2_Y}. \qquad (12.1.17)$$

Dabei gilt:

(i) $R^2 \in [0, 1]$ und

(ii) $R^2 = r^2_{XY} = \hat{\beta}^2_1 \frac{\tilde{s}^2_X}{\tilde{s}^2_Y}$,

wobei r_{XY} die empirische Korrelation ist.

> Zum Nachweis von $R^2 = r^2_{XY}$ siehe Lehrbuch S. 660

> Nachweis von $r^2_{XY} = \hat{\beta}^2_1 \frac{\tilde{s}^2_X}{\tilde{s}^2_Y}$:

$$r^2_{XY} = \left(\frac{\tilde{s}_{XY}}{\sqrt{\tilde{s}^2_X \tilde{s}^2_Y}}\right)^2 = \frac{\tilde{s}^2_{XY}}{\tilde{s}^2_X \tilde{s}^2_Y} = \frac{\tilde{s}^2_{XY}}{\tilde{s}^4_X} \times \frac{\tilde{s}^2_X}{\tilde{s}^2_Y} = \hat{\beta}^2_1 \frac{\tilde{s}^2_X}{\tilde{s}^2_Y}$$

> $R^2 \in [0,1]$ folgt aus $r_{XY} \in [-1,1]$

• Beispiel 12.1.1 •

Wir betrachten nochmals das Rechenbeispiel 5.2.9 (vgl. Folie 226).

Beobachtungswerte: (1, 1), (2, 2), (3, 1), (4, 3), (5, 2)

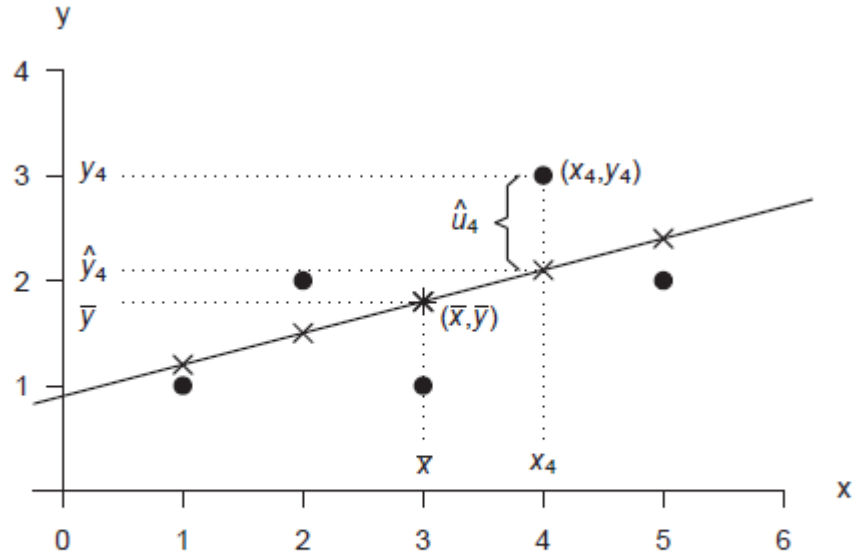


Abb. 5.2.18: L_2-Regression durch den Schwerpunkt

> Berechnung der KQ-Gerade, der gefitteten Werte und der Residuen

12.1 Einfaches lineares Regressionsmodell

Tab. 5.2.4: Arbeitstabelle zur Bestimmung einer KQ-Geraden

i	x_i	y_i	$x_i y_i$	x_i^2	$\hat{y}_i$	$\hat{u}_i$	$\lvert\hat{u}_i\rvert$	$\hat{u}_i^2$
1	1	1	1	1	1.2	−0.2	0.2	0.04
2	2	2	4	4	1.5	0.5	0.5	0.25
3	3	1	3	9	1.8	−0.8	0.8	0.64
4	4	3	12	16	2.1	0.9	0.9	0.81
5	5	2	10	25	2.4	−0.4	0.4	0.16
Summe	15	9	30	55	9	0	2.8	1.9

wird zur Berechnung der KQ-Gerade nicht benötigt

$$\bar{x} = \frac{15}{5} = 3$$

$$\bar{y} = \frac{9}{5} = 1.8$$

$$\hat{\beta}_1 = \frac{\tilde{s}_{XY}}{\tilde{s}_X^2} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2} = \frac{30/5 - 3 \times 1.8}{55/5 - 3^2} = 0.3$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 1.8 - 0.3 \times 3 = 0.9$$

$$\Rightarrow \quad \hat{y}(x) = \hat{\beta}_0 + \hat{\beta}_1 x = 0.9 + 0.3x$$

$$\Rightarrow \quad \hat{y}_1 = 0.9 + 0.3 \times 1 = 1.2, \quad \text{usw.}$$

$$\Rightarrow \quad \hat{u}_1 = y_1 - \hat{y}_1 = 1 - 1.2 = -0.2, \text{ usw.}$$

> Berechnung des Bestimmtheitsmaßes

$$R^2 = \hat{\beta}_1^2 \frac{\tilde{s}_X^2}{\tilde{s}_Y^2} = \hat{\beta}_1^2 \times \frac{1/n \sum_{i=1}^n x_i^2 - \bar{x}^2}{1/n \sum_{i=1}^n y_i^2 - \bar{y}^2}$$

$$= 0.3^2 \times \frac{55/5 - 3^2}{19/5 - 1.8^2} = 0.3^2 \times \frac{2}{0.56}$$

$$\approx 0.3214$$

Alternativ:

$$\tilde{s}_{\hat{U}}^2 = \frac{1}{n} \sum_{i=1}^n \left(\hat{u}_i - \bar{\hat{u}}\right)^2 = \frac{1}{n} \sum_{i=1}^n \hat{u}_i^2 = \frac{1}{5} \times 1.9 = 0.38$$

$$\tilde{s}_Y^2 = \frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2 = 19/5 - 1.8^2 = 0.56$$

$$\Rightarrow \quad R^2 = 1 - \frac{\tilde{s}_{\hat{U}}^2}{\tilde{s}_Y^2} = 1 - \frac{0.38}{0.56} \approx 0.3214$$

● Standardfehler der Regression ●

> Definition: $SER = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2} = \sqrt{\frac{n}{n-2} \times \tilde{s}_{\hat{U}}^2}$

Standard Error of Regression

> Interpretation: ≈ Standardabweichung der KQ-Residuen

⇒ ca. 95% aller Beobachtungen sollten innerhalb eines Schlauches der Breite $4 \times SER$ um die KQ-Gerade herum liegen

> Zusammenhang zum Bestimmtheitsmaß:

Aus $SER^2 = \frac{n}{n-2} \times \tilde{s}_{\hat{U}}^2$ und $R^2 = 1 - \frac{\tilde{s}_{\hat{U}}^2}{\tilde{s}_Y^2}$ folgt

$$SER^2 = \frac{n}{n-2} \times \tilde{s}_Y^2 \times (1 - R^2) \quad \text{bzw.}$$

$$R^2 = 1 - \frac{n-2}{n} \times \frac{SER^2}{\tilde{s}_Y^2}$$

● Beispiel 12.1.2 ●

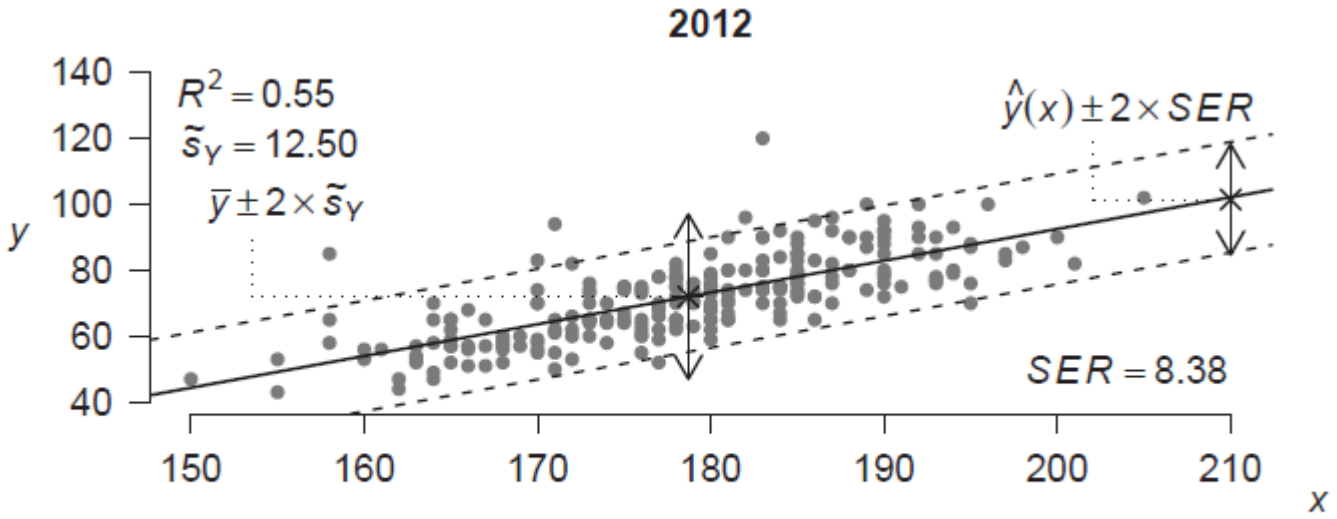


Abb. 12.1.3: Standardfehler der Regression

Sofern man den geschätzten Zusammenhang als wahr unterstellt, kann man das Gewicht mit einer Sicherheit von 95% auf

$$2 \times 8.38kg = 16.76kg$$

genau prognostizieren.

12.1.2 Modellannahmen und theoretische KQ-Regression

Ensembles von Modellannahmen

- **Modell KN: Nichtstochastischer Regressor** -

Definition 12.1.2: Klassisches Modell mit nichtstochastischem Regressor

Wir sprechen von einem **klassischen linearen Regressionsmodell mit nichtstochastischem Regressor**, falls folgende Annahmen gelten:

(A0*) $Y_i = \beta_0 + \beta_1 x_i + U_i$ für $i = 1, \ldots, n$ und

(A1*) die Fehler $U_1, \ldots, U_n$ sind unabhängig $N(0, \sigma_U^2)$-verteilt.

Dabei sind $x_1, \ldots, x_n$ nichtstochastisch und $\tilde{s}_X^2 > 0$.

> (A0*) Speziallfall von (A0) mit $P(X_i = x_i) = 1$

> $(x_1, Y_1), \ldots, (x_n, Y_n)$ sind unabhängig heterogen verteilt

> $(Y_1), \ldots, (Y_n)$ sind unabhängig heterogen verteilt

12.1 Einfaches lineares Regressionsmodell

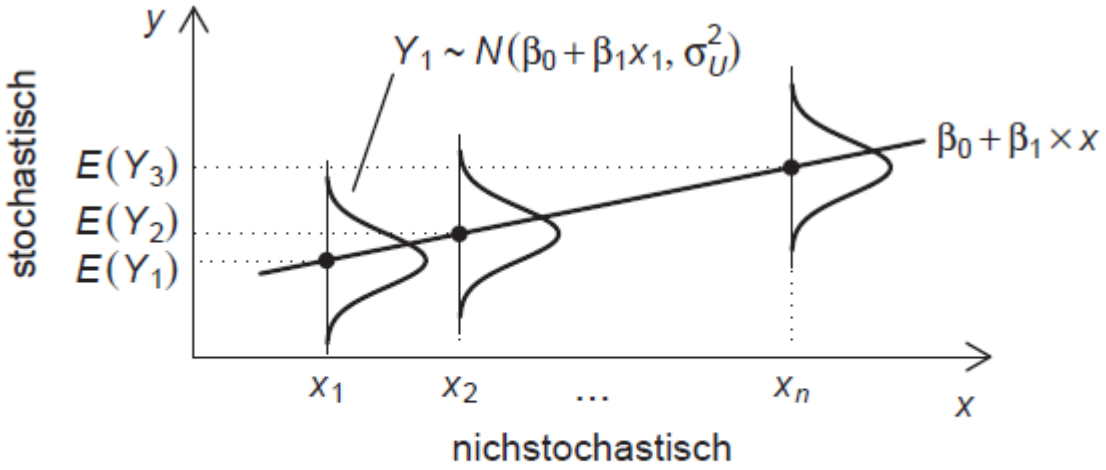


Abb. 12.1.4: Klassisches Modell mit nichtstochastischem Regressor

> $E(Y_i) = E(\beta_0 + \beta_1 x_i + U_i) = E(\beta_0) + E(\beta_1 x_i) + E(U_i)$

$= \beta_0 + \beta_1 x_i$

> $Var(Y_i) = Var(\beta_0 + \beta_1 x_i + U_i) = Var(U_i) = \sigma_U^2$

> $U_i \sim N(0, \sigma_U^2)$

$\Rightarrow \beta_0 + \beta_1 x_i + U_i = Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma_U^2)$

12.1 Einfaches lineares Regressionsmodell

• Modell KS: Stochastischer Regressor •

Definition 12.1.3: Klassisches Modell mit stochastischem Regressor

Wir sprechen von einem **klassischen linearen Regressionsmodell mit stochastischem Regressor**, falls folgende Annahmen gelten:

(A0) $Y_i = \beta_0 + \beta_1 X_i + U_i$ für $i = 1, \ldots, n$,

(A1**) $U_i | X_i = x_i \sim N(0, \sigma_U^2)$ für $i = 1, \ldots, n$,

(A2) $(X_1, Y_1), \ldots, (X_n, Y_n)$ sind u.i.v. und

(A3) alle Zufallsvariablen genügen Annahme (M) aus Definition 8.2.1 und besitzen positive Varianzen.

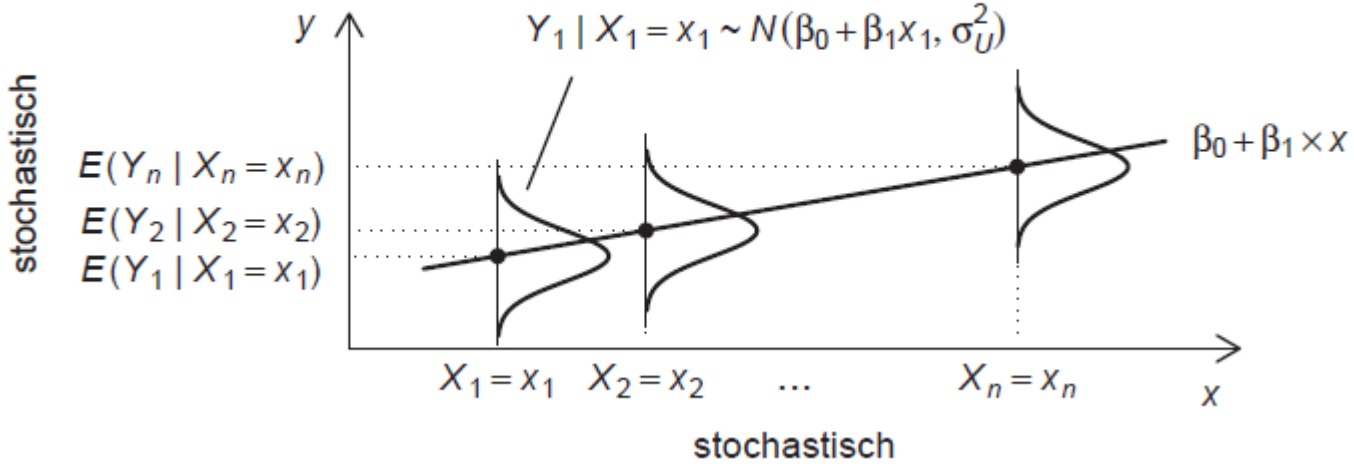
> $U_i | X_i = x_i \sim N(0, \sigma_U^2)$

$\Rightarrow U_i, X_i$ unabhängig

$\Rightarrow \; Cov(X_i, U_i) = 0$

12.1 Einfaches lineares Regressionsmodell

Abb. 12.1.5: Klassisches Modell mit stochastischem Regressor



> Da $U_i = Y_i - \beta_0 - \beta_1 X_i$ folgt aus (A2):

$\Rightarrow$ $(X_1, U_1), \ldots, (X_n, U_n)$ u. i. v.

$\Rightarrow$ $U_1, \ldots, U_n$ u. i. v. mit $U_i \sim N(0, \sigma_U^2)$

> $E(U_i|X_i = x_i) = E(U_i) = 0$

> $Var(U_i|X_i = x_i) = Var(U_i) = \sigma_U^2$

bedingte und **unbedingte Homoskedastizität**

> $E(Y_i|X_i = x_i) = E(\beta_0 + \beta_1 X_i + U_i|X_i = x_i)$

$= E(\beta_0|X_i = x_i) + E(\beta_1 X_i|X_i = x_i) + E(U_i|X_i = x_i)$

$= \beta_0 + \beta_1 x_i + E(U_i|X_i = x_i) = \beta_0 + \beta_1 x_i$

> $Var(Y_i|X_i = x_i) = Var(\beta_0 + \beta_1 X_i + U_i|X_i = x_i)$

$= Var(U_i|X_i = x_i) = \sigma_U^2$

> $U_i \mid X_i = x_i \sim N(0, \sigma_U^2)$

$\Rightarrow \quad Y_i \mid X_i = x_i \sim N(\beta_0 + \beta_1 x_i, \sigma_U^2)$

● Beispiel 12.1.3: Bivariate Normalverteilung ●

Annahmen von Modell KS, wobei (A2) konkretisiert wird zu

$$\text{(A2*) } (X_1, Y_1), \ldots, (X_n, Y_n) \text{ sind u.i.v. wie } (X, Y) \text{ mit } \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix}\right).$$

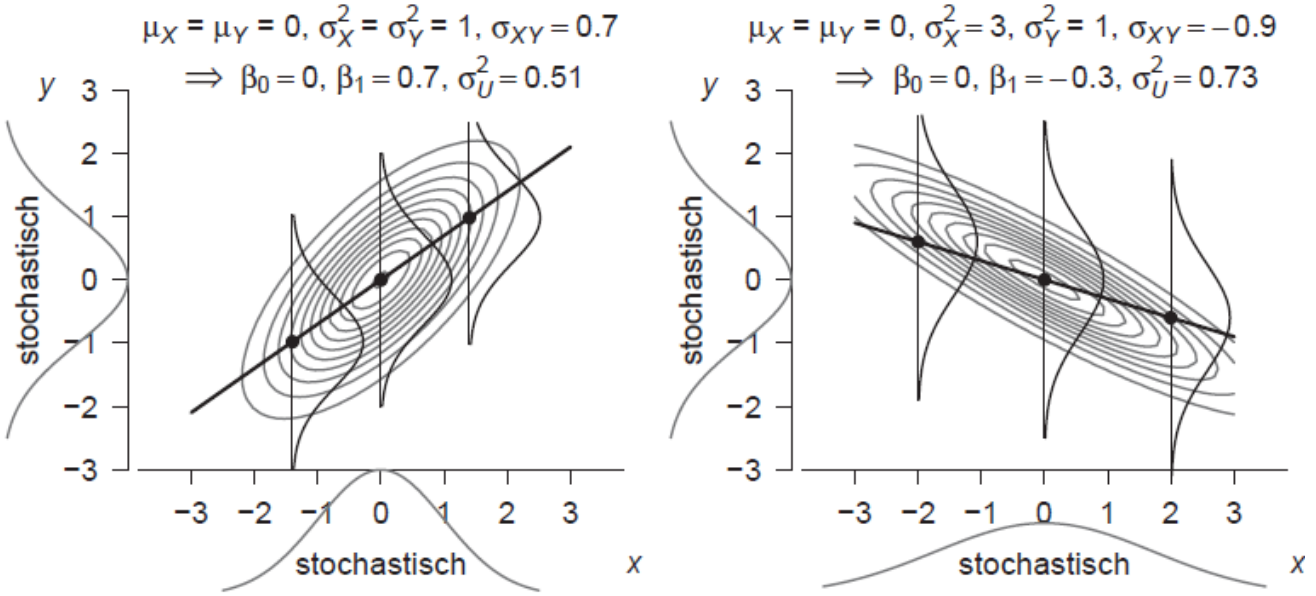


Abb. 12.1.6: Klassisches Modell bei bivariater Normalverteilung

• Modell BH: Bedingt heteroskedastischer Fehler •

Definition 12.1.4: Modell mit bedingt heteroskedastischem Fehler

Wir sprechen von einem **linearen Regressionsmodell mit bedingt heteroskedastischem Fehler**, falls gilt:

(A0) $Y_i = \beta_0 + \beta_1 X_i + U_i$ für $i = 1, \ldots, n$,

(A1) $E(U_i | X_i = x_i) = 0$ für $i = 1, \ldots, n$,

(A2) $(X_1, Y_1), \ldots, (X_n, Y_n)$ sind u.i.v. und

(A3) alle Zufallsvariablen genügen Annahme (M) aus Definition 8.2.1 und besitzen positive Varianzen.

> X_i, U_i nicht zwingend unabhängig, jedoch unkorreliert (siehe Folie 877)

> Da $U_i = Y_i - \beta_0 - \beta_1 X_i$ folgt aus (A2):

$\Rightarrow$ $(X_1, U_1), \ldots, (X_n, U_n)$ u. i. v.

$\Rightarrow$ $U_1, \ldots, U_n$ u. i. v.

