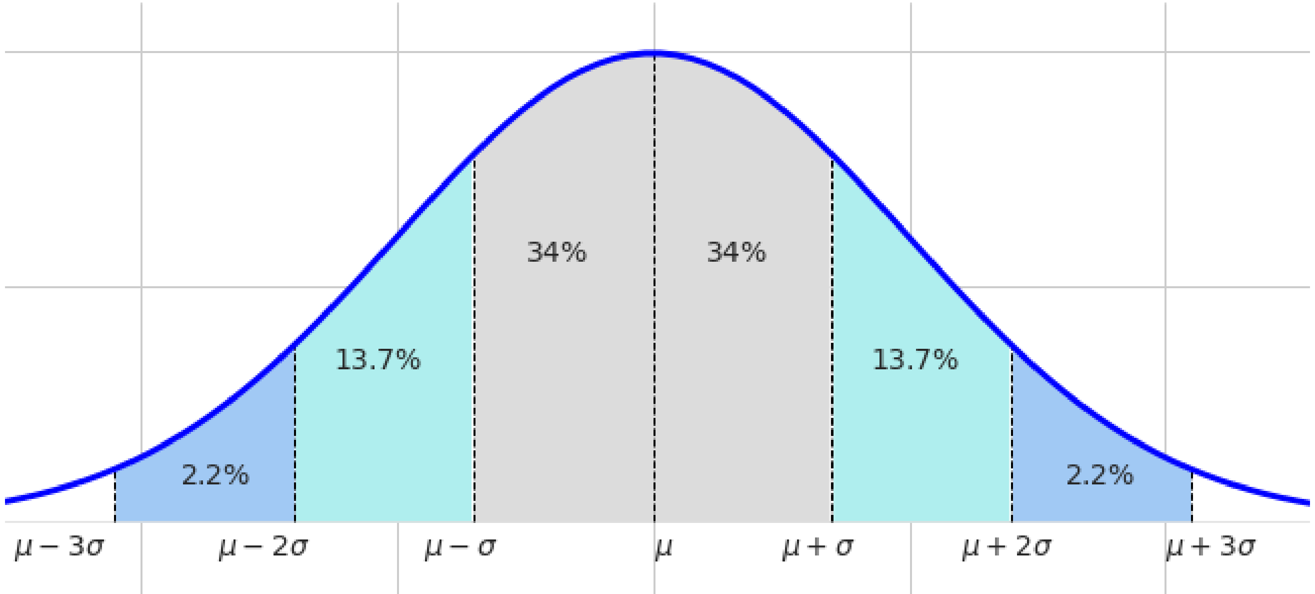
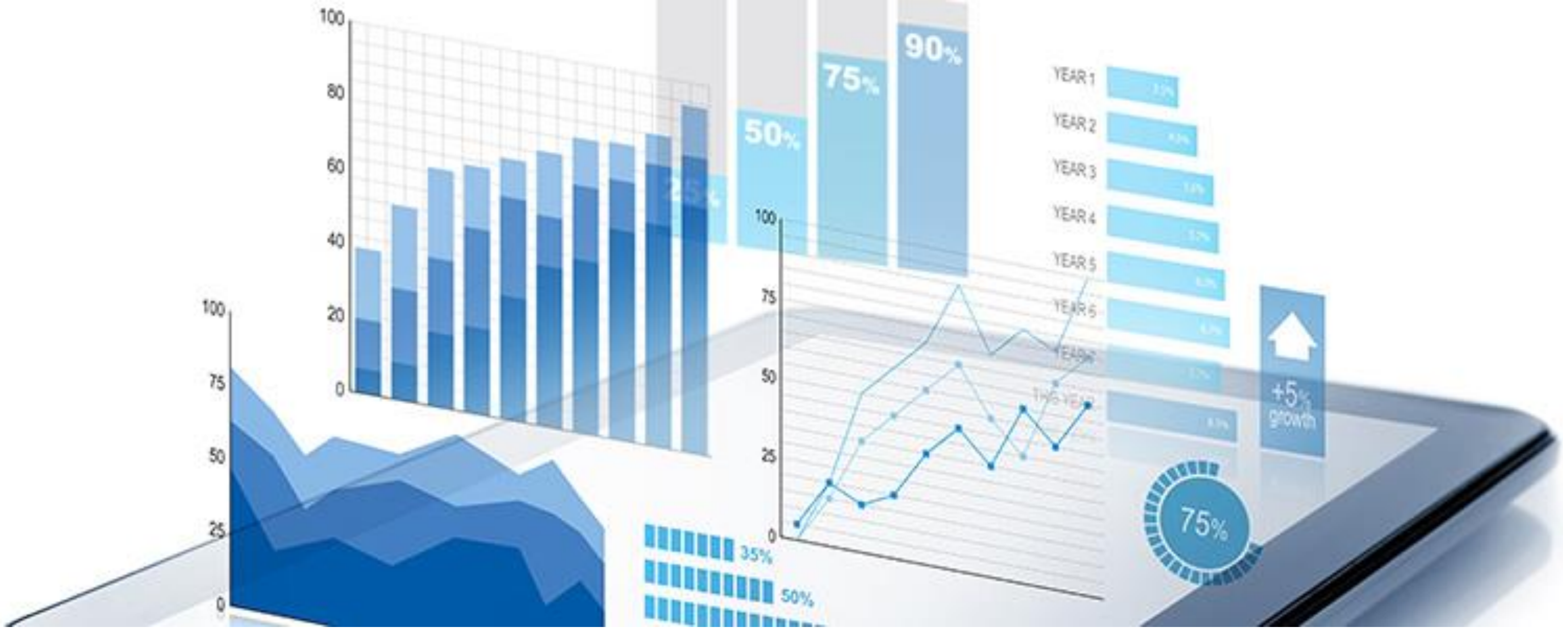
HOCHSCHULE HEILBRONN



› STATISTIK

Statistik | Prof. Dr. Florian Kauffeldt

HOCHSCHULE HEILBRONN



› EINFÜHRUNG

Statistik | Prof. Dr. Florian Kauffeldt

VORLESUNG:
BEDEUTUNG BERUF

Warum ist Wissen über Datenanalyse wichtig für ihr künftiges Arbeitsleben?

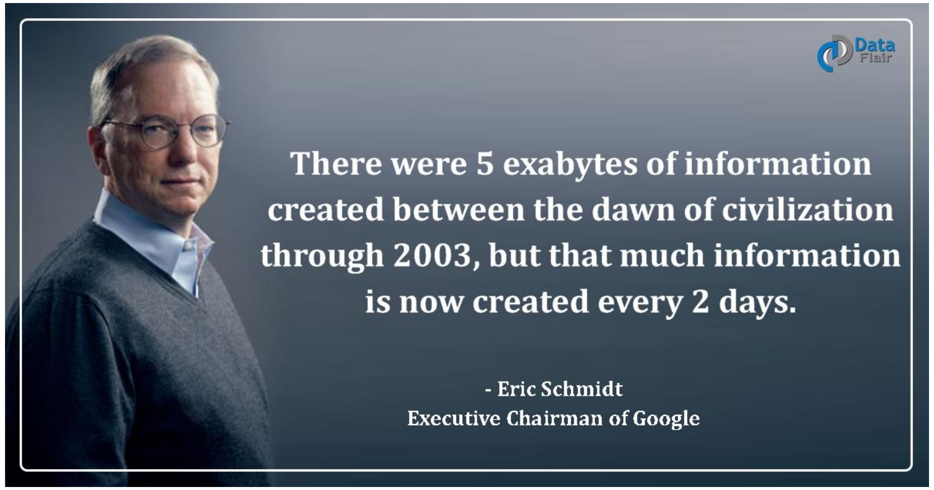
1. Grundlagenwissen Statistik ist obligatorisch für Führungspositionen
2. Tendenz steigend aufgrund von neue Technologien wie z.B. Machine Learning

Warum?

> Unternehmen: Bestmöglichen Entscheidungen basierend auf vorliegenden Daten

> Zunehmende Computerisierung: Datenverarbeitung in Echtzeit

ZITATE GESCHÄFTSWELT



VORLESUNG: VERBINDUNG ZU ANDEREN KURSEN

HOCHSCHULE HEILBRONN

- Sozialwissenschaftliche Forschung ist häufig empirisch (basiert auf Daten)
- Unabdingbar, um wissenschaftliche Forschungsarbeiten zu verstehen
- Methodologische Verbindung zu den meisten anderen Vorlesungen
- Fundamentale Instrumente für empirisches Arbeiten

WARUM STATISTIK?

HOCHSCHULE HEILBRONN

Gender Pay Gap

= Differenz durchschnittlicher Verdienst Frauen und Männer geteilt durch Verdienst Männer

> Fakt:

Frauen verdienen 18 % weniger je Stunde als Männer (2020)

(Quelle: https://www.destatis.de/DE/Themen/Arbeit/Arbeitsmarkt/Qualitaet-Arbeit/Dimension-1/gender-pay-gap.html)

> Interpretation: "Systematische Benachteiligung von Frauen"

Ist das zwangsläufig so?

WARUM STATISTIK?

HOCHSCHULE HEILBRONN

- Fakt:

Heilbronn ist die Stadt mit dem höchsten Durchschnittseinkommen in Deutschland

(siehe z.B.: https://de.wikipedia.org/wiki/Liste_der_Landkreise_nach_Einkommen)

- Interpretation: "Heilbronn ist die reichste Stadt Deutschlands"

Stimmt das?

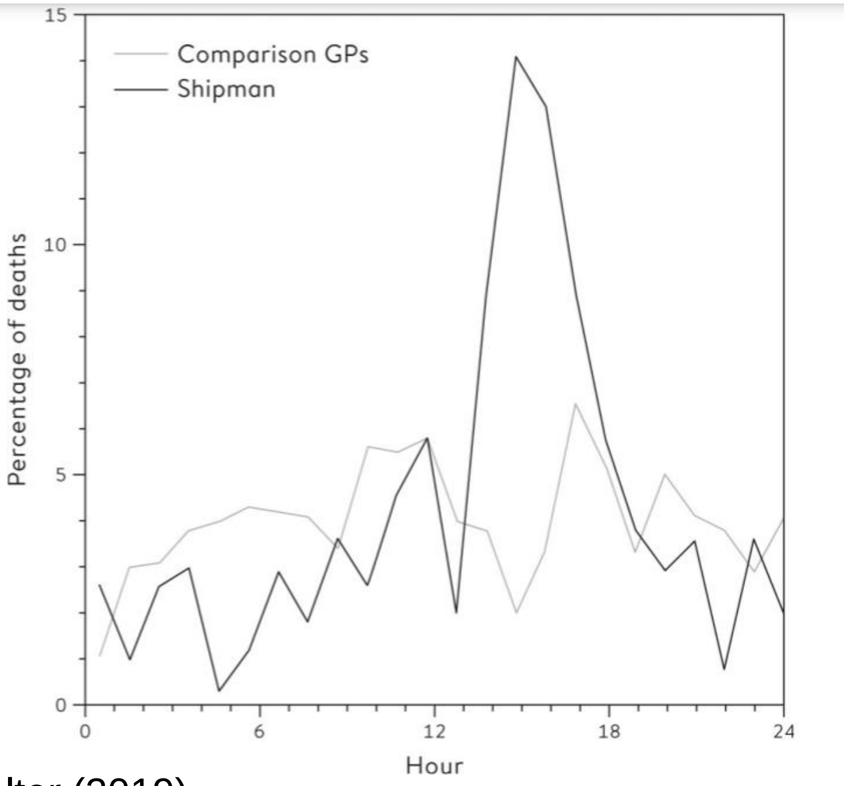
WARUM STATISTIK?

Anhand von Statistik konnte ein Doktor (Shipman) überführt werden

> Er hat mindestens 215 seiner Patienten mit einer Opiat-Überdosis ermordet

> Meiste Tote am frühen Nachmittag

> Nach seinen Hausbesuchen…



Quelle:
Spiegelhalter (2019)

WARUM STATISTIK?

HOCHSCHULE HEILBRONN

Basierend auf einer wahren Geschichte (anderer Zeit, Ort und Partei)

(siehe https://www.rwi-essen.de/unstatistik/)

- Texas: 38% der Population fahren einen SUV
- 57% sind Republikaner
- Eine Umfrage unter SUV Fahrern ergibt, dass 78% für die Republikaner gestimmt haben

Zeitungsartikel: „78% der Republikaner sind nicht umweltbewusst"

Was ist falsch?

WARUM STATISTIK?

HOCHSCHULE HEILBRONN

- Fakt:

Fakultät IB: Ca. 7 der 10 besten Studierenden in Statistik waren weiblich

Sind Frauen besser in Statistik als Männer?

VORLESUNG: DOKUMENTE

Alle Dokumente erhalten Sie über den Download-Link:

VORLESUNG: ZUSATZMATERIAL

- *BasicMath.pdf* enthält mathematische Grundlagen
- *StatisticalTables.pdf* enthält statistische Tabellen (z.B. Standardnormalverteilung)
- *AppendixStatistics.pdf* enthält wichtige statistische Formeln

VORLESUNG: VORKENNTNISSE MATHEMATIK

HOCHSCHULE HEILBRONN

- Erforderlichen mathematischen Kenntnisse: Gymnasialniveau Mathematik
- Es wird vorausgesetzt, dass Sie diese Konzepte kennen und verstehen
- Ansonsten: <u>Eigenständige</u> Nacharbeit der Konzepte
- Hilfreiches Dokument *Basic Math.pdf*!


HOCHSCHULE HEILBRONN

MATHEMATICS FOR STATISTICS

©Florian Kauffeldt

VORLESUNG: QUELLEN

HOCHSCHULE HEILBRONN

AUER, B. & ROTTMANN, H. (2015). Statistik und Ökonometrie für Wirtschaftswissenschaftler. Leipzig: Springer

HÄRDLE, W.K., KLINKE, S., & RÖNZ, B. (2015). Introduction to Statistics. Heidelberg: Springer

KOSFELD, R., ECKEY, H. & TÜRCK, M. (2019). Wahrscheinlichkeitsrechnung und induktive Statistik. Wiesbaden: Springer

MATHAI, A.M. & HAUBOLD, H.J. (2018). Probability and Statistics. De Gruyter: Berlin

VORLESUNG: INHALT

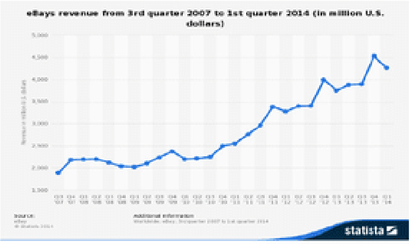
HOCHSCHULE HEILBRONN

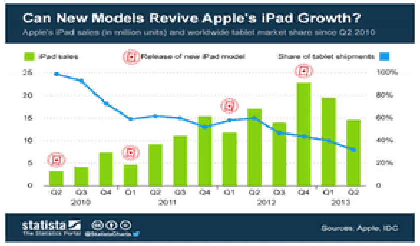
1. Deskriptive Statistik
2. Wahrscheinlichkeitstheorie
3. Inferenzstatistik / Schließende Statistik

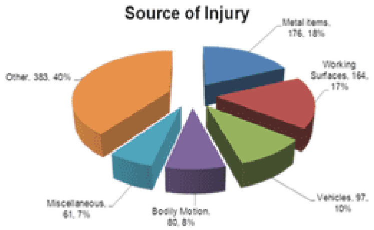
HOCHSCHULE HEILBRONN

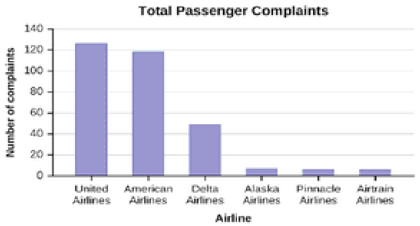
THE END!

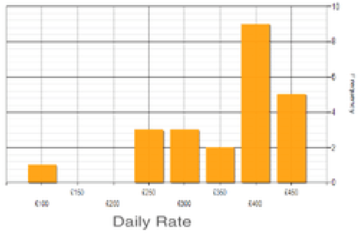
Please refer any questions to:
Prof. Dr. Florian Kauffeldt
Faculty of International Business
florian.kauffeldt@hs-heilbronn.de

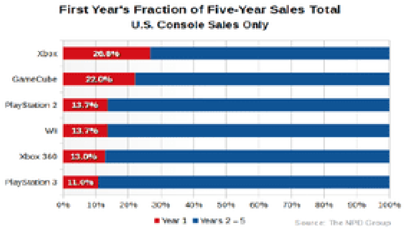












HOCHSCHULE HEILBRONN

› DESKRIPTIVE STATISTIK

Statistik | Prof. Dr. Florian Kauffeldt

TEILGEBIETE DER STATISTIK

Statistik

- **Deskriptive Statistik**
  > Beschreibung von Daten
- **Inferenzstatistik**
  > Induktive Rückschlüsse basierend auf Daten

WAS IST DESKRIPTIVE STATISTIK?

HOCHSCHULE HEILBRONN

- Ziele Deskriptive Statistik: Beschreibung von Daten (keine induktiven Schlüsse)
- Deskriptive Statistik: Methoden, um Daten zusammenzufassen und zu visualisieren

DESKRIPTIVE STATISTIK
INHALT

1. Daten und Messniveaus

2. Statistiken und Parameter

3. Univariate Statistiken
   > Lagemaße (Modus, Median, Mittelwert, Quantile)
   > Streuungsmaße (Standardabweichung, Varianz)

4. Multivariate Statistiken
   > Kovarianz
   > Korrelationskoeffizient (Pearson, Spearman Rho)

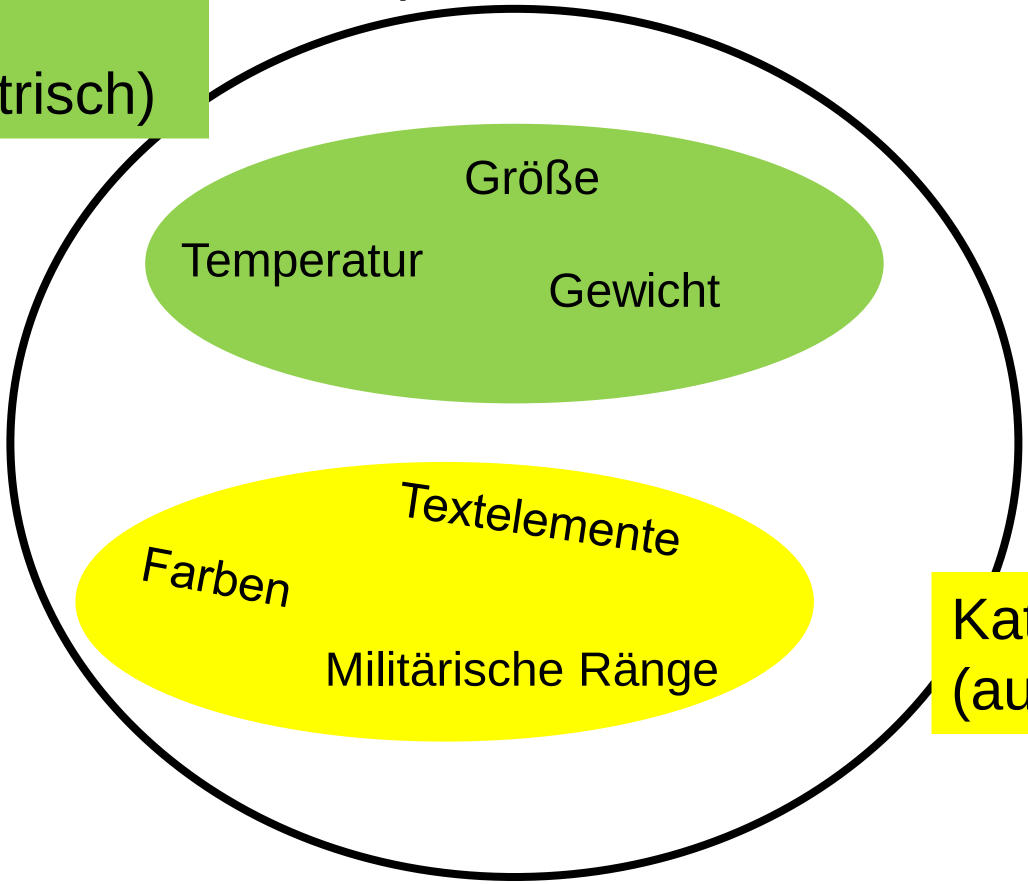
› DATEN

WAS SIND DATEN?

Daten sind Merkmale von Objekten, die durch Beobachtung erhoben werden

Beispiele Daten

Numerische Daten
(auch: quantitativ oder metrisch)



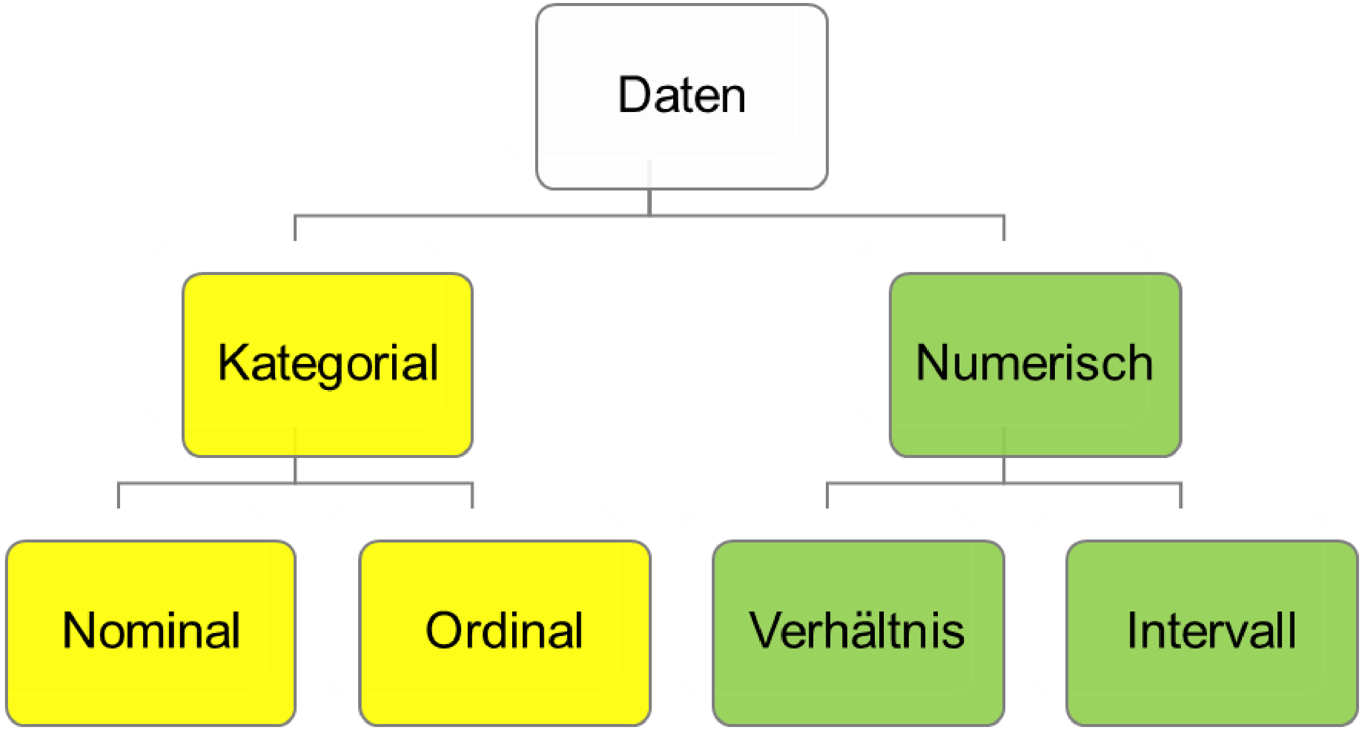
Kategoriale Daten
(auch: qualitative Daten)

DATEN: MESSNIVEAUS

HOCHSCHULE HEILBRONN

Art	Skala	Messung von	Unter-scheiden	Sor-tieren	Ab-stände	Verhält-nisse	Beispiel
Kategorial	Nominal	Kategorien ohne Rangfolge	✓	X	X	X	Farben
	Ordinal	Kategorien mit Rangfolge	✓	✓	X	X	Militär-ränge
Numerisch	Intervall	Zahlen ohne natürlichen Nullpunkt	✓	✓	✓	X	Temp. °C
	Verhältnis	Zahlen mit natürlichen Nullpunkt	✓	✓	✓	✓	Temp. Kelvin

DATEN: MESSNIVEAUS



DATEN: MESSNIVEAUS

HOCHSCHULE HEILBRONN

Beispiel. Wettrennen.

Erfassung auf Skala

Teilnehmer	Nominal	Ordinal	Verhältnis
A	> Teilgenommen	> 2. Platz	> 16.2 Sek.
B	> Teilgenommen	> 1. Platz	> 11.3 Sek.
C	> Disqualifiziert	> -	> -
D	> Teilgenommen	> 3. Platz	> 16.9 Sek.

DATEN: MESSNIVEAUS

HOCHSCHULE HEILBRONN

Daten	Art	Skala
Körpergröße (182cm, 167cm, …)		
Geschlecht		
IQ (100, 99, 116, 89,…)		
Prozentsatz korrekter Testantworten		
Platzierung in einem Schönheitswettbewerb		
Reiseziele (New York, Berlin)		
Zustimmungsgrade (stimmte nicht zu, …, stimme voll zu)		
Jahreszahlen (1640, 1920, 2020,…)		

DATENARTEN: DISKRET UND STETIG

Numerische Daten

Diskret	Stetig
Bestimmte Zwischenwerte	**Beliebige** Zwischenwerte
Beispiel: Anzahl Studenten zw. 63 und 65 64	Beispiel: Größe zwischen 1,70 und 1,80 1,7000001; 1,751111; 1,78787878; ….