12.1 Einfaches lineares Regressionsmodell

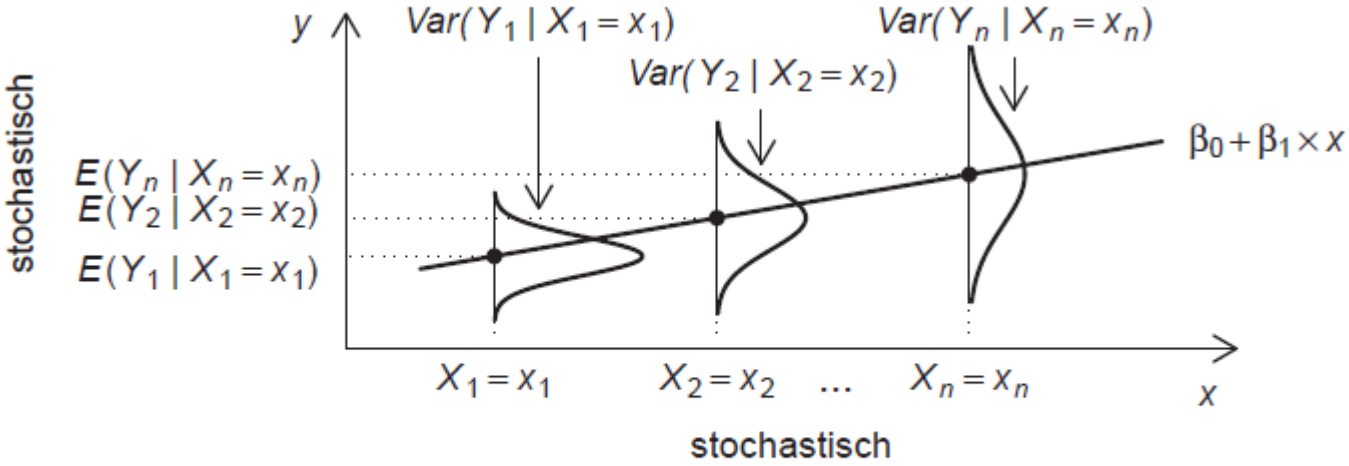


Abb. 12.1.7: Modell mit bedingt heteroskedastischem Fehler

> $E(U_i) = E[E(U_i|X_i)] = E(0) = 0$

> Keine Aussage über $Var(U_i|X_i = x_i)$:

⇒ $Var(U_i|X_i = x_i) = const.$ kann nicht o. W. angenommen werden

⇒ **bedingte Homoskedastizität** nicht zwingend gegeben,
stattdessen **bedingte Heteroskedastizität** möglich

⇒

> $E(Y_i|X_i = x_i) = E(\beta_0 + \beta_1 X_i + U_i|X_i = x_i)$

$= \cdots = \beta_0 + \beta_1 x_i$

> Keine Verteilungsaussage über U_i oder Y_i

> $Var(U_i) = E[Var(U_i|X_i)] + Var[E(U_i|X_i)]$

$= E[Var(U_i|X_i)] =: \sigma_U^2$ **Unbedingte Homoskedastizität**

> Fehler und Regressor sind zwar nicht zwingend unabhängig, jedoch unkorreliert:

$$Cov(X_i, U_i) = E(X_i U_i) - E(X_i)\underbrace{E(U_i)}_{=0}$$
$$= E[E(X_i U_i)|X_i]$$
$$= E[X_i E(U_i|X_i)] = 0$$

> $Var(Y_i|X_i = x_i) = Var(\beta_0 + \beta_1 X_i + U_i|X_i = x_i)$

$= \cdots = Var(U_i|X_i = x_i)$

12.1 Einfaches lineares Regressionsmodell

• Zusammenfassung •

Tab. 12.1.1: Modellimmanente Eigenschaften der verschiedenen Annahmeensembles

Eigenschaft	KN	KS	BH	UHV
(1) $Y_i = \beta_0 + \beta_1 X_i + U_i$	✓	✓	✓	✓
(2) $(X_1, Y_1), \ldots, (X_n, Y_n)$ sind unabhängig	✓	✓	✓	✓
(3) $(X_1, Y_1), \ldots, (X_n, Y_n)$ sind identisch verteilt	✗	✓	✓	✗
(4) $U_1, \ldots, U_n$ sind unabhängig	✓	✓	✓	✓
(5) $U_1, \ldots, U_n$ sind identisch verteilt	✓	✓	✓	✗
(6) $U_1, \ldots, U_n$ sind $N(0, \sigma_U^2)$-verteilt	✓	✓	✗	✗
(7) U_i und X_i sind unabhängig	✓	✓	✗	✗
(8) $Cov(X_i, U_i) = E(X_i U_i) = 0$	✓	✓	✓	✓
(9) $E(U_i) = E(U_i \mid X_i = x_i) = 0$	✓	✓	✓	✓
(10) $E(Y_i \mid X_i = x_i) = \beta_0 + \beta_1 x_i$	✓	✓	✓	✓
(11) $Var(U_i) = E[Var(U_i \mid X_i)]$	✓	✓	✓	✓
(12) $Var(U_i) = \sigma_U^2$ (unbedingte Homoskedastizität)	✓	✓	✓	✗

(13) $Var(U_i\|X_i = x_i) = \sigma_U^2$	✓	✓	✗	✗
(bedingte Homoskedastizität)				
(14) $U_i\|X_i = x_i \sim N(0, \sigma_U^2)$	✓	✓	✗	✗
(15) $Y_i\|X_i = x_i \sim N(\beta_0 + \beta_1 x_i, \sigma_U^2)$	✓	✓	✗	✗
(16) $\beta_0 = \mu_Y - \beta_1\mu_X, \beta_1 = \sigma_{XY}/\sigma_X^2$	✗	✓	✓	✗
(17) $\sigma_U^2 = \sigma_Y^2 - \sigma_{XY}^2/\sigma_X^2$	✗	✓	✓	✗

zu (16) und (17) siehe Satz 12.1.2 (Folie 884)

- **Alternativ auch (X_i, U_i) statt (X_i, Y_i)**
- **Sonstige Verallgemeinerungen**

Theoretische KQ-Regression

● Überblick ●

● Definition und Eigenschaften ●

> Theoretische KQ-Regression für 2 Zufallsvariablen X und Y:

$$\min_{\beta_0,\beta_1} \bar{\bar{Q}}(\beta_0,\beta_1) \quad \text{mit} \quad \bar{\bar{Q}}(\beta_0,\beta_1) = E[(Y-\beta_0-\beta_1 X)^2]$$

> Ziel: Minimierung des ***MSE*** oder ***MSFE*** (Mean Squared Forecast Error) durch besten linearen Prädiktor

> Zunächst einmal gilt:

$$E[(Y-\beta_0-\beta_1 X)^2]$$

$$= Var(Y-\beta_0-\beta_1 X) + [E(Y-\beta_0-\beta_1 X)]^2$$

$$= \sigma_Y^2 - 2\beta_1\sigma_{XY} + \beta_1^2\sigma_X^2 + (\mu_Y-\beta_0-\beta_1\mu_X)^2$$

(i) $$\frac{\partial \bar{\bar{Q}}}{\partial \beta_0} = -2(\mu_Y - \beta_0 - \beta_1 \mu_X) \overset{!}{=} 0$$

$$\Leftrightarrow \quad \beta_0 = \mu_Y - \beta_1 \mu_X$$

(ii) $$\frac{\partial \bar{\bar{Q}}}{\partial \beta_1} = -2\sigma_{XY} + 2\beta_1 \sigma_X^2 - 2\mu_X(\mu_Y - \beta_0 - \beta_1 \mu_X) \overset{!}{=} 0$$

Setze (i) in (ii) ein:

$$-2\sigma_{XY} + 2\beta_1 \sigma_X^2 = 0$$

$$\Leftrightarrow \quad \beta_1 = \frac{\sigma_{XY}}{\sigma_X^2}$$

Die theoretischen KQ-Koeffizienten (Lösungen):

$$\bar{\bar{\beta}}_0 = \mu_Y - \bar{\bar{\beta}}_1 \mu_X, \quad \bar{\bar{\beta}}_1 = \frac{\sigma_{XY}}{\sigma_X^2}$$

Theoretische KQ-Regression

Gegeben seien zwei Zufallsvariablen X und Y mit

$$\sigma_Y^2 < \infty \quad \text{und} \quad 0 < \sigma_X^2 < \infty.$$

Dann sind die ***theoretischen KQ-Koeffizienten*** einer **theoretischen KQ-Regression** von Y auf X die Lösungen für das Problem

$$\min_{\beta_0,\beta_1} \overline{\widehat{Q}}(\beta_0, \beta_1) \quad \text{mit } \overline{\widehat{Q}}(\beta_0, \beta_1) = E[(Y - \beta_0 - \beta_1 X)^2].$$

Dabei gilt

$$\overline{\widehat{\beta}}_1 = \sigma_{XY}/\sigma_X^2 \quad \text{und} \quad \overline{\widehat{\beta}}_0 = \mu_Y - \overline{\widehat{\beta}}_1 \mu_X, \text{ wobei}$$

$$\overline{\widehat{y}}(x) = \overline{\widehat{\beta}}_0 + \overline{\widehat{\beta}}_1 x$$

als **theoretische KQ-Gerade** bezeichnet wird. Der **beste lineare MSE-Prädiktor** von Y auf Basis von X ist gegeben durch

$$\overline{\widehat{Y}} = \overline{\widehat{y}}(X) = \overline{\widehat{\beta}}_0 + \overline{\widehat{\beta}}_1 X.$$

Dabei gilt:

$$E(Y - \overline{\widehat{Y}}) = 0 \quad \text{und} \tag{12.1.38}$$

$$Var(Y - \overline{\widehat{Y}}) = \sigma_Y^2(1 - \varrho_{XY}^2). \tag{12.1.39}$$

Analog zu den Eigenschaften der empirischen KQ-Regression (Folie 855) gelten folgende Eigenschaften:

(1) $\bar{\hat{y}}(\mu_X) = \mu_Y$

(2) $E\left(\bar{\hat{Y}}\right) = E(Y) = \mu_Y$

(3) $E\left(Y - \bar{\hat{Y}}\right) = 0$

(4) $Cov\left(Y - \bar{\hat{Y}}, X\right) = 0$

(5) $Cov\left(Y - \bar{\hat{Y}}, \bar{\hat{Y}}\right) = 0$

(6) $\underbrace{Var(Y) = Var\left(\bar{\hat{Y}}\right) + Var\left(Y - \bar{\hat{Y}}\right)}$

Theoretische Streuungszerlegungsformel der KQ-Regression

$$\Rightarrow \quad \Re^2 = \frac{Var\left(\bar{\hat{Y}}\right)}{Var(Y)} = 1 - \frac{Var\left(Y - \bar{\hat{Y}}\right)}{Var(Y)}$$

Theoretisches Bestimmtheitsmaß

● Zusammenhang zur Modellgeraden in den Modellen KS und BH ●

Satz 12.1.2: Theoretische Regressionsgerade = KQ-Gerade

Die in den Modellen KS und BH (Definitionen 12.1.3 und 12.1.4) enthaltene theoretische Regressionsgerade ist die KQ-Gerade einer theoretischen Regression von Y auf X. Dabei gilt:

$$\beta_0 = \mu_Y - \beta_1 \mu_X, \quad \beta_1 = \sigma_{XY}/\sigma_X^2 \quad \text{und} \quad \sigma_U^2 = \sigma_Y^2(1 - \varrho_{XY}^2),$$

wobei

$$\mu_Y = E(Y_i), \quad \mu_X = E(X_i), \quad \sigma_X^2 = Var(X_i), \quad \sigma_Y^2 = Var(Y_i),$$
$$\sigma_{XY} = Cov(X_i, Y_i), \quad \varrho_{XY} = Corr(X_i, Y_i) \quad \text{und} \quad Var(U_i) = \sigma_U^2$$

für $i = 1, \ldots, n$ sind.

> Vgl. Eigenschaften (16) und (17) auf Folie 879

> Es gelten die Eigenschaften (9) und (8)

$$E(U_i) = 0 \quad \text{und} \quad Cov(X_i, U_i) = 0$$

> Mit Annahme (A2), d. h. insbesondere der identischen Verteilung, folgt

$$\mu_Y = E(Y_i) = E(\beta_0 + \beta_1 X_i + U_i)$$

$$= E(\beta_0) + \beta_1 E(X_i) + E(U_i)$$

$$= \beta_0 + \beta_1 \mu_X$$

$$\Leftrightarrow \quad \beta_0 = \mu_Y - \beta_1 \mu_X$$

> Andererseits folgt:

$$\sigma_{XY} = Cov(X_i, Y_i) = Cov(X_i, \beta_0 + \beta_1 X_i + U_i)$$

$$= \beta_1 Cov(X_i, X_i) + Cov(X_i, U_i) = \beta_1 \sigma_X^2$$

$$\Leftrightarrow \quad \beta_1 = \frac{\sigma_{XY}}{\sigma_X^2}$$

> Weiter gilt:

$$\sigma_U^2 = Var(U_i) = Var(Y_i - \beta_0 - \beta_1 X_i)$$

$$= Var\left(Y_i - \frac{\sigma_{XY}}{\sigma_X^2} X_i\right)$$

$$= Var(Y_i) + \left(\frac{\sigma_{XY}}{\sigma_X^2}\right)^2 Var(X_i) - 2\frac{\sigma_{XY}}{\sigma_X^2} Cov(X_i, Y_i)$$

$$= \sigma_Y^2 + \frac{\sigma_{XY}^2}{\sigma_X^2} - 2\frac{\sigma_{XY}^2}{\sigma_X^2} = \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2} = \sigma_Y^2\left(1 - \frac{\sigma_{XY}^2}{\sigma_Y^2 \sigma_X^2}\right)$$

$$= \sigma_Y^2(1 - \rho_{XY}^2)$$

- **Zusammenhang zur Modellgeraden in den Modellen KN und UHV** -

Zusätzliche Konvergenzannahmen benötigt...

12.1.3 Verteilungstheoretische Grundlagen

Verteilungen der KQ-Schätzer

● **Alternative Darstellungen der KQ-Schätzer** ●

Zur Herleitung der Verteilungen der KQ-Schätzer

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \quad \text{und} \quad \hat{\beta}_1 = \tilde{S}_{XY} / \tilde{S}_X^2$$

erweisen sich gewisse alternative Darstellungen für die Formeln der KQ-Schätzer als hilfreich.

Satz 12.1.3: Alternative Darstellungen der KQ-Schätzer

Im einfachen linearen Regressionsmodell gilt, wenn $\tilde{S}_X^2 > 0$:

$$\hat{\beta}_0 = \beta_0 - \overline{X}(\hat{\beta}_1 - \beta_1) + \overline{U} \quad \text{und} \tag{12.1.48}$$

$$\hat{\beta}_1 = \beta_1 + V \quad \text{mit } V = \frac{1}{\tilde{S}_X^2} \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X}) U_i. \tag{12.1.49}$$

> Nachweis von (12.1.48):

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n}(\beta_0 + \beta_1 X_i + U_i) = \boxed{\beta_0 + \beta_1 \bar{X} + \bar{U}}$$

$$\Rightarrow \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = \boxed{\beta_0 + \beta_1 \bar{X} + \bar{U}} - \hat{\beta}_1 \bar{X}$$

$$= \beta_0 - \bar{X}\left(\hat{\beta}_1 - \beta_1\right) + \bar{U}$$

> Nachweis von (12.1.49):

$$\tilde{S}_{XY} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(\beta_0 + \beta_1 X_i + U_i - \beta_0 - \beta_1 \bar{X} - \bar{U})$$

$$= \beta_1 \times \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 + \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(U_i - \bar{U})$$

$$= \beta_1 \times \tilde{S}_X^2 + \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})U_i - \bar{U} \times \underbrace{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})}_{=0}$$

$$= \beta_1 \times \tilde{S}_X^2 + \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})U_i$$

$$\Rightarrow \quad \hat{\beta}_1 = \tilde{S}_{XY}/\tilde{S}_X^2 = \beta_1 + \frac{1}{\tilde{S}_X^2} \times \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})U_i$$

● Verteilung der KQ-Schätzer im Modell KN ●

> Herleitung der Verteilung von $\hat{\beta}_1$:

$$\hat{\beta}_1 - \beta_1 = \frac{1}{\tilde{s}_X^2} \times \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})U_i = \sum_{i=1}^{n} c_i U_i$$

$\Rightarrow$ gewichtete Summe der Fehler mit Gewichten

$$c_i = \frac{x_i - \bar{x}}{n\tilde{s}_X^2}$$

Da die Fehler U_i unabhängig $N(0, \sigma_U^2)$-verteilt sind, gilt:

$$c_i U_i \sim N\left(0, c_i^2 \sigma_U^2\right)$$

$$\Rightarrow \quad \hat{\beta}_1 - \beta_1 = \sum_{i=1}^{n} c_i U_i \;\sim\; N\left(0, \sum_{i=1}^{n} c_i^2 \sigma_U^2\right)$$

Mit

$$\sum_{i=1}^{n} c_i^2 \sigma_U^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n^2 \tilde{s}_X^4} \sigma_U^2 = \frac{\sigma_U^2}{n \tilde{s}_X^4} \times \underbrace{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}_{= \tilde{s}_X^2} = \frac{\sigma_U^2}{n \tilde{s}_X^2}$$

erhält man daraus dann

$$\hat{\beta}_1 - \beta_1 \;\sim\; N\left(0, \frac{\sigma_U^2}{n \tilde{s}_X^2}\right) \quad \text{bzw.} \quad \hat{\beta}_1 \;\sim\; N\left(\beta_1, \frac{\sigma_U^2}{n \tilde{s}_X^2}\right)$$

> Herleitung der Verteilung von $\hat{\beta}_0$:

$$\hat{\beta}_0 - \beta_0 = -\bar{x}\left(\hat{\beta}_1 - \beta_1\right) + \bar{U} = -\bar{x}\sum_{i=1}^{n} c_i U_i + \frac{1}{n}\sum_{i=1}^{n} U_i$$

$$= \sum_{i=1}^{n} -\bar{x} c_i U_i + \sum_{i=1}^{n} \frac{1}{n} \times U_i = \sum_{i=1}^{n} k_i U_i$$

$\Rightarrow$ gewichtete Summe der Fehler mit Gewichten

$$k_i = -\bar{x} c_i + \frac{1}{n}$$

$$\Rightarrow \quad \hat{\beta}_0 - \beta_0 = \sum_{i=1}^{n} k_i U_i \;\sim\; N\left(0, \sum_{i=1}^{n} k_i^2 \sigma_U^2\right)$$

Mit

$$\sum_{i=1}^{n} k_i^2 = \sum_{i=1}^{n} \bar{x}^2 c_i^2 - \frac{2\bar{x}}{n}\underbrace{\sum_{i=1}^{n} c_i}_{=0} + \underbrace{\sum_{i=1}^{n} \frac{1}{n^2}}_{\frac{n}{n^2} = \frac{1}{n}}$$

$$\uparrow$$

$$= \frac{(x_i - \bar{x})^2}{n^2 \tilde{s}_X^4}$$

$$= \frac{\bar{x}^2}{n^2 \tilde{s}_X^4} \sum_{i=1}^{n} (x_i - \bar{x})^2 + \frac{1}{n} = \frac{\bar{x}^2}{n\tilde{s}_X^4} \times \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 + \frac{1}{n}$$

$$= \frac{\bar{x}^2}{n\tilde{s}_X^4} \times \tilde{s}_X^2 + \frac{\tilde{s}_X^2}{n\tilde{s}_X^2} = \frac{\bar{x}^2 + \tilde{s}_X^2}{n\tilde{s}_X^2}$$

Wegen $\tilde{s}_X^2 + \bar{x}^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2$

erhalten wir schließlich

$$\hat{\beta}_0 - \beta_0 \sim N\left(0, \frac{\frac{1}{n}\sum_{i=1}^{n} x_i^2 \, \sigma_U^2}{n\tilde{s}_X^2}\right) \quad \text{bzw.} \quad \hat{\beta}_0 \sim N\left(\beta_0, \frac{\frac{1}{n}\sum_{i=1}^{n} x_i^2 \, \sigma_U^2}{n\tilde{s}_X^2}\right)$$

12.1 Einfaches lineares Regressionsmodell

> Einige verteilungstheoretische Implikationen

$$E(\hat{\beta}_0) = \beta_0, \quad E(\hat{\beta}_1) = \beta_1 \qquad \text{(Erwartungstreue)}$$

$$Var(\hat{\beta}_0) = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i^2 \sigma_U^2}{n\tilde{s}_X^2} \qquad \text{(vgl. Abb. 12.1.9)}$$

$$Var(\hat{\beta}_1) = \frac{\sigma_U^2}{n\tilde{s}_X^2} \qquad \text{(vgl. Abb. 12.1.8)}$$

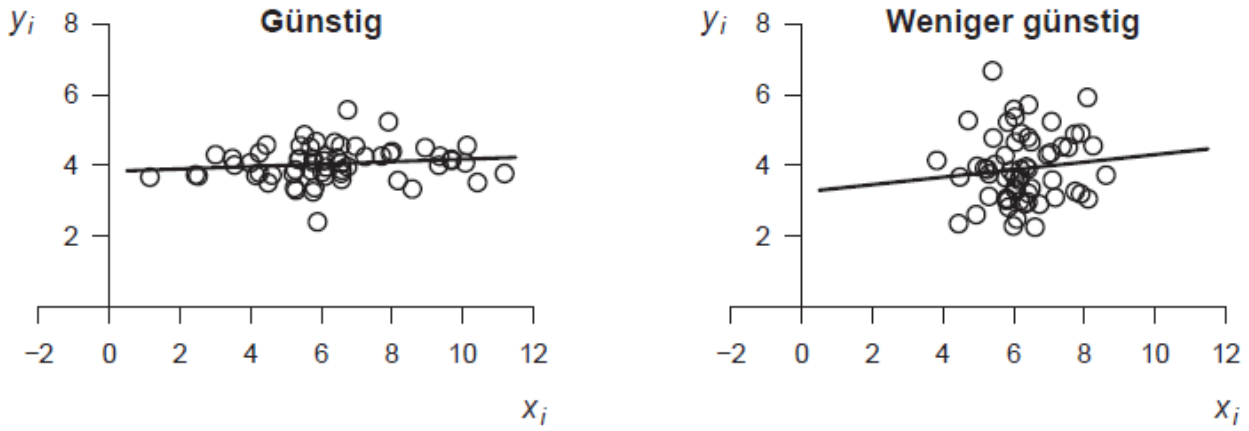


Abb. 12.1.8: Schätzung des Steigungskoeffizienten

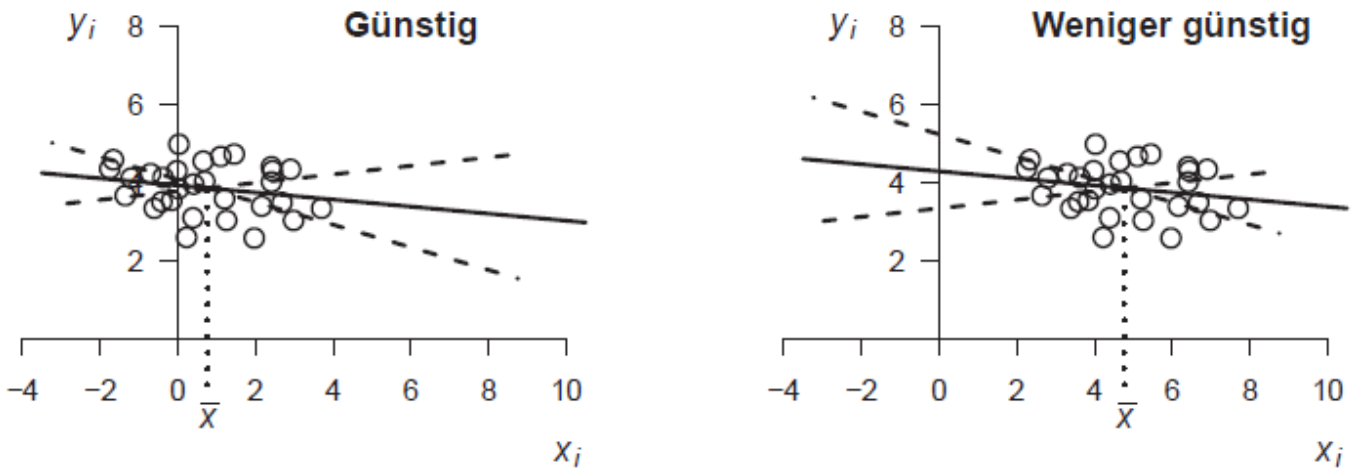


Abb. 12.1.9: Schätzung des Achsenabschnitts

● Verteilung der KQ-Schätzer im Modell KS ●

> Da Regressoren stochastisch, jetzt Groß- statt Kleinschreibweise:

$$\hat{\beta}_1 - \beta_1 = \sum_{i=1}^{n} C_i U_i \quad \text{mit} \quad C_i = \frac{X_i - \bar{X}}{n\tilde{S}_X^2}$$

> Da die Gewichte C_i nun stochastisch sind, sind die Terme $C_i U_i$ i. A. nicht mehr normalverteilt.