DATENARTEN:
UNIVARIAT UND MULTIVARIAT

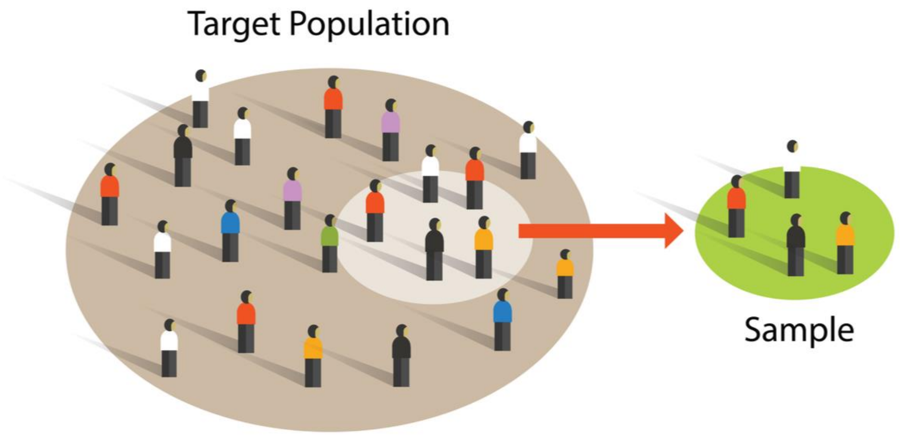
HOCHSCHULE HEILBRONN

Daten

Univariat	Multivariat
Messen ein **einziges** Merkmal Beispiel: Größe	Messen **mehrere** Merkmale Beispiel: Größe + Gewicht

DATENERHEBUNG: ZUFALLSSTICHPROBEN

Erhebung von Daten über **Zufallsstichproben:**



- Stichprobe = Teilmenge einer Population (= Menge der Merkmalsträger)

- Zufall = Jedes Element der Population wird mit gleicher Wahrscheinlichkeit gezogen

› STATISTIKEN UND PARAMETER

STATISTIKEN UND PARAMETER

HOCHSCHULE HEILBRONN

!Wichtige Unterscheidung!

Statistik	Parameter
- Kennzahl einer Stichprobe - Bekannt	- Kennzahl einer Population - I.d.R. unbekannt - Schätzung basierend auf Stichprobe

STATISTIKEN UND PARAMETER: NOTATIONELLE UNTERSCHEIDUNG

- **Lateinische** Symbole → **Statistiken**
- **Griechische** Symbole → **Parameter**

Beispiel. Mittelwert

- $\bar{\boldsymbol{x}}$ = Stichprobenmittelwert
- $\boldsymbol{\mu}$ = Populationsmittelwert

STATISTIKEN UND PARAMETER: ARTEN (AUSBLICK)

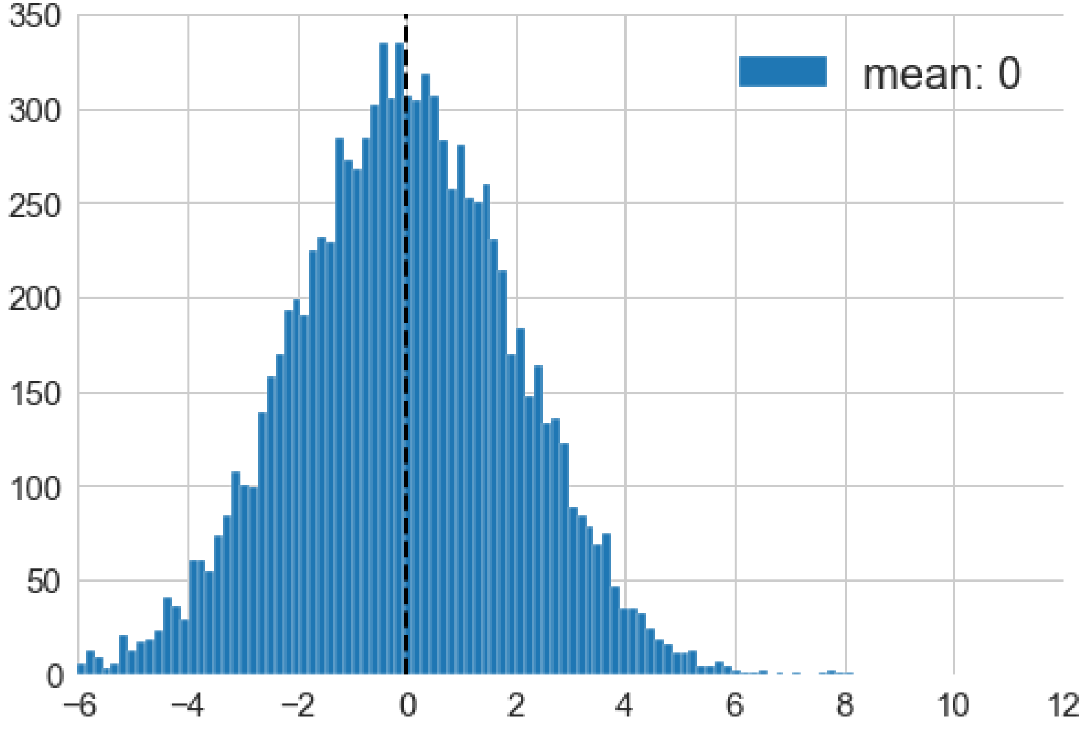
HOCHSCHULE HEILBRONN

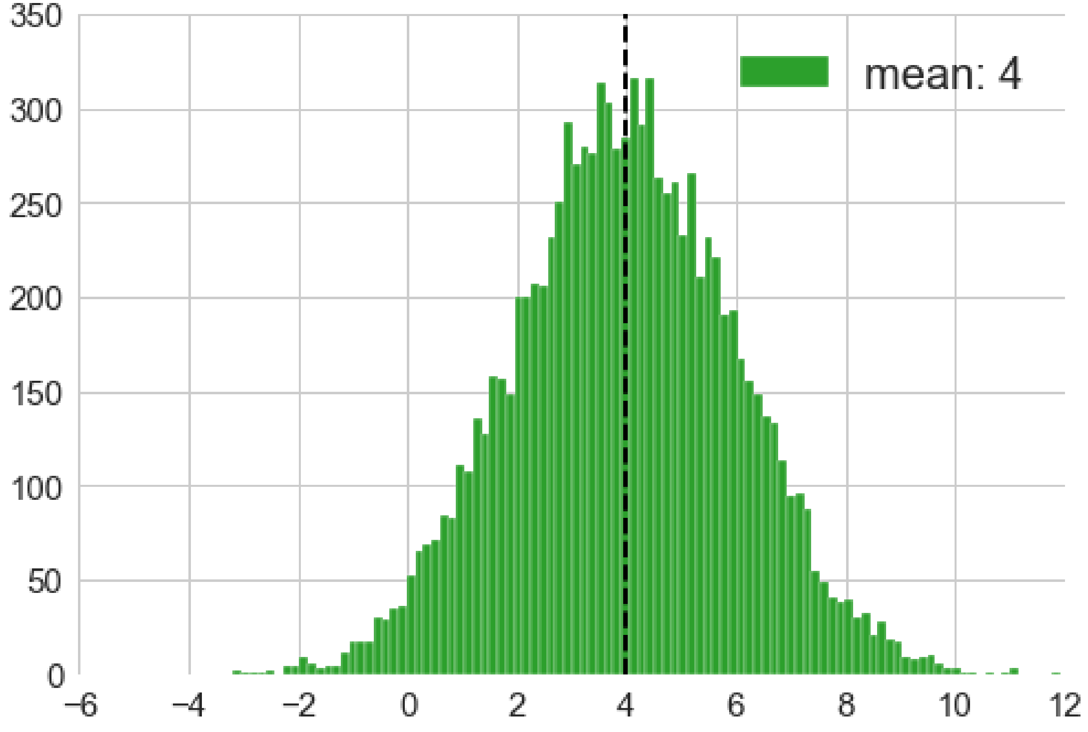
Datenart	**Messung von**	**Statistik / Parameter**
Univariat	Lage	Mittelwert, Modus, Median, Quantile
	Streuung	Varianz, Standardabweichung
Multivariat	Zusammenhang	Kovarianz, Korrelationskoeffizient (Pearson, Spearman Rho)

› UNIVARIATE STATISTIKEN / PARAMETER

Lagemaße = Typische Werte einer Verteilung

LAGEMAßE CHARAKTERISIEREN
DIE LAGE DER VERTEILUNG





LAGEMAßE: MODUS, MEDIAN, MITTELWERT

Gegeben ein Datensatz $(x_1, \dots, x_n)$ der Länge n:

- $Modus$ (auch Modalwert) = Wert, der am häufigsten vorkommt
- $Median$ = Wert in der Mitte, wenn die Werte von klein nach groß sortiert wurden
- $Mittelwert$ = Arithmetisches Mittel:

$$Mittelwert = \frac{x_1 + \cdots + x_n}{n}$$

LAGEMAßE: QUANTILE

$i - Quantil =$

> Wert unter dem $i\%$ der Werte einer Stichprobe liegen (und $(1 - i)\%$ darüber),

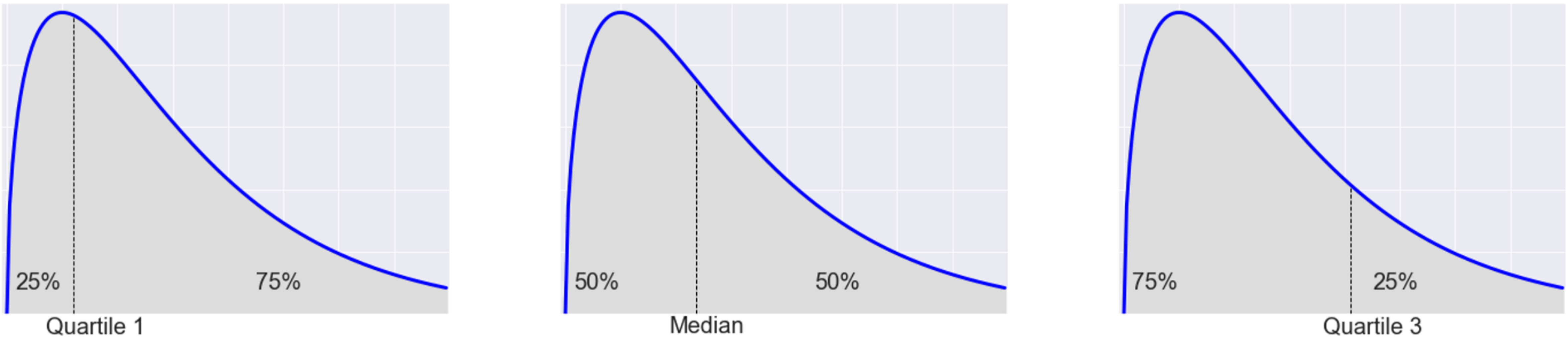
wenn die Werte von klein nach groß sortiert wurden

Beispiele.

- 50-Quantil → 50% darunter, 50% darüber → Median
- 25-Quantil → 25% darunter, 75% darüber → 1. Quartil

LAGEMAßE: QUARTILE (25-, 50- UND 75-QUANTIL)

HOCHSCHULE HEILBRONN



LAGEMAẞE: ERFORDERLICHE MESSNIVEAUS

HOCHSCHULE HEILBRONN

> Der häufigste Wert (Modalwert) kann immer berechnet werden

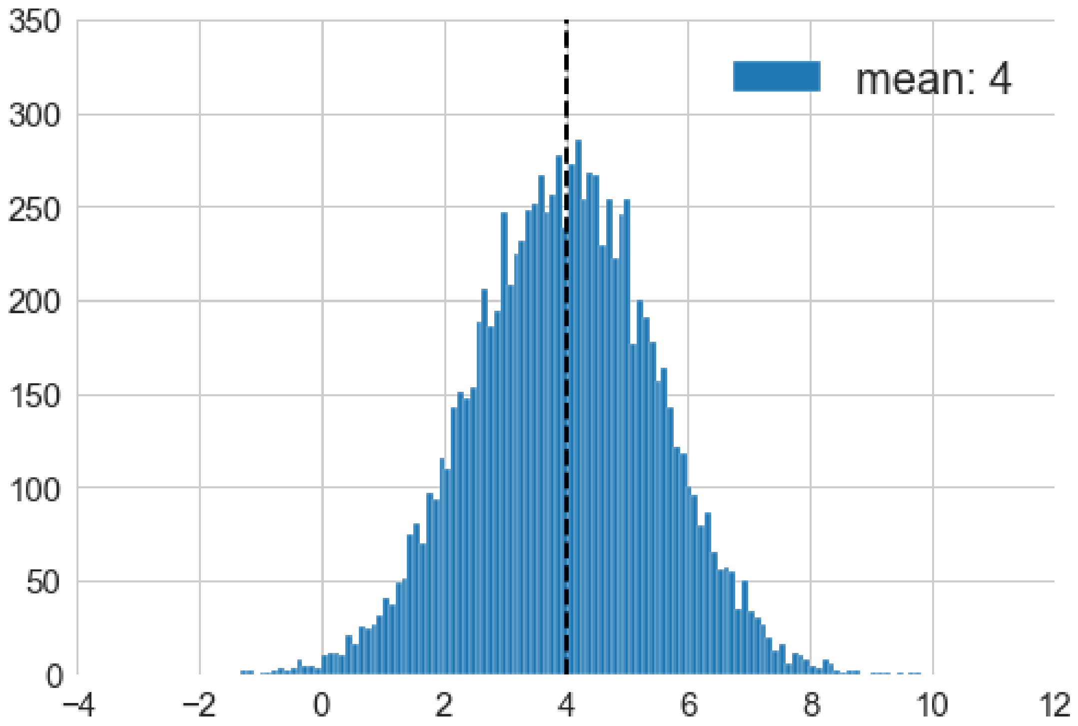
> Quantile setzen eine Ordnung des Datensatzes voraus → Mindestens ordinale Daten

> Der Mittelwert kann nur für numerische Daten berechnet werden

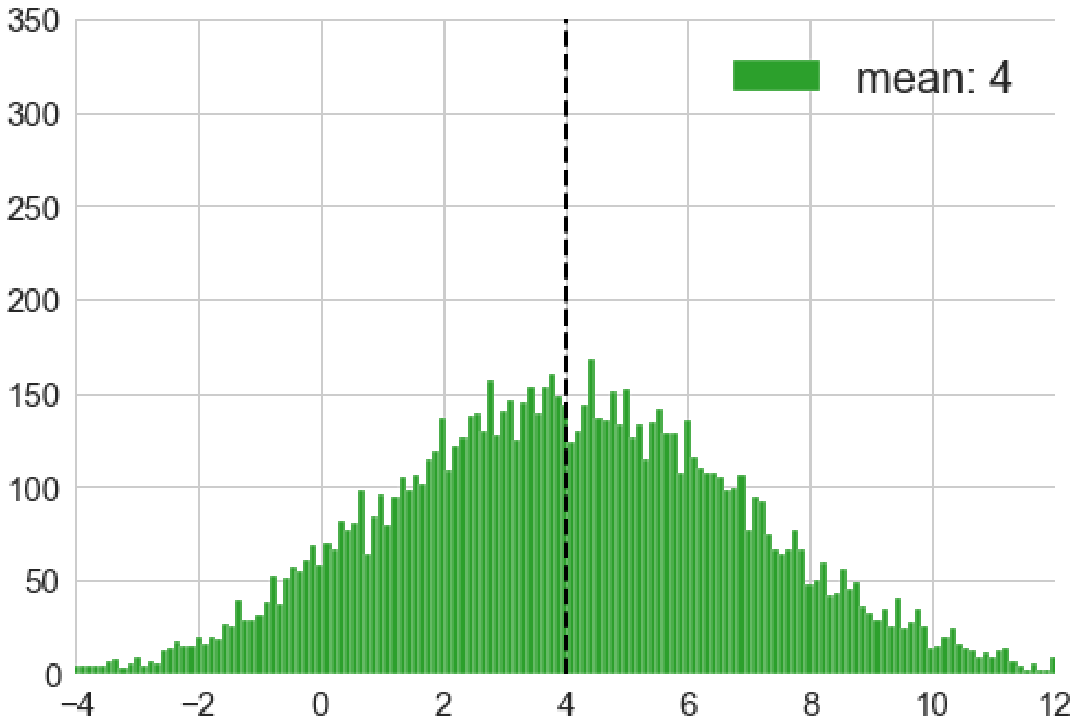
WARUM STREUUNGSMAẞE?

HOCHSCHULE HEILBRONN

Niedrige Streuung um den Mittelwert



Hohe Streuung um den Mittelwert



Beide Verteilungen haben den gleichen Mittelwert (4), aber streuen unterschiedlich

STREUUNGSMAßE

Streuungsmaße = Schwankungsbreite Verteilung

STREUUNGSMAẞE:
MESSEN STREUUNG UM DEN MITTELWERT

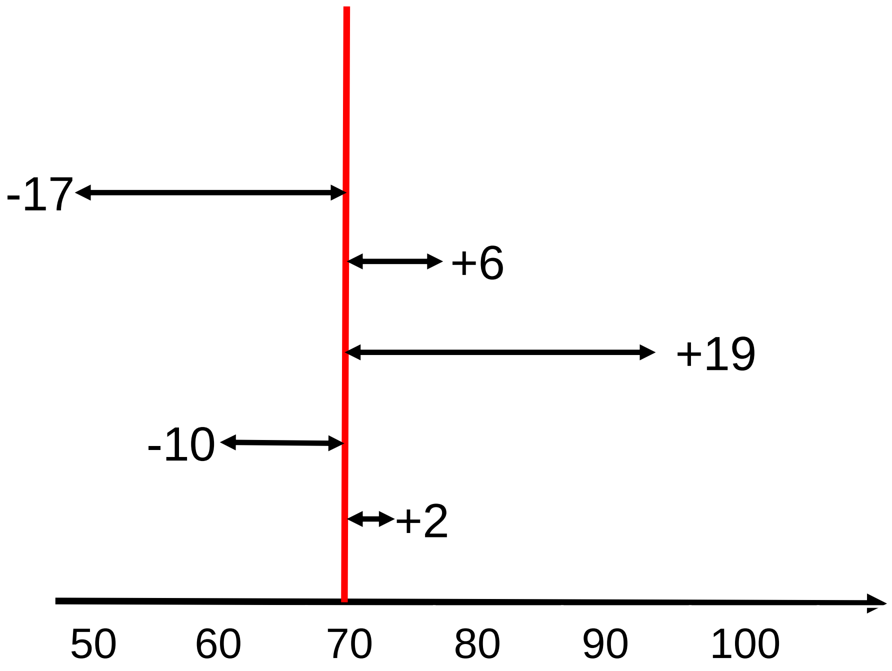
HOCHSCHULE HEILBRONN

Maß für die Streuung der Daten um ihren Mittelwert

> Durchschnitt der Summe der Abweichungen vom Mittelwert?

Beispiel.

	$Gewicht_i$	$Gewicht_i - Mittelwert$
	55	-17
	78	6
	91	19
	62	-10
	74	2
Summe / n	$Mittelwert = 72$	$0/5 = 0$



$\boldsymbol{Mittelwert = 72}$

STREUUNGSMAßE: MESSEN STREUUNG UM DEN MITTELWERT

HOCHSCHULE HEILBRONN

Varianz = durchschnittliche **quadratische** Abweichung vom Mittelwert

Beispiel.

	Beobachtung: $Gewicht_i$	Abweichungen: $Gewicht_i - Mittelwert$	Quadrierte Abweichungen: $(Gewicht_i - Mittelwert)^2$
	55	-17	289
	78	6	36
	91	19	361
	62	-10	100
	74	2	4
Summe / n	$Mw. = 72$	0	$\frac{(Gewicht_1 - Mittelw.)^2 + \cdots + (Gewicht_5 - Mittelw.)^2}{n} = 158$

Varianz

STREUUNGSMAßE: INTERPRETATION

Varianz = 158 Kg^2

> Was bedeutet das?

Standardabweichung = $\sqrt{158}\ Kg$ = 12,6 Kg

STREUUNGSMAßE: FORMELN

Sei $(x_1, \ldots, x_n)$ ein Datensatz der Länge n:

- $Varianz$ = durchschnittliche quadratische Abweichung vom Mittelwert:

$$Varianz = \frac{(x_1 - Mittelwert)^2 + \cdots + (x_n - Mittelwert)^2}{n}.$$

- $Standardabweichung$ = positive Quadratwurzel der Varianz:

$$Standardabweichung = \sqrt{Varianz}$$

STREUUNGSMAẞE:
ERFORDERLICHE MESSNIVEAUS

HOCHSCHULE HEILBRONN

Für die Berechnung von Varianz und Standardabweichung braucht man den Mittelwert

→ können nur für numerische Daten berechnet werden

Übung zu Streuungsmaßen (Excel)

› MULTIVARIATE STATISTIKEN / PARAMETER

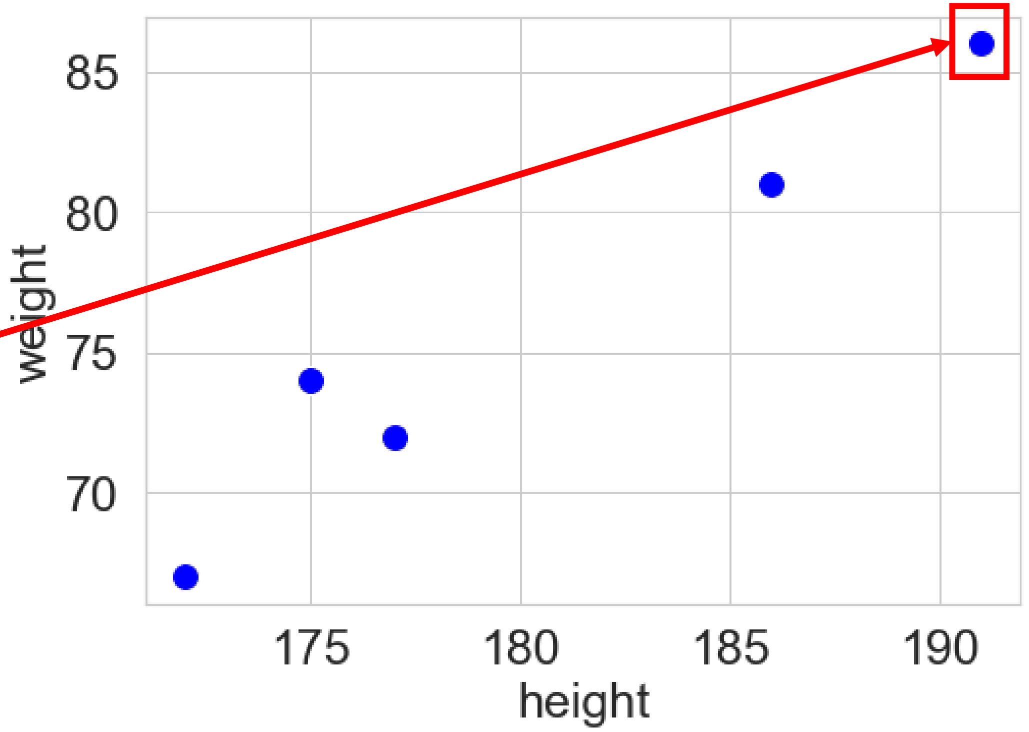
ZUSAMMENGHANGSMAßE

HOCHSCHULE HEILBRONN

Häufig möchte man messen, ob zwei Merkmale eine Beziehung haben

Beispiel. Zusammenhang zwischen Größe und Gewicht.