> Deshalb jetzt bedingte Betrachtung („Trick“):

$$C_i U_i | X_1, \dots, X_n \sim N\left(0, C_i^2 \sigma_U^2\right) = N\left(0, \frac{(X_i - \bar{X})^2 \sigma_U^2}{n^2 \left(\tilde{S}_X^2\right)^2}\right) \quad \text{für} \quad \tilde{S}_X^2 > 0$$

$$\Rightarrow \quad \hat{\beta}_1 - \beta_1 | X_1, \dots, X_n \sim N\left(0, \frac{\sigma_U^2}{n \tilde{S}_X^2}\right) \quad \text{und}$$

$$\hat{\beta}_0 - \beta_0 | X_1, \dots, X_n \sim N\left(0, \frac{\frac{1}{n} \sum_{i=1}^n X_i^2 \, \sigma_U^2}{n \tilde{S}_X^2}\right)$$

bzw.

$$\hat{\beta}_1 | X_1, \dots, X_n \sim N\left(\beta_1, \frac{\sigma_U^2}{n \tilde{S}_X^2}\right) \quad \text{und}$$

$$\hat{\beta}_0 | X_1, \dots, X_n \sim N\left(\beta_0, \frac{\frac{1}{n} \sum_{i=1}^n X_i^2 \, \sigma_U^2}{n \tilde{S}_X^2}\right)$$

> Verteilungstheoretische Implikationen

$$E(\hat{\beta}_0) = E\left[E(\hat{\beta}_0|X_1, \dots, X_n)\right] = E[\beta_0] = \beta_0$$

$$E(\hat{\beta}_1) = E\left[E(\hat{\beta}_1|X_1, \dots, X_n)\right] = E[\beta_1] = \beta_1$$

} erwartungstreu (bedingt und unbedingt)

$$Var(\hat{\beta}_0|X_1, \dots X_n) = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i^2 \, \sigma_U^2}{n\tilde{S}_X^2}$$

$$Var(\hat{\beta}_1|X_1, \dots X_n) = \frac{\sigma_U^2}{n\tilde{S}_X^2}$$

• Verteilung des KQ-Schätzers für β_1 im Modell BH •

> Da $(X_1, Y_1), \ldots, (X_n, Y_n)$ u. i. v. und $U_i = Y_i - \beta_0 - \beta_1 X_i$, sind auch $(X_1, U_1), \ldots, (X_n, U_n)$ u.i.v.

$\Rightarrow$ $(X_1 - \mu_X)U_1, \ldots, (X_n - \mu_X)U_n$ sind u.i.v. mit

$$E[(X_i - \mu_X)U_i] = E\big[E\big((X_i - \mu_X)U_i | X_i\big)\big] = E[(X_i - \mu_X)\underbrace{E(U_i|X_i)}_{=0}]$$
$$= E[(X_i - \mu_X) \times 0] = 0$$

und

$$Var[(X_i - \mu_X)U_i] = E\big[(X_i - \mu_X)^2 U_i^2\big] = E\left[E\left((X_i - \mu_X)^2 U_i^2 | X_i\right)\right]$$
$$= E\big[(X_i - \mu_X)^2 E\big(U_i^2 | X_i\big)\big]$$
$$= E[(X_i - \mu_X)^2 \times Var(U_i|X_i)]$$

$(X_1 - \mu_X)U_1, \ldots, (X_n - \mu_X)U_n$ sind u.i.v. mit Erwartungswert 0.

> Anwendung des ZGWS für u.i.v.-Zufallsvariablen (Satz 7.4.3, Folie 549):

$$Y_n = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_X)U_i}{\sqrt{\frac{Var\left((X_i - \mu_X)U_i\right)}{n}}} \overset{a}{\sim} N(0,1)$$

Ebenso gilt:

$$(*)\ Z_n = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})U_i}{\sqrt{\frac{Var\left((X_i - \mu_X)U_i\right)}{n}}} \overset{a}{\sim} N(0,1)$$

> Nachweis von $(*)$:

Mit der Zerlegung des Zählerausdrucks

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})\,U_i = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_X + \mu_X - \bar{X})\,U_i$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_X)\,U_i + \frac{1}{n}\sum_{i=1}^{n}(\mu_X - \bar{X})U_i$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_X)\,U_i + (\mu_X - \bar{X})\bar{U}$$

folgt die Zerlegung von $(*)$ in

$$\underbrace{\frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})U_i}{\sqrt{\frac{Var\left((X_i - \mu_X)U_i\right)}{n}}}}_{= Z_n \overset{a}{\sim} N(0,1)} = \underbrace{\frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_X)U_i}{\sqrt{\frac{Var\left((X_i - \mu_X)U_i\right)}{n}}}}_{= Y_n \overset{a}{\sim} N(0,1)} + \underbrace{\frac{(\mu_X - \bar{X})\bar{U}}{\sqrt{\frac{Var\left((X_i - \mu_X)U_i\right)}{n}}}}_{= A_n \xrightarrow{p} 0}$$

Das Verteilungsresultat für Z_n resultiert aus Slutsky‘s Theorem (Folie 583)

Damit wäre $(*)$ nachgewiesen. Allerdings fehlt noch der ...

> Nachweis von $A_n \xrightarrow{p} 0$:

$$A_n = \underbrace{\frac{(\mu_X - \bar{X})\sigma_U}{Var\big((X_i - \mu_X)U_i\big)}}_{= B_n \xrightarrow{p} 0} \times \underbrace{\frac{\bar{U}}{\sigma_U/n}}_{= W_n \overset{a}{\sim} N(0,1)}$$

Slutsky‘s Theorem $\Rightarrow A_n \xrightarrow{p} 0$

Da $U_1, \dots, U_n$ u.i.v. sind mit $E(U_i) = 0$ und $Var(U_i) = \sigma_U^2$, folgt $W_n \overset{a}{\sim} N(0,1)$ mit dem ZGWS. Andererseits sind $X_1, \dots, X_n$ u.i.v. mit $E(X_i) = \mu_X$ und $Var(X_i) = \sigma_X^2$. Deshalb folgt mit dem GGZ:

$$\bar{X} \xrightarrow{p} \mu_X \text{ bzw. } \bar{X} - \mu_X \xrightarrow{p} 0$$

und schließlich $B_n \xrightarrow{p} 0$.

Mit Slutsky‘s Theorem folgt dann $A_n \xrightarrow{p} 0$.

> Für Resultat $(*)$ auf Folie 898 können wir alternativ auch schreiben

$$(**) \quad \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \bar{X})U_i \overset{a}{\sim} N\left(0, Var\left((X_i - \mu_X)U_i\right)\right)$$

> Kombiniere Satz 12.1.3 (Folie 887) mit $(**)$ + Slutsky's Theorem:

$$\sqrt{n}\left(\hat{\beta}_1 - \beta_1\right) = \underbrace{\frac{1}{\tilde{S}_X^2}}_{\substack{\text{Satz 10.1.4,} \\ \text{Folie 657}}} \times \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \bar{X})U_i}_{(**)} \overset{a}{\sim} N\left(0, \frac{Var\left((X_i - \mu_X)U_i\right)}{(\sigma_X^2)^2}\right)$$

$$\xrightarrow{p} 1/\sigma_X^2$$

$$\Rightarrow \quad \sqrt{n}\left(\hat{\beta}_1 - \beta_1\right) \overset{a}{\sim} N\left(0, \frac{Var\left((X_i - \mu_X)U_i\right)}{(\sigma_X^2)^2}\right)$$

$$\Leftrightarrow \quad \hat{\beta}_1 - \beta_1 \overset{approx.}{\sim} N\left(0, \frac{Var\left((X_i - \mu_X)U_i\right)}{n(\sigma_X^2)^2}\right)$$

$$\Leftrightarrow \quad \hat{\beta}_1 \overset{approx.}{\sim} N\left(\beta_1, \frac{Var\left((X_i - \mu_X)U_i\right)}{n(\sigma_X^2)^2}\right)$$

> Verteilungstheoretische Implikationen

$$E(\hat{\beta}_1) = \beta_1 + E\left[\frac{1}{\tilde{S}_X^2} \times \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})U_i\right]$$

$$= \beta_1 + E\left[E\left(\frac{1}{\tilde{S}_X^2} \times \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})U_i \,|X_1, \dots, X_n\right)\right]$$

$$= \beta_1 + E\left[\frac{1}{\tilde{S}_X^2} \times \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})\underbrace{E(U_i|X_1, \dots, X_n)}_{=E(U_i|X_i) = 0}\right]$$

$= \beta_1$ Annahme (A1)

$\Rightarrow$ erwartungstreu (unbedingt)

$$Var\left(\sqrt{n}(\hat{\beta}_1 - \beta_1)\right) \approx \frac{1}{(\sigma_X^2)^2}\underbrace{E[(X_i - \mu_X)^2 Var(U_i|X_i)]}_{\text{(Folie 897)} \quad = Var\left((X_i - \mu_X)U_i\right)}$$

> Deutung der approximativen Varianz

$$E[(X_i - \mu_X)^2 Var(U_i|X_i)]$$

$$= Cov\big((X_i - \mu_X)^2, Var(U_i|X_i)\big) + \sigma_X^2 \underbrace{E[Var(U_i|X_i)]}_{= \mathrm{Var}(U_i) = \sigma_U^2}$$

Man beachte hierzu (vgl. Satz 7.2.7)

$$E[Var(U_i|X_i)] = Var(U_i) - Var[\underbrace{E(U_i|X_i)}_{=0}] = Var(U_i)$$

Interpretation:

Bei gegebenen Varianzen von Regressor und Fehler (σ_X^2 bzw. σ_U^2) wird die Varianz des KQ-Schätzers $\hat{\beta}_1$ umso größer, desto stärker die quadratische Abweichung des Regressorwerts vom Schwerpunkt mit der bedingten Fehlervarianz positiv korreliert ist. Sie wird kleiner bei negativer Korrelation (vgl. Abb. 12.1.10).

12.1 Einfaches lineares Regressionsmodell

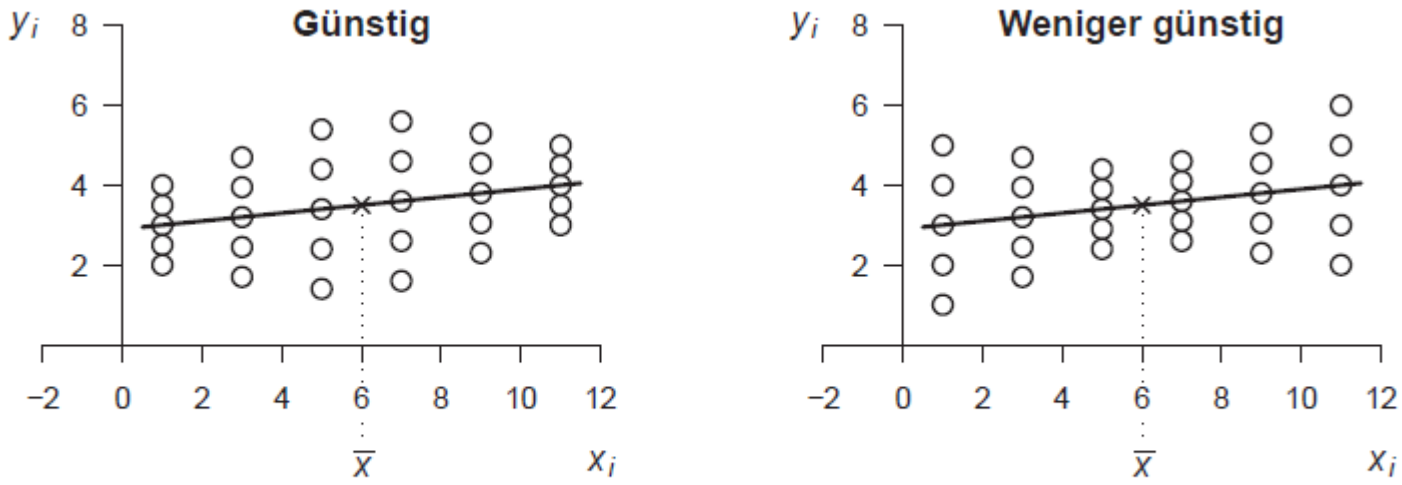


Abb. 12.1.10: Schätzung des Steigungskoeffizienten im Modell BH

> Spezialfall: Homoskedatischer Fehler

$$Var\left(\sqrt{n}(\hat{\beta}_1 - \beta_1)\right) \approx \frac{1}{(\sigma_X^2)^2} E[(X_i - \mu_X)^2 \sigma_U^2] = \frac{\sigma_U^2}{\sigma_X^2}$$

$$\Rightarrow \quad Var(\hat{\beta}_1) \approx \frac{\sigma_U^2}{n\sigma_X^2}$$

enger Bezug zu Modell KS (vgl. Folie 895)

● Verteilung des KQ-Schätzers für β_0 im Modell BH ●

> Resultate ohne Herleitung und ohne Deutung:

$$\Rightarrow \quad \sqrt{n}\left(\hat{\beta}_0 - \beta_0\right) \overset{a}{\sim} N\left(0, \frac{E\left(H_i^2 U_i^2\right)}{\left(E\left(H_i^2\right)\right)^2}\right) \quad \text{wobei} \quad H_i = 1 - \frac{\mu_X X_i}{E\left(X_i^2\right)}$$

$$\Leftrightarrow \quad \hat{\beta}_0 - \beta_0 \overset{approx.}{\sim} N\left(0, \frac{E\left(H_i^2 U_i^2\right)}{n\left(E\left(H_i^2\right)\right)^2}\right)$$

$$\Leftrightarrow \quad \hat{\beta}_0 \overset{approx.}{\sim} N\left(\beta_0, \frac{E\left(H_i^2 U_i^2\right)}{n\left(E\left(H_i^2\right)\right)^2}\right)$$

> Exakte Erwartungstreue liegt auch hier vor

● Verteilung der KQ-Schätzer im Modell UHV ●

Konsistenz und Effizienz der KQ-Schätzer

● Konsistenz der KQ-Schätzer ●

> In den Modellen KS und BH gilt (vgl. Folie 879 bzw. 884):

$$\beta_0 = \mu_Y - \beta_1 \mu_X \text{ und } \beta_1 = \sigma_{XY}/\sigma_X^2$$

Wegen $\bar{X} \xrightarrow{p} \mu_X$, $\bar{Y} \xrightarrow{p} \mu_Y$, $\tilde{S}_{XY} \xrightarrow{p} \sigma_{XY}$ und $\tilde{S}_X^2 \xrightarrow{p} \sigma_X^2$

(vgl. etwa Satz 10.1.4, 8.2.3 (Stetigkeitssatz), Beispiel 8.3.6) folgt

$$\underbrace{\frac{\tilde{S}_{XY}}{\tilde{S}_X^2}}_{=\hat{\beta}_1} \xrightarrow{p} \underbrace{\frac{\sigma_{XY}}{\sigma_X^2}}_{=\beta_1} \text{ und } \underbrace{\bar{Y} - \frac{\tilde{S}_{XY}}{\tilde{S}_X^2} \times \bar{X}}_{=\hat{\beta}_0} \xrightarrow{p} \underbrace{\mu_Y - \frac{\sigma_{XY}}{\sigma_X^2}\mu_X}_{=\beta_0}$$

d. h. $\hat{\beta}_1 \xrightarrow{p} \beta_1$ und $\hat{\beta}_0 \xrightarrow{p} \beta_0$

> In Modellen KN und UHV zusätzliche Konvergenzannahmen nötig

● Effizienz der KQ-Schätzer und Gauß-Markov-Theorem ●

Schätzung der Varianzen der KQ-Schätzer

● Hintergrund ●

> Zur Erinnerung:

Falls $X_1, \ldots, X_n$ unabhängig $N(\mu, \sigma^2)$, dann gilt:

$$\hat{\mu} = \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad Z = \frac{\hat{\mu}-\mu}{\sqrt{\sigma^2/n}} \sim N(0,1) \text{ und } T = \frac{\hat{\mu}-\mu}{\sqrt{S^2/n}} \sim t(n-1)$$

Mit $\sigma_{\hat{\mu}}^2 = Var(\hat{\mu}) = \sigma^2/n$ bzw. $\hat{\sigma}_{\hat{\mu}}^2 = \widehat{Var(\hat{\mu})} = S^2/n$

können wir KI's und Teststatistiken konstruieren und notieren

$$\left[\hat{\mu} \pm z_{1-\alpha/2}\sqrt{\sigma_{\hat{\mu}}^2}\right] \text{ bzw. } \left[\hat{\mu} \pm t_{n-1,1-\alpha/2}\sqrt{\hat{\sigma}_{\hat{\mu}}^2}\right]$$

und $Z = \frac{\hat{\mu}-\mu}{\sqrt{\sigma_{\hat{\mu}}^2}}$ bzw. $T = \frac{\hat{\mu}-\mu}{\sqrt{\hat{\sigma}_{\hat{\mu}}^2}}$

bekannter bzw. geschätzter **Standardfehler**

> Jetzt analog: KI's und Tests für β_0 und β_1

● Schätzung der Varianzen im klassischen Fall ●

> Es gelten folgende Verteilungsresultate (vgl. Folien 890, 892 und 895):

Modell KN: $\sigma^2_{\hat{\beta}_0} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i^2 \, \sigma_U^2}{n\tilde{s}_X^2}$ $\sigma^2_{\hat{\beta}_1} = \frac{\sigma_U^2}{n\tilde{s}_X^2}$

Modell KS: $\sigma^2_{\hat{\beta}_0|X_1,\ldots,X_n} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i^2 \, \sigma_U^2}{n\tilde{S}_X^2}$ $\sigma^2_{\hat{\beta}_1|X_1,\ldots,X_n} = \frac{\sigma_U^2}{n\tilde{S}_X^2}$

> Die dazu korrespondierenden (konsistenten) Schätzer lauten:

$\tilde{\sigma}^2_{\hat{\beta}_0} = \tilde{\sigma}^2_{\hat{\beta}_0|X_1,\ldots,X_n} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i^2 \, \hat{\sigma}_U^2}{n\tilde{S}_X^2}$ bzw. $\tilde{\sigma}^2_{\hat{\beta}_1} = \tilde{\sigma}^2_{\hat{\beta}_1|X_1,\ldots,X_n} = \frac{\hat{\sigma}_U^2}{n\tilde{S}_X^2}$

mit $\hat{\sigma}_U^2 = SER^2 = \frac{1}{n-2}\sum_{i=1}^{n} \hat{U}_i^2$ (vgl. Folie 867)

Nur-Homoskedastizitäts-konsistente Varianzschätzer

● Schätzung der Varianzen im Modell BH ●

> Es gelten folgende Verteilungsresultate (vgl. Folie 901 und 905):

$$\sigma^2_{\hat{\beta}_0} \approx \frac{E\left(H_i^2 U_i^2\right)}{n\left(E\left(H_i^2\right)\right)^2} \text{ mit } H_i = 1 - \frac{\mu_X X_i}{E\left(X_i^2\right)} \quad \text{und} \quad \sigma^2_{\hat{\beta}_1} \approx \frac{Var\left((X_i - \mu_X)U_i\right)}{n\left(\sigma_X^2\right)^2}$$

Man beachte (Folie 897): $Var\left((X_i - \mu_X)U_i\right) = E\left[(X_i - \mu_X)^2 U_i^2\right]$

> Die dazu korrespondierenden (konsistenten) Schätzer lauten:

$$\hat{\sigma}^2_{\hat{\beta}_0} = \frac{\frac{1}{n}\sum_{i=1}^n \hat{H}_i^2 \hat{U}_i^2}{n\left(\frac{1}{n}\sum_{i=1}^n \hat{H}_i^2\right)^2} \quad \text{mit} \quad \hat{H}_i = 1 - \frac{\bar{X} X_i}{\frac{1}{n}\sum_{i=1}^n X_i^2}$$

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2 \hat{U}_i^2}{n\left(\tilde{S}_X^2\right)^2}$$

Heteroskedastizitäs-konsistente oder **Heteroskedastizitäts-robuste Varianzschätzer**

● Schätzung der Varianzen im Modell UHV ●

Verteilungen der Inferenzstatistiken

- **Hintergrund**
- **Verteilungen im klassischen Modell**

Verteilung der maßgeblichen Inferenzstatistik (ohne Nachweis):

$$\tilde{T}_{\beta_j} = \frac{\hat{\beta}_j - \beta_j}{\tilde{\sigma}_{\hat{\beta}_j}} \sim t(n-2) \quad \text{für } j = 0{,}1 \text{ und } n \geq 3$$

und $\tilde{\sigma}^2_{\hat{\beta}_0}$ bzw. $\tilde{\sigma}^2_{\hat{\beta}_1}$ wie auf Folie 908