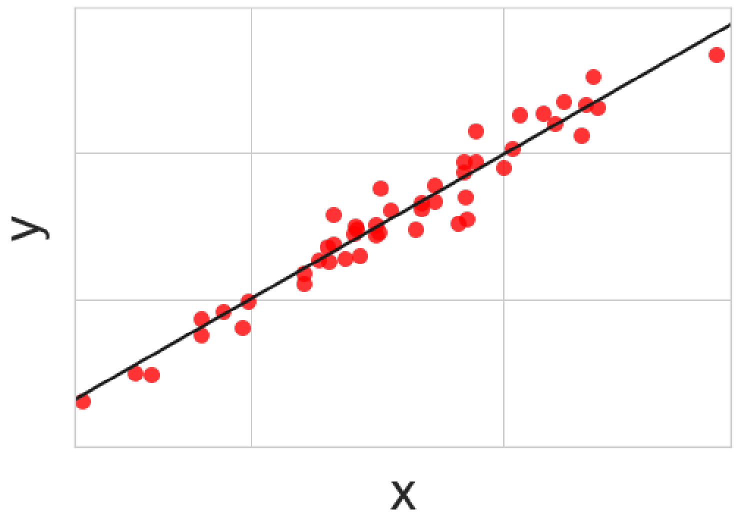
Person i	1	2	3	4	5
Height in cm (x_i)	172	175	177	186	191
Weight in kg (y_i)	67	74	72	81	86



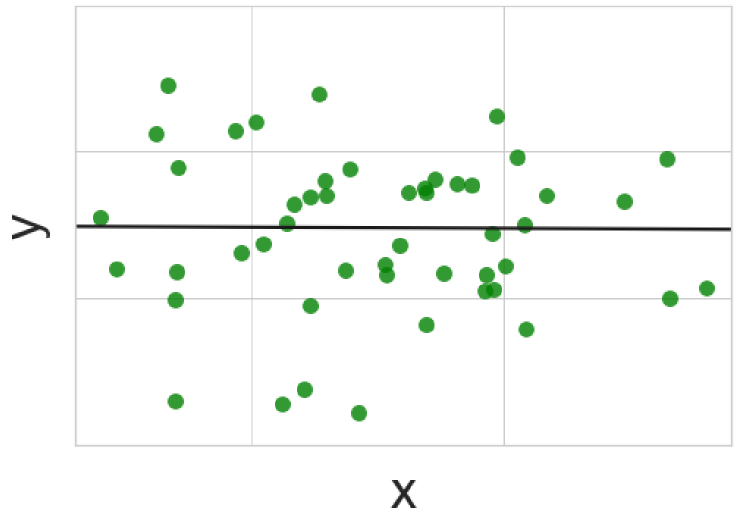
ZUSAMMENHANGSMAß: KOVARIANZ

Kovarianz: Misst Stärke der linearen Beziehung zwischen 2 Merkmalen X und Y

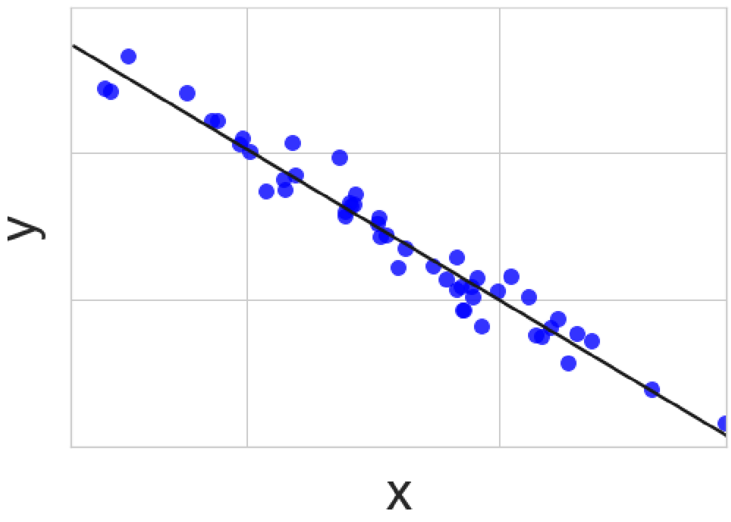
Positive Kovarianz:
$X \uparrow \rightarrow\ Y \uparrow$



Kovarianz nahe 0:
Keine lin. Beziehung



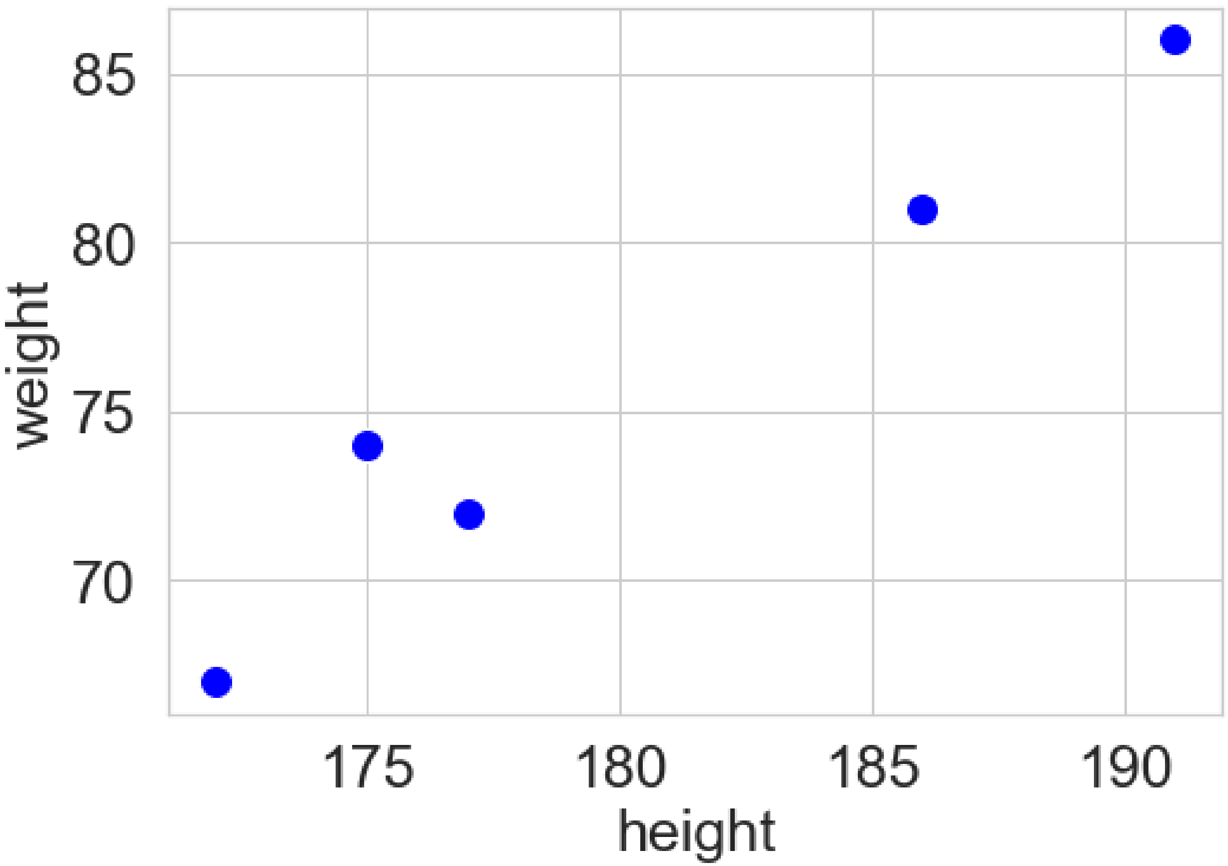
Negative Kovarianz:
$X \uparrow \rightarrow\ Y \downarrow$



ZUSAMMENHANGSMAß: KOVARIANZ

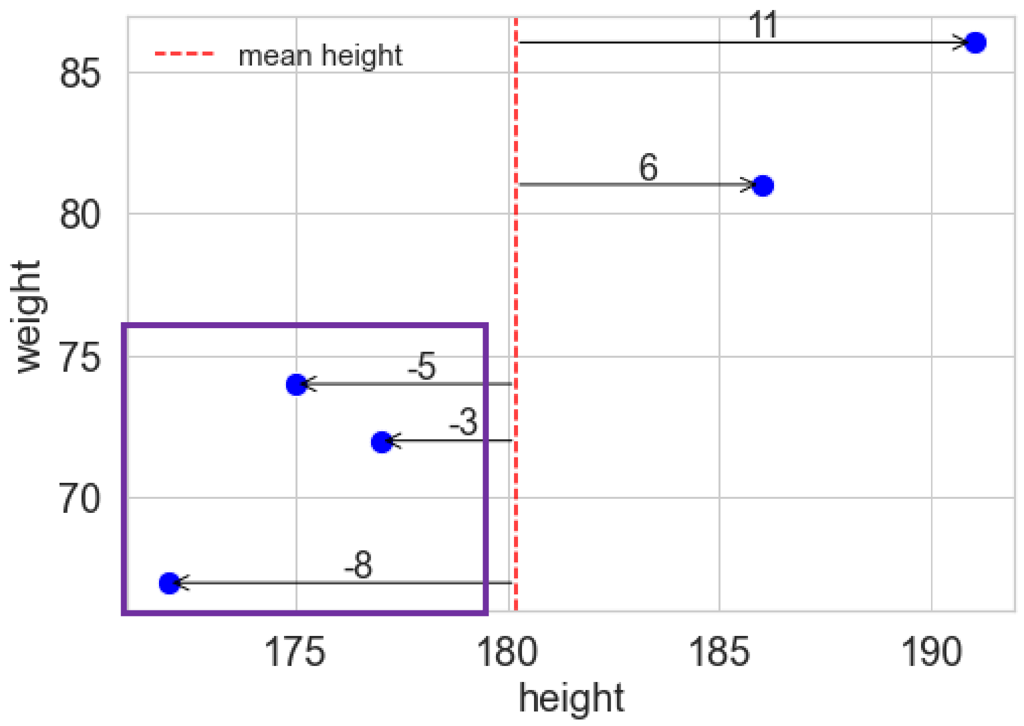
HOCHSCHULE HEILBRONN

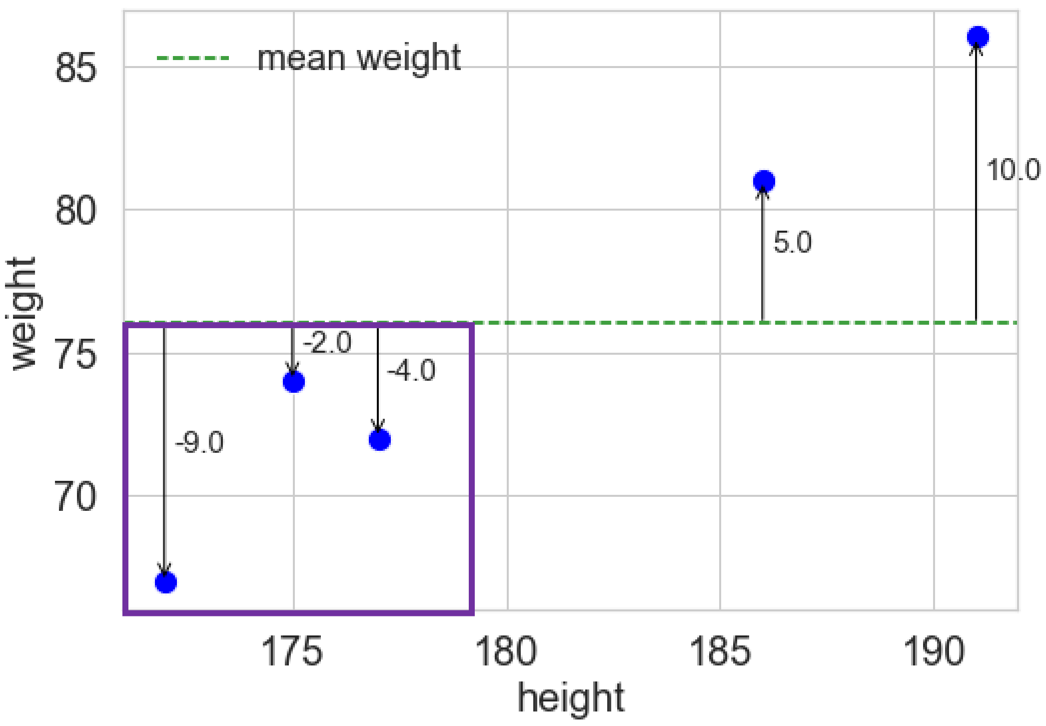
Beispiel (Fortsetzung). Beziehung zw. Größe und Gewicht.



ZUSAMMENHANGSMAß: KOVARIANZ

HOCHSCHULE HEILBRONN

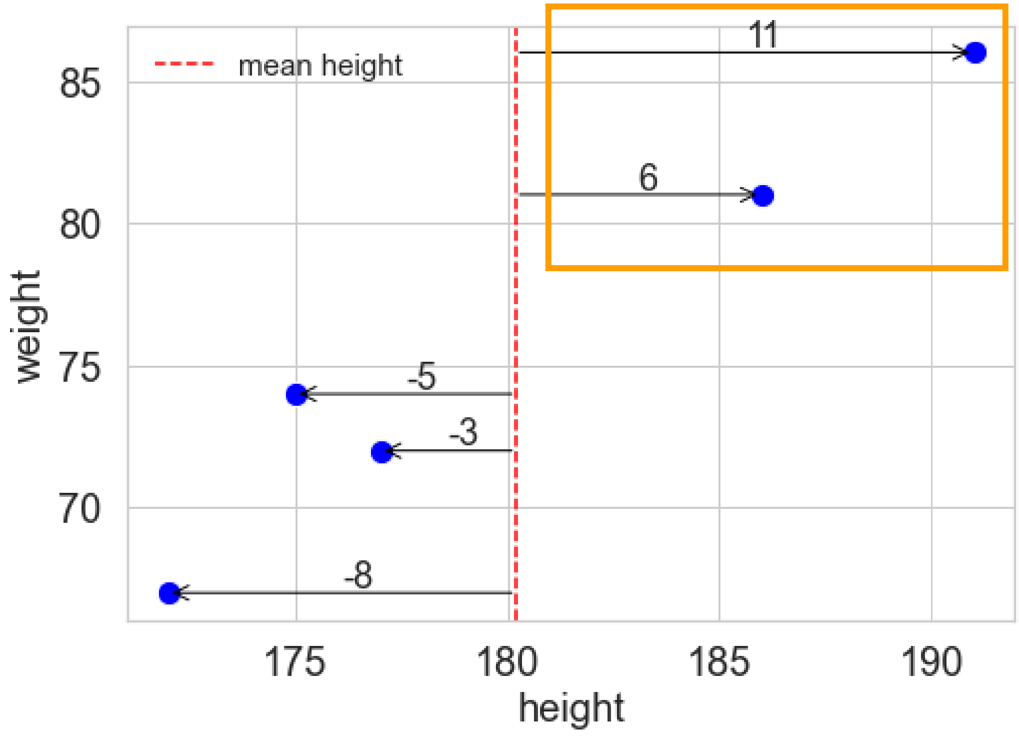


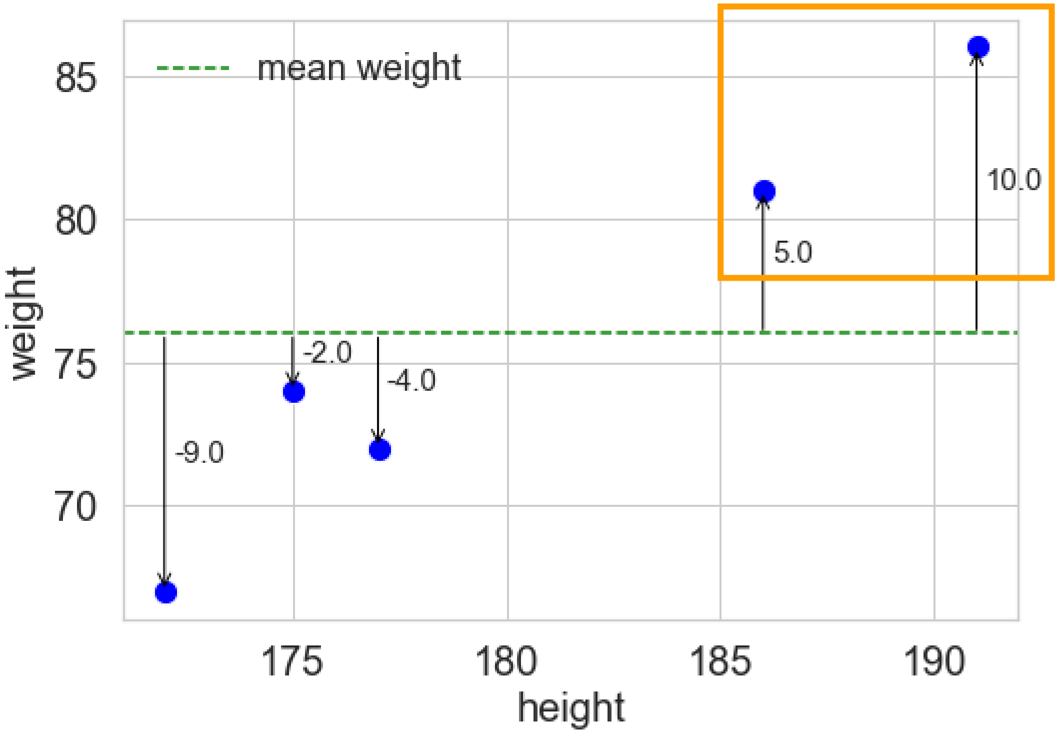


Gewicht und Größer der **ersten 3 Personen** liegen **unter dem Durchschnitt**

ZUSAMMENHANGSMAß: KOVARIANZ

HOCHSCHULE HEILBRONN



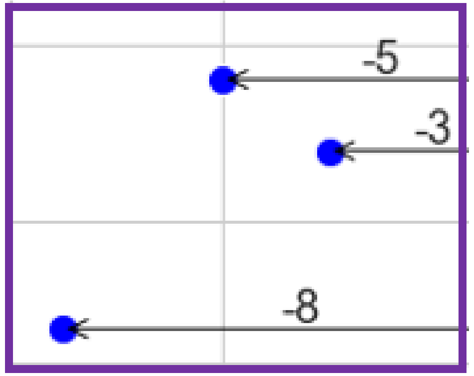


Gewicht und Größer der letzten **beiden Personen** liegen **über dem Durchschnitt**

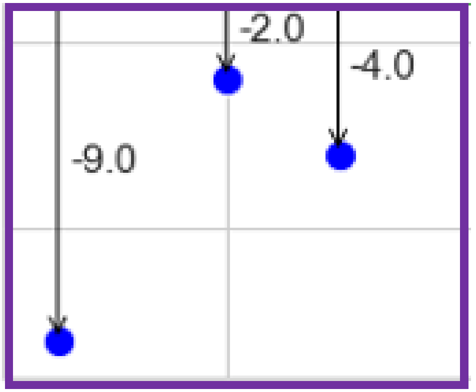
ZUSAMMENHANGSMAß: KOVARIANZ

HOCHSCHULE HEILBRONN

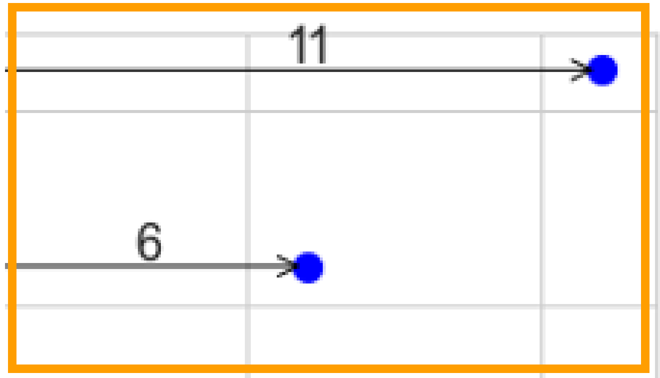
Berechnung.



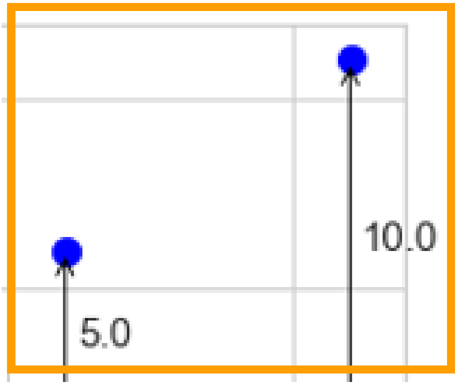
×



=



×



=

	Produkt Abweichungen: $(height_i - MW_h) \cdot (weight_i - MW_w)$
	72
	10
	12
	30
	110
Summe/ n	$cov(height, weight) = 234/5 = 46.8$

ZUSAMMENHANGSMAß: KOVARIANZ

HOCHSCHULE HEILBRONN

Sei $(x_1, y_1), \ldots, (x_n, y_n)$ ein Datensatz mit zwei Merkmalen der Länge n:

> $Kovarianz$ der Merkmale = Durchschnitt Produkt Abweichungen vom Mittelwert:

$$cov(x,y) = \frac{(x_1 - MW_x) \cdot (y_1 - MW_y) + \cdots + (x_n - MW_x) \cdot (y_n - MW_y)}{n}.$$

ZUSAMMENHANGSMAß: KOVARIANZ

HOCHSCHULE HEILBRONN

Beispiel (Fortsetzung). Ausführliche Berechnung.

> Was passiert mit der Kovarianz, wenn man Größe (height) in m statt in cm misst?

Beobachtung: height (in cm)	Abweichungen: $height_i - MW_{height}$	Beobachtung: weight (in kg)	Abweichungen: $weight_i - MW_{weight}$	Produkt Abweichungen: $(height_i - MW_{height}) \cdot (weight_i - MW_{weight})$
172	-8	67	-9	72
175	-5	74	-2	10
177	-3	72	-4	12
186	6	81	5	30
191	11	86	10	110
$MW_{height} = 180$		$MW_{weight} = 76$		$cov(height, weight) = 46.8$

ZUSAMMENHANGSMAß: PEARSON KORRELATIONSKOEFFIZIENT

HOCHSCHULE HEILBRONN

> Wir möchten ein normiertes Maß, das nicht von der Einheit abhängt

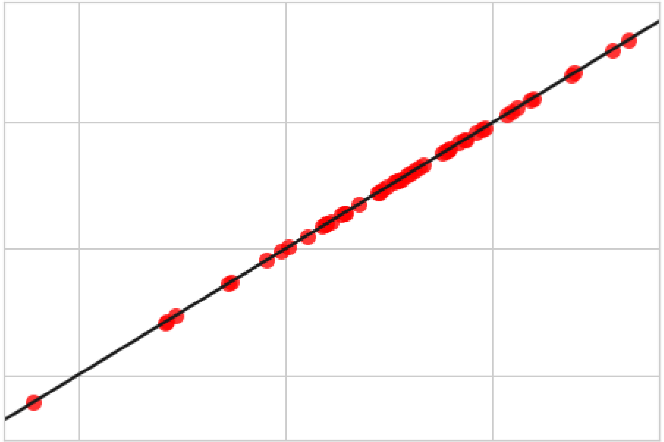
Sei $(x_1, y_1), \dots, (x_n, y_n)$ ein Datensatz mit zwei Merkmalen der Länge n:

> $Pearson\ Korrelationskoeffizient\ r$ = Kovarianz durch Produkt Standardabweichungen:

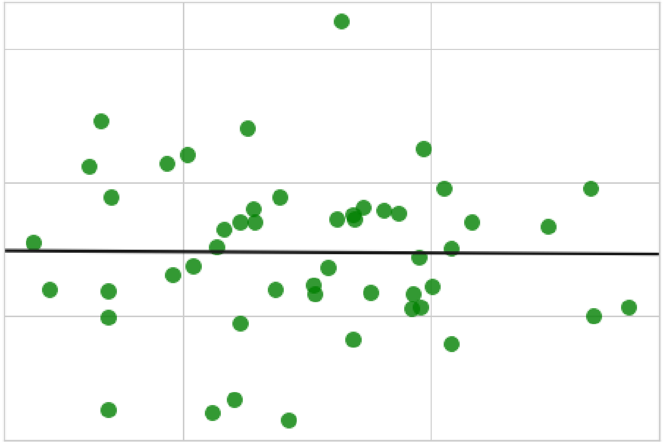
$$r = \frac{cov(x, y)}{sd(x) \times sd(y)}.$$

ZUSAMMENHANGSMAß:
PEARSON KORRELATIONSKOEFFIZIENT

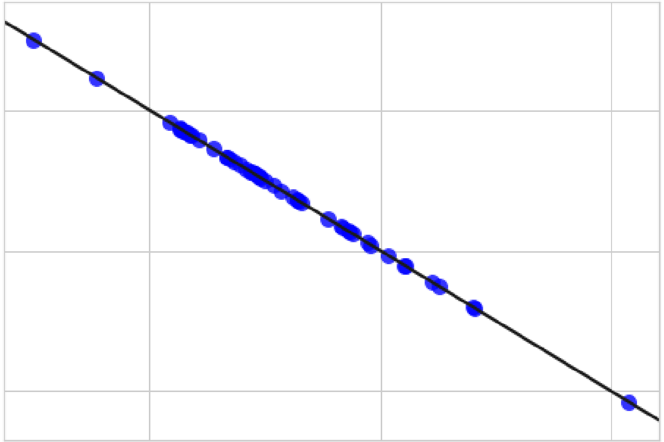
$r_{x,y} = 1$
perfekt positiv linear



$r_{x,y} = 0$
nicht linear



$r_{x,y} = -1$
perfekt negativ linear

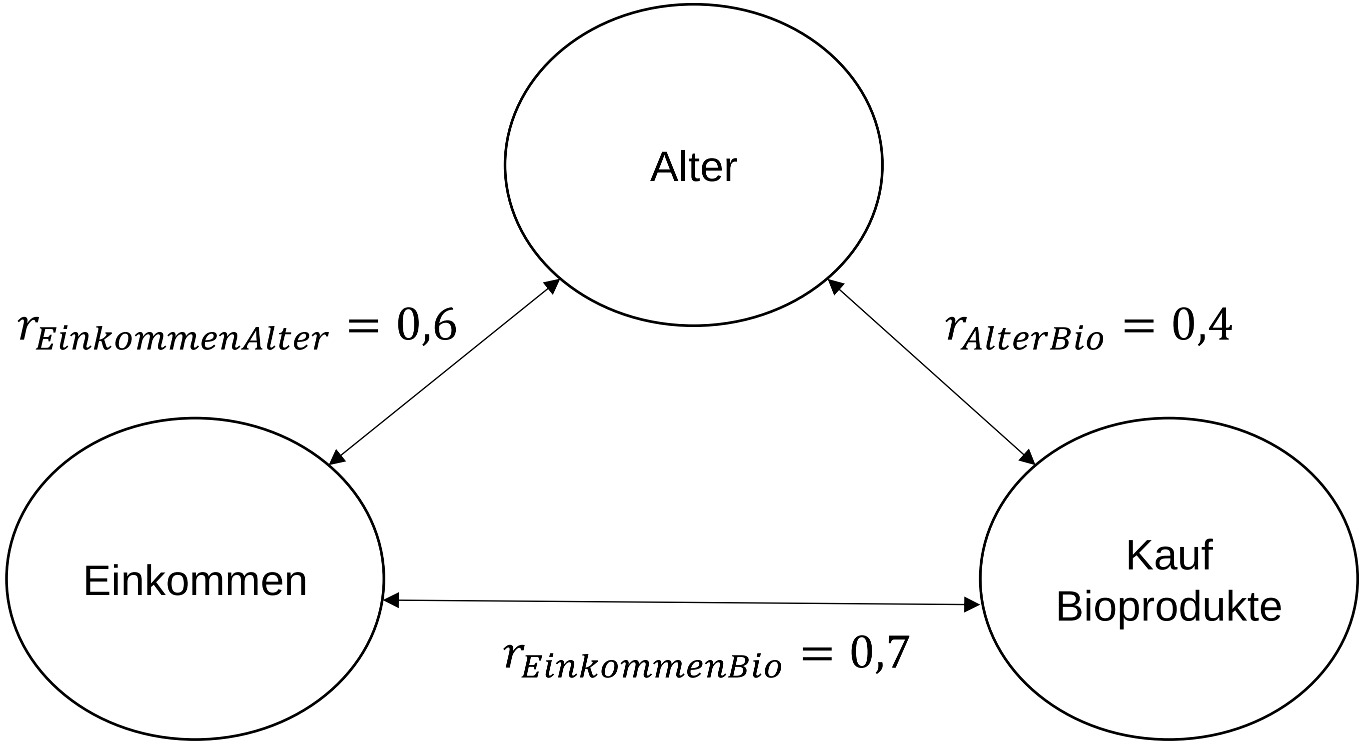


KOVARIANZ UND PEARSON KORRELATION: EIGENSCHAFTEN

HOCHSCHULE HEILBRONN

- Kovarianz nimmt Werte zwischen $-\infty$ und ∞ an
- Korrelationskoeffizient nimmt Werte zwischen -1 und +1 an
- Kovarianz hängt von der Einheit der Messungen ab, Korrelationskoeffizient nicht
- Je negativer (positiver), desto stärker ist der negative (positive) lineare Zusammenhang