- **Verteilungen in den Modellen BH und UHV**

Verteilung der maßgeblichen Inferenzstatistik (ohne Nachweis):

$$\hat{T}_{\beta_j} = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \overset{a}{\sim} N(0,1) \quad \text{für } j = 0{,}1 \text{ und } n \geq 60$$

und $\hat{\sigma}^2_{\hat{\beta}_0}$ bzw. $\hat{\sigma}^2_{\hat{\beta}_1}$ wie auf Folie 909

12.1.4 Schätzen und Testen

Konfidenzintervalle und Tests

● Herleitung von Konfidenzintervallen ●

> Ansatz in den klassischen Modellen:

$$1-\alpha = P\left(-t_{n-2,1-\alpha/2} \leq \tilde{T}_{\beta_j} \leq t_{n-2,1-\alpha/2}\right) = \cdots$$

$$= P\left(\hat{\beta}_j - t_{n-2,1-\alpha/2} \times \tilde{\sigma}_{\hat{\beta}_j} \leq \beta_j \leq \hat{\beta}_j + t_{n-2,1-\alpha/2} \times \tilde{\sigma}_{\hat{\beta}_j}\right)$$

> Ansatz in den Modellen BH und UHV:

$$1-\alpha = P\left(-z_{1-\alpha/2} \leq \hat{T}_{\beta_j} \leq z_{1-\alpha/2}\right) = \cdots$$

$$= P\left(\hat{\beta}_j - z_{1-\alpha/2} \times \hat{\sigma}_{\hat{\beta}_j} \leq \beta_j \leq \hat{\beta}_j + z_{1-\alpha/2} \times \hat{\sigma}_{\hat{\beta}_j}\right)$$

● Konstruktion von Tests ●

• Zusammenfassung •

Satz 12.1.4: Konfidenzintervalle und Tests im Regressionsmodell

Gegeben sei ein einfaches lineares Regressionsmodell gemäß Definition 12.1.1. Dann gilt:

1. ein $(1-\alpha)$-Konfidenzintervall für β_j mit $j = 0$ oder 1 ist
   (i) in den Modellen KN und KS für $n \geq 3$ gegeben durch

$$\left[\widehat{\beta}_j - t_{n-2,1-\alpha/2} \cdot \widetilde{\sigma}_{\widehat{\beta}_j},\ \widehat{\beta}_j + t_{n-2,1-\alpha/2} \cdot \widetilde{\sigma}_{\widehat{\beta}_j}\right],$$

   (ii) im Modell BH für großes n gegeben durch

$$\left[\widehat{\beta}_j - z_{1-\alpha/2} \cdot \widehat{\sigma}_{\widehat{\beta}_j},\ \widehat{\beta}_j + z_{1-\alpha/2} \cdot \widehat{\sigma}_{\widehat{\beta}_j}\right],$$

   wobei gilt:

Schlange-Schätzer

$$\widetilde{\sigma}^2_{\widehat{\beta}_1} = \frac{\widehat{\sigma}^2_U}{n\widetilde{S}^2_X}, \qquad \widetilde{\sigma}^2_{\widehat{\beta}_0} = \frac{\frac{1}{n}\sum_{i=1}^n X_i^2 \widehat{\sigma}^2_U}{n\widetilde{S}^2_X},$$

Dach-Schätzer

$$\widehat{\sigma}^2_{\widehat{\beta}_1} = \frac{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2 \widehat{U}_i^2}{n(\widetilde{S}^2_X)^2}, \qquad \widehat{\sigma}^2_{\widehat{\beta}_0} = \frac{\frac{1}{n}\sum_{i=1}^n \widehat{H}_i^2 \widehat{U}_i^2}{n(\frac{1}{n}\sum_{i=1}^n \widehat{H}_i^2)^2}$$

mit $\widetilde{S}_X^2 > 0$,

$$\widehat{\sigma}_U^2 = \frac{1}{n-2}\sum_{i=1}^{n}\widehat{U}_i^2 \quad \text{und} \quad \widehat{H}_i = 1 - \frac{\overline{X}X_i}{\frac{1}{n}\sum_{i=1}^{n}X_i^2}.$$

2. die Entscheidungsregeln eines Tests zum Niveau α für das Testproblem
   a) $H_0 : \beta_j \geq \beta_{j,0}$ vs. $H_1 : \beta_j < \beta_{j,0}$,
   b) $H_0 : \beta_j \leq \beta_{j,0}$ vs. $H_1 : \beta_j > \beta_{j,0}$,
   c) $H_0 : \beta_j = \beta_{j,0}$ vs. $H_1 : \beta_j \neq \beta_{j,0}$,

   (i) basieren in den Modellen KN und KS auf

$$\widetilde{T}_{\beta_j} = \frac{\widehat{\beta}_j - \beta_{j,0}}{\widetilde{\sigma}_{\widehat{\beta}_j}} \sim t(n-2) \quad \text{für } \beta_j = \beta_{j,0} \text{ und } j = 0, 1.$$

   und lauten: Verwerfe H_0 in
   a), falls $\widetilde{t}_{\beta_j} < -t_{n-2,1-\alpha}$,
   b), falls $\widetilde{t}_{\beta_j} > t_{n-2,1-\alpha}$,
   c), falls $|\widetilde{t}_{\beta_j}| > t_{n-2,1-\alpha/2}$,

(ii) basieren im Modell BH auf

$$\widehat{T}_{\beta_j} = \frac{\widehat{\beta}_j - \beta_{j,0}}{\widehat{\sigma}_{\widehat{\beta}_j}} \overset{a}{\sim} N(0,1) \quad \text{für } \beta_j = \beta_{j,0} \text{ und } j = 0, 1.$$

und lauten für großes n: Verwerfe H_0 in

a), falls $\widehat{t}_{\beta_j} < -z_{1-\alpha}$,

b), falls $\widehat{t}_{\beta_j} > z_{1-\alpha}$,

c), falls $|\widehat{t}_{\beta_j}| > z_{1-\alpha/2}$,

3. mit entsprechenden zusätzlichen Konvergenzannahmen sind die Konfidenzintervalle und Tests für Modell BH auch für Modell UHV gültig.

Im Falle von Modell BH handelt es sich in (ii) um approximative Konfidenzintervalle bzw. approximative Tests für $n \geq 60$.

● Äquivalenz der Tests auf $\beta_{1,0} = 0$ und $\rho_{XY} = 0$ ●

> Die Teststatistik bezüglich β_1 in den Modellen KN und KS für den Nullhypothesenwert $\beta_{1,0} = 0$ ist mit derjenigen des **Korrelationstests** (Folie 835) identisch, d. h. es gilt:

$$\tilde{T}_{\beta_1} = T_R \text{ für } \beta_{1,0} = 0$$

> Nachweisskizze:

$$\tilde{T}_{\beta_1} = \frac{\hat{\beta}_1 - 0}{\sqrt{\tilde{\sigma}_{\hat{\beta}_1}}} = \frac{\tilde{S}_{XY}/\tilde{S}_X^2}{\sqrt{\hat{\sigma}_U^2/(n\tilde{S}_X^2)}} = \cdots = \sqrt{n-2} \times \frac{R_{XY}}{\sqrt{1 - R_{XY}^2}} = T_R$$

> Korrelationstest und Tests in den klassischen Modellen mit $\beta_{1,0} = 0$ sind äquivalent (führen stets zu identischen Testentscheidungen)

Spezialfall: Binärer Regressor

● Hintergrund ●

> Ein Spezialfall eines einfachen linearen Regressionsmodells liegt vor, falls der Regressor **binär** (0-1-Variable) ist. Grundsätzlich können damit auch kategoriale Merkmale als Regressoren verwendet werden.

> Inferenz bei binärem Regressor ist äquivalent zur Inferenz über Erwartungswertdifferenzen

● Interpretation der Regressionskoeffizienten ●

> Ausgehend vom Modell $Y_i = \beta_0 + \beta_1 X_i + U_i$ mit binärem X_i erhalten wir in allen Modellvarianten mit Annahme (A1), (A1*) und (A1**) dann

$$E(Y_i|X_i = 0) = \beta_0 \quad \text{und} \quad E(Y_i|X_i = 1) = \beta_0 + \beta_1$$

> Setzen wir

$$\mu_0 = E(Y_i|X_i = 0) \quad \text{und} \quad \mu_1 = E(Y_i|X_i = 1)$$

erhalten wir die Beziehungen

$$\mu_0 = \beta_0 \quad \text{und} \quad \beta_1 = \mu_1 - \mu_0$$

• KQ-Schätzer bei binärem Regressor •

Satz 12.1.5: KQ-Schätzer bei binärem Regressor

Gegeben sei eine Stichprobe $(X_1, Y_1), \ldots, (X_n, Y_n)$ mit binärem X_i und $\tilde{S}_X^2 > 0$. Dann gilt:

$$\hat{\beta}_0 = \overline{Y}_0 \quad \text{und} \quad \hat{\beta}_1 = \overline{Y}_1 - \overline{Y}_0.$$

Dabei bezeichnet $\overline{Y}_0$ das (bedingte) Stichprobenmittel aller Y_i mit $X_i = 0$ und $\overline{Y}_1$ das (bedingte) Stichprobenmittel aller Y_i mit $X_i = 1$.

> Nachweis umständlich jedoch mathematisch eigentlich nicht anspruchsvoll

> Ergebnisse sind intuitiv nachvollziehbar!

• Varianzschätzer bei binärem Regressor •

Satz 12.1.6: Varianzschätzer bei binärem Regressor

Gegeben sei eine Stichprobe $(X_1, Y_1), \ldots, (X_n, Y_n)$ mit binärem X_i und $\widetilde{S}_X^2 > 0$. Dann gilt:

(i) $\widetilde{\sigma}^2_{\hat{\beta}_0} = \dfrac{S_p^2}{N_0}$ und $\widetilde{\sigma}^2_{\hat{\beta}_1} = \dfrac{S_p^2}{N_0} + \dfrac{S_p^2}{N_1}$, wobei

$$S_p^2 = \frac{1}{N_0 + N_1 - 2}(N_0 \widetilde{S}_0^2 + N_1 \widetilde{S}_1^2) \quad \text{für } n \geq 3,$$

(ii) $\widehat{\sigma}^2_{\hat{\beta}_0} = \dfrac{\widetilde{S}_0^2}{N_0}$ und $\widehat{\sigma}^2_{\hat{\beta}_1} = \dfrac{\widetilde{S}_0^2}{N_0} + \dfrac{\widetilde{S}_1^2}{N_1}$ für $n \geq 2$.

Dabei bezeichnet $\widetilde{S}_0^2$ die (nichtkorrigierte) Stichprobenvarianz aller Y_i mit $X_i = 0$ und $\widetilde{S}_1^2$ die entsprechende Stichprobenvarianz für $X_i = 1$. Die stochastischen Stichprobenumfänge N_0 und N_1 geben die betreffende Anzahl von Nullen bzw. Einsen an.

• Äquivalenz der Inferenz bezüglich β_1 und $\mu_1 - \mu_0$ •

Äquivalenz der Inferenz bezüglich β_1 und $\mu_1 - \mu_0$

Konfidenzintervalle und Teststatistiken sind

(i) in den Modellen KN und KS zu denen über Erwartungswertdifferenzen bei Normalverteilung und homoskedastischer Varianz identisch,

(ii) in Modell BH zu denen über Erwartungswertdifferenzen ohne Verteilungsannahme und unbekannter heteroskedastischer Varianz identisch.

> Ad (i):

KI für β_1: $\left[\hat{\beta}_1 - t_{n-2,1-\alpha/2} \times \tilde{\sigma}_{\hat{\beta}_1}, \hat{\beta}_1 + t_{n-2,1-\alpha/2} \times \tilde{\sigma}_{\hat{\beta}_1}\right]$

identisch mit dem für $\mu_1 - \mu_0$:

$$\left[\bar{Y}_1 - \bar{Y}_0 - t_{n-2,1-\alpha/2}\sqrt{\frac{S_P^2}{n_0} + \frac{S_P^2}{n_1}}, \bar{Y}_1 - \bar{Y}_0 + t_{n-2,1-\alpha/2}\sqrt{\frac{S_P^2}{n_0} + \frac{S_P^2}{n_1}}\right]$$

Tests über β_1 identisch mit denen über $\mu_1 - \mu_0$:

$$\tilde{T}_{\beta_1} = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\tilde{\sigma}^2_{\hat{\beta}_1}}} = \frac{\bar{Y}_1 - \bar{Y}_0 - \delta_0}{\sqrt{S_P^2/n_0 + S_P^2/n_1}} = \tilde{T}_D$$

falls $\beta_{1,0} = \delta_0$

> Ad (ii):

KI für β_1: $\left[\hat{\beta}_1 - z_{1-\alpha/2} \times \hat{\sigma}_{\hat{\beta}_1}, \hat{\beta}_1 + z_{1-\alpha/2} \times \hat{\sigma}_{\hat{\beta}_1}\right]$

identisch mit dem für $\mu_1 - \mu_0$:

$$\left[\bar{Y}_1 - \bar{Y}_0 - z_{1-\alpha/2}\sqrt{\frac{\tilde{S}_0^2}{N_0} + \frac{\tilde{S}_1^2}{N_1}}, \bar{Y}_1 - \bar{Y}_0 + z_{1-\alpha/2}\sqrt{\frac{\tilde{S}_0^2}{N_0} + \frac{\tilde{S}_1^2}{N_1}}\right]$$

Tests über β_1 identisch mit denen über $\mu_1 - \mu_0$:

$$\hat{T}_{\beta_1} = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\hat{\sigma}^2_{\hat{\beta}_1}}} = \frac{\bar{Y}_1 - \bar{Y}_0 - \delta_0}{\sqrt{\tilde{S}_0^2/N_0 + \tilde{S}_1^2/N_1}} = \hat{T}_D$$

falls $\beta_{1,0} = \delta_0$

● Tests über $\mu_1 - \mu_0$ auch bei stochastischen Gruppenumfängen ●

Da in den Modellen KS und BH der Regressor stochastisch modelliert wird, folgt, dass die Konfidenzintervalle und Teststatistiken für Erwartungswert-differenzen auch bei stochastischen Gruppenumfängen N_0 und N_1 verwendet werden können.

Adäquatheit bestimmter Modellannahmen

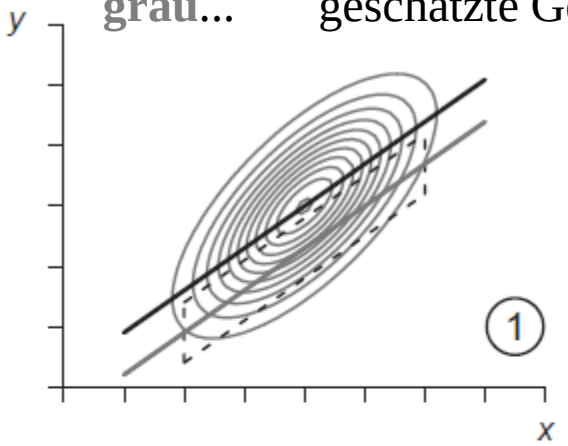
• Hintergrund •

• Noch vor den eigentlichen Annahmen: Repräsentativität •

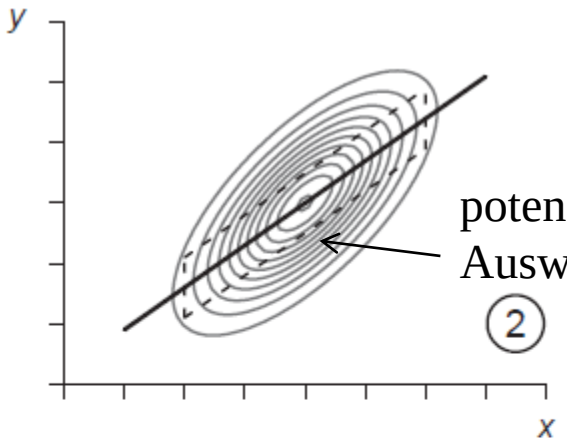
Stellen wir uns vor, ein Immobilienexperte möchte den Zusammenhang zwischen Wohnfläche und Nettomiete in einer Stadt untersuchen. Dazu nimmt er s ich kurzerhand die Wochenendausgabe der ansässigen Lokalzeitung zur Hand und notiert alle Angebote des Wohnungsmarktes bezüglich Wohnfläche und Miete. Eine auf solche Weise gewonnene Stichprobe könnten wir dann als **Quasi-Stichprobe** erachten (Abschnitt 9.2). Abbildung 12.1.11 illustriert verschiedene (stark stilisierte) Situationen, in denen der **potenzielle Auswahlbereich** (gestrichelt) die Grundgesamtheit (Dichtekonturen) nicht vollständig abdeckt. Infolgedessen führt dies zu einer mehr oder weniger stark ausgeprägten **Stichprobenverzerrung**. Im Einzelnen könnte man sich dazu folgende Szenarien (1) bis (4) vorstellen:

12.1 Einfaches lineares Regressionsmodell

schwarz... wahre Gerade
grau... geschätzte Gerade

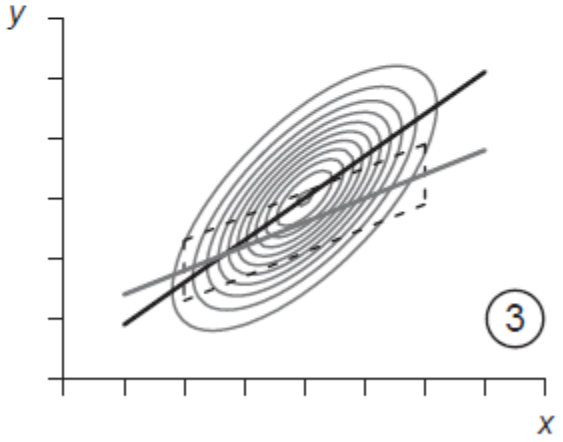


Geschätzte Gerade zu tief

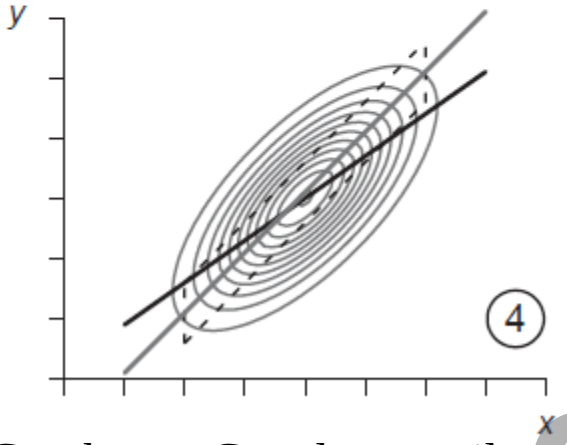


potenzieller Auswahlbereich

Geschätzte Gerade dennoch richtig



Geschätzte Gerade zu flach



Geschätzte Gerade zu steil

Abb. 12.1.11: Potenzielle Auswahlbereiche und Stichprobenverzerrungen

- **Linearitätsannahme**
- **Stochastischer oder nichtstochastischer Regressor**
- **Messfehlerprobleme**
- **Unabhängigkeitsannahme**
- **$E(U_i \mid X_i) = 0$ und OVB-Problem**

Annahme impliziert, dass...

... Resteinflüsse auf die abhängige Variable „gleichmäßig um die Gerade herum streuen“

... dass die Eigenschaft (vgl. Folie 877 bzw. 878 Eigenschaft 8)

$$Cov(X_i, U_i) = E(X_i U_i) = 0$$

erfüllt ist, welche anhand folgender Leitfrage geprüft wird:

Leitfrage für das OVB-Problem

Gibt es neben dem Regressor X, also der beobachteten Einflussgröße, eine weitere maßgebliche Einflussgröße auf Y, die mit X korreliert ist?

> Bejahung der Leitfrage impliziert $Cov(X_i, U_i) \neq 0$ und damit auch $E(U_i|X_i) \neq 0$. Die Umkehrung gilt nicht zwingend! Dies bedeutet, dass $E(U_i|X_i) \neq 0$ nicht zwingend $Cov(X_i, U_i) \neq 0$ impliziert.

> $Cov(X_i, U_i) \neq 0$ führt zu einer inkonsistenten Schätzung von β_1, was sich im Rahmen der späteren Überlegungen zu **systematischen Verzerrungen** auch formal fassen lässt (vgl. etwa Satz 12.2.3)

> Die durch $Cov(X_i, U_i) \neq 0$ verursachte **asymptotische Verzerrung** wird auch als **OVB** (**O**mitteld-**V**ariable-**B**ias) bezeichnet und das damit korrespondierende Problem als **OVB-Problem**.

> In Modell KN $Cov(x_i, U_i) = 0$ per se erfüllt, da x_i nichtstochastisch. Leitfrage dennoch von Relevanz (vgl. Fallbeispiel 1 auf Folie 930)

> Das multiple lineare Regressionsmodell stellt einen möglichen Lösungansatz für das OVB-Problem dar (vgl. Abschnitt 12.2)

- **$E(U_i \mid X_i) = 0$, Messfehler- und Endogenitätsproblem**

  Auch andere Probleme wie **Messfehler** oder **simultane Kausalität** können zu $E(U_i|X_i) \neq 0$ führen (weiterführende Themen)

- **$Cov(X_i, U_i) = 0$ lässt sich nicht anhand der KQ-Residuen prüfen**

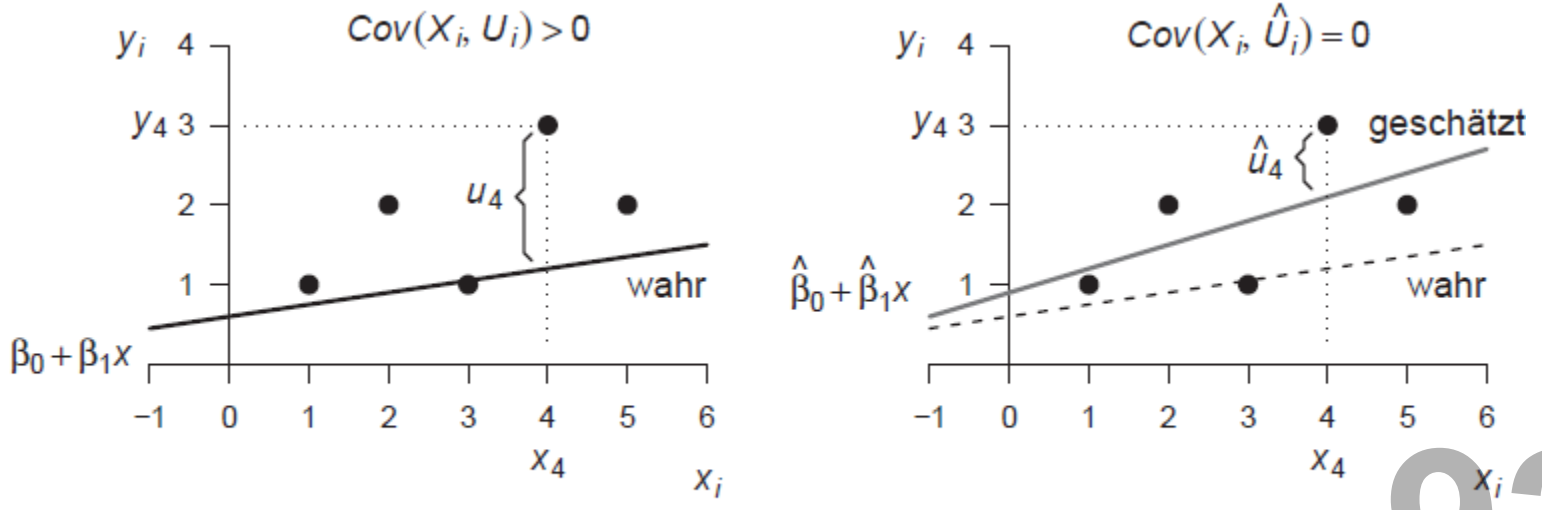


Abb. 12.1.12: Verzerrung durch Korrelation von Regressor und Fehler

● Homoskedastischer oder heteroskedastischer Fehler ●

> Rein optische Prüfung kann in die Irre führen (vgl. Abb. 12.1.13), insbesondere bei schiefer Verteilung des Regressors

> Statistische Tests auf Heteroskedastizität wie etwa **White-Test**

> KQ-Methode bei Heteroskedastizität eigentlich ineffizient, effiziente **verallgemeinerte KQ-Methode** jedoch schwer umsetzbar

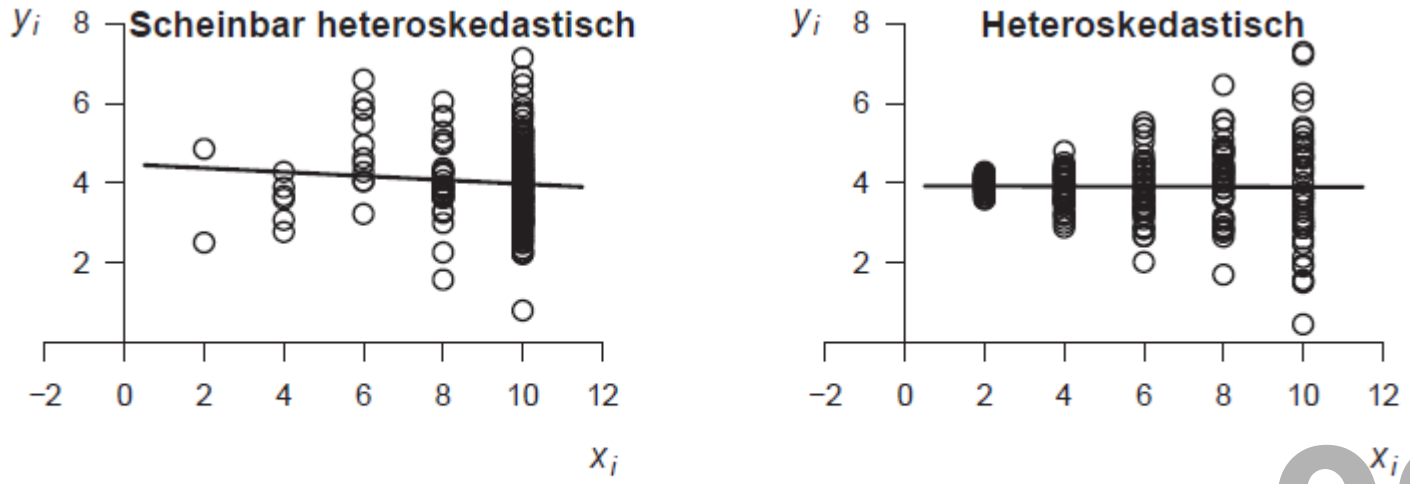


Abb. 12.1.13: Scheinbare und tatsächliche Heteroskedastizität

> Bei Vorliegen von Homoskedastizität sollten die Nur-Homoskedastizitätskonsistenten Varianzschätzer (Folie 908) verwendet werden.

> Die fälschliche Unterstellung von Homoskedastizität führt zu inkonsistenten Varianzschätzungen, was dann verfälschte Sicherheits- und Irrtumswahscheinlichkeiten zur Folge hat (verfälschte Inferenz)

> Die unnötige Verwendung der Heteroskedastizitäts-konsistenten Varianzschätzer (Folie 909) in Modell BH (d. h. es liegt tatsächlich Homoskedasizität vor), führt hingegen nur zu einem Effizienzverlust bei der Varianzschätzung. Tendenziell führt dies zu längeren Konfidenzintervallen und zu Güteverlusten beim Testen. Da die Schätzer jedoch auch unter Homoskedastizität konsistent sind, bleibt die Inferenz gültig.

● Normalverteilungsannahme ●

● Identisch oder heterogen verteilt ●

● Technische Annahmen ●

● Fazit und Empfehlung ●

> Einschließlich der Repräsentativität stellen alle mit der Annahme

$$E(U_i|X_i) = 0$$

in Verbindung stehenden Aspekte die wichtigsten Prüfsteine dar, da sie bei Verletzung zu ernsthaften Verfälschungen (inkonsistenten Schätzungen) führen.

> Modell BH mit den heteroskedastizitäts-konsistenten Varianzschätzern stellt für viele Fälle in der Praxis, in denen ein u. i. v.-Schema adäquat erscheint, eine gute Wahl dar. Dabei gilt: Lieber unnötig von Heteroskedastizität ausgehen als fälschlich Homoskedasitzität unterstellen.

> Modelle KN, KS und UHV als Spezialfälle in Modell BH enthalten

Die für Modell BH vorgestellten Konfidenzintervalle und Tests sind auch in Modell KS und unter Hinzunahme zusätzlicher Konvergenzannahmen in den Modellen KN und UHV verwendbar, sofern $n \geq 60$.

Fallbeispiel 1: Bewässerung und Wachstum

● Hintergrund ●

Problemstellung fiktiv (vgl. Abb. 12.1.14, Folie 932):

Ein Botaniker untersucht den Zusammenhang zwischen Bewässerungsmenge (in Liter) und Wuchshöhe (in Meter) einer bestimmten Pflanze. Dazu züchtet er 5 Exemplare in einem Labor unter identischen Bedingungen an. Lediglich die wöchentliche Bewässerungsmenge wählt er unterschiedlich und variiert diese zwischen 1 und 5 Liter. Sein Modellansatz lautet:

$$Y_i = \beta_0 + \beta_1 x_i + U_i \quad \text{für } i = 1, \ldots, 5,$$

wobei Y_i die nach 4 Wochen gemessene Wuchshöhe und x_i die festgesetzte wöchentliche Bewässerungsmenge für die i-te Pflanze ist.

● Modell KN: Diskussion der Modellannahmen ●

> Linearitätsannahme (A0):
Fachkenntnisse nötig, approximative Linearität, beschränkt auf bestimmten Wertebereich

> Annahme (A1*): $U_1, \ldots, U_n$ unabhängig $N(0, \sigma_U^2)$

1. Teilaspekt: $E(U_i) = 0$

Angenommen, Szenario wie in Abb. 12.1.14:
Bewässerungsmenge steigt von links nach rechts. Gleichzeitig wirke sich Tageslicht wachstumssteigernd aus. Dann wären die Bewässerungsmengen mit den Realisationen der Resteinflüsse (darin ist das Licht enthalten) tendenziell positiv korreliert (höhere Mengen gehen mit mehr Licht einher und geringere Mengen mit weniger Licht). Formal gilt zwar $Cov(x_i, U_i) = 0$, jedoch erscheint $E(U_i) = 0$ nicht mehr adäquat. Vielmehr sollten die Erwartungswerte der Fehler von links nach rechts ansteigen. Weitere Szenarien denkbar...

12.1 Einfaches lineares Regressionsmodell

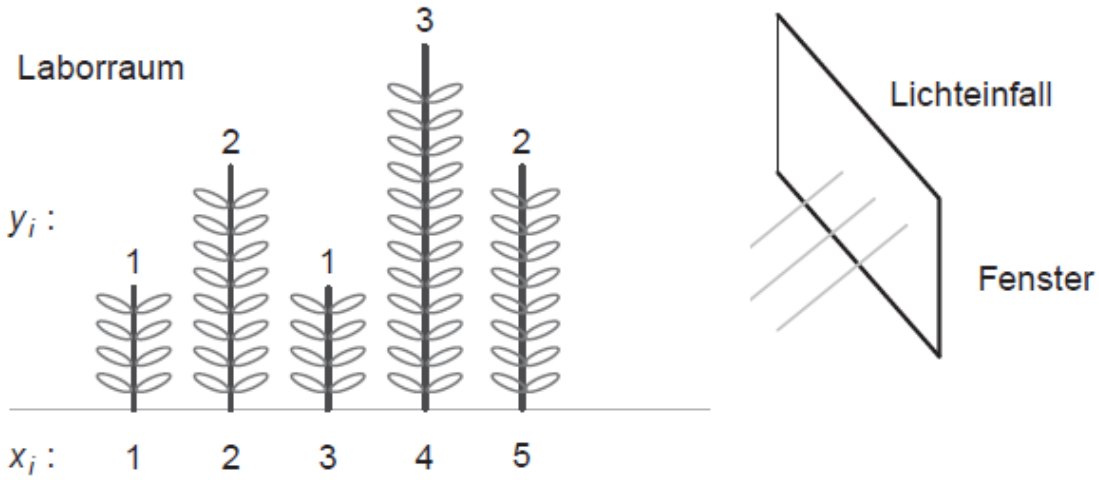


Abb. 12.1.14: Laborexperiment: Bewässerung und Wuchshöhe

2. Teilaspekt: $Var(U_i) = \sigma_U^2$ (Homoskedastizität)

Nicht ganz realistisch, da „natürlicherweise“ mit steigendem Niveau meist die Variabilität zunimmt.

3. Teilaspekt: $U_1, \ldots, U_n$ unabhängig

Könnte etwa verletzt sein, wenn Pflanzen zu dicht beieinander stehen und sich gegenseitig beeinflussen. Ansonsten realistische Annahme.

4. Teilaspekt: Normalverteilte Fehler

In kontrollierten Experimenten erscheint Normalverteilung als die „natürliche Verteilung“ häufig realistisch.

• Berechnung •

> Beobachtungswerte (x_i, y_i): (1, 1), (2, 2), (3, 1), (4, 3), (5, 2)

> Berechnung der KQ-Schätzer und von R^2 (vgl. Folien 864-866):

$$\hat{\beta}_0 = 0.9, \quad \hat{\beta}_1 = 0.3, \quad R^2 = 0.3214$$

> Berechnung des *SER* (vgl. Tabelle 5.2.4 auf Folie 865):

$$SER^2 = \frac{1}{n-2}\sum_{i=1}^{n} \hat{u}_i^2 = \frac{1.9}{5-2} \approx 0.6333$$

$$\Rightarrow SER \approx 0.7958$$

> Berechnung der Nur-Homoskedastizitätskonsistenten Varianzschätzer (vgl. Folien 867, 908):

$$\tilde{s}_X^2 = \frac{1}{5}\sum_{i=1}^{5} x_i^2 - \bar{x}^2 = \frac{55}{5} - 3^2 = 2$$

$$\hat{\sigma}_U^2 = SER^2 = 0.6333$$

$$\Rightarrow \tilde{\sigma}^2_{\hat{\beta}_0} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i^2\,\hat{\sigma}_U^2}{n\tilde{s}_X^2} = \frac{11 \times 0.6333}{5 \times 2} \approx 0.6966$$

(„Schlangeschätzer“)

$$\Rightarrow \tilde{\sigma}^2_{\hat{\beta}_1} = \frac{\hat{\sigma}_U^2}{n\tilde{S}_X^2} = \frac{0.6333}{5 \times 2} \approx 0.0633$$

> Berechnung der **Standardfehler** (SE… Standard Error):

$$\Rightarrow \widetilde{SE}(\hat{\beta}_0) = \sqrt{\tilde{\sigma}^2_{\hat{\beta}_0}} = \tilde{\sigma}_{\hat{\beta}_0} \approx 0.8346$$

$$\Rightarrow \widetilde{SE}(\hat{\beta}_1) = \sqrt{\tilde{\sigma}^2_{\hat{\beta}_1}} = \tilde{\sigma}_{\hat{\beta}_1} \approx 0.2516$$

● Ergebnisse und Interpretation ●

$$\widehat{Wuchshöhe} = \underset{(0.83)}{0.9} + \underset{(0.25)}{0.3 \cdot Wassermenge}, \quad R^2 \approx 0.32, \quad SER \approx 0.8.$$

> 0.95-KI für β_0 und β_1 gemäß Satz 12.1.4 (Folie 912):

$$[\hat{\beta}_0 - t_{3,0.975}, \times \tilde{\sigma}_{\hat{\beta}_0}, \hat{\beta}_0 + t_{3,0.975}, \times \tilde{\sigma}_{\hat{\beta}_0}]$$

$$[\hat{\beta}_1 - t_{3,0.975}, \times \tilde{\sigma}_{\hat{\beta}_1}, \hat{\beta}_1 + t_{3,0.975}, \times \tilde{\sigma}_{\hat{\beta}_1}]$$

$$\Rightarrow \quad \beta_0: \ [0.9 \pm 3.18 \times 0.83] = [-1.74, 3.54]$$

$$\Rightarrow \quad \beta_1: \ [0.3 \pm 3.18 \times 0.25] = [-0.50, 1.10]$$

Interpretation:
In beiden Intervallen 0 enthalten. Beide Koeffizienten sind bei einem Niveau von 5% nicht signifikant von 0 verschieden.

> Test zum Niveau 5% bezüglich β_1 gemäß Satz 12.1.4 (Folie 913):

Entscheidungsproblem: Wirkt sich höhere Bewässerung signifikant steigernd auf das Wachstum aus?

Testproblem: $H_0\colon\ \beta_1 \leq 0$ vs. $H_1\colon\ \beta_1 > 0$

Teststatistik: $\tilde{T}_{\beta_1} = \dfrac{\hat{\beta}_1 - 0}{\tilde{\sigma}_{\hat{\beta}_1}} \overset{\beta_1 = 0}{\sim} t(3)$

Testniveau (Signifikanzniveau): $\alpha = 0.05$

Kritischer Wert: $c = t_{3,95} = 2.35$

Testprozedur: Falls $\tilde{t}_{\beta_1} > 2.35$, Entscheidung für H_1 (Verwerfung von H_0), sonst Beibehaltung von H_0

Testergebnis hier: $\tilde{t}_{\beta_1} = \dfrac{0.3}{0.25} = 1.2$

$\Rightarrow$ Kein signifikanter positiver Effekt

Fallbeispiel 2: Klassengröße und Lernerfolg

● Hintergrund ●

Realer Datensatz

Merkmal	Variable	Erläuterung
Testergebnis	testscr	Erreichte Punktezahl im Test
Klassengröße	str	Anzahl von Schülern pro Lehrer
Englisch-Lerner-Anteil	elpct	Prozentualer Anteil von Schülern mit Sprachproblemen
Begünstigten-Anteil	mealpct	Prozentualer Anteil von Schülern, die vergünstigtes Mensaessen beziehen

Daten von Schülern der 5. Klasse aus einer standardisierten Prüfung in Kalifornien aus den Jahren 1998, 1999.
Umfang: $n = 420$ Schuldistrikte

12.1 Einfaches lineares Regressionsmodell

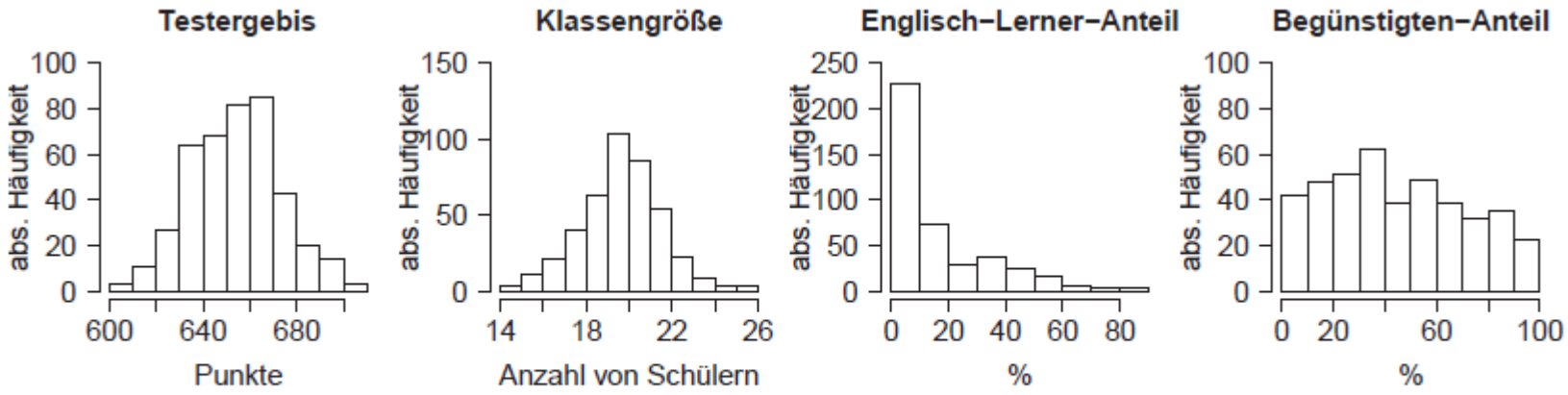


Abb. 12.1.15: Empirische Verteilung einzelner Variablen

	testscr	str	elpct	mealpct
testscr	1.0000	-0.2264	-0.6441	-0.8688
str	-0.2264	1.0000	0.1876	0.1352
elpct	-0.6441	0.1876	1.0000	0.6531
mealpct	-0.8688	0.1352	0.6531	1.0000

12.1 Einfaches lineares Regressionsmodell

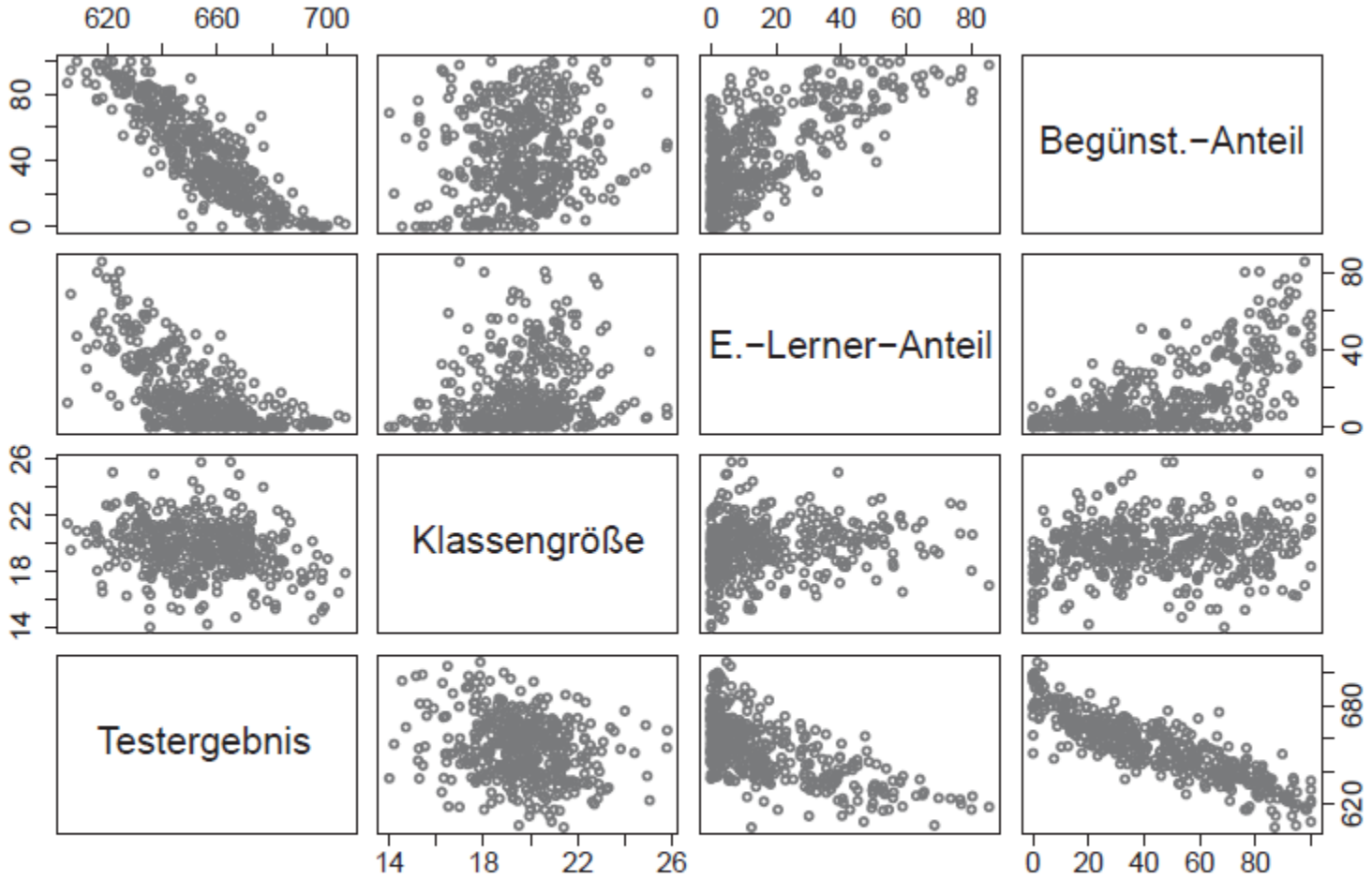


Abb. 12.1.16: Paarweise empirische Abhängigkeit der Variablen

Problemstellung:
Angenommen, man möchte untersuchen, ob es sich lohnt mehr Lehrer einzustellen, um den Lernerfolg der Schüler zu verbessern. Der grundsätzliche Modellansatz laute dabei zunächst einmal

$$Y_i = \beta_0 + \beta_1 X_i + U_i \quad \text{für } i = 1, \ldots, 420,$$

wobei Y_i das (durchschnittlich) Testergebnis im i-ten Distrikt ist und X_i die (durchschnittliche) Klassengröße

● Modell BH: Diskussion der Modellannahmen ●

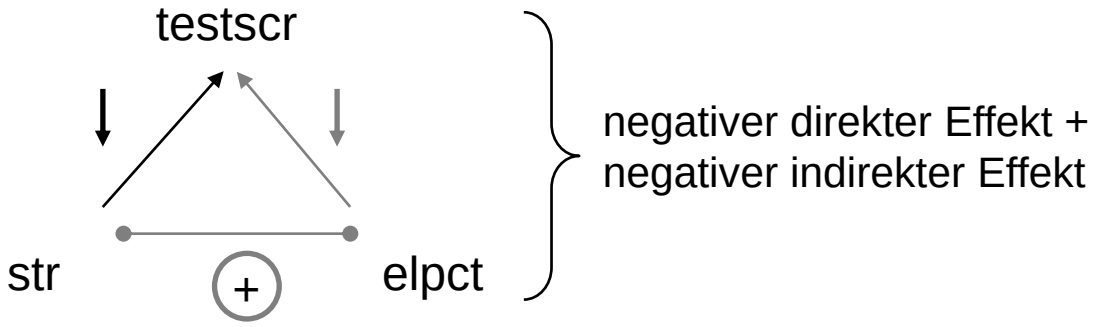
> Linearitätsannahme (A0):
Scheint bei Betrachtung des Streudiagramms (testscr vs. str) in Ordnung zu gehen. Fachspezifische Kenntnisse weiter hilfreich…

> Annahme (A1): $E(U_i|X_i) = 0$

Wir stellen hierzu die Leitfrage für das OVB-Problem (Folie 925):
Gibt es neben dem Regressor X, also der beobachteten Einflussgröße, eine weitere maßgebliche Einflussgröße auf Y, die mit X korreliert ist?