ZUSAMMENHANGSMAß: PEARSON PARTIELLE KORRELATION



ZUSAMMENHANGSMAß: PEARSON PARTIELLE KORRELATION

$Partielle\ Korrelation$ = Herausrechnen des Effekts von weiteren Variablen:

> Korrelation von x und y während für z kontrolliert wird:

$$r_{xy\,|\,z} = \frac{r_{xy} - r_{xz} \times r_{yz}}{\sqrt{1 - r_{xz}^2} \times \sqrt{1 - r_{yz}^2}}$$

KOVARIANZ UND PEARSON KORRELATION: ERFORDERLICHE MESSNIVEAUS

Für die Berechnung von Kovarianz und Pearson Korrelationen braucht man den Mittelwert

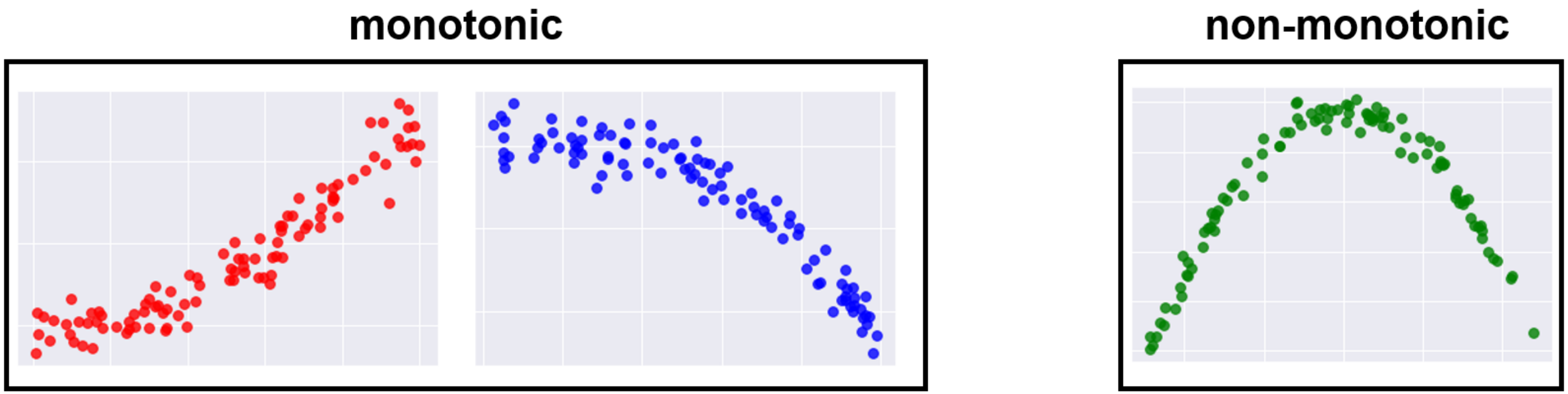
→ können nur für numerische Daten berechnet werden

ZUSAMMENHANGSMAß: SPEARMAN RANGKORRELATION

$Spearman\ Rho$ = Korrelationskoeffizient für ordinale kategoriale Daten

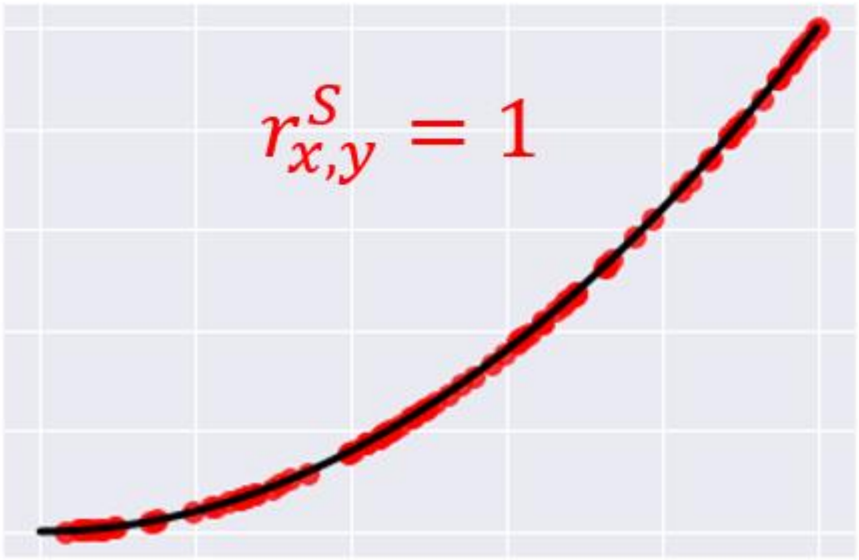
- basiert auf Rängen

- misst nicht nur lineare, sondern montone Zusammenhänge:



ZUSAMMENHANGSMAß: SPEARMAN RANGKORRELATION

Perfekter positiver monotoner Zusammenhang:



ZUSAMMENHANGSMAß: SPEARMAN RANGKORRELATION

HOCHSCHULE HEILBRONN

Berechnung Spearman Rho :

> wenn es keine verbundenen (gleiche) Ränge gibt:

$$r_{xy}^{S} = 1 - \frac{6 \times (d_1^2 + \cdots + d_n^2)}{n(n^2 - 1)}$$

wobei $d_i = rang(x_i) - rang(y_i)$

> wenn es verbundene Ränge gibt:

$$r_{xy}^{S} = r_{rang(x)rang(y)}$$

ZUSAMMENHANGSMAß:
SPEARMAN RANGKORRELATION

HOCHSCHULE HEILBRONN

Beispiel. Bestimmen Sie den Spearman Rangkorrelationskoeffizienten:

Studium	**Noten**
Sehr zufrieden	Gut
Geht so	Sehr gut
Sehr unzufrieden	Schlecht
zufrieden	Geht so

ZUSAMMENHANGSMAß: INTERPRETATION KORRELATION

Interpretation	Correlation value
Small correlation	0.10 to 0.29
Medium correlation	0.30 to 0.49
Large correlation	0.50 to 1.0

Außerdem:

$$r^2 = Bestimmtheitsmaß$$

> Gibt an, wie viel % der Varianz durch die untersuchte Beziehung erklärt werden.

ZUSAMMENFASSUNG: STATISTIKEN UND ERFORDERLICHE MESSNIVEAUS

HOCHSCHULE HEILBRONN

Art	Statistik	Messniveau		
		Numerisch	Kategorial Ordinal	Kategorial Nominal
Lagemaß	Modalwert	✓	✓	✓
	Median	✓	✓	X
	Mittelwert	✓	X	X
Streuungsmaß	Alle	✓	X	X
Zusammenhang	Kovarianz, Pearson	✓	X	X
	Spearman Rho	✓	✓	X

THE END!

Please refer any questions to:
Prof. Dr. Florian Kauffeldt
Faculty of International Business
florian.kauffeldt@hs-heilbronn.de

› WAHRSCHEINLICHKEITS-THEORIE

Statistik | Prof. Dr. Florian Kauffeldt

WAHRSCHEINLICHKEITSTHEORIE
INHALT

HOCHSCHULE HEILBRONN

› WAHRSCHEINLICHKEIT

WAS SIND WAHRSCHEINLICHKEITEN?

HOCHSCHULE HEILBRONN

In Alltag sagen wir Sätze wie:

- Ich bin sicher, dass Bayern München morgen gewinnt
- Wahrscheinlich werde ich Statistik bestehen
- Möglicherweise wird es morgen regnen

→ Ausdruck unterschiedlicher Grade von Unsicherheit über künftige Ereignisse

- Wahrscheinlichkeiten formalisieren dieses Konzept

WAS SIND WAHRSCHEINLICHKEITEN?

> Wahrscheinlichkeit = Maß zwischen 0 und 1, das Unsicherheit repräsentiert

> Je höher die Wahrscheinlichkeit eines Ereignisses, desto eher wird es eintreten

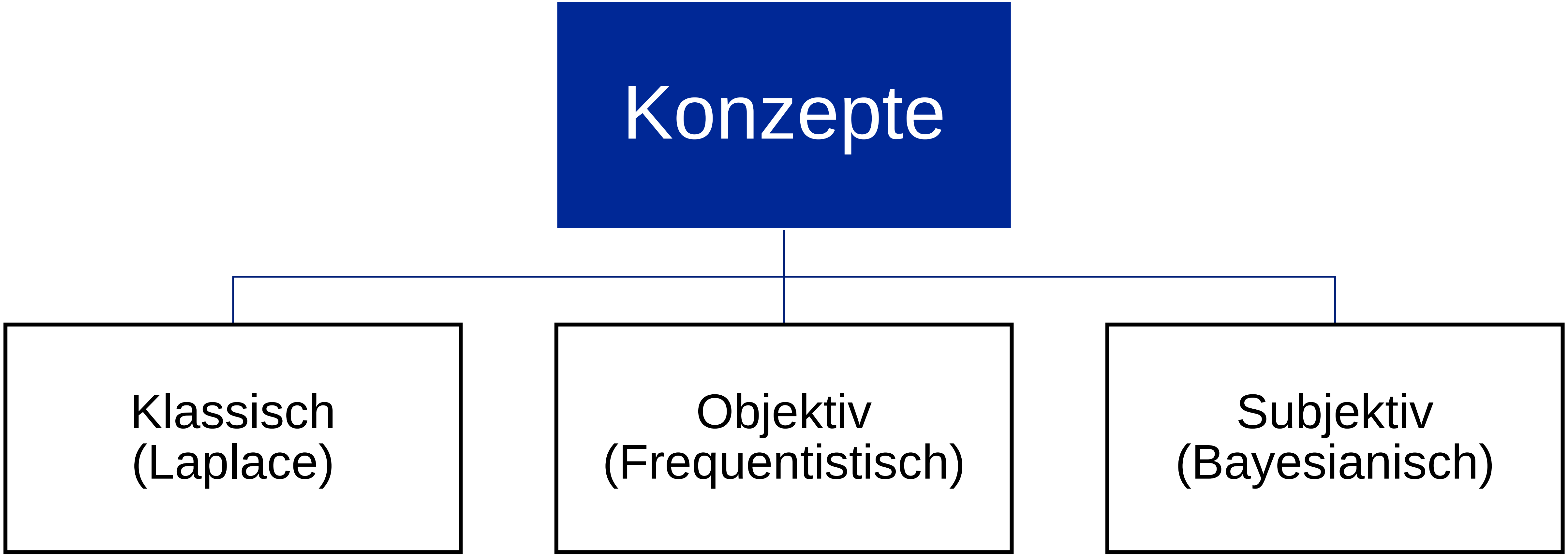
Notation. Die Wahrscheinlichkeit eines Ereignisses E bezeichnen wir mit $P(E)$

> Ein Ereignis E mit $P(E) = 0$ heißt *unmöglich*

> Ein Ereignis E mit $P(E) = 1$ heißt *sicher*

WAHRSCHEINLICHKEITS-BEGRIFFE

HOCHSCHULE HEILBRONN



WAHRSCHEINLICHKEITSBEGRIFFE: LAPLACE

HOCHSCHULE HEILBRONN

Pierre-Simon Laplace (1749 – 1827)

- Reduktion Ereignisse auf gleichwahrscheinliche symmetrische Fälle

- **Prinzip des unzureichenden Grundes**: Kein Grund für andere Wahrscheinlichkeit

- Wahrscheinlichkeit Ereignisses E:

$$P(E) = \frac{Anzahl\ Fälle\ bei\ denen\ E\ eintritt}{Gesamtanzahl\ Fälle}$$

WAHRSCHEINLICHKEITSBEGRIFFE: FREQUENTISTISCH

HOCHSCHULE HEILBRONN

Richard von Mises, Egon Pearson, John Venn,…

> Häufige Wiederholung Zufallsexperiment:

Relative Häufigkeiten → "Wahre" Wahrscheinlichkeiten

> Die Wahrscheinlichkeit eines Ereignisses E kann wie folgt geschätzt werden:

$$P(E) = \frac{Anzahl\,Versuche\;bei\;denen\;E\;eingetreten\;ist}{Gesamtanzahl\,Versuche}$$

WAHRSCHEINLICHKEITSBEGRIFFE: BAYESIANISCH (BEDINGTE WAHRSCHEINLICHKEIT)

HOCHSCHULE HEILBRONN

Thomas Bayes (1701 – 1761)

> Wahrscheinlichkeiten = subjektive Einschätzungen

> Bedingte Wahrscheinlichkeiten = Aktualisierte Wahrscheinlichkeit bei neuer Information

> Regel = Bayes-Theorem (wird später behandelt → Multivariate Verteilungen)

WAHRSCHEINLICHKEIT: OPERATION MIT EREIGNISSEN

Linda ist 31 Jahre alt, Single, offen und sehr intelligent. Sie hat Philosophie studiert. Als Studentin hat sie sich intensive mit Diskriminierung und sozialer Gerechtigkeit beschäftigt sowie an Anti-Atomkraft-Demonstrationen teilgenommen.

Was ist wahrscheinlicher?

1. Linda ist eine Bankangestellte.
2. Linda ist eine Bankangestellte und aktive Feministin.

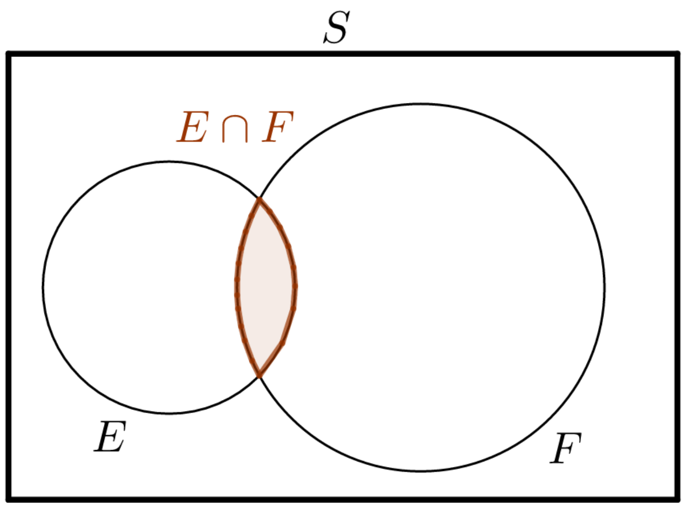
WAHRSCHEINLICHKEIT: OPERATIONEN MIT EREIGNISSEN (SCHNITT ∩)

HOCHSCHULE HEILBRONN

- S = Stichprobenraum (hier: Menge aller Frauen)

- Ereignis E = Bankangestellte, Ereignis F = Aktive Feministin

Schnitt: $\boldsymbol{E \cap F} = \text{Bankangestellte } \mathit{UND} \text{ aktive Feministin}$



Wahrscheinlichkeit Ereignis
= Fläche Ereignis / Fläche von S

- Fläche von E ist größer als die Fläche von $E \cap F$ $\rightarrow P(E) > P(E \cap F)$

WAHRSCHEINLICHKEIT: OPERATION MIT EREIGNISSEN

Linda ist 31 Jahre alt, Single, offen und sehr intelligent. Sie hat Philosophie studiert. Als Studentin hat sie sich intensive mit Diskriminierung und sozialer Gerechtigkeit beschäftigt sowie an Anti-Atomkraft-Demonstrationen teilgenommen.

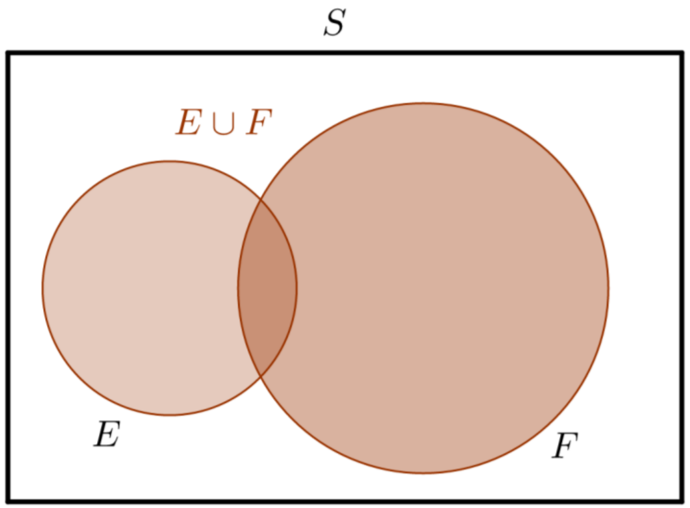
Was ist wahrscheinlicher?

1. Linda ist eine Bankangestellte.
2. Linda ist eine Bankangestellte oder aktive Feministin.

WAHRSCHEINLICHKEIT: OPERATIONEN MIT EREIGNISSEN (VEREINIGUNG ∪)

- S = Stichprobenraum (hier: Menge aller Frauen)
- Ereignis E = Bankangestellte, Ereignis F = Aktive Feministin

Vereinigung: $\boldsymbol{E} \cup \boldsymbol{F} = \text{Bankangestellte } \mathit{ODER/UND} \text{ aktive Feministin}$



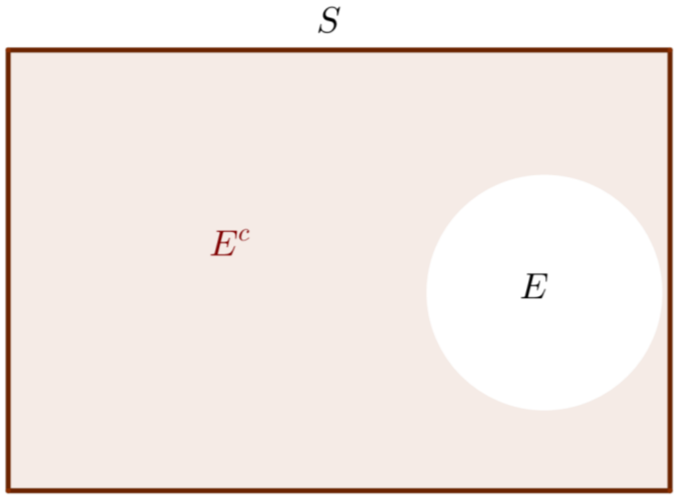
Wahrscheinlichkeit Ereignis
= Fläche Ereignis / Fläche von S

- Fläche von E ist kleiner als die Fläche von $E \cup F$ $\rightarrow P(E) < P(E \cup F)$

WAHRSCHEINLICHKEIT: OPERATIONEN MIT EREIGNISSEN (KOMPLEMENTÄR)

> Ereignis E = Bankangestellte

Komplement: $E^c = \textcolor{red}{keine}$ Bankangestellte



Weiteres Beispiel. Münzwurf.

> Ereignis E = ⚁ oder ⚂ oder ⚄ oder ⚅

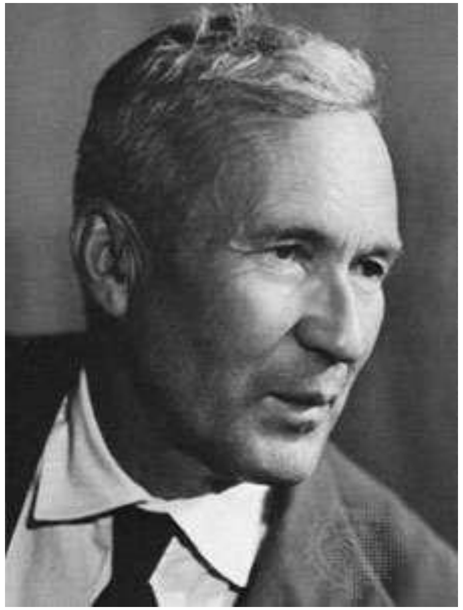
> Komplementärereignis = E^c = ⚀ oder ⚃

WAHRSCHEINLICHKEIT: AXIOME VON KOLMOGOROV

HOCHSCHULE HEILBRONN

Andrey Kolmogorov (1903 – 1987)

> hat Wahrscheinlichkeiten anhand von mathematischen Eigenschaften (Axiome) definiert



WAHRSCHEINLICHKEIT: AXIOME VON KOLMOGOROV

Axiome der Wahrscheinlichkeitsrechnung (Kolmogorov, 1928):

(A1) Nichtnegativität. Die Wahrscheinlichkeit jedes Ereignisses E ist größer oder gleich 0:

$$P(E) \geq 0.$$

(A2) Normalisierung. Irgendetwas muss passieren: Der Stichprobenraum S ist sicher:

$$P(S) = 1.$$

(A3) Additivität. Die Wahrscheinlichkeit von disjunkten Ereignissen E, F ist:

$$P(E \cup F) = P(E) + P(F).$$

Zwei Ereignisse sind disjunkt, wenn sie sich nicht überschneiden ($E \cap F = \emptyset$)

WAHRSCHEINLICHKEIT: IMPLIKATIONEN DER AXIOME VON KOLMOGOROV

HOCHSCHULE HEILBRONN

- Irgendetwas muss passieren → Wahrscheinlichkeit, dass nichts passiert ist 0:

$$P(\emptyset) = 0$$

- Die Wahrscheinlichkeit jedes Ereignisses liegt zwischen 0 und 1:

$$0 \leq P(E) \leq 1$$

- "Kleinere" Ereignisse haben kleinere Wahrscheinlichkeiten:

$$Wenn\ E\ Teilmenge\ von\ F\ ist, dann\ P(E) \leq P(F)$$

WAHRSCHEINLICHKEIT: IMPLIKATIONEN DER AXIOME VON KOLMOGOROV

- Die Vereinigungswahrscheinlichkeit von zwei Ereignissen E und F ist:

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

AXIOME VON KOLMOGOROV: EXKURS (PARTITION)

Eine *Partition* des Stichprobenraums S ist eine Menge von Ereignissen,

- die **paarweise disjunkt (= leere Schnittmenge)** sind und
- deren **Vereinigung der Stichprobenraum** ist

Beispiele. Würfelwurf. $Stichprobenraum = \{$⚀,⚁,⚂,⚃,⚄,⚅$\}$. Partitionen:

- $E_1 = \{$⚀,⚁,⚂$\}$ und $E_2 = \{$⚃,⚄,⚅$\}$
- $E_1 = \{$⚀,⚂$\}$ und $E_2 = \{$⚁,⚃,⚄,⚅$\}$
- Weitere?

WAHRSCHEINLICHKEIT: IMPLIKATIONEN DER AXIOME VON KOLMOGOROV

HOCHSCHULE HEILBRONN

- Die Summe der Wahrscheinlichkeiten jeder Partition von S ist 1:

$$P(E_1) + \cdots + P(E_k) = 1 \text{ für jede Partition } \{E_1, \ldots, E_k\} \text{ von } S.$$

› ZUFALLSEXPERIMENTE UND -VARIABLEN

ZUFALLSEXPERIMENTE: DEFINITION

Ein *Zufallsexperiment* ist ein Vorgang der

- unter identischen Bedingungen wiederholt werden kann und
- dessen mögliche Ergebnisse im Vorhinein bekannt und unsicher sind

MEHRSTUFIGE ZUFALLSEXPERIMENTE

Mehrstufiges Zufallsexperiment = fixierten Anzahl von Wiederholungen eines Experiments.

> Mehrstufige Experimente haben größere Stichprobenräume

Beispiel. Münzwurf.