Antwort hierzu:
OVB-Leitfrage kann bejaht werden, da *str* mit *elpct* und *mealpct* positiv korreliert ist und beide, d. h. *elpct* und *mealpct*, sich auf den Lernerfolg, gemessen über *testscr*, (negativ) auswirken sollten. Dafür sprechen die negativen Korrelationen von –0.64 bzw. –0.87.

OVB-Szenario beschränkt auf *testscr*, *str* und *elpct* (Kap. 5, Folie 260 ff.):



Im einfachen Regressionsmodell sollte β_1 unterschätzt werden, da der negative **direkte Effekt** von *str* durch den negativen **indirekten Effekt** von *elpc*t verstärkt wird. Dies führt zu einem noch stärkeren negativen **Gesamteffekt** von *str.*

> Annahme (A2): $(X_1, Y_1), \ldots, (X_n, Y_n)$ u. i.v.

1. Teilaspekt: Art der Stichprobe

Hypothetische Stichprobe (aus **hypothetischer Grundgesamtheit**), da Vollerhebung der Schuldistrikte in Kalifornien zu bestimmter Zeit.

2. Teilaspekt: $(X_1, Y_1), \ldots, (X_n, Y_n)$ sind unabhängig

Es kann wohl von gewissen räumlichen Korrelationen ausgegangen werden. Insbesondere könnten Distrikte, die geographisch näher beieinander liegen, ähnlicher zueinander sein (Expertenwissen nötig).

3. Teilaspekt: $(X_1, Y_1), \ldots, (X_n, Y_n)$ identisch verteilt

Distrikte, die in einem Jahr besonders gut (schlecht) sind, sind es mit hoher Wahrscheinlichkeit auch im Folgejahr. Deshalb erscheinen identische Verteilungen unrealistisch. Ein u.h.v.-Schema wäre eigentlich adäquat. Mit zusätzlichen technischen Annahmen wäre die Inferenz jedoch völlig identisch.

> Annahme (A3): Endliche Momente, positive Varianzen (vgl. Folie 875):
Rein technische Annahmen, keine praktische Relevanz

• Berechnung •

> Händische Berechnung theoretisch möglich; relevante Formeln gemäß Folien 865 ff. und 909. Soll hier nicht exerziert werden, da zu aufwendig ($n = 420$!)

> Berechnung mit Hilfe einer geeigneten Software (R):

$$\hat{\beta}_0 = 698.93, \quad \hat{\beta}_1 = -2.28, \quad SER = 18.58, \quad R^2 = 0.05$$

$$\hat{\sigma}^2_{\hat{\beta}_0} = 106.91, \quad \hat{\sigma}^2_{\hat{\beta}_1} = 0.27$$ („Dachschätzer“)

$$\Rightarrow \quad \widehat{SE}(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_0}} = \hat{\sigma}_{\hat{\beta}_0} = 10.34$$

$$\Rightarrow \quad \widehat{SE}(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_1}} = \hat{\sigma}_{\hat{\beta}_1} = 0.52$$

● Ergebnisse und Interpretation ●

$$\widehat{Testergebnis} = \underset{(10.34)}{693.93} \underset{(0.52)}{-2.28} \cdot Klassengröße, \quad R^2 \approx 0.05, \quad SER \approx 18.58$$

> 0.95-KI für β_0 und β_1 gemäß Satz 12.1.4 (Folie 912):

$$[\hat{\beta}_0 - z_{0.975}, \times \hat{\sigma}_{\hat{\beta}_0}, \hat{\beta}_0 + z_{0.975}, \times \hat{\sigma}_{\hat{\beta}_0}]$$
$$[\hat{\beta}_1 - z_{0.975}, \times \hat{\sigma}_{\hat{\beta}_1}, \hat{\beta}_1 + z_{0.975}, \times \hat{\sigma}_{\hat{\beta}_1}]$$

$$\Rightarrow \quad \beta_0: \ [698.93 \pm 1.96 \times 10.36] = [678.62, 719.24]$$
$$\Rightarrow \quad \beta_1: \ [-2.28 \pm 1.96 \times 0.52] = [-3.30, -1.26]$$

Interpretation:
Da beide Intervalle die 0 nicht enthalten, sind beide Koeffizienten bei einem Niveau von 5% signifikant von 0 verschieden.

> Test zum Niveau 5% bezüglich β_1 gemäß Satz 12.1.4 (Folie 913):

Entscheidungsproblem: Wirken sich größere Klassen signifikant negativ auf den Lernerfolg aus?

Testproblem: $H_0\colon\ \beta_1 \geq -1$ vs. $H_1\colon\ \beta_1 < -1$

Teststatistik: $\hat{T}_{\beta_1} = \dfrac{\hat{\beta}_1 - (-1)}{\hat{\sigma}_{\hat{\beta}_1}} \overset{a}{\underset{\beta_1 = -1}{\sim}} N(0,1)$

Testniveau (Signifikanzniveau): $\alpha = 0.05$

Kritischer Wert: $c = z_{0.05} = -1.64$

Testprozedur: Falls $\hat{t}_{\beta_1} < -1.64$, Entscheidung für H_1 (Verwerfung von H_0), sonst Beibehaltung von H_0

Testergebnis hier: $\hat{t}_{\beta_1} = \dfrac{-2.28 + 1}{0.52} = -2.46$

$\Rightarrow$ Signifikanter negativer Effekt um mehr als -1

12.1 Einfaches lineares Regressionsmodell

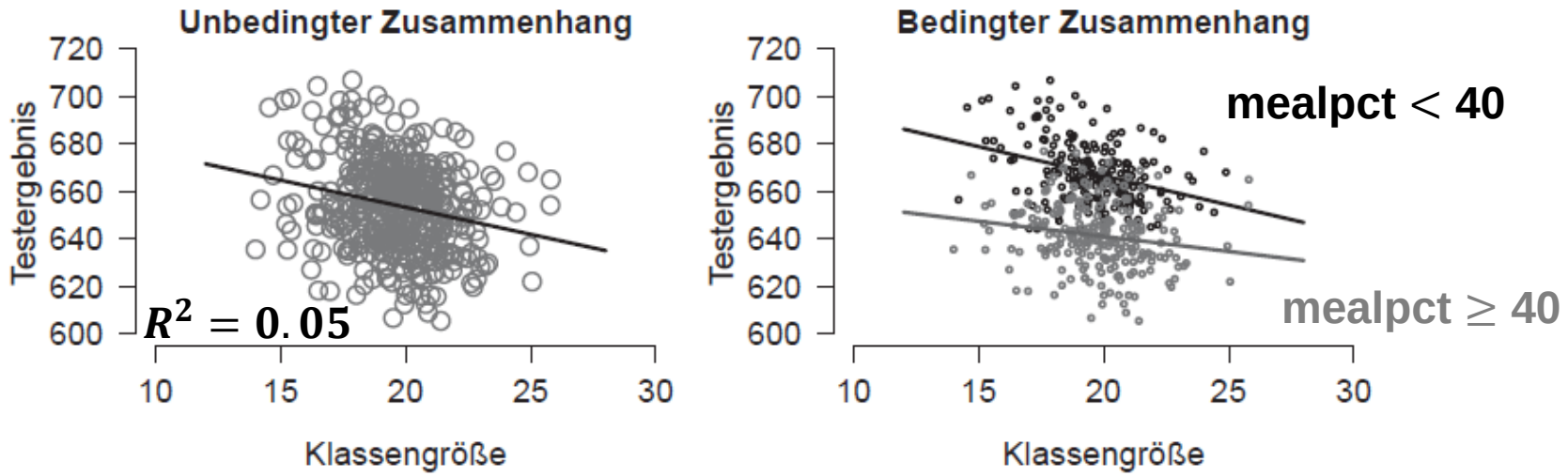


Abb. 12.1.17: Zusammenhang von Klassengröße und Testergebnis

Interpretation:

+ Eine um 5 Schüler vergrößerte Klasse verschlechtert die Testscores um mehr als 10 Punkte
+ Klassengröße signifikant, jedoch wenig relevant
+ OVB-Problematik vorhanden (*elpct*, *mealpct*)
+ Untersuchung in homogeneren Untergruppen legt nahe, dass sozialer Hintergrund relevanter ist; durchschnittlicher cher Abstand von etwa 30 Punkten (vgl. Abb. 12.1.17)

12.2.1 Partielle lineare KQ-Regression

Empirische partielle Regression

● Hintergrund ●

> Verletzung der Annahmen
$E(U_i) = 0$ oder $E(U_i|X_i) = 0$
kann zu inkonsistenten Schätzungen führen

> In enger Verbindung damit steht die OVB-Problematik
(vgl. Fallbeispiele 1 und 2 des vorhergehenden Abschnitts)

> 1. Lösungsansatz: Analyse in homogeneren Untergruppen
(häufig jedoch inpraktikabel, insbesondere bei kleinem n)

> 2. Lösungsansatz: Partielle oder multiple Regression

⇒ „Lineares Herausrechnen“ störender (OVB verursachender Variablen (Thema von Abschnitt 12.2)

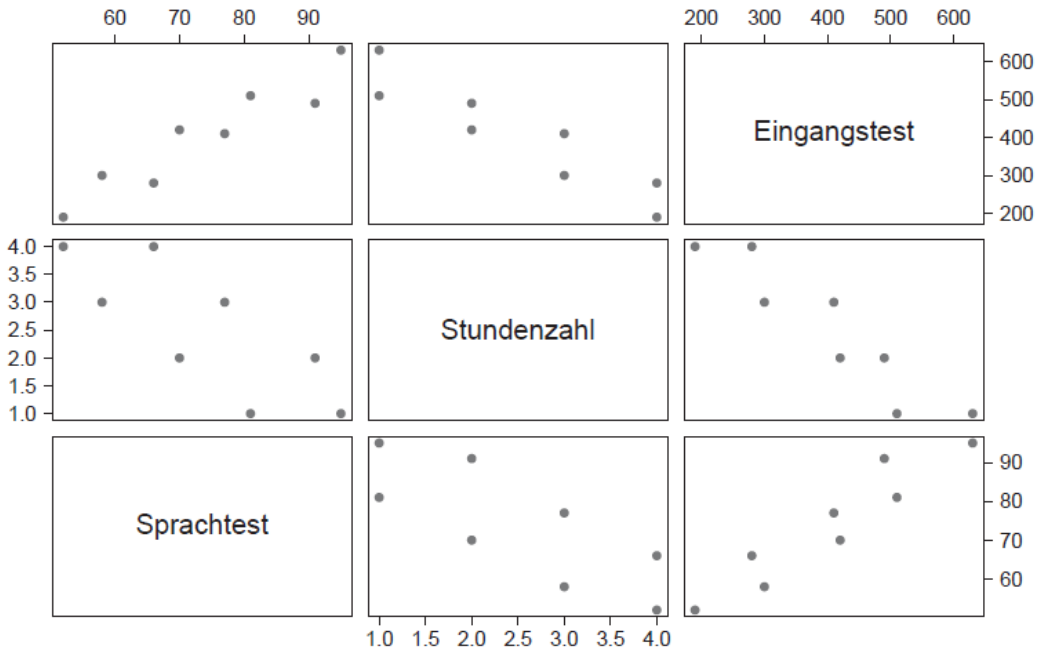
• Beispiel 12.2.1: Empirische partielle Regression •

Tab. 12.2.1: Nachhilfeunterricht an einer Sprachschule

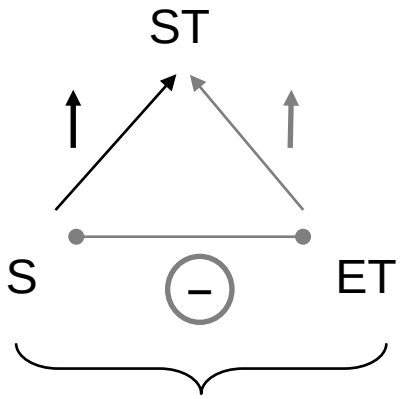
Schüler	Stundenzahl	Punkteanzahl im Eingangstest	Punkteanzahl im Sprachtest
1	4	190	52
2	4	280	66
3	3	300	58
4	3	410	77
5	2	420	70
6	2	490	91
7	1	510	81
8	1	630	95

$$\widehat{Sprachkenntnisse} = \underset{(8.71)}{98.75} \underset{(3.18)}{- 10.0 \cdot Stundenzahl}, \quad R^2 \approx 0.62, \quad SER \approx 10.06.$$

Ist Nachhilfe kontraproduktiv oder liegt hier möglicherweise ein OVB-Problem vor?



OVB-Szenario beschränkt auf *testscr*, *str* und *elpct*



positiver direkter Effekt +
negativer indirekter Effekt

Abb. 12.2.1: Nachhilfeunterricht an einer Sprachschule

	ST	S	ET
ST	1.0000	-0.7887	0.9444
S	-0.7887	1.0000	-0.9273
ET	0.9444	-0.9273	1.0000

Im einfachen Regressionsmodell sollte β_1 unterschätzt werden, da der positive **direkte Effekt** von *S* durch den negativen **indirekten Effekt** von *ET* kompensiert wird. Überkompensation führt hier sogar zu einem negativen **Gesamteffekt** von *S*.

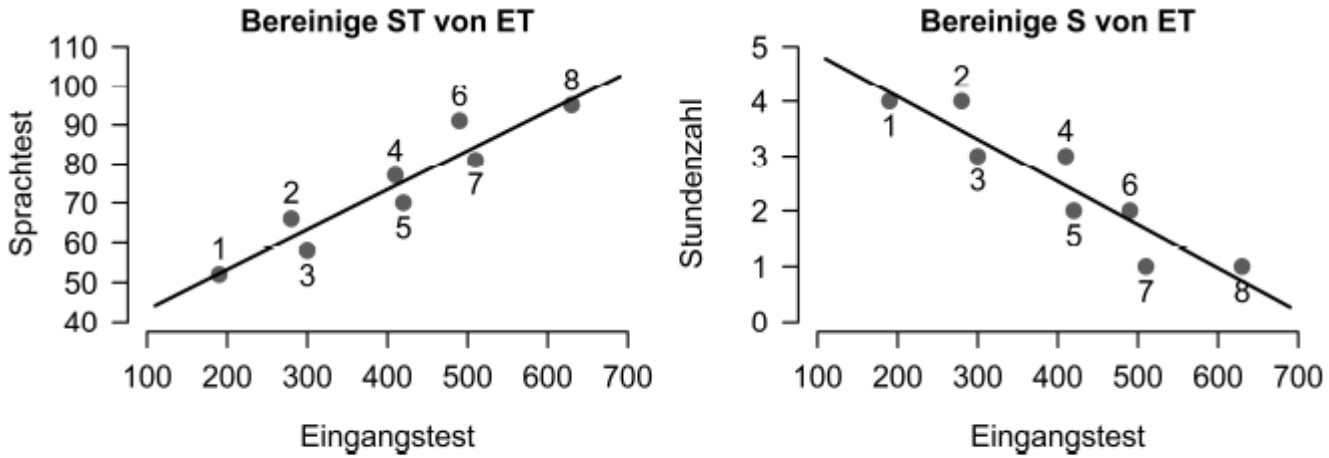
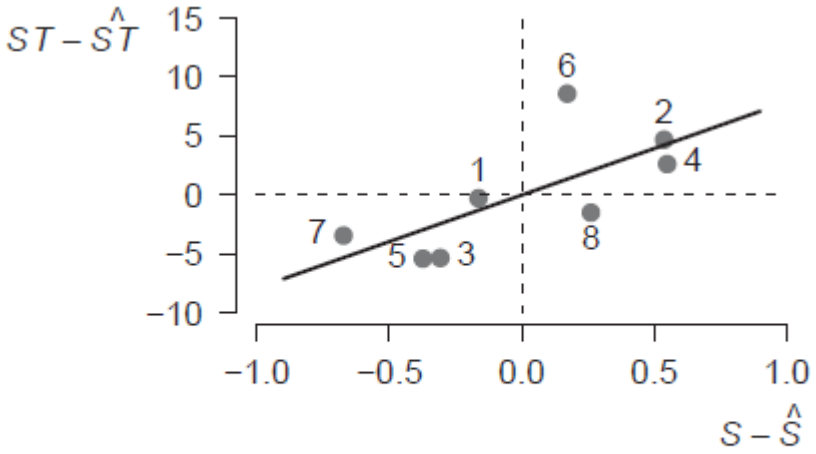


Abb. 12.2.2: Lineares Herausrechnen von ET aus ST und S

Interpretationsbeispiele:

Schüler 2: … bezogen auf sein Eingangsniveau
… überdurchschnittliches Abschneiden
… überdurchschnittlich viel Nachhilfe

Schüler 2: … bezogen auf sein Eingangsniveau
… unterdurchschnittliches Abschneiden
… unterdurchschnittlich viel Nachhilfe



Interpretation:
Schüler, die über- bzw. unter durchschnittlich viel Nachhilfe nehmen schneiden über- bzw. unterdurchschnittlich ab

Abb. 12.2.3: Partieller Regressionsplot von S und ST unter ET

> **Partielle Korrelation** zwischen *S* und *ST* unter *ET*:

$$r_{S,ST \bullet ET} = 0.70$$

> **Partieller Effekt** von *S* auf *ST* unter *ET*:

$$b_{ST \sim S \bullet ET} = 7.88$$

> Unter bestimmten Rahmenbedingungen entspricht der partielleEffekt dem direkten Effekt (theoretische Fundierung später durch Satz 12.2.3, Folie 979)

• Zusammenfassung und Formelapparat •

Empirische partielle Regression im 3-Variablenfall

Gegeben seien dreidimensionale Beobachtungswerte $(x_1, z_1, y_1), \ldots, (x_n, z_n, y_n)$ zu den Merkmalen X, Z und Y. Außerdem sei die (empirische) Kovarianzmatrix der Beobachtungen von X und Z invertierbar. Man führe nun folgende einfache lineare Regressionen und Berechnungen durch:

1. KQ-Regression von Y auf Z und Extraktion der KQ-Residuen.
2. KQ-Regression von X auf Z und Extraktion der KQ-Residuen.
3. KQ-Regression der Residuen aus Schritt 1 auf die Residuen aus Schritt 2.

Dann wird die im 3. Schritt durchgeführte Regression als **partielle Regression von** Y **auf** X **unter** Z bezeichnet. Der korrespondierende Steigungskoeffizient wird als **partieller Regressionskoeffizient** bezeichnet und mit

$$b_{Y\sim X\bullet Z}$$

notiert. Die Korrelation der Residuen aus Schritt 1 und Schritt 2 heißt **partielle Korrelation zwischen** X **und** Y **unter** Z und wird mit

$$r_{XY\bullet Z}$$

notiert. Unter Verwendung der üblichen Notation gilt dann:

$$b_{Y\sim X\bullet Z} = \frac{\tilde{s}_{XY}\tilde{s}_Z^2 - \tilde{s}_{YZ}\tilde{s}_{XZ}}{\tilde{s}_X^2\tilde{s}_Z^2(1-r_{XZ}^2)} \quad \text{und} \tag{12.2.1}$$

$$r_{XY\bullet Z} = \frac{r_{XY} - r_{YZ}r_{XZ}}{\sqrt{(1-r_{YZ}^2)(1-r_{XZ}^2)}}, \tag{12.2.2}$$

wenn die entsprechenden Ausdrücke definiert sind.

● Empirische Verzerrung ●

> Gegeben seien die drei Variablen X, Y und Z. Wie unterscheidet sich der empirische Regressionskoeffizient einer einfachen linearen Regression von Y auf X, notiert, $b_{Y\sim X}$ von demjenigen einer partiellen Regression von Y auf X unter Z, notiert mit $b_{Y\sim X\bullet Z}$?

> Es lässt sich zeigen, dass gilt:

$$b_{Y\sim X} = b_{Y\sim X\bullet Z} + \underbrace{b_{Y\sim Z\bullet X}\frac{\tilde{s}_{XZ}}{\tilde{s}_X^2}}_{\text{Empirische Verzerrung}} \quad \text{mit} \quad b_{Y\sim X} = \frac{\tilde{s}_{XY}}{\tilde{s}_X^2}$$

● Erweiterung auf höherdimensionale Fälle ●

Möchte man den störenden (OVB-verursachenden) Effekt von 2 oder mehr Variablen herausrechnen, benötigt man die Technik der **multiplen linearen Regression** (Unterabschnitt 12.2.2).

Theoretische partielle Regression

• Zusammenfassung und Formelapparat •

Theoretische partielle Regression im 3-Variablenfall

Gegeben sei ein dreidimensionaler Zufallsvektor (X, Z, Y). Die (theoretische) Kovarianzmatrix von $(X, Z)^T$ sei invertierbar. Man führe nun folgende theoretische Regressionen und Berechnungen durch:

1. Regression von Y auf Z und Extraktion des Prognosefehlers $Y - \overline{\widehat{Y}}$.
2. Regression von X auf Z und Extraktion des Prognosefehlers $X - \overline{\widehat{X}}$.
3. Regression des Fehlers aus Schritt 1 auf den Fehler aus Schritt 2.

Dann wird die im 3. Schritt durchgeführte Regression als **theoretische partielle Regression von Y auf X unter Z** bezeichnet. Der korrespondierende Steigungskoeffizient wird als **theoretischer partieller Regressionskoeffizient** bezeichnet und mit

$$\beta_{Y\sim X\bullet Z}$$

notiert. Die Korrelation der beiden Fehler aus Schritt 1 und Schritt 2 heißt **theoretische partielle Korrelation zwischen X und Y unter Z** und wird mit

$$\varrho_{XY \bullet Z}$$

notiert. Unter Verwendung der üblichen Notation gilt dann:

$$\beta_{Y \sim X \bullet Z} = \frac{\sigma_{XY}\sigma_Z^2 - \sigma_{YZ}\sigma_{XZ}}{\sigma_X^2\sigma_Z^2(1-\varrho_{XZ}^2)} \quad \text{und} \tag{12.2.5}$$

$$\varrho_{XY \bullet Z} = \frac{\varrho_{XY} - \varrho_{YZ}\varrho_{XZ}}{\sqrt{(1-\varrho_{YZ}^2)(1-\varrho_{XZ}^2)}}, \tag{12.2.6}$$

wenn die entsprechenden Ausdrücke definiert sind.

Formeln der theoretischen Quantitäten völlig analog zu denen der empirischen Quantitäten (vgl. Folie 953)

● Bedingte und partielle Korrelation ●

Im Allgemeinen unterscheiden sich bedingte und partielle Korrelationen. Unter bestimmten Bedingungen können diese jedoch identisch sein (etwa bei gemeinsamer **multivariater Normalverteilung**.

● Erweiterung auf höherdimensionale Fälle ●

Theoretische multiple Regression als Technik nötig…

● Theoretische Verzerrung ●

> Gegeben seien die drei Zufallsvariablen X, Y und Z. Wie unterscheidet sich der theoretische Regressionskoeffizient einer einfachen Regression von Y auf X, notiert, $\beta_{Y\sim X}$ von demjenigen einer partiellen Regression von Y auf X unter Z, notiert mit $\beta_{Y\sim X\bullet Z}$?

> Es lässt sich zeigen, dass gilt:

$$\beta_{Y\sim X} = \beta_{Y\sim X\bullet Z} + \underbrace{\beta_{Y\sim Z\bullet X}\frac{\sigma_{XZ}}{\sigma_X^2}}_{\text{Theoretische Verzerrung}} \quad \text{mit} \quad \beta_{Y\sim X} = \frac{\sigma_{XY}}{\sigma_X^2}$$

Verbindung von Empirie und Theorie

● Grundlagen ●

> Wir fassen nun dreidimensionale (empirische) Beobachtungswerte

$$(x_1, y_1, z_1), (x_2, y_2, z_2), \dots, (x_n, y_n, z_n)$$

als Realisationen dreidimensionaler Zufallsvektoren (Stichprobe)

$$(X_1, Y_1, Z_1), (X_2, Y_2, Z_2), \dots, (X_n, Y_n, Z_n)$$

auf und interpretieren das Ganze als Schätzproblem:

> Die theoretischen Quantitäten werden als zu schätzende Parameter aufgefasst und die stochastischen Varianten der empirischen Quantitäten als dazu korrespondierende Schätzer.

● Konsistente Schätzungen bei der partiellen Regression ●

> Wir definieren

$$\hat{\beta}_{Y\sim X\bullet Z} = \frac{\tilde{S}_{XY}\tilde{S}_Z^2 - \tilde{S}_{YZ}\tilde{S}_{XZ}}{\tilde{S}_X^2\tilde{S}_Z^2(1-R_{XZ}^2)} \quad \text{und} \quad \hat{\rho}_{XY\bullet Z} = \frac{R_{XY} - R_{YZ}^2 R_{XZ}^2}{(1-R_{YZ}^2)(1-R_{XZ}^2)}$$

> Aufgrund der vorhergehenden Resultate gilt dann unter u.i.v.-Schema:

$$\hat{\beta}_{Y\sim X\bullet Z} \xrightarrow{p} \beta_{Y\sim X\bullet Z} \text{ und } \hat{\rho}_{XY\bullet Z} \xrightarrow{p} \rho_{XY\bullet Z}$$

> Dies ergibt sich mit dem GGZ und dem multivariaten Stetigkeitssatz (man beachte auch Beispiel 8.3.6)

● Asymptotische Verzerrung ●

Definieren wir weiter

$$\hat{\beta}_{Y\sim X} = \frac{\tilde{S}_{XY}}{\tilde{S}_X^2} \text{ und } \beta_{Y\sim X} = \frac{\sigma_{XY}}{\sigma_X^2}$$

erhalten wir (vgl. hierzu Folien 944 und 947):

$$\hat{\beta}_{Y\sim X} \xrightarrow{p} \beta_{Y\sim X\bullet Z} + \underbrace{\beta_{Y\sim Z\bullet X}\frac{\sigma_{XZ}}{\sigma_X^2}}_{\text{Asymptotische Verzerrung}}$$

● Erweiterung auf höherdimensionale Fälle ●

12.2.2 Multiple lineare KQ-Regression

Empirische multiple Regression

● Motivation und Überblick ●

> Abhängige Variable hängt von mehreren Einflussgrößen linear ab

> Direkte (kausale) Effekte einzelner Variablen lassen sich simultan schätzen (unter bestimmten Modellannahmen)

> Erhöhung des Erklärungsgehalts des Modells im Vergleich zur einfachen linearen Regression; Erhöhung der Prognosegüte

> Lösungsansatz für das OVB-Problem (systematische Verzerrungen); Äquivalenz zur **partiellen Regression**

● Definition ●

siehe Folie 961

Definition 12.2.1: Empirische multiple Regression

Angenommen, es liegen $(p+1)$-dimensionale metrische Beobachtungswerte $(p \geq 1)$

$$(x_{11}, x_{12}, \ldots, x_{1p}, y_1), (x_{21}, x_{22}, \ldots, x_{2p}, y_2), \ldots, (x_{n1}, x_{n2}, \ldots, x_{np}, y_n)$$

korrespondierend zu den Merkmalen $X_1, X_2, \ldots, X_p$ und Y vor. Dabei sei die zu den X-Merkmalen korrespondierende (empirische) Kovarianzmatrix invertierbar. Dann bezeichnen wir die Lösungen zum Minimierungsproblem

$$\min_{b_0, b_1, \ldots, b_p} \widehat{Q}(b_0, b_1, \ldots, b_p) \quad \text{mit}$$

$$\widehat{Q}(b_0, b_1, \ldots, b_p) := \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \cdots - b_p x_{ip})^2$$

als **KQ-Regressionskoeffizienten** und notieren diese mit

$$\hat{b}_0, \hat{b}_1, \ldots, \hat{b}_p.$$

Die Berechnungsmethode heißt **multiple lineare KQ-Regression von Y auf $X_1, \ldots, X_p$.** Im Falle $p = 1$ spricht man von einer **einfachen linearen KQ-Regression.** Die Funktion

$$\hat{y}(x_1, x_2, \ldots, x_p) = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \cdots + \hat{b}_p x_p$$

12.2 Einführung in das multiple lineare Regressionsmodell

definiert im Falle $p \geq 2$ eine **KQ-Regressionsebene**, im Falle $p = 1$ eine **KQ-Gerade**. Die Werte

$$\hat{y}_i = \hat{y}(x_{i1}, x_{2i}, \ldots, x_{ip}) = \hat{b}_0 + \hat{b}_1 x_{i1} + \hat{b}_2 x_{i2} + \cdots + \hat{b}_p x_{ip} \quad \text{für } i = 1, \ldots, n$$

heißen **gefittete Werte** und sind als durch die Regressionsebene prognostizierte oder linear approximierte Werte interpretierbar. Die Abweichungen

$$\hat{u}_i = y_i - \hat{y}_i \quad \text{für } i = 1, \ldots, n$$

entsprechen den Prognose- bzw. Approximationsfehlern und werden als **KQ-Residuen** bezeichnet.

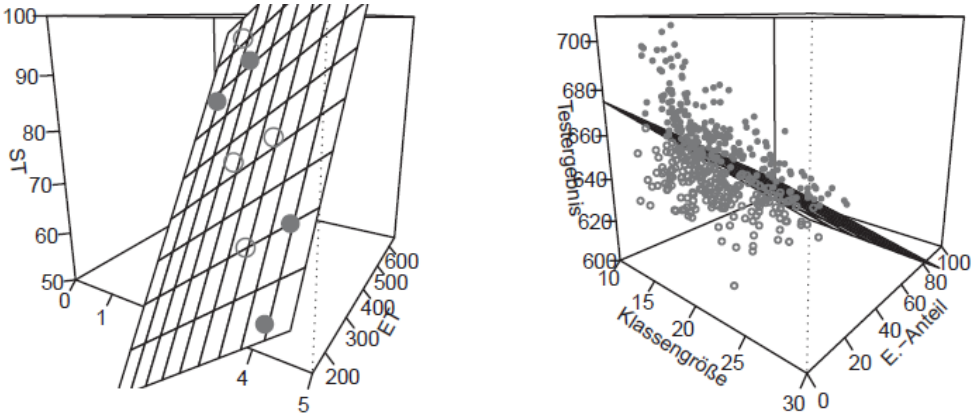


Abb. 12.2.4: KQ-Regressionsebenen im 3-Variablen-Fall

● Eigenschaften der empirischen multiplen KQ-Regression ●

Es gelten weiterhin die 6 Eigenschaften aus Satz 12.1.1 (Folie 855) mit entsprechender Anpassung für die Eigenschaften 1 und 4:

Eigenschaft Nr. 1: $\hat{y}(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p) = \bar{y}$

Eigenschaft Nr. 4: $r_{\hat{U}X_1} = 0, r_{\hat{U}X_2} = 0, \ldots, r_{\hat{U}X_p} = 0$

● Bestimmtheitsmaß und Standardfehler der Regression ●

> Mit Eigenschaft 6 (**Streuungszerlegungsformel**) gilt weiterhin:

$$R^2 = \frac{\tilde{s}_{\hat{Y}}^2}{\tilde{s}_Y^2} = 1 - \frac{\tilde{s}_{\hat{U}}^2}{\tilde{s}_Y^2}$$

> Der **Standardfehler der multiplen Regression** (***SER***) ist definiert als

$$SER = \sqrt{\frac{1}{n-p-1}\sum_{i=1}^{n} \hat{u}_i^2}$$

Dabei ist p die Anzahl der Regressoren ohne Konstante.

● Zusammenhang zur empirischen partiellen Regression ●

> Die Regressionskoeffizienten $\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_p$ bezüglich $X_1, X_2, \ldots, X_p$ einer einer multiplen Regression von Y auf $X_1, X_2, \ldots, X_p$ stimmen mit denen entsprechender partieller Regressionen überein, d. h.

$$\hat{b}_j = b_{Y \sim X_j \bullet \mathbf{X}_{-j}} \quad \text{für} \quad j = 1, \ldots, p.$$

Dabei ist $b_{Y \sim X_j \bullet \mathbf{X}_{-j}}$ der **partielle Regressionskoeffizient** einer partiellen Regression von Y auf X_j unter $X_1, X_2, \ldots, X_p$ ohne X_j

> Spezialfall $p = 2$ mit Y, X_1 und X_2:

Für die Regressionskoeffizienten $\hat{b}_1$ und $\hat{b}_2$ der multiplen Regression von Y auf X_1 und X_2 gilt:

$$\hat{b}_1 = b_{Y \sim X_1 \bullet X_2} = \frac{\tilde{s}_{X_1 Y} \tilde{s}^2_{X_2} - \tilde{s}_{X_2 Y} \tilde{s}_{X_1 X_2}}{\tilde{s}^2_{X_1} \tilde{s}^2_{X_2} \left(1 - r^2_{X_1 X_2}\right)}$$

und

$$\hat{b}_2 = b_{Y \sim X_2 \bullet X_1} = \frac{\tilde{s}_{X_2 Y} \tilde{s}^2_{X_1} - \tilde{s}_{X_1 Y} \tilde{s}_{X_1 X_2}}{\tilde{s}^2_{X_1} \tilde{s}^2_{X_2} \left(1 - r^2_{X_1 X_2}\right)}$$

Die KQ-Regressionskoeffizienten einer multiplen Regression stimmen mit den Koeffizienten einer entsprechenden partiellen Regression überein.

> Dies bedeutet dass zur Ermittlung partieller Regressionskoeffizienten das 3-stufige Verfahren gemäß Folie 952 mittels multipler Regression abgekürzt werden kann. Außerdem erhält man simultan bezüglich aller Variablen die korrespondierenden partiellen Koeffizienten

> Berechnung mit Hilfe statistischer Software (etwa R)

● Beispiel 12.2.1 fortgesetzt ●

> Gegeben nochmals die Situation von Folie 948 mit dem OVB-Problem.

> Eine multiple lineare Regression von *ST* auf *S* und *ET* ergibt:

$$\widehat{ST} = -11.28 + 7.88 \cdot S + 0.16 \cdot ET, \quad R^2 \approx 0.95, \quad SER \approx 4.17.$$

$\hat{b}_S = b_{ST \sim S \bullet ET}$

Vergleich mit Folien 951: Übereinstimmung

$\hat{b}_{ET} = b_{ST \sim ET \bullet S}$

Vergleich mit Folie 948: Anstieg bzw. Rückgang

Theoretische multiple Regression

• Definition •

Definition 12.2.2: Theoretische multiple Regression

Gegeben seien $(p+1)$ Zufallsvariablen $X_1, X_2, \ldots, X_p$ und Y mit $p \geq 1$, welche Annahme (M) aus Definition 8.2.1 genügen. Außerdem sei die (theoretische) Kovarianzmatrix von $(X_1, \ldots, X_p)^T$ invertierbar. Dann bezeichnen wir die Lösungen zum Minimierungsproblem

$$\min_{\beta_0,\beta_1,\ldots,\beta_p} \overline{\overline{Q}}(\beta_0, \beta_1, \ldots, \beta_p) \quad \text{mit}$$

$$\overline{\overline{Q}}(\beta_0, \beta_1, \ldots, \beta_p) := E[(Y_i - \beta_0 - \beta_1 X_1 - \beta_2 X_2 - \cdots - \beta_p X_p)^2]$$

als **theoretische KQ-Regressionskoeffizienten** und notieren diese mit

$$\overline{\overline{\beta}}_0, \overline{\overline{\beta}}_1, \ldots, \overline{\overline{\beta}}_p.$$

Die Funktion

$$\bar{\hat{y}}(x_1, x_2, \ldots, x_p) = \bar{\hat{\beta}}_0 + \bar{\hat{\beta}}_1 x_1 + \cdots + \bar{\hat{\beta}}_p x_p$$

definiert im Falle $p \geq 2$ eine **theoretische KQ-Regressionsebene**, im Falle $p = 1$ eine **theoretische KQ-Gerade**. Die stochastische Statistik

$$\bar{\hat{Y}} = \bar{\hat{\beta}}_0 + \bar{\hat{\beta}}_1 X_1 + \cdots + \bar{\hat{\beta}}_p X_p$$

heißt **bester linearer MSE Prädiktor** von Y auf Basis von $X_1, \ldots, X_p$.

Kompakte formelmäßige Darstellung der KQ-Koeffizienten erfolgt üblicherweise in matrixalgebraischer Form (Inhalt fortgeschrittener Kurse)

• Eigenschaften der theoretischen multiplen KQ-Regression •

Analog zu den Eigenschaften der empirischen multiplen Regression (Folie 963) gelten folgende Eigenschaften:

(1) $\bar{\hat{y}}\left(\mu_{X_1}, \mu_{X_2}, \ldots, \mu_{X_p}\right) = \mu_Y$

(2) $E\left(\bar{\hat{Y}}\right) = E(Y) = \mu_Y$

(3) $E\left(Y - \bar{\hat{Y}}\right) = 0$

(4) $Cov\left(Y-\bar{\hat{Y}}, X_1\right)=0, \ldots, Cov\left(Y-\bar{\hat{Y}}, X_p\right)=0$

(5) $Cov\left(Y-\bar{\hat{Y}}, \bar{\hat{Y}}\right)=0$

(6) $\underbrace{Var(Y)=Var\left(\bar{\hat{Y}}\right)+Var\left(Y-\bar{\hat{Y}}\right)}$

Theoretische Streuungszerlegungsformel der Regression

- **Theoretisches Bestimmtheitsmaß und theoretischer Standardfehler** -

$$\Re^2=\frac{Var\left(\bar{\hat{Y}}\right)}{Var(Y)}=1-\frac{Var\left(Y-\bar{\hat{Y}}\right)}{Var(Y)}$$

Theoretisches Bestimmtheitsmaß

$$\Rightarrow \quad \sqrt{Var\left(Y-\bar{\hat{Y}}\right)}=\sqrt{Var(Y)(1-\Re^2)}$$

Theoretischer Standardfehler

● Zusammenhang zur theoretischen partiellen Regression ●

> Die Regressionskoeffizienten $\bar{\hat{\beta}}_1, \bar{\hat{\beta}}_2, \ldots, \bar{\hat{\beta}}_p$ bezüglich $X_1, X_2, \ldots, X_p$ einer theoretischen multiplen Regression von Y auf $X_1, X_2, \ldots, X_p$ stimmen mit denen entsprechender theoretischer partieller Regressionen überein, d. h.

$$\bar{\hat{\beta}}_j = \beta_{Y \sim X_j \bullet \mathbf{X}_{-j}} \quad \text{für} \quad j = 1, \ldots, p.$$

Dabei ist $\beta_{Y \sim X_j \bullet \mathbf{X}_{-j}}$ der **partielle Regressionskoeffizient** einer theoretischen partiellen Regression von Y auf X_j unter $X_1, X_2, \ldots, X_p$ ohne X_j

> Spezialfall $p = 2$ mit Y, X_1 und X_2:

Für die Regressionskoeffizienten $\bar{\hat{\beta}}_1$ und $\bar{\hat{\beta}}_1$ der theoretischen multiplen Regression von Y auf X_1 und X_2 gilt:

$$\bar{\hat{\beta}}_1 = \beta_{Y \sim X_1 \bullet X_2} = \frac{\sigma_{X_1 Y}\sigma_{X_2}^2 - \sigma_{X_2 Y}\sigma_{X_1 X_2}}{\sigma_{X_1}^2 \sigma_{X_2}^2 \left(1 - \rho_{X_1 X_2}^2\right)}$$

und

$$\bar{\hat{\beta}}_2 = \beta_{Y \sim X_2 \bullet X_1} = \frac{\sigma_{X_2 Y}\sigma_{X_1}^2 - \sigma_{X_1 Y}\sigma_{X_1 X_2}}{\sigma_{X_1}^2 \sigma_{X_2}^2 \left(1 - \rho_{X_1 X_2}^2\right)}$$

Die KQ-Regressionskoeffizienten einer theoretischen multiplen Regression stimmen mit den Koeffizienten einer entsprechenden theoretischen partiellen Regression überein.

Verbindung von Empirie und Theorie

Da nun die KQ-Regressionskoeffizienten der empirischen und der theoretischen multiplen Regression mit den Koeffizienten der empirischen bzw. theoretischen partiellen Regressionen übereinstimmen, können die im vorhergehenden Abschnitt hergeleiteten Konsistenzeigenschaften direkt übertragen werden. Demnach können im Zusammenhang von Stichproben die aus der multiplen Regression gewonnenen empirischen Regressionskoeffizienten als Schätzer der korrespondierenden theoretischen Größen aufgefasst werden. Unter bestimmten Annahmen wie etwa unter einem u.i.v.-Schema lassen sich hierbei entsprechende Konsistenzeigenschaften begründen.

Statistische Modelle und Inferenz

- Hintergrund
- Statistisches Grundmodell

Definition 12.2.3: Multiples lineares Regressionsmodell

Gegeben sei eine Stichprobe aus n beobachtbaren Zufallsvektoren

$$(X_{11}, X_{12}, \ldots, X_{1p}, Y_1), \ldots, (X_{n1}, X_{n2}, \ldots, X_{np}, Y_n)$$

mit $p \geq 1$ und n nicht beobachtbaren Zufallsvariablen $U_1, \ldots, U_n$. Gilt dann

(A0M) $Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + U_i$ für $i = 1, \ldots, n$,

so bezeichnet man (A0M) als **multiples lineares Regressionsmodell**. Die Parameter $\beta_0, \beta_1, \ldots, \beta_p$ werden als **theoretische Regressionskoeffizienten** und die Variablen $U_1, \ldots, U_n$ als **stochastische Fehler** oder **stochastische Residuen** bezeichnet. Die Funktion

$$y(x) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

definiert die **theoretische Regressionsebene**, wobei der Definitionsbereich von $x_1, \ldots, x_p$ fallabhängig einzuschränken ist.

● Interpretation ●

> Y_i … i-te Beobachtung der abhängigen (kausal beeinflussten) Variablen, auch **Regressand**

> X_{ij} … i-te Beobachtung der j-ten Einflussvariable, auch **Regressor**

> Regressoren üben mehr oder weniger direkte kausale Effekte auf Regressanden aus

> In den Fehlern U_i stecken alle weiteren (kausalen) Einflüsse auf Y_i

> Abhängigkeit linear:
Veränderung der j-ten Einflussvariable um Δx Einheiten führt zu einer Veränderung des abhängigen Variable um $\beta_j \Delta x$ Einheiten, gegeben, dass andere Einflussvariablen und Fehler konstant bleiben.

> Deutung der Regressionskoeffizienten als direkte (kausale) Effekte der einzelnen Einflussvariablen

• Ensembles von Modellannahmen •

Definition 12.2.4: Modell BH als multiples Regressionsmodell

Wir sprechen von einem **multiplen linearen Regressionsmodell mit bedingt heteroskedastischem Fehler**, falls gilt:

(A0M) $Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + U_i$ für $i = 1, \ldots, n$,

(A1M) $E(U_i | X_{i1} = x_{i1}, \ldots, X_{ip} = x_{ip}) = 0$ für $i = 1, \ldots, n$,

(A2M) $(X_{11}, X_{12}, \ldots, X_{1p}, Y_1), \ldots, (X_{n1}, X_{n2}, \ldots, X_{np}, Y_n)$ sind u.i.v.,

(A3M) alle Zufallsvariablen genügen Annahme (M) aus Definition 8.2.1 und besitzen positive Varianzen und

(A4M) die Kovarianzmatrix von $(X_{i1}, X_{i2}, \ldots, X_{ip})^T$ ist invertierbar.