MEHRSTUFIGE ZUFALLSEXPERIMENTE

Mehrstufige Zufallsexperimente können in Baumdiagrammen dargestellt werden

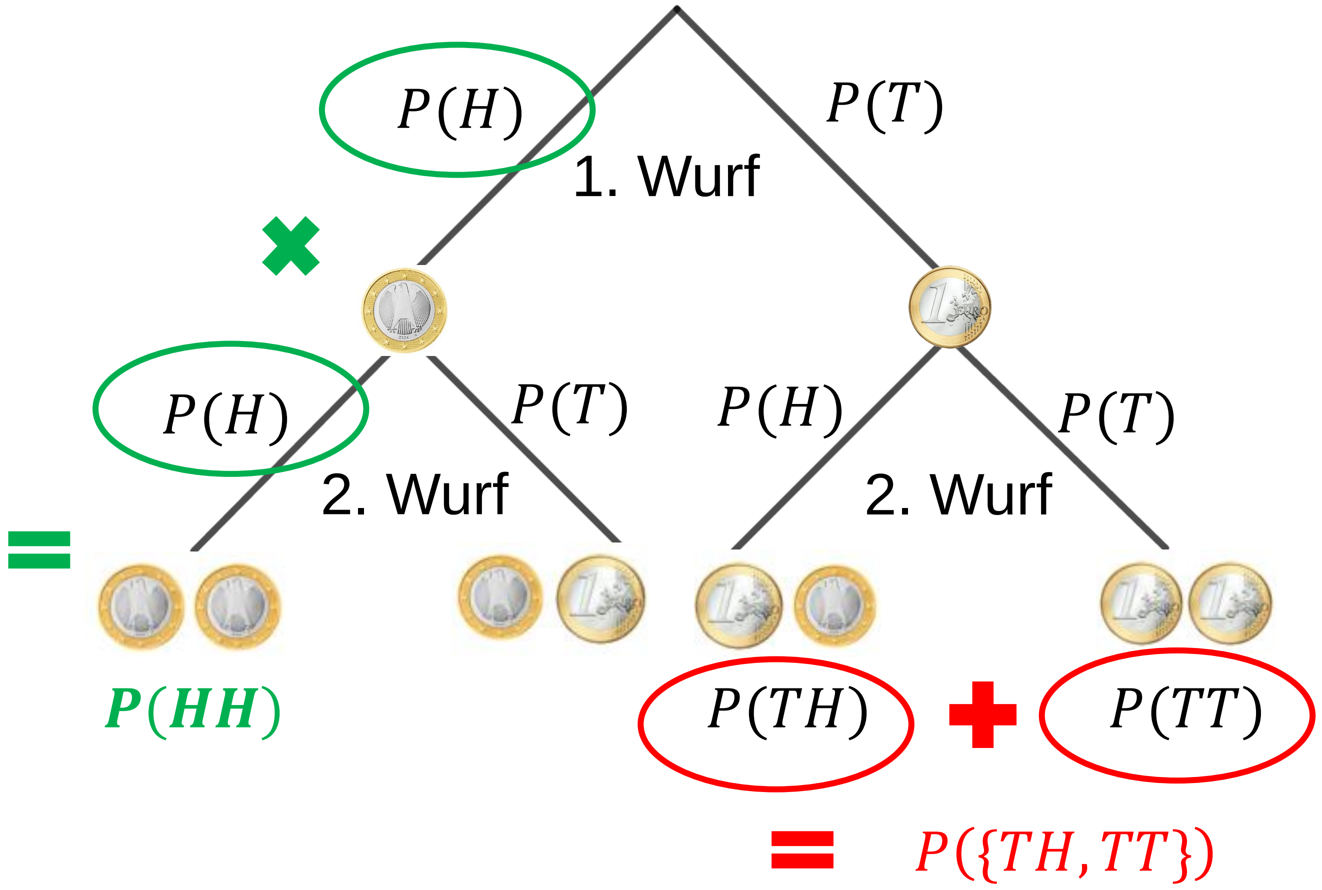
Beispiel. Zweimaliger Münzwurf.

> Produktregel:

$$P(HH) = P(H) \times P(H) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

> Summenregel:

$$P(\{TH, TT\}) = P(TH) + P(H) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$



WAS SIND ZUFALLSVARIABLEN?

Meistens interessieren wir uns nicht für alle Details eines Zufallsexperiments.

Eine $Zufallsvariable$

> hängt von den Ergebnissen eines Zufallsexperiments ab

> misst relevante Charakteristika des Experiments

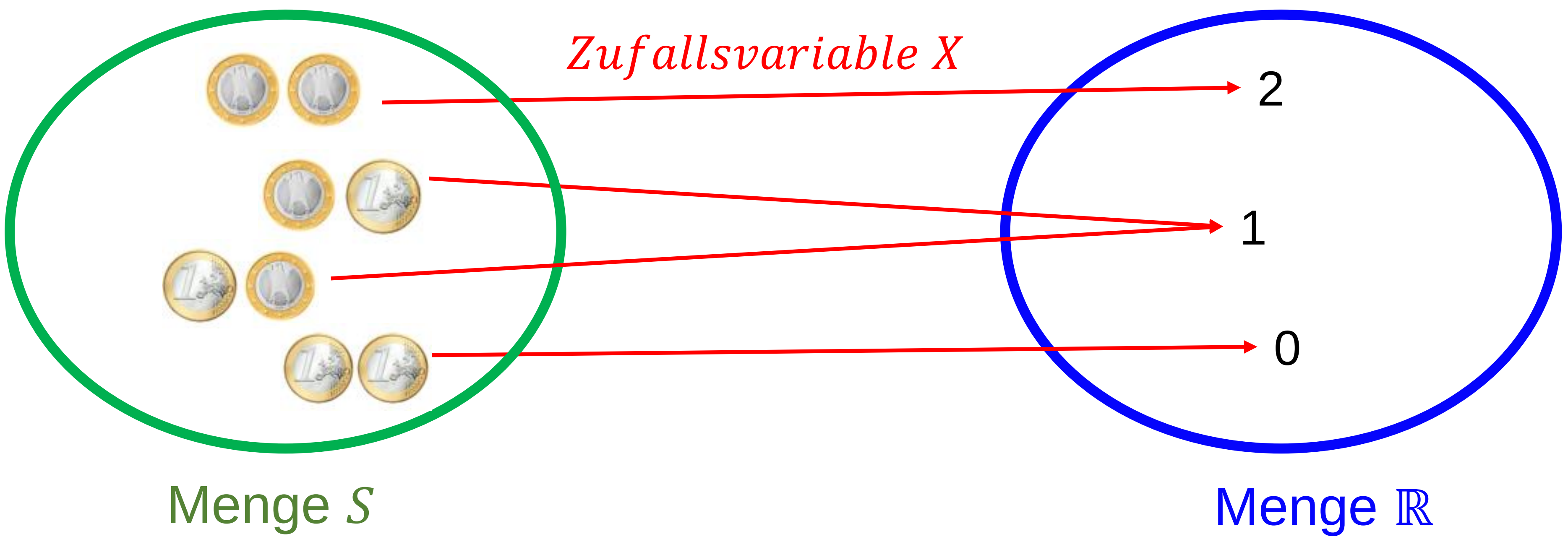
Beispiel. Mehrfacher Münzwurf.

> Wie häufig wird "Kopf" geworfen?

> Zufallsvariable, die zählt wie häufig "Kopf" geworfen wurde

WAS SIND ZUFALLSVARIABLEN?

Beispiel. Zweifacher Münzwurf. Zufallsvariable: Funktion X = Anzahl "Kopf".



WAS SIND ZUFALLSVARIABLEN?

HOCHSCHULE HEILBRONN

Zufallsvariable

Eine $Zufallsvariable$ X ist eine Funktion, die jedem Ergebnis $s \in S$ eines Zufalls-experiments ein Element s' einer Menge S' zuordnet (häufig eine Zahl: $S' = \mathbb{R}$):

$$X: S \to S'.$$

Beispiel. Zweimaliges Werfen einer Münze. X = Anzahl Kopf.

> X ist eine Stufenfunktion, die wie folgt geschrieben werden kann:

$$X(s) = \begin{cases} 2, & wenn\ s = HH \\ 1, & wenn\ s = HT\ oder\ TH \\ 0, & wenn\ s = TT \end{cases}$$

ZUFALLSVARIABLEN UND ZUFALLSEXPERIMENTE

Beispiel. Zweifacher Münzwurf.

Zufallsexperiment

$s \in S$	$P(s)$
	¼
	¼
	¼
	¼

X

Zufallsvariable X = Anzahl "Kopf"

$x \in S'$	$P'(X = x)$
2	¼
1	½
0	¼

ZUFALLSVARIABLEN UND ZUFALLSEXPERIMENTE

HOCHSCHULE HEILBRONN

Das Konzept einer Zufallsvariablen = Natürliche Erweiterung Zufallsexperiment

$$Zufallsvariable\colon X\colon S \to S'$$

- Wahl $S' = S \to$ Ursprüngliches Zufallsexperiment

ZUFALLSVARIABLEN: ANMERKUNGEN

- Kombinationen von Zufallsvariablen

$$X + Y, X \cdot Y, \ldots$$

UND

- Kombinationen von Zufallsvariablen mit einer Konstanten

$$X + Konstante, X \cdot Konstante, \ldots$$

sind ebenfalls Zufallsvariablen

ZUFALLSVARIABLEN: DISKRET UND STETIG

HOCHSCHULE HEILBRONN

Diskrete Zufallsvariable

Nimmt eine **abzählbare** Anzahl (endlich oder unendlich) von Werten an.

Stetige Zufallsvariable

Nimmt eine **nicht abzählbare** unendliche Anzahl von Werten an.

Beispiel. Anzahl an Sechsen bei vierfachem Würfelwurf:

$$0, 1, 2, 3, 4$$

Beispiel. Größe einer zufällig ausgewählten Person:

$$145\ cm \ \leq\ X\ \leq\ 220\ cm$$

ZUFALLSVARIABLEN: DISKRET UND STETIG

Zufallsvariable	Art
Anzahl Verkäufe eines Verkäufers in einer Woche	
Die Größe eines zufällig ausgewählten Erwachsenen	
Die Zeit zwischen der Ankunft zweier Touristen an einem Urlaubsort	
Anzahl Kunden in einer Stichprobe, die ein bestimmtes Produkt gegenüber allen Wettbewerbern bevorzugen	
Neuer Wohnkomplex – die Zeit zwischen der Fertigstellung und dem Verkauf der letzten Wohnung	
Erwartete Anzahl Studierende in einer Vorlesung	
Die Tiefe, bis zu der gebohrt werden muss, bis man auf Öl stößt	

ZUFALLSVARIABLEN: ERWARTUNGSWERT

HOCHSCHULE HEILBRONN

Beispiel 1. Wette: Gewinn 2€ mit 70%, Verlust 3 € mit 30%.

> Was würde man im Schnitt „erwarten“, wenn man diese Wette immer wieder annimmt?

Beispiel 2. Würfelwurf.

> Welche durchschnittliche Augenzahl würde man erwarten bei vielen Wiederholungen?

ZUFALLSVARIABLEN: ERWARTUNGSWERT

Erwartungswert $E[X]$ Zufallsvariable X = wahrscheinlichkeitsgewichtet Summe Werte:

$$E[X] = P(x_1) \cdot x_1 + \cdots + P(x_n) \cdot x_n$$

> Der Erwartungswert kann als "wahrer" langfristiger Durchschnittswert gesehen werden

ZUFALLSVARIABLEN: ERWARTUNGSWERT UND MAẞE

Beispiel 3. Durchschnittsnoten in Statistik der letzten Jahre:

Jahr	14	15	16	17	18	19	20	21	22	23
Durchschnittsnote	2,3	2,3	2,7	3,3	2,3	2,7	2,7	1,7	3,3	2,3

> Mittelwert der Durchschnittsnoten?

> Erwartungswert der Durchschnittsnoten?

ZUFALLSVARIABLEN: ERWARTUNGSWERT UND MAẞE

Sei X eine Zufallsvariable.

Es gilt:

$$Mittelwert\ X = E[X]$$

und

$$Varianz\ X = E[(X - E[X])^2] = E[(X - Mittelwert\ X)^2]$$

und

$$Kovarianz\ X, Y = E[(X - Mittelwert\ X) \times (Y - Mittelwert\ Y)]$$

HOCHSCHULE HEILBRONN

› UNIVARIATE WAHRSCHEIN-LICHKEITSVERTEILUNGEN

UNIVARIATE VERTEILUNGEN: DISKRET UND STETIG

- $Diskrete\ Verteilung$ = Verteilung einer diskreten Zufallsvariablen
- $Stetige\ Verteilung$ = Verteilung einer stetigen Zufallsvariablen

UNIVARIATE DISKRETE VERTEILUNGEN: WAHRSCHEINLICHKEITSFUNKTION

HOCHSCHULE HEILBRONN

Eine diskrete Verteilung

- ist definiert durch eine $Wahrscheinlichkeitsfunktion\ P(X = x)$
- Liste aller Werte der Zufallsvariablen und Ihrer Wahrscheinlichkeiten

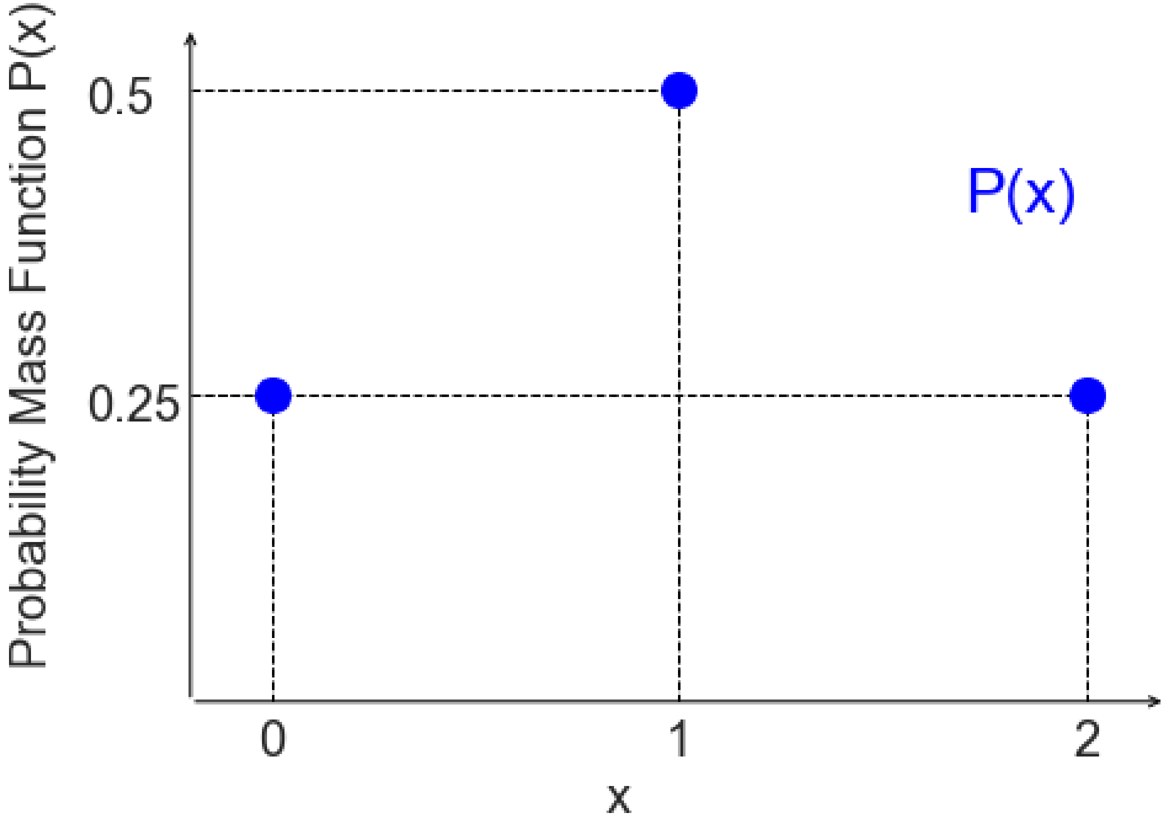
Beispiel. Zweifacher Münzwurf. X = Anzahl "Kopf". Wahrscheinlichkeitsverteilung:

x	$P(X = x)$
2	¼
1	½
0	¼

UNIVARIATE DISKRETE VERTEILUNGEN: WAHRSCHEINLICHKEITSFUNKTION

Wahrscheinlichkeitsfunktion – graphisch:

x	$P(X = x)$
2	¼
1	½
0	¼



UNIVARIATE DISKRETE VERTEILUNGEN: BERECHNUNG VON WAHRSCHEINLICHKEITEN

HOCHSCHULE HEILBRONN

x	$P(X = x)$
2	¼
1	½
0	¼

> Was ist die Wahrscheinlichkeit, dass mindestens 1x "Kopf" geworden wird $P(X \geq 1)$?

UNIVARIATE DISKRETE VERTEILUNGEN: BERECHNUNG VON WAHRSCHEINLICHKEITEN

Die Wahrscheinlichkeit, dass X einen Wert aus einer Menge M annimmt ist:

$$P(X \in M) = \sum_{x \in M} P(X = x)$$

UNIVARIATE DISKRETE VERTEILUNGEN: GLEICHVERTEILUNG

Eine diskrete Zufallsvariable heißt $gleichverteilt$,

- wenn alle ihre Werte gleichwahrscheinlich sind.

Beispiele?

UNIVARIATE DISKRETE VERTEILUNGEN: BERNOULLI-VERTEILUNG

Eine diskrete Zufallsvariable heißt $Bernoulli - verteilt$,

- wenn sie nur zwei Werte hat (0 and 1). Wert 1 heißt $Erfolg$ und Wert 0 $Misserfolg$.

Beispiele?

UNIVARIATE DISKRETE VERTEILUNGEN: BERNOULLI-VERTEILUNG

HOCHSCHULE HEILBRONN

Die Wahrscheinlichkeitsfunktion der Bernoulli-Verteilung ist

$$P(X) = \begin{cases} \pi & für\, X = 1 \\ (1-\pi) & für\, X = 0 \end{cases},$$

- wobei $\pi =$ Erfolgswahrscheinlichkeit

UNIVARIATE DISKRETE VERTEILUNGEN: BERNOULLI-VERTEILUNG

HOCHSCHULE HEILBRONN

Vereinfachte Formel Varianz Bernoulli-verteilte Zufallsvariable X:

$$Varianz\,X = \pi \times (1 - \pi),$$

- wobei $\pi =$ Erfolgswahrscheinlichkeit

UNIVARIATE DISKRETE VERTEILUNGEN: BINOMIALVERTEILUNG

HOCHSCHULE HEILBRONN

Eine diskrete Zufallsvariable heißt $binomialverteilt$,

> wenn sie die Anzahl der Erfolge in einem Bernoulli-Prozess der Länge n zählt.

$Bernoulli - Prozess$ = Folge von n unabhängig und identisch verteilten Bernoulli-Variablen:

1. Fixierte Länge n
2. Nur 2 mögliche Ergebnisse pro Versuch: Erfolg und Misserfolg (Bernoulli-Variable)
3. Erfolgsw'keit π ist in allen Versuchen identisch (unabhängig und identisch verteilt)

UNIVARIATE DISKRETE VERTEILUNGEN: BINOMIALVERTEILUNG

HOCHSCHULE HEILBRONN

Ist dies ein Bernoulli-Prozess?

> Ein kleines Reisebüro hat 20 Reisekataloge bestellt.

> Aufgrund schlechter Erfahrungen mit Fehldrucken möchte das Reisebüro 3 Kataloge auf Fehldruck prüfen, bevor es die Lieferung annimmt.

> 3 Kataloge werden sequentiell zufällig ausgewählt und auf Fehldruck geprüft.

> Was das Reisebüro nicht weiß: 2 Kataloge sind Fehldrucke

UNIVARIATE DISKRETE VERTEILUNGEN: BINOMIALVERTEILUNG

HOCHSCHULE HEILBRONN

Beispiel. Fünfacher Münzwurf.

- Die Münze ist leicht unfair: W'keit Kopf = 2/5 (W'keit Zahl = 3/5)
- $X = Anzahl\ Kopf$

Was ist die Wahrscheinlichkeit 0-mal Kopf zu werfen $P(X = 0)$?

- 0-mal Kopf (H) = 5-mal Zahl (T): 1 Ereignis: $TTTTT$

$$P(X = 0) = \frac{3}{5} \cdot \frac{3}{5} \cdot \frac{3}{5} \cdot \frac{3}{5} \cdot \frac{3}{5} = \frac{3}{5}^5 = 1 \cdot \frac{2}{5}^0 \cdot \frac{3}{5}^5 \approx 7.78\%$$

UNIVARIATE DISKRETE VERTEILUNGEN: BINOMIALVERTEILUNG

HOCHSCHULE HEILBRONN

Beispiel (Fortsetzung). Fünfacher Münzwurf.

Was ist die Wahrscheinlichkeit exakt 1-mal Kopf zu werfen ($X = 1$)?

- 5 Ereignisse: $HTTTT, THTTT, TTHTT, TTTHT, TTTTH$

$$P(X = 1) = \frac{2}{5} \cdot \frac{3}{5} \cdot \frac{3}{5} \cdot \frac{3}{5} \cdot \frac{3}{5} + \cdots + \frac{3}{5} \cdot \frac{3}{5} \cdot \frac{3}{5} \cdot \frac{3}{5} \cdot \frac{2}{5} = 5 \cdot \frac{2}{5}^1 \cdot \frac{3}{5}^4 \approx 25.92\%$$

UNIVARIATE DISKRETE VERTEILUNGEN: BINOMIALVERTEILUNG

HOCHSCHULE HEILBRONN

Beispiel (Fortsetzung). Fünfacher Münzwurf.

Was ist die Wahrscheinlichkeit exakt 2-mal Kopf zu werfen ($X = 2$)?

- 10 Ereignisse: $HHTTT, HTHTT, \dots, TTTHH$

$$P(X=2) = \frac{2}{5}\cdot\frac{2}{5}\cdot\frac{3}{5}\cdot\frac{3}{5}\cdot\frac{3}{5} + \cdots + \frac{3}{5}\cdot\frac{3}{5}\cdot\frac{3}{5}\cdot\frac{2}{5}\cdot\frac{2}{5} = 10\cdot\frac{2}{5}^{2}\cdot\frac{3}{5}^{3} \approx 34.56\%$$

UNIVARIATE DISKRETE VERTEILUNGEN: BINOMIALVERTEILUNG

HOCHSCHULE HEILBRONN

Es existiert ein Muster!

Anzahl Erfolge (x)

Erfolgswahrscheinlichkeit (π)

$$P(X = 0) = 1 \cdot \frac{2}{5}^0 \cdot \frac{3}{5}^5$$

Anzahl Misserfolge ($n - x$)

$$P(X = 1) = 5 \cdot \frac{2}{5}^1 \cdot \frac{3}{5}^4$$

Misserfolgs-wahrscheinlichkeit ($1 - \pi$)

Anzahl Ereignisse

$$P(X = 2) = 10 \cdot \frac{2}{5}^2 \cdot \frac{3}{5}^3$$

> Wie können wir die Anzahl der Ereignisse berechnen?