• Keine Multikollinearität •

(A4M): Regressoren dürfen nicht perfekt linear abhängig sein

● Modellimmanente Eigenschaften ●

> Eigenschaften (1) bis (17) von Tabelle 12.1.1 (Folie 878) können analog auf den multiplen Fall übertragen werden. Insbesondere gilt dann etwa:

$$Cov(X_{ij}, U_i) = E(X_{ij}U_i) = 0 \text{ für } i = 1, \dots, n \text{ und } j = 1, \dots, p$$

⇒ Fehler ist mit allen Regressoren unkorreliert

> Zu den Eigenschaften (16) und (17) beachte man nachfolgenden Satz.

● Theoretische Regressionsebene = theoretische KQ-Regressionsebene ●

Satz 12.2.1: Theoretische Regressionsebene = KQ-Regressionsebene

Die in den multiplen Modellen KS und BH enthaltene theoretische Regressionsebene ist die KQ-Regressionsebene einer theoretischen Regression von Y auf $X_1, \dots, X_p$. Dabei gilt:

$$\beta_0 = \mu_Y - \beta_1\mu_{X_1} - \cdots - \beta_p\mu_{X_p}, \tag{12.2.29}$$

$$\beta_j = \beta_{Y\sim X_j \bullet \mathbf{X}_{-j}} \quad \text{für } j = 1, \dots, p \quad \text{und} \tag{12.2.30}$$

$$\sigma_U^2 = \sigma_Y^2(1 - \Re^2). \tag{12.2.31}$$

● Schätzen und Testen ●

> Formeln für die Regressionskoeffizienten aus der empirischen multiplen Regression werden in stochastische Varianten überführt und als Schätzer interpretiert

> In Modellen KS und BH (u.i.v.-Schema) ergibt sich dann etwa im Spezialfall $p = 2$:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2 \xrightarrow{p} \beta_0$$

$$\hat{\beta}_1 = \frac{\tilde{S}_{X_1 Y} \tilde{S}_{X_2}^2 - \tilde{S}_{X_2 Y} \tilde{S}_{X_1 X_2}}{\tilde{S}_{X_1}^2 \tilde{S}_{X_2}^2 \left(1 - R_{X_1 X_2}^2\right)} \xrightarrow{p} \beta_1$$

$$\hat{\beta}_2 = \frac{\tilde{S}_{X_2 Y} \tilde{S}_{X_1}^2 - \tilde{S}_{X_1 Y} \tilde{S}_{X_1 X_2}}{\tilde{S}_{X_1}^2 \tilde{S}_{X_2}^2 \left(1 - R_{X_1 X_2}^2\right)} \xrightarrow{p} \beta_2$$

> In Modellen KN und UHV zusätzliche Konvergenzannahmen nötig zum Erhalt obiger Konsistenzresultate

> Entwicklung der Inferenzverfahren ansonsten völlig analog wie im einfachen Regressionsmodell:

1. Herleitung der Verteilung der KQ-Schätzer
2. Schätzung der Varianzen (Standardfehler) der KQ-Schätzer
3. Herleitung der Verteilung der maßgeblichen Inferenzstatistiken

> Keine näheren Ausführungen hierzu (Inhalt fortgeschrittener Kurse)

> Ergebnisse (ohne Herleitung) lauten wie folgt:

Satz 12.2.2: Inferenz im multiplen Regressionsmodell

Gegeben sei ein multiples lineares Regressionsmodell gemäß Definition 12.2.3. Dann gilt für $j = 0, 1, \ldots, p$ und $n \geq p + 2$:

1. ein $(1 - \alpha)$-Konfidenzintervall für β_j ist
   (i) in den Modellen KN und KS gegeben durch

$$[\hat{\beta}_j - t_{n-p-1,1-\alpha/2}\tilde{\sigma}_{\hat{\beta}_j}, \hat{\beta}_j + t_{n-p-1,1-\alpha/2}\tilde{\sigma}_{\hat{\beta}_j}],$$

   (ii) in Modell BH für großes n gegeben durch

$$[\hat{\beta}_j - z_{1-\alpha/2}\hat{\sigma}_{\hat{\beta}_j}, \hat{\beta}_j + z_{1-\alpha/2}\hat{\sigma}_{\hat{\beta}_j}],$$

wobei $\widetilde{\sigma}^2_{\hat{\beta}_j}$ bzw. $\widehat{\sigma}^2_{\hat{\beta}_j}$ **nur-homoskedasizitäts-konsistente Varianzschätzer** bzw. **heteroskedastizitäts-robuste Varianzschätzer** bezeichnen,

2. die Entscheidungsregeln der üblichen Tests basieren
   (i) in den Modellen KN und KS auf

$$\widetilde{T}_{\beta_j} = \frac{\hat{\beta}_j - \beta_{j,0}}{\widetilde{\sigma}_{\hat{\beta}_j}} \sim t(n-p-1) \quad \text{für } \beta_j = \beta_{j,0},$$

   (ii) in Modell BH für großes n auf

$$\widehat{T}_{\beta_j} = \frac{\hat{\beta}_j - \beta_{j,0}}{\widehat{\sigma}_{\hat{\beta}_j}} \overset{a}{\sim} N(0,1) \quad \text{für } \beta_j = \beta_{j,0},$$

3. mit entsprechenden zusätzlichen Konvergenzannahmen sind die Konfidenzintervalle und Tests für Modell BH auch für Modell UHV gültig.

> Kompakte formelmäßige Darstellung erfolgt üblicherweise in matrixalgebraischer Form (Inhalt fortgeschrittener Kurse)

> Berechnung mit Hilfe statistischer Software (wie etwa R)

● Adäquatheit von Modellannahmen und OVB ●

> Prüfung erfolgt analog wie im einfachen Regressionsmodell (vgl. Folie 922 ff.) in erweitertem Sinne

> Aus $E(U_i|X_i = x_i) = 0$ wird $E\left(U_i|X_{i1} = x_{i1}, X_{i2} = x_{i2}, \ldots, X_{ip} = x_{ip}\right) = 0$ mit der korrespondierenden Leitfrage…

Leitfrage für das OVB-Problem

Gibt es neben den Regressoren $X_1, \ldots, X_p$ noch eine weitere maßgebliche Einflussgröße auf Y, die mit mindestens einem Regressor korreliert ist?

● OVB im 3-Variablen-Fall ●

> OVB-Leitfrage erhält für $p = 2$ folgende theoretische Fundierung, welche auch die Daumenregeln zu systematischen Verzerrungen aus Unterkapitel 5.3 begründen (zur Herleitung siehe LB, S. 739):

Satz 12.2.3: OVB im 3-Variablen-Fall

Angenommen, Modell KS oder BH mit der Grundgleichung

$$Y = \beta_0 + \beta_X \cdot X + \beta_Z \cdot Z + V$$

sei in einer bestimmten Situation adäquat. Schätzt man

$$Y = \beta_0 + \beta_X \cdot X + U$$

mit den gewöhnlichen KQ-Schätzern, so beträgt der OVB für β_X

$$OVB_X = \beta_Z \frac{\sigma_{XZ}}{\sigma_X^2}, \quad \text{wobei} \quad \beta_Z \sigma_{XZ} = Cov(X, U). \qquad (12.2.39)$$

● Direkte und indirekte Effekte und Merkregeln ●

> Schätzen wir den direkten Effekt von X auf Y, notiert mit β_X, (fälschlicherweise) ohne Berücksichtigung von Z nur mit einer einfachen Regression von Y auf X, so ist die Schätzung wegen $Cov(X, U) \neq 0$ asymptotisch verzerrt mit

$$OVB_X = \frac{Cov(X,U)}{\sigma_X^2} \text{ oder } OVB_X = \beta_Z \beta_{Z\sim X}$$

> Man erhält dann stattdessen eine Schätzung für den Gesamteffekt mit

$$\beta_{Y\sim X} = \beta_X + OVB_X$$

oder $\beta_{Y\sim X} = \beta_X + \beta_Z\beta_{Z\sim X}$ mit $\beta_{Z\sim X} = \frac{\sigma_{XZ}}{\sigma_X^2}$

Gesamteffekt = Direkter Effekt + Indirekter Effekt.

von X auf Y | von X auf Y | von X über Z auf Y

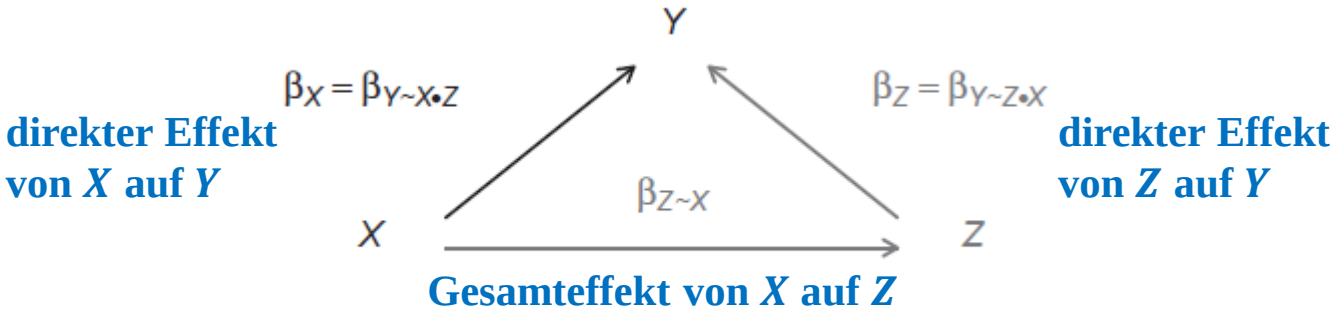


Abb. 12.2.5: Direkte und indirekte Effekte

12.2.3 Fallbeispiele

Fallbeispiel 2 fortgesetzt: Determinanten des Lernerfolgs

Tab. 12.2.2: Geschätzte Varianten für Modell BH

Modell / Regressor	1	2	3	4
Konstante	698.93	686.03	702.91	700.15
	(10.34)	(8.70)	(5.50)	(5.54)
Klassengröße	−2.28	−1.10	−1.12	−1.00
	(0.52)	(0.43)	(0.27)	(0.27)
E.-Anteil		−0.65		−0.12
		(0.03)		(0.03)
B.-Anteil			−0.60	−0.55
			(0.02)	(0.02)
R^2	0.0512	0.4264	0.7668	0.7745
SER	18.58	14.46	9.22	9.08

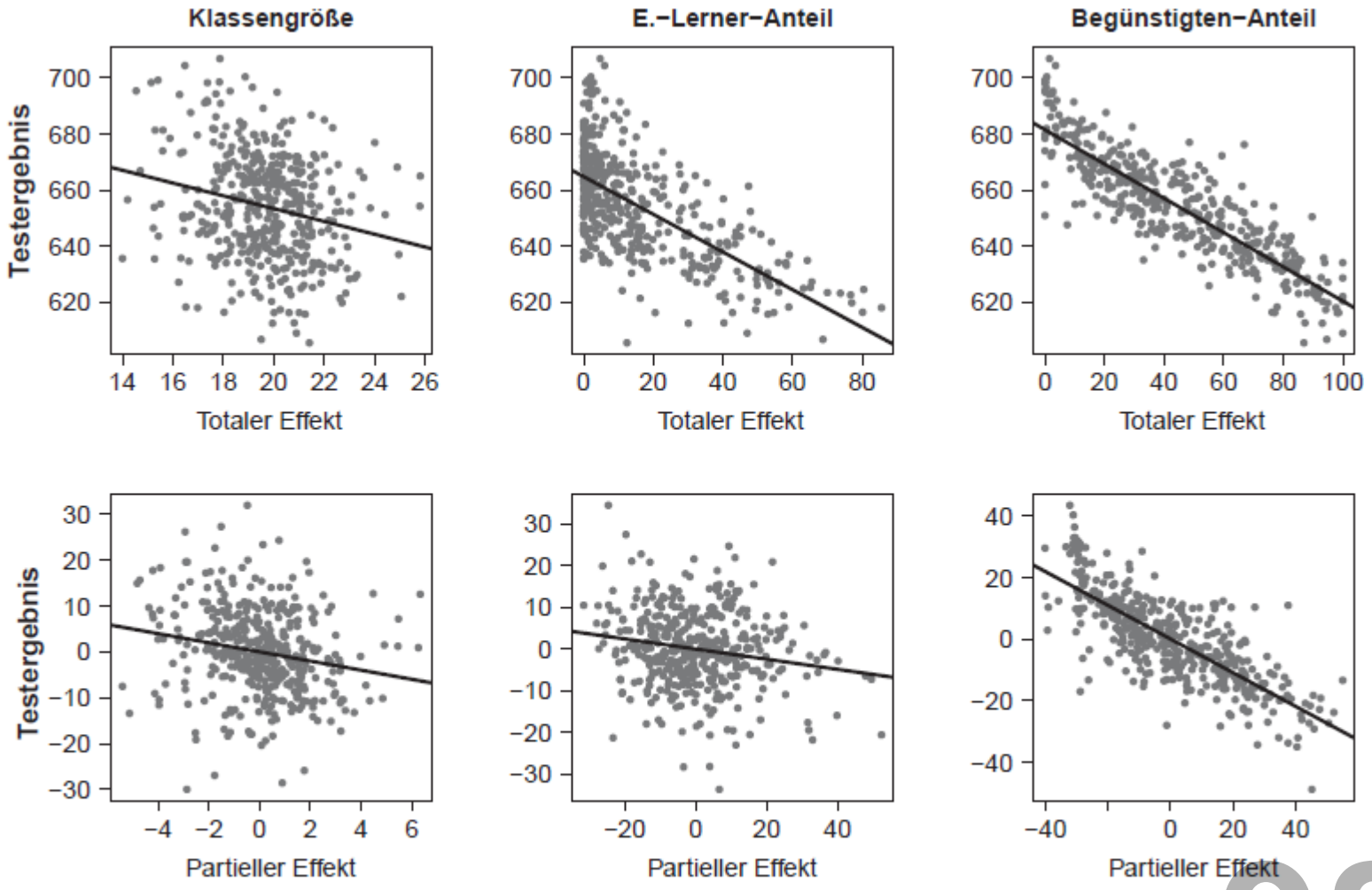


Abb. 12.2.6: Totale und direkte Effekte der Einflussvariablen

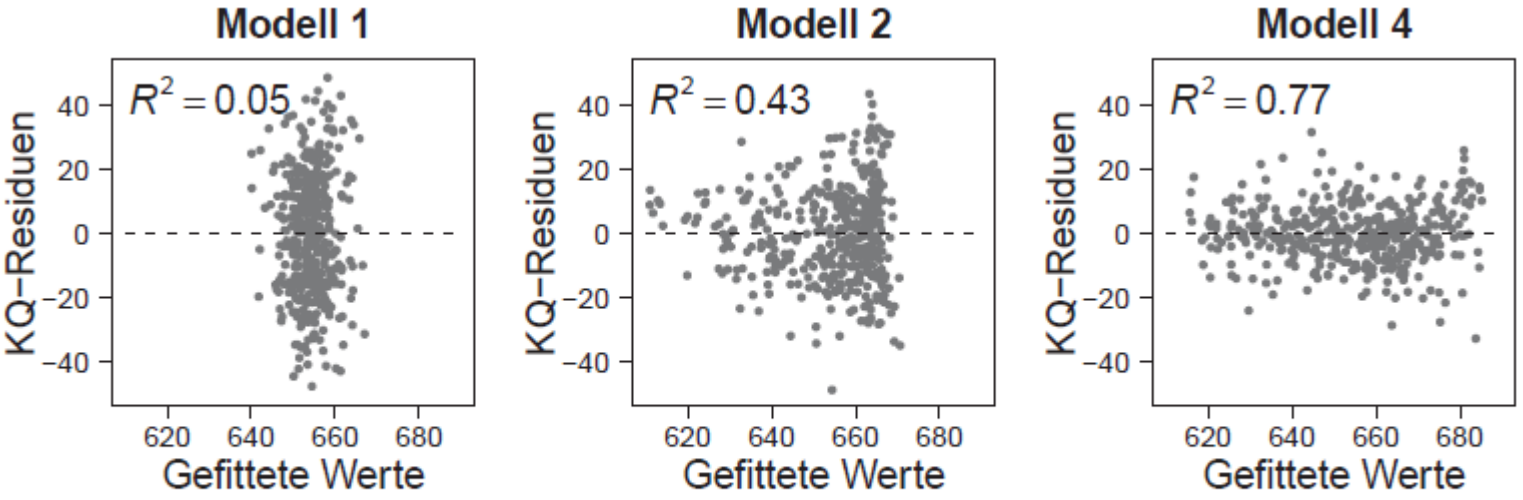


Abb. 12.2.7: Residualplots der Modelle 1, 2 und 4

Fallbeispiel 3: Gewicht und Geschlecht

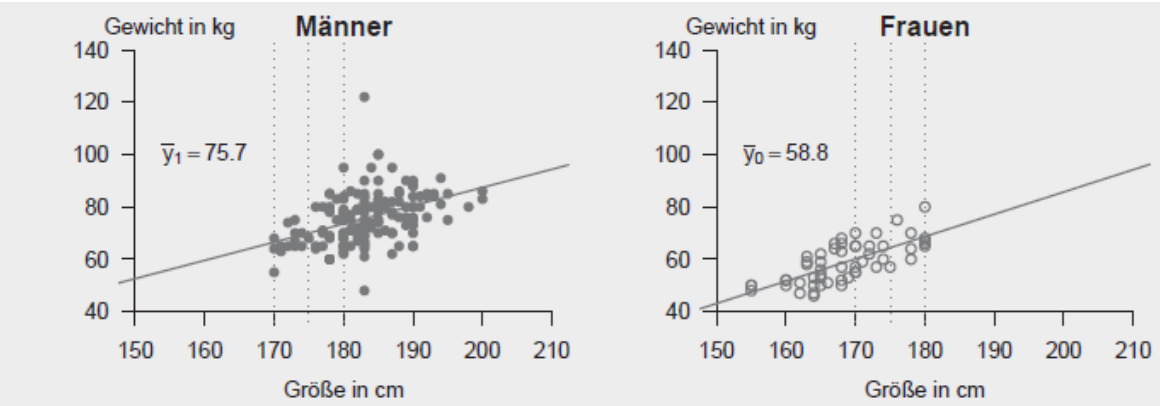


Abb. 5.3.18: Zusammenhang von Größe und Gewicht nach Geschlecht

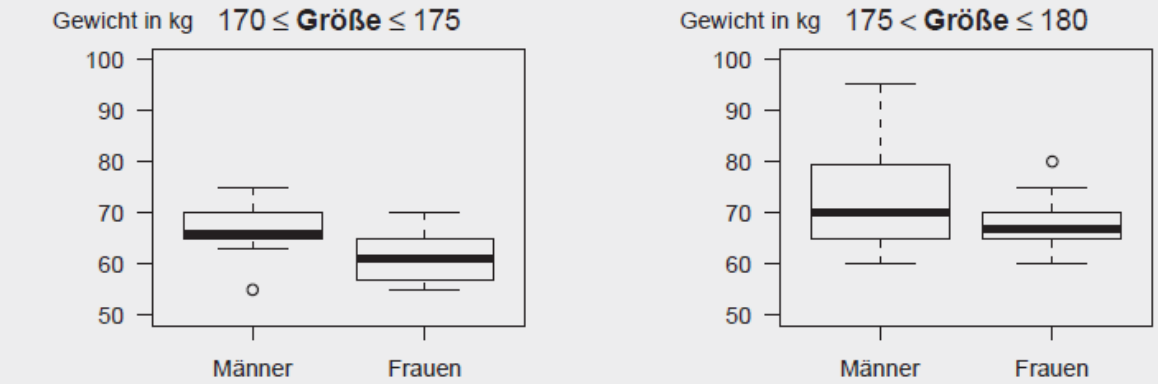


Abb. 5.3.19: Zusammenhang von Geschlecht und Gewicht nach Größe

$$\widehat{Gewicht} = \underset{(0.78)}{75.68} - \underset{(1.50)}{16.88} \cdot Geschlecht, \quad R^2 \approx 0.39, \quad SER \approx 9.45.$$

$$\widehat{Größe} = \underset{(0.51)}{183.17} - \underset{(0.99)}{14.74} \cdot Geschlecht, \quad R^2 \approx 0.53, \quad SER \approx 6.20.$$

$$\widehat{Gewicht} = \underset{(0.78)}{-60.36} - \underset{(1.50)}{5.93} \cdot Geschlecht + \underset{(0.09)}{0.74} \cdot Größe, \; R^2 \approx 0.53, \; SER \approx 8.23.$$

Der Gesamteffekt von Geschlecht auf Gewicht ist zerlegbar in

$$\hat{\beta}_{Gewicht \sim Geschlecht} = \hat{\beta}_{Gewicht \sim Geschlecht \bullet Größe} + \hat{\beta}_{Gewicht \sim Größe \bullet Geschlecht} \cdot \hat{\beta}_{Größe \sim Geschlecht}$$

In Zahlen ergibt sich dabei im vorliegenden Fall:

$$-16.88 = -5.93 + 0.74 \cdot (-14.74)$$

Fallbeispiel 4: Binäre Regressoren und ANOVA-Modelle

● Einfache Regression mit binärem Regressor ●

> Statistische Inferenz in Bezug auf Vergleich zweier Gruppen

> **General-F-Test** kann für Erwartungswertvergleich herangezogen werden.

● Einfache ANOVA-Modelle ●

> Statistische Inferenz in Bezug auf Mehrgruppenvergleiche

> **General-F-Test** kann für Erwartungswertvergleich herangezogen werden:

$$H_0\colon \mu_0 = \mu_1 = \cdots = \mu_p \quad vs. \quad H_1\colon \text{ nicht } H_0$$

> ANOVA-Modelle und sind jedoch noch deutlich flexibler in Bezug auf statistische Problemstellungen.

> General-F-Test nur Spezialfall aus der vielseitig anwendbaren „F-Test-Familie“ (Inhalt fortgeschrittener Kurse)