> Anzahl Möglichkeiten x Objekte (Erfolge) aus n Objekten (Versuche) zu ziehen:

= Binomialkoeffizient

UNIVARIATE DISKRETE VERTEILUNGEN: BINOMIALVERTEILUNG

$Binomialkoeffizient$ $\binom{n}{x}$ (lies "*n* über k"):

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}.$$

- wobei $n!$ (lies "*n* Fakultät") $= n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1$
- Beispiel: $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$

UNIVARIATE DISKRETE VERTEILUNGEN: BINOMIALVERTEILUNG

HOCHSCHULE HEILBRONN

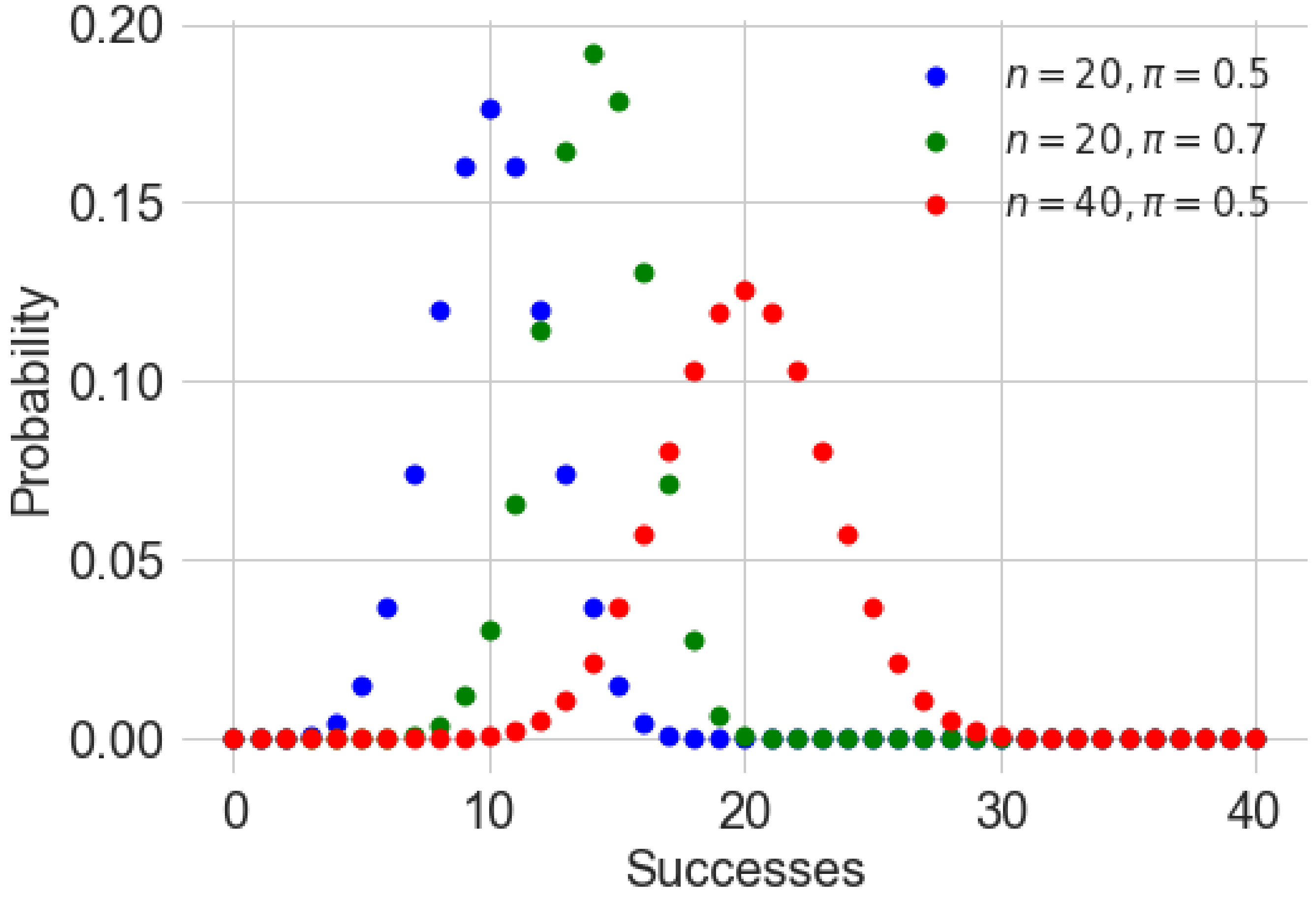
Die Wahrscheinlichkeitsfunktion der Binomialverteilung ist

$$P_{\pi,n}(X = x) = \binom{n}{x} \cdot \pi^x \cdot (1 - \pi)^{n-x},$$

- π = Erfolgswahrscheinlichkeit und n = Länge des Bernoulli-Prozesses

UNIVARIATE DISKRETE VERTEILUNGEN: BINOMIALVERTEILUNG

Binomiale Wahrscheinlichkeitsfunktion für unterschiedliche Paramterkombinationen:



UNIVARIATE DISKRETE VERTEILUNGEN: BINOMIALVERTEILUNG

HOCHSCHULE HEILBRONN

Beispiel. Einfluss von Frauen und Männern auf familiäre Kaufentscheidungen

- Studie: In 70% der Fällen haben Männer einen maßgeblichen Einfluss beim Autokauf
- 4 Familien sind im Begriff ein neues Auto zu kaufen

a) Bestimmen Sie die Erfolgswahrscheinlichkeit und n

b) Wahrscheinlichkeit, dass der Mann in **genau** zwei Fällen die Entscheidung bestimmt?

c) W'keit, dass der Mann in höchstens 3 Familien die Kaufentscheidung bestimmt?

d) W'keit, dass der Mann in **mindestens** 2 Familien die Kaufentscheidung bestimmt?

Lösung: $P(X \geq 2) = 91.63\%$

UNIVARIATE STETIGE VERTEILUNGEN: WAHRSCHEINLICHKEITSDICHTEFUNKTION

HOCHSCHULE HEILBRONN

Bei stetigen Verteilungen ist die Wahrscheinlichkeit einzelner Werte (Atome) 0

→ Wir können hier also keine Wahrscheinlichkeitsfunktion verwenden

UNIVARIATE STETIGE VERTEILUNGEN: WAHRSCHEINLICHKEITSDICHTEFUNKTION

HOCHSCHULE HEILBRONN

Lösung: $Wahrscheinlichkeitsdichtefunktion\ f(x)$

- identifiziert Regionen mit höherer bzw. niedrigerer Wahrscheinlichkeit

  $f(2) > f(8)$: Wahrscheinlichkeit, dass Zufallsvariable Wert nahe 2 annimmt ist höher

- Wahrscheinlichkeitsdichten $\neq$ Wahrscheinlichkeiten! Es ist z.B. möglich, dass $f(x) > 1$

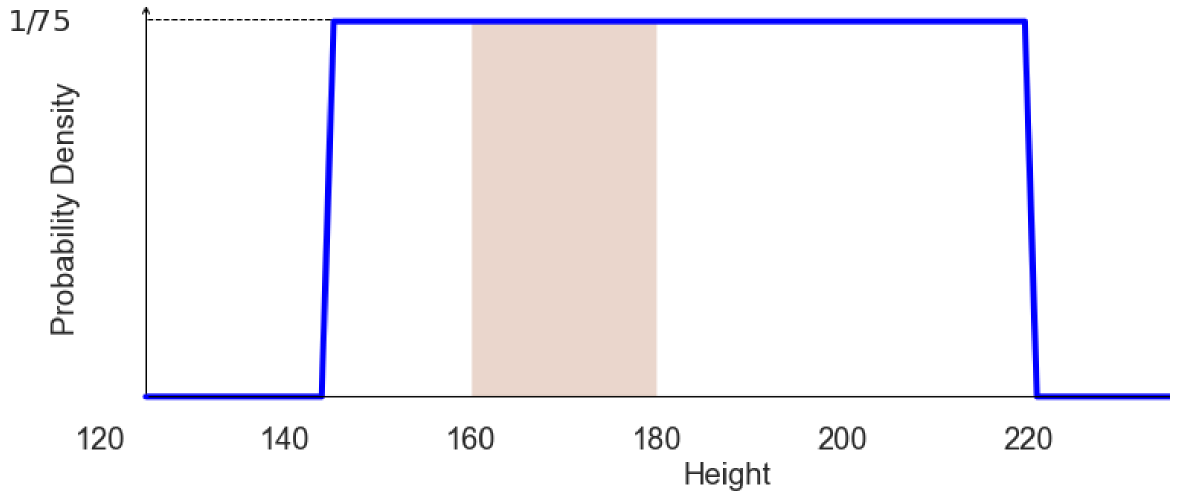
UNIVARIATE STETIGE VERTEILUNGEN: BERECHNUNG VON WAHRSCHEINLICHKEITEN

HOCHSCHULE HEILBRONN

Wahrscheinlichkeit eines Bereichs = Fläche unter Wahrscheinlichkeitsdichtefunktion $f(x)$

Beispiel. X = Größe ($145\ cm \leq X \leq 220\ cm$) eine zufällig ausgewählten Person.

- Annahme: Alle Größen sind gleichwahrscheinlich ($f(x) = \frac{1}{75}$ für alle Größen x)
- P(Größe zwischen 160 und 180)?



UNIVARIATE STETIGE VERTEILUNGEN: BERECHNUNG VON WAHRSCHEINLICHKEITEN

HOCHSCHULE HEILBRONN

Die Wahrscheinlichkeit, dass X einen Wert im zwischen x_1 und x_2 annimmt:

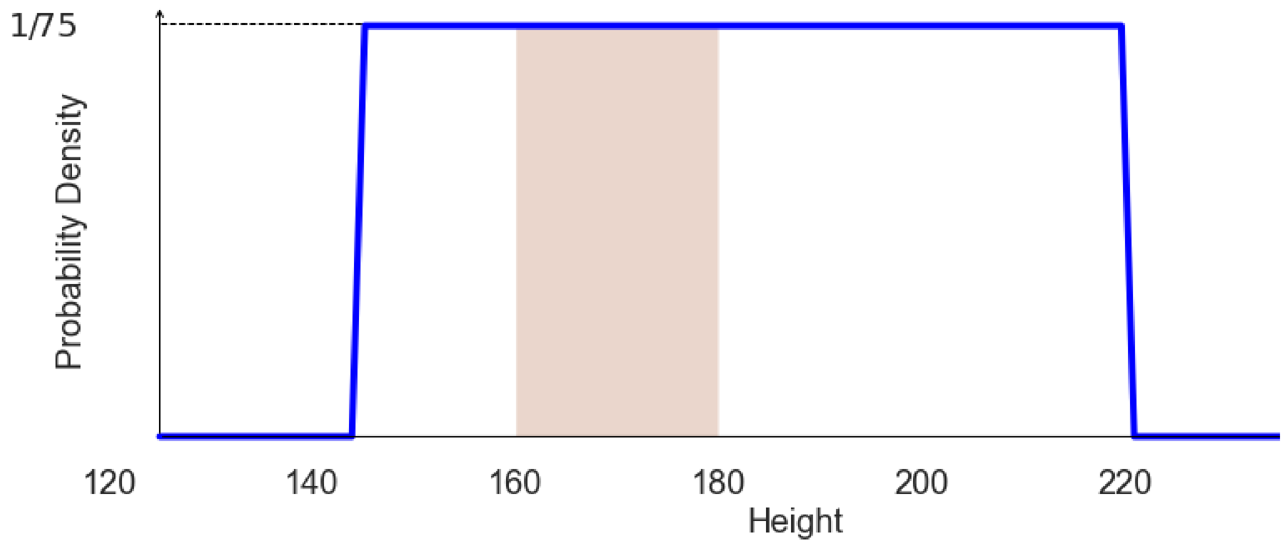
$$P(X \in [x_1, x_2]) = \int_{x_1}^{x_2} f(x)\,\mathrm{d}x$$

UNIVARIATE STETIGE VERTEILUNGEN: GLEICHVERTEILUNG

Eine stetige Zufallsvariable X heißt $gleichverteilt$,

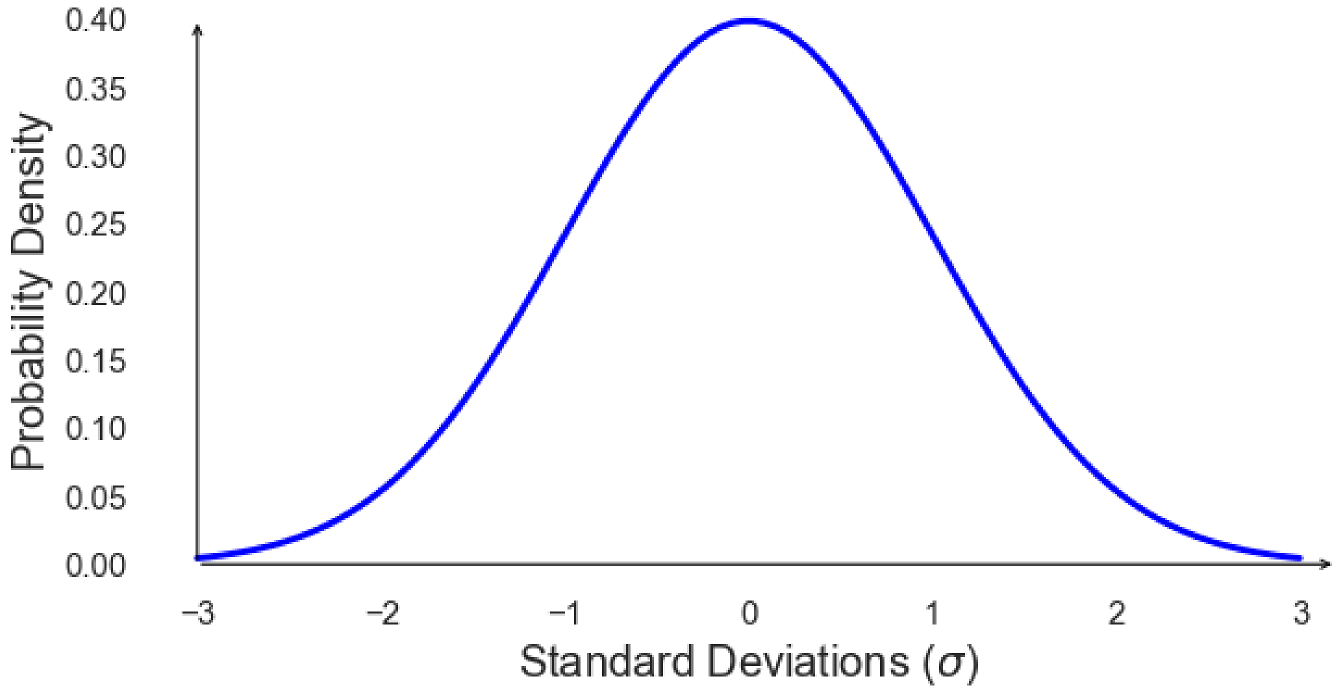
> wenn alle Werte x die gleiche Wahrscheinlichkeitsdichte $f(x)$ haben.

Beispiel.



UNIVARIATE STETIGE VERTEILUNGEN: NORMALVERTEILUNG

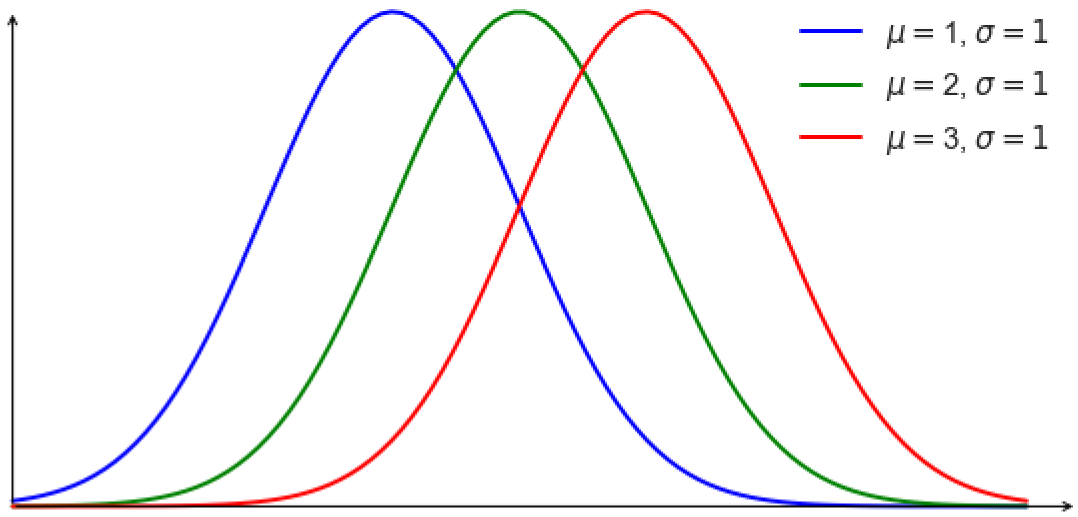
In der Realität haben die meisten Verteilungen eine Glockenform:

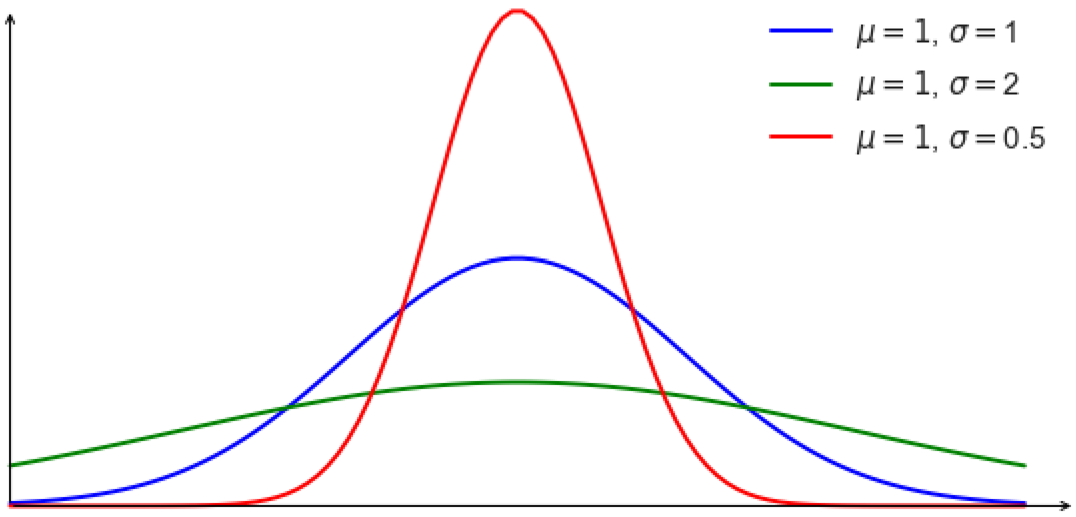


- Derartige Verteilungen heißen *Normal*- oder *Gaußverteilung*
- Glockenform: Extremwerte (Ränder) sind unwahrscheinlicher als mittlere Werte

UNIVARIATE STETIGE VERTEILUNGEN: NORMALVERTEILUNG

Normalverteilung für verschiedene Erwartungswerte μ und Standardabweichungen σ:



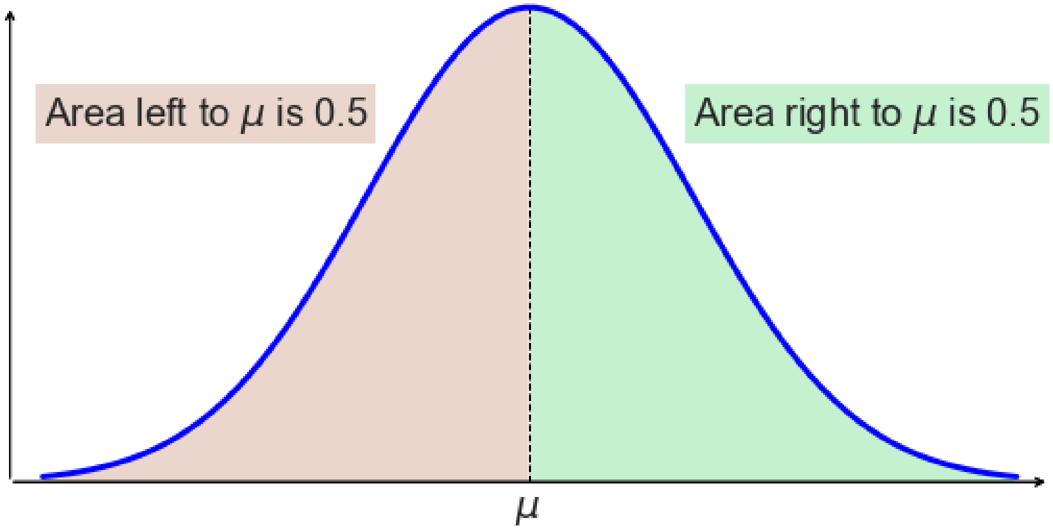


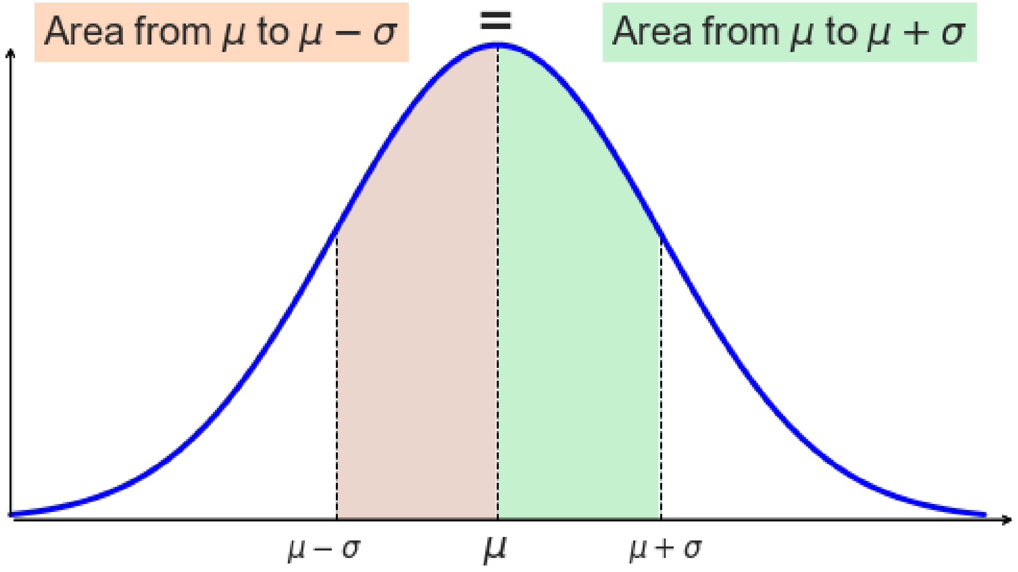
UNIVARIATE STETIGE VERTEILUNGEN: NORMALVERTEILUNG

HOCHSCHULE HEILBRONN

Symmetrie der Normalverteilung:

> Die Normalverteilung ist perfekt symmetrisch um Ihren Erwartungswert μ:





UNIVARIATE STETIGE VERTEILUNGEN: NORMALVERTEILUNG

HOCHSCHULE HEILBRONN

Eine stetige Zufallsvariable X heißt $normalverteilt$,

- wenn sie folgende Dichtefunktion hat:

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

- e ($\approx$ 2.7183) und π ($\approx$ 3.1416) sind mathematische Konstanten
- μ = Erwartungswert und σ = Standardabweichung der Verteilung

Die Normalverteilung heißt $Standardnormalverteilung$, wenn $\mu = 0$ und $\sigma = 1$

UNIVARIATE STETIGE VERTEILUNGEN: NORMALVERTEILUNG

Berechnung Wahrscheinlichkeiten Normalverteilung:

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

$=$ 

> Gute Nachrichten! Wir können Integralrechnung vermeiden:

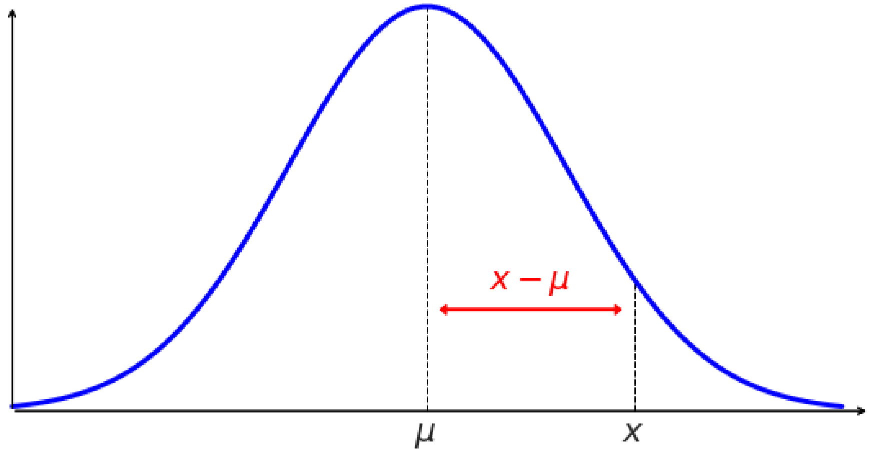
Tabelle mit Wahrscheinlichkeiten (Flächen) für Standardnormalverteilung

UNIVARIATE STETIGE VERTEILUNGEN: Z-TRANSFORMATION NORMALVERTEILUNG

Gegeben eine Zufallsvariable X mit Erwartungswert μ und Standardabweichung σ.

> Der $z - Wert$ eines Wertes x von X ist:

$$z(x) = \frac{x - \mu}{\sigma}.$$



> $z(x)$ = Distanz zwischen μ und x in Standardabweichungen σ

> Wenn X normalverteilt ist, dann ist $Z(X)$ standardnormalverteilt

UNIVARIATE STETIGE VERTEILUNGEN: Z-TRANSFORMATION NORMALVERTEILUNG

Beispiel. Z-Transformation.

Wahrscheinlichkeitsverteilung:

x	p	$p \cdot x$
2	0.1	0.2
4	0.2	0.8
6	0.4	2.4
8	0.2	1.6
10	0.1	1
Summe		$\boldsymbol{\mu_x = 6}$
		$\boldsymbol{\sigma_X = 2.19}$

Z-transformierte Wahrscheinlichkeitsverteilung:

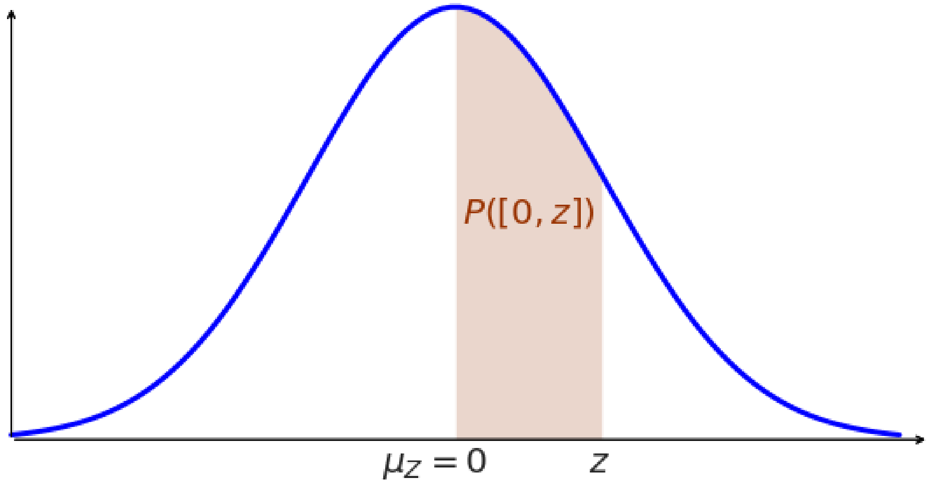
$z(x)$	p	$p \cdot z(x)$
-1.83	0.1	-0.18
-0.91	0.2	-0.18
0	0.4	0
0.91	0.2	0.18
1.83	0.1	0.18
Summe		$\boldsymbol{\mu_Z = 0}$
		$\boldsymbol{\sigma_Z = 1}$

$$z(x) = \frac{x - \mu_x}{\sigma_x}$$

UNIVARIATE STETIGE VERTEILUNGEN: TABELLE STANDARDNORMALVERTEILUNG

HOCHSCHULE HEILBRONN

Z-Tabelle: Wahrscheinlichkeit $P([0,z])$



Z	0.00	0.01	0.02	0.03	0.04
0.00	0.0000	0.0040	0.0080	0.0120	0.0160
0.10	0.0398	0.0438	0.0478	0.0517	0.0557
0.20	0.0793	0.0832	0.0871	0.0910	0.0948
0.30	0.1179	0.1217	0.1255	0.1293	0.1331
0.40	0.1554	0.1591	0.1628	0.1664	0.1700
0.50	0.1915	0.1950	0.1985	0.2019	0.2054
0.60	0.2257	0.2291	0.2324	0.2357	0.2389
0.70	0.2580	0.2611	0.2642	0.2673	0.2704
0.80	0.2881	0.2910	0.2939	0.2967	0.2995
0.90	0.3159	0.3186	0.3212	0.3238	0.3264
1.00	0.3413	0.3438	0.3461	0.3485	0.3508
1.10	0.3643	0.3665	0.3686	0.3708	0.3729
1.20	0.3849	0.3869	0.3888	0.3907	0.3925
1.30	0.4032	0.4049	0.4066	0.4082	0.4099
1.40	0.4192	0.4207	0.4222	0.4236	0.4251
1.50	0.4332	0.4345	0.4357	0.4370	0.4382

> Z-Wert bis 1. Nachkommastelle

> 2. Nachkommastelle

Beispiel. $P([0{,}1.22])?$ $= 0.3888$

UNIVARIATE STETIGE VERTEILUNGEN: TABELLE STANDARDNORMALVERTEILUNG

HOCHSCHULE HEILBRONN

Beispiel. Eine Person wird zufällig ausgewählt. $X = Größe$.

> Größe ist normalverteilt mit $\mu_X = 167cm$ und $\sigma_X = 10cm$.

a) Berechne die W'keit, dass die Person zwischen 167 und 178cm ist. $P_X([167,178])$

b) Wahrscheinlichkeit, dass X zwischen 150 und 181 cm ist?: $P_X([150,181])$

c) Bestimme die Wahrscheinlichkeit, dass die Person größer als 175 ist: $P_X(X > 175)$

d) Wahrscheinlichkeit, dass X zwischen 170 und 175 cm ist?: $P_X([170,175])$

UNIVARIATE STETIGE VERTEILUNGEN: TABELLE STANDARDNORMALVERTEILUNG

Beispiel. Aggresivitäts-Score. $X = Score\ Frauen, Y = Score\ Männer$

- X ist normalverteilt mit $\mu_X = 38.82$ und $\sigma_X = 7.91$ (*)
- Y ist normalverteilt mit $\mu_Y = 40.86$ und $\sigma_Y = 8.69$ (*)

(*) Siehe “Gender Differences in Aggression-related Responses on EEG and ECG” in *Exp Neurobiol.* 27

Eine Frau und ein Mann werden zufällig ausgewählt.

- Wie wahrscheinlich ist es, dass der Mann aggressiver ist als die Frau?

ABSCHÄTZUNG STETIGER VERTEILUNGEN: EMPIRISCHE REGEL (NORMALVERTEILUNG)

Empirische Regel: Abschätzung der Normalverteilung

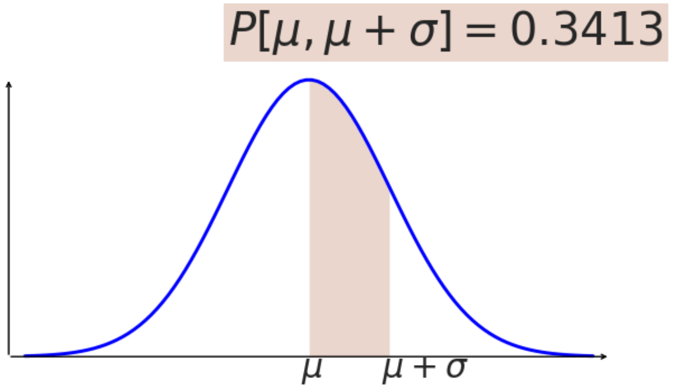
Wahrscheinlichkeit, dass eine normalverteilte Variable im Intervall

Mittelwert (μ) bis μ +1 Standardabweichung? +2 St.abw.? +3 St.abw.?

> Z-Werte für $[\mu, \mu + \sigma]$?

$$z(\mu) = 0$$

$$z(\mu + \sigma) = \frac{\mu + \sigma - \mu}{\sigma} = 1$$

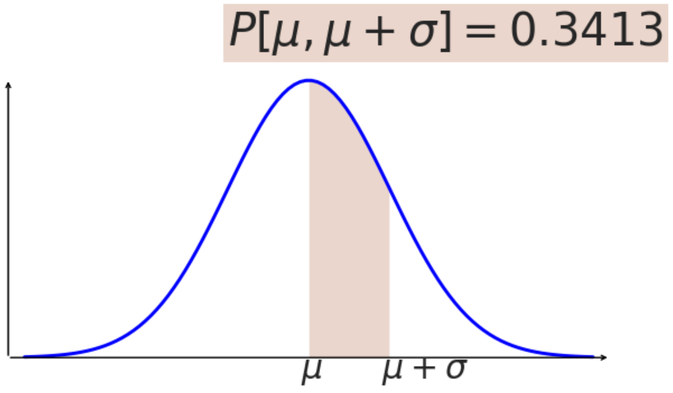


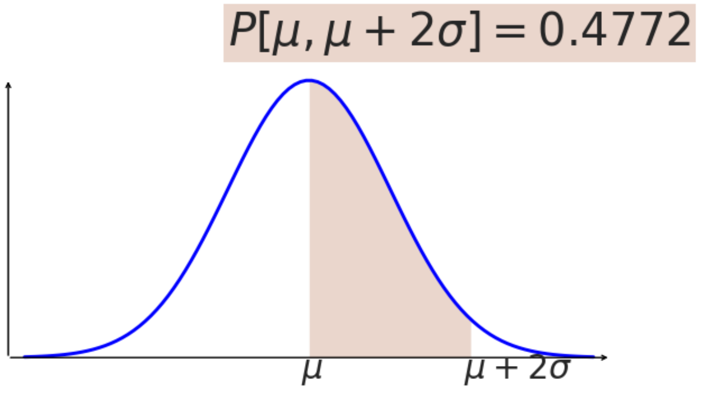
ABSCHÄTZUNG STETIGER VERTEILUNGEN: EMPIRISCHE REGEL (NORMALVERTEILUNG)

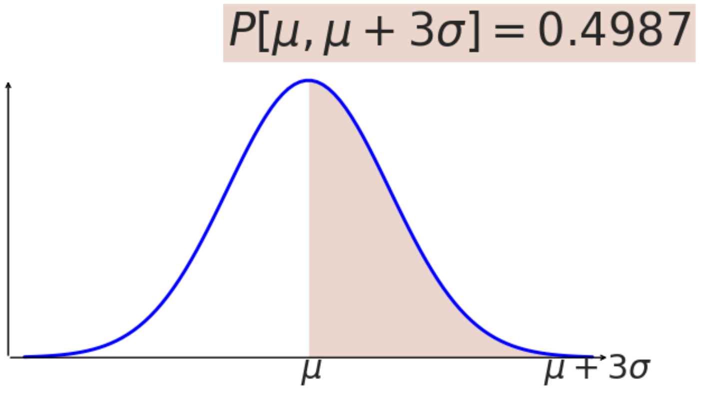
Empirische Regel: Abschätzung der Normalverteilung

Analog:

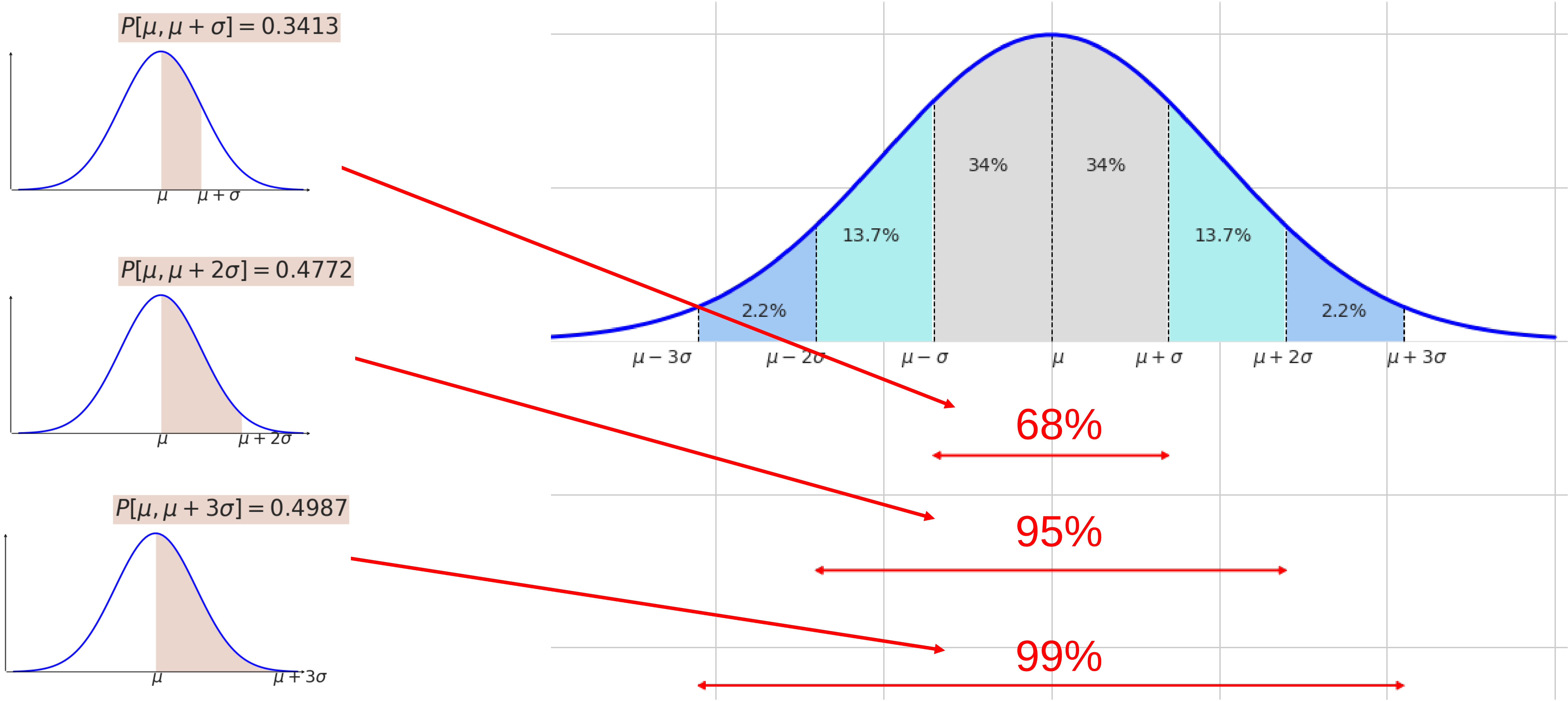
1.00	0.3413
2.00	0.4772
3.00	0.4987







ABSCHÄTZUNG STETIGER VERTEILUNGEN: EMPIRISCHE REGEL (NORMALVERTEILUNG)



ABSCHÄTZUNG STETIGER VERTEILUNGEN: CHEBYSHEVS UNGLEICHUNG

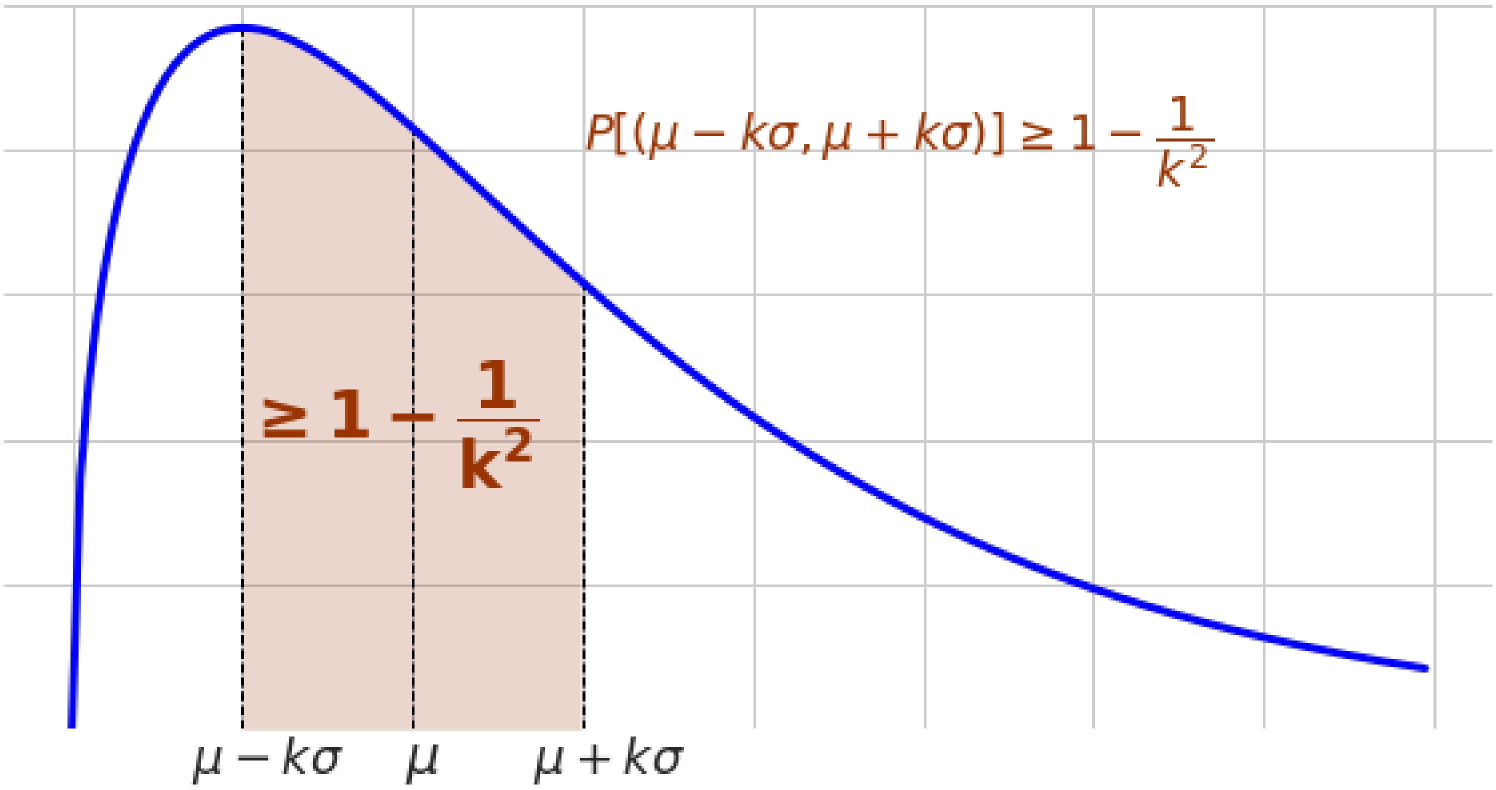
HOCHSCHULE HEILBRONN

> Die empirische Regel gilt nur für die Normalverteilung

> Gibt es eine Regel für andere Verteilungen

Ja! Chebyshevs Ungleichung:

> k = positive reelle Zahl



ABSCHÄTZUNG STETIGER VERTEILUNGEN: CHEBYSHEVS UNGLEICHUNG

HOCHSCHULE HEILBRONN

Anwendung von Chebyshevs Ungleichung

- k = Anzahl Standardabweichungen, die die Daten höchsten vom Mittelwert entfernt sind

k	Wahrscheinlichkeit
1	$\geq 0\%$
2	$\geq 75\%$
3	$\geq 88.89\%$

ABSCHÄTZUNG STETIGER VERTEILUNGEN: CHEBYSHEVS UNGLEICHUNG

HOCHSCHULE HEILBRONN

Beispiel. Nahrungsmittelunternehmen geben im Schnitt 135 € (Standardabweichung: 16 €) für Ads pro Internet-Nutzer aus. Die Art der Verteilung ist unbekannt.

a) Bestimmen Sie die Mindestwahrscheinlichkeit, dass eine zufällig ausgewählte Firma zwischen 115 € und 155 € ausgibt.

b) Bestimmen Sie die Ausgabenspanne (Intervall) in welche mindestens 50% der Unternehmen fallen

› MULTIVARIATE WAHRSCHEIN-LICHKEITSVERTEILUNGEN

MULTIVARIATE VERTEILUNGEN: BEDINGTE WAHRSCHEINLICHKEITEN

HOCHSCHULE HEILBRONN

Thomas Bayes (1701 – 1761)

> Wahrscheinlichkeiten = subjektive Einschätzungen

MULTIVARIATE VERTEILUNGEN: BEDINGTE WAHRSCHEINLICHKEITEN

HOCHSCHULE HEILBRONN

Bayesianischer Ansatz. Unterscheidung zwischen:

A−Priori Wahrscheinlichkeit:
Subjektive Wahrscheinlichkeit $P(E)$

Beobachtung
Ereignis F

A-Posteriori Wahrscheinlichkeit:
Bedingte Wahrscheinlichkeit $P(E \mid F)$

> Bedingte Wahrscheinlichkeit $P(E \mid F)$: W'keit von E nachdem F beobachtet wurde

MULTIVARIATE VERTEILUNGEN: BEDINGTE WAHRSCHEINLICHKEITEN

Beispiel. Bayesianischer Ansatz: Würfelwurf: $E = \{$⚁, ⚃$\}$ und $F = \{$⚁$\}$.

- Wahrscheinlichkeiten von E und F?

- Wahrscheinlichkeit von E unter der Bedingung, dass F eingetreten ist?

- Wahrscheinlichkeit von F unter der Bedingung, dass E eingetreten ist?

MULTIVARIATE VERTEILUNGEN: BEDINGTE WAHRSCHEINLICHKEITEN

HOCHSCHULE HEILBRONN

Bayes-Theorem und totale Wahrscheinlichkeit.

Bayes-Theorem

Die bedingte Wahrscheinlichkeit von Ereignis E nachdem F beobachtet wurde ist:

$$P(E \mid F) = \frac{P(F \mid E) \times P(E)}{P(F)} = \frac{P(E \cap F)}{P(F)}.$$

- $Bayes - Theorem$: Anpassung A-priori Wahrscheinlichkeiten im Lichte neuer Information
- Anwendung Bayes-Theorem erfordert häufig den $Satz\ der\ totalen\ Wahrscheinlichkeit$

MULTIVARIATE VERTEILUNGEN: BEDINGTE WAHRSCHEINLICHKEITEN

HOCHSCHULE HEILBRONN

Beispiel aus der Einführung. Erinnerung

- Texas: 38% fahren einen SUV, 57% wählen die Republikaner
- Umfrage unter SUV-Fahrern: 78% sind Republikaner

Wir müssen folgende Wahrscheinlichkeiten unterscheiden:

- Bedingte W'keit Republikaner gegeben man ist SUV-Fahrer:

$$P(Republikaner \mid SUV) = 78\%$$ **Das misst die Studie**

- Bedingte W'keit SUV-Fahrer gegeben man ist Republikanter:

$$P(SUV \mid Republikaner) = ?$$ **Das wollte die Studie messen**

MULTIVARIATE VERTEILUNGEN: BEDINGTE WAHRSCHEINLICHKEITEN

Wir können nun das Bayes-Theorem anwenden, um den Machern der Studie zu helfen

MULTIVARIATE VERTEILUNGEN: BEDINGTE WAHRSCHEINLICHKEITEN

HOCHSCHULE HEILBRONN

Betrachten Sie einen Corona-Test der 90% sensitiv ist und 80% spezifisch.

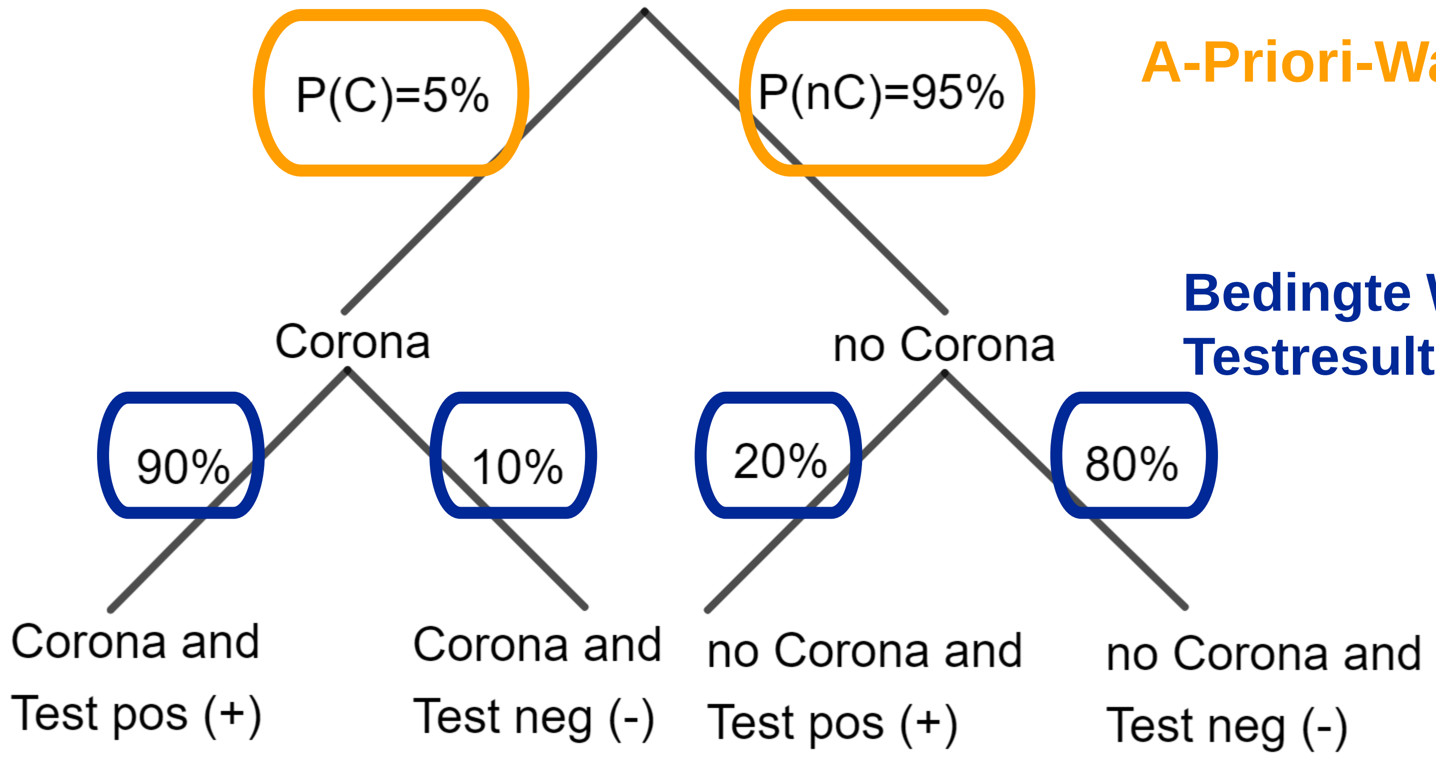
- 90% sensitiv: 90% der Leute, die Corona haben, erhalten ein positives Testresultat
- 80% spezifisch: 80% der Leute, die gesund sind, erhalten ein negatives Testresultat
- Außerdem hat ca. 5% der Bevölkerung Corona

Frage: Wahrscheinlichkeit, dass man Corona hat gegeben ein positives Testresultat?

MULTIVARIATE VERTEILUNGEN: BEDINGTE WAHRSCHEINLICHKEITEN

HOCHSCHULE HEILBRONN

Beispiel (Fortsetzung).



A-Priori-Wahrscheinlichkeiten

Bedingte Wahrscheinlichkeiten der Testresultate gegeben Corona / kein Corona

MULTIVARIATE VERTEILUNGEN: BEDINGTE WAHRSCHEINLICHKEITEN

HOCHSCHULE HEILBRONN

Beispiel (Fortsetzung).

Wir müssen folgende Wahrscheinlichkeiten unterscheiden:

> Wahrscheinlichkeit einer Corona-Infektion (C):

$$P(C) = 5\%$$

> Bedingte Wahrscheinlichkeit eines positiven Tests (+) gegeben man hat Corona (C):

$$P(+ \mid C) = 90\%$$

> Bedingte Wahrscheinlichkeit einer Corona-Infektion (C) gegeben ein positiver Test (+):

$$P(C \mid +) = ?$$ Diese möchten wir wissen

MULTIVARIATE VERTEILUNGEN: BEDINGTE WAHRSCHEINLICHKEITEN

HOCHSCHULE HEILBRONN

Beispiel (Fortsetzung).

- Bayes-Theorem:

$$P(Corona \mid +) = \frac{P(+ \mid Corona) \times P(Corona)}{P(+)}$$

MULTIVARIATE VERTEILUNGEN: BEDINGTE WAHRSCHEINLICHKEITEN

HOCHSCHULE HEILBRONN

Beispiel (Fortsetzung).

- Satz der totalen Wahrscheinlichkeit:

$$P(+) = P(+ \mid Corona) \times P(Corona) + P(+ \mid KeinCorona) \times P(KeinCorona)$$

MULTIVARIATE VERTEILUNGEN: BEDINGTE WAHRSCHEINLICHKEITEN

HOCHSCHULE HEILBRONN

Beispiel (Fortsetzung).

> Bayes-Theorem:

$$P(Corona \mid +) = \frac{\overbrace{\boxed{P(+ \mid Corona)}}^{90\%} \cdot \overbrace{\boxed{P(Corona)}}^{5\%}}{\underbrace{\boxed{P(+)}}_{23.5\%}} = 19.15\%$$

→ Die W'keit Corona zu haben bei einem positiven Test ist 19.15%

Beispiel. 200 Menschen. 47 (= 23.5%) werden positiv getestet → 10 (= 5%) haben Corona.

MULTIVARIATE VERTEILUNGEN: GEMEINSAME VERTEILUNG DISKRETER VARIABLEN

HOCHSCHULE HEILBRONN

$Gemeinsame\ Wahrscheinlichkeitsfunktion$ von 2 Zufallsvariablen X und Y

- gibt die gemeinsamen Wahrscheinlichkeiten an: $P(X = x, Y = y)$

MULTIVARIATE VERTEILUNGEN: GEMEINSAME VERTEILUNG DISKRETER VARIABLEN

HOCHSCHULE HEILBRONN

Beispiel. Betrachten Sie folgenden Datensatz:

X	1	1	0	1	0
Y	0	0	1	1	0

> Bestimmen Sie gemeinsame Verteilung von X und Y

MULTIVARIATE VERTEILUNGEN: STOCHASTISCHE UNABHÄNGIGKEIT

HOCHSCHULE HEILBRONN

Zwei Zufallsvariablen sind $stochastisch\ unabhängig$, wenn

> die Realisierung einer Zufallsvariable keinen Einfluss auf die W‘keiten der anderen hat

> Also: bedingte Wahrscheinlichkeit = unbedingte Wahrscheinlichkeit:

$$P(X = x \mid Y = y) = P(X = x)\ f\text{ü}r\ alle\ x, y.$$

> Einsetzen in Bayes-Theorem $P(X = x \mid Y = y) = \frac{P(X=x, Y=y)}{P(Y=y)}$:

$$\boldsymbol{P(X = x, Y = y) = P(X = x) \times P(Y = y)\ f\text{ü}r\ alle\ x, y}.$$

MULTIVARIATE VERTEILUNGEN: STOCHASTISCHE UNABHÄNGIGKEIT

Beispiel. Sind X und Y stochastisch unabhängig?

X	1	1	0	1	0
Y	0	0	1	1	0

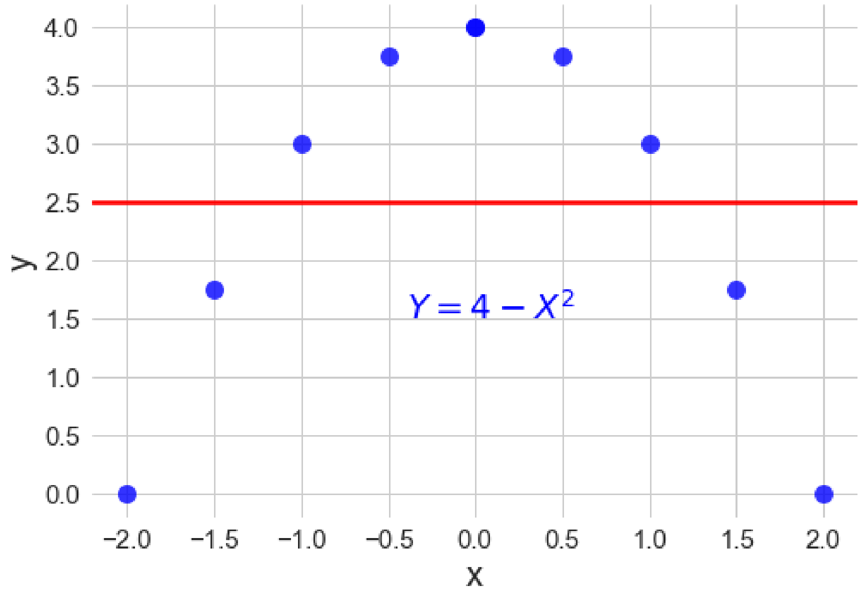
MULTIVARIATE VERTEILUNGEN: UNABHÄNGIGKEIT UND KORRELATION

HOCHSCHULE HEILBRONN

- Zwei Zufallsvariablen X und Y sind korreliert, wenn deren Kovarianz ungleich 0 ist
- **Korrelierte** Zufallsvariablen sind immer **stochastisch abhängig**
- **Stochastisch abhängige** Zufallsvariablen können aber **unkorreliert sind**

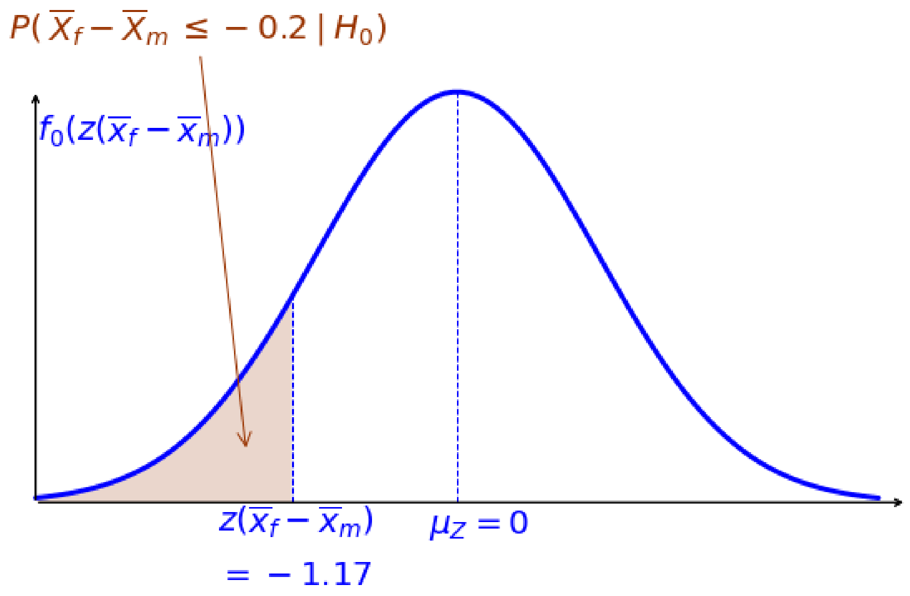
Beispiel.

- X und Y sind perfekt abhängig aber unkorreliert
- Erinnerung: Korrelation = Lineare Beziehung
- Nicht alle Beziehungen sind linear!



THE END!

Please refer any questions to:
Prof. Dr. Florian Kauffeldt
Faculty of International Business
florian.kauffeldt@hs-heilbronn.de



› INFERENZSTATISTIK

Statistik | Prof. Dr. Florian Kauffeldt

TEILGEBIETE DER STATISTIK

Statistik

Deskriptive Statistik

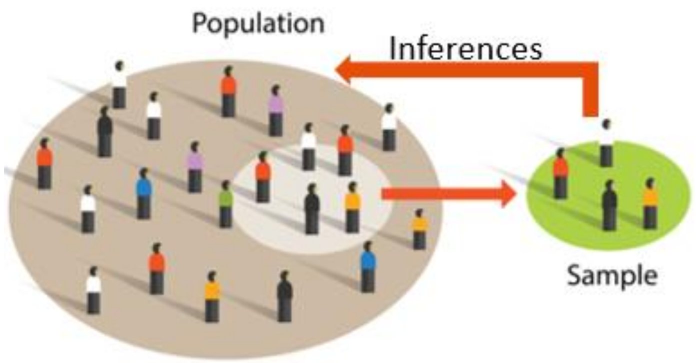
> Beschreibung von Daten

Inferenzstatistik

> Induktive Rückschlüsse basierend auf Daten

WAS IST INFERENZSTATISTIK?

> Inferenzstatistik: Rückschlüsse von Stichproben auf die Population



Beispiel. In einer Stichprobe verdienen Frauen im Schnitt 211 € mehr als Männer.

> Können wir daraus schließen, dass Frauen im Allgemeinen mehr verdienen?

Solche und ähnliche Fragen werden mit inferenzstatistischen Methoden beantwortet.

INFERENZSTATISTIK
INHALT

1. Statistische Schätzer

> Punktschätzer und Schätzwertverteilung

> Zentraler Grenzwertsatz

> Intervallschätzer

2. Hypothesentests

> Komponenten

> Parametrische Tests (Ein- und Zweistichproben Z-Test, Korrelationstest)

> Nicht-Parametrische Tests (Mann-Whitney U, Spearman Rho, Chi2)

› STATISTISCHE SCHÄTZER

WAS IST EIN SCHÄTZER?

HOCHSCHULE HEILBRONN

Täglich werden Entscheidungen basierend auf Daten getroffen:

> Die Regierung möchte Güterströme von Auslandsmärkten vorhersagen

> Aktienhändler möchten Aktienentwicklungen vorhersagen

> Konsumenten möchten eine Einschätzung über Güter erlangen

I.d.R. sind die Verteilungsparameter (Mittelwert, Standardabweichung,…) unbekannt

$Schätzer$ = Regel zur Schätzung eines Parameters basierend auf einer Stichprobe

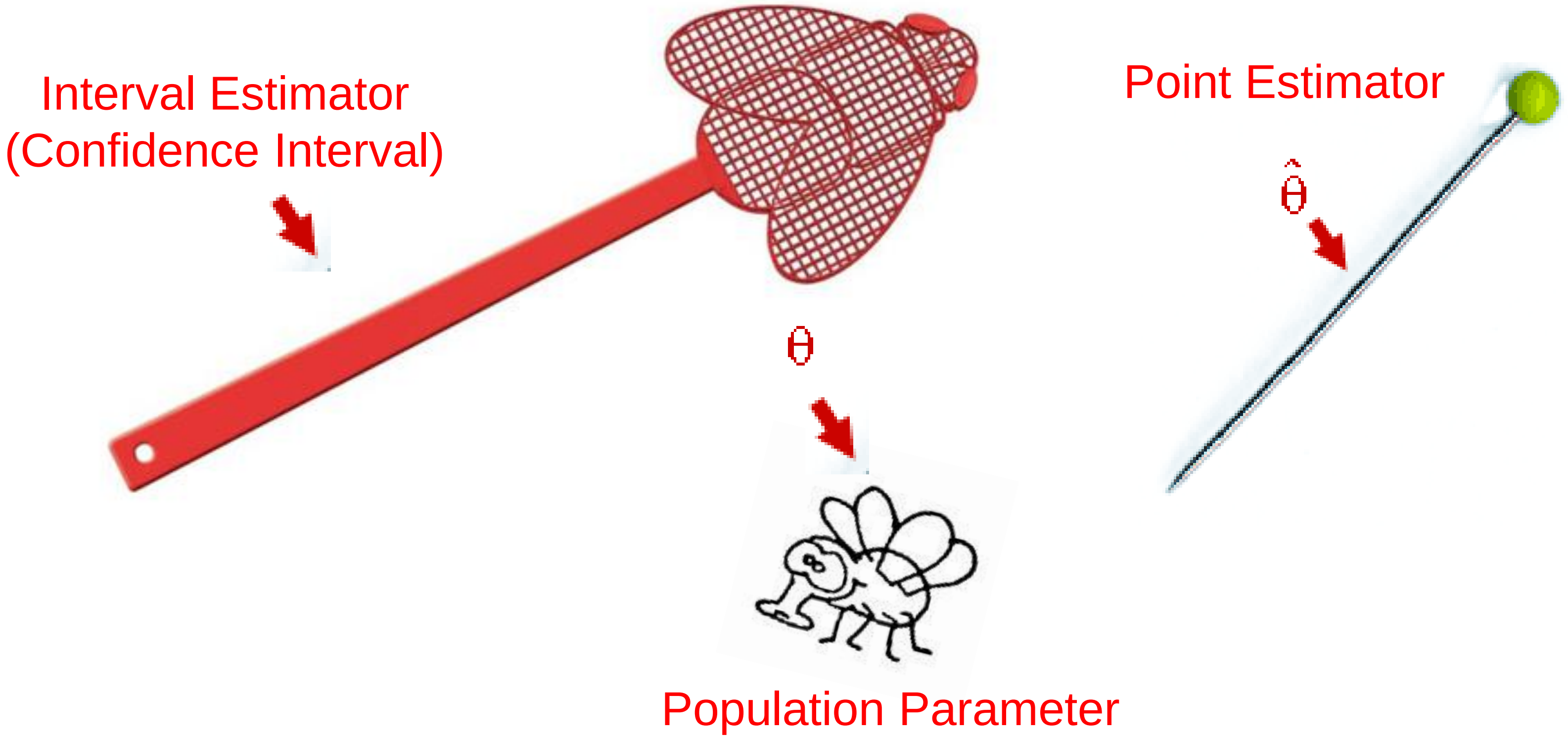
ARTEN VON SCHÄTZERN

HOCHSCHULE HEILBRONN

Schätzer

Punktschätzer	Intervallschätzer (Konfidenzintervall)
> Einzige Zahl als Schätzung Beispiel. Durchschnittsgröße = 176 cm	> Intervall: Obere und untere Schätzgrenze Beispiel. Durchschnittsgröße liegt zwischen 175 cm und 180 cm

ARTEN VON SCHÄTZERN



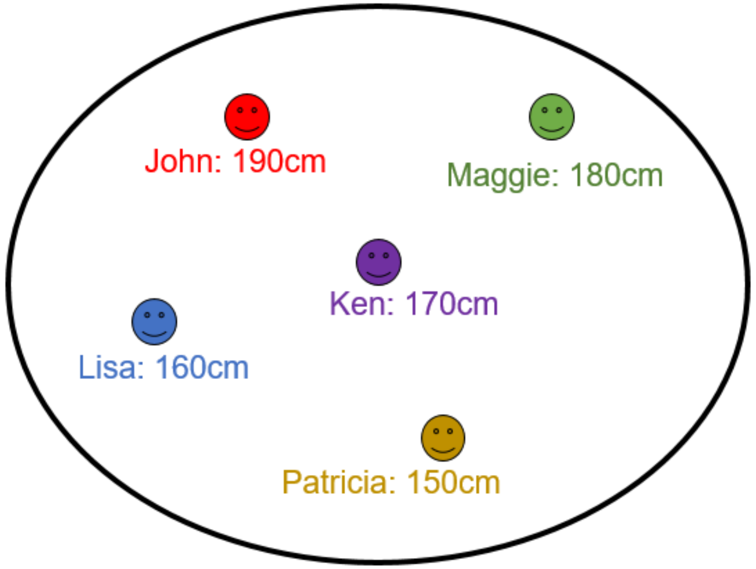
SCHÄTZER: NOTATION

Notation. Schätzwerte werden mit überdachten griechischen Buchstaben $\widehat{\ }$ bezeichnet.

Beispiel. $\hat{\mu}$ ist der Schätzwert für den Erwartungswert μ.

PUNKTSCHÄTZER: MITTELWERT

Beispiel. Betrachten Sie folgende Population:



> Der Populationsmittelwert ist:

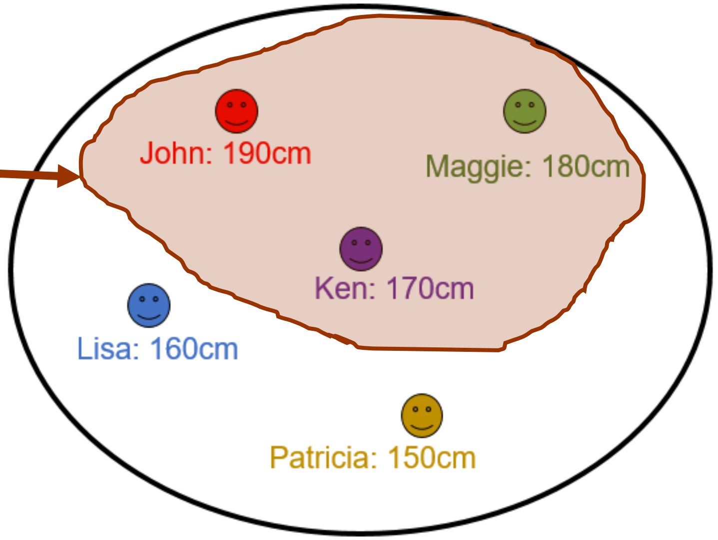
$$\mu = \frac{150 + 160 + 170 + 180 + 190}{5} = \mathbf{170}$$

PUNKTSCHÄTZER: MITTELWERT

HOCHSCHULE HEILBRONN

Beispiel (Fortsetzung). Der Mittelwert $\mu = 170$ sei **unbekannt.**

- Schätzer Populationsmittelwert (μ) = Stichprobenmittelwert ($\bar{x}$)
- Ziehe eine Stichprobe, z.B. Größe $n = 3$
- Dann Schätzwert Populationsmittelwert:



$$\hat{\mu} = \bar{x} = \frac{170 + 180 + 190}{3} = 180$$

VERTEILUNG DER SCHÄTZWERTE

HOCHSCHULE HEILBRONN

> Unsere Stichprobe überschätzt den wahren Mittelwert

> Andere Stichproben unterschätzen den wahren Mittelwert

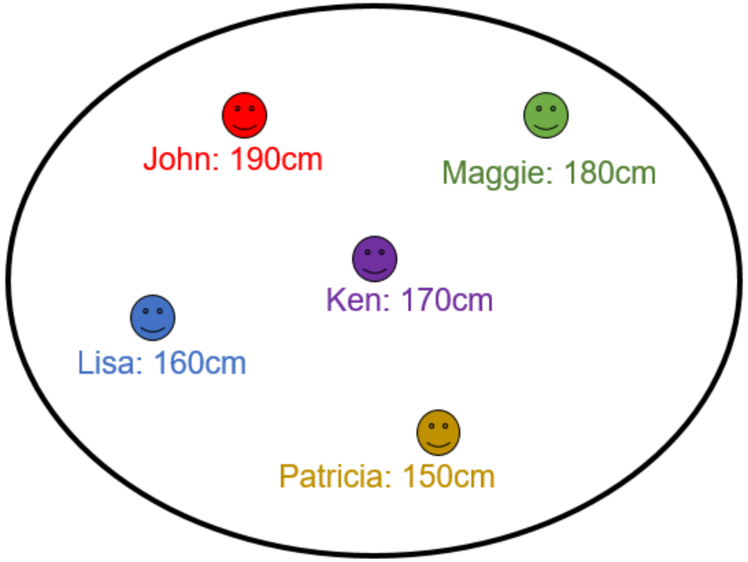
> Können wir etwas über den möglichen Schätzfehler sagen?

Ja → Wir müssen die Verteilung der Schätzwerte untersuchen

VERTEILUNG DER SCHÄTZWERTE: ANZAHL STICHPROBEN

HOCHSCHULE HEILBRONN

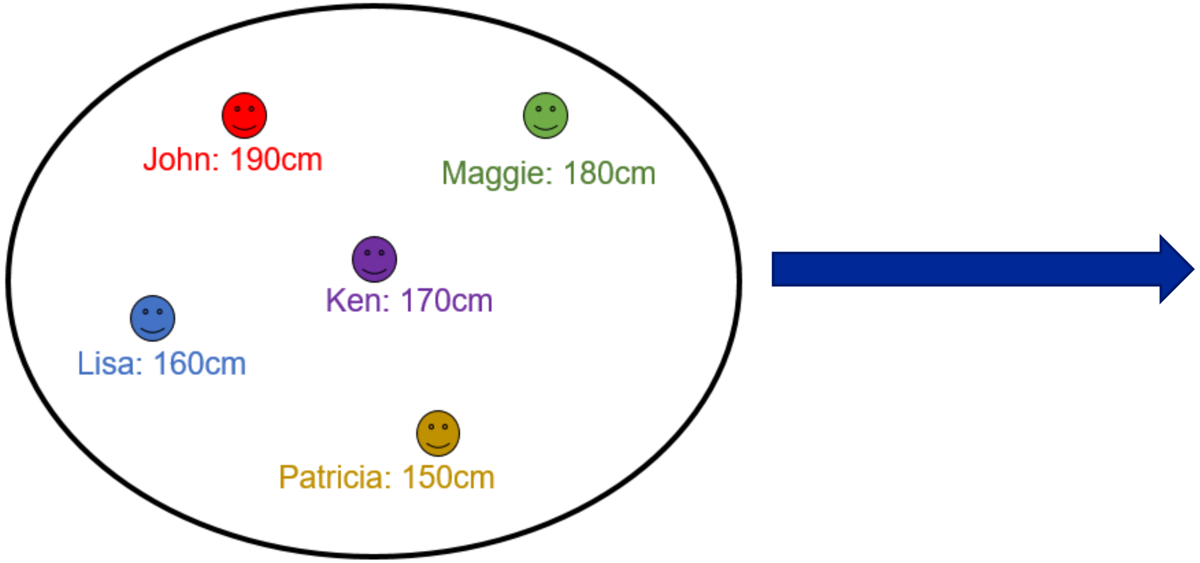
Beispiel.



> Wie viele unterschiedliche Stichproben mit $n = 3$ können wir ziehen?

$$\binom{Populationsgröße}{Stichprobengröße} = \binom{5}{3} = \frac{5!}{3!\,2!} = \frac{5 \cdot 4}{2} = 10$$

VERTEILUNG DER SCHÄTZWERTE: WAHRSCHEINLICHKEIT EINER STICHPROBE



	Verteilung Stichproben	
i	Stichprobe i	$P(i)$
1	150, 160, 170	10%
2	150, 160, 180	10%
3	150, 160, 190	10%
4	150, 170, 180	10%
5	150, 170, 190	10%
6	150, 180, 190	10%
7	160, 170, 180	10%
8	160, 170, 190	10%
9	160, 180, 190	10%
10	170, 180, 190	10%

Zufallsstichprobe = Jedes Element der Population wird mit gleicher W'keit ausgewählt

VERTEILUNG DER SCHÄTZWERTE: SCHÄTZER MITTELWERT

HOCHSCHULE HEILBRONN

Stichprobenmittelwert = Schätzer Populationsmittelwert

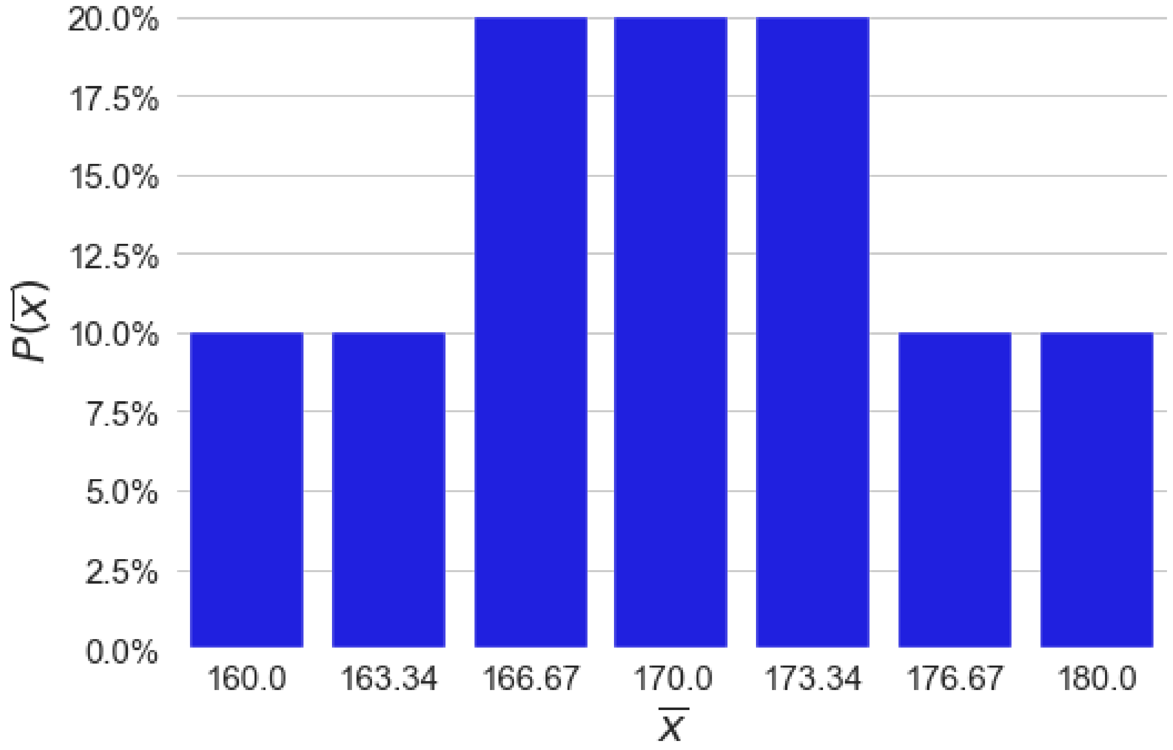
	Verteilung Stichproben		
i	**Stichprobe i**	$P(i)$	**Mittelw. $\bar{x}_i$**
1	150, 160, 170	10%	160
2	150, 160, 180	10%	163.34
3	150, 160, 190	10%	166.67
4	150, 170, 180	10%	166.67
5	150, 170, 190	10%	170
6	150, 180, 190	10%	173.34
7	160, 170, 180	10%	170
8	160, 170, 190	10%	173.34
9	160, 180, 190	10%	176.67
10	170, 180, 190	10%	180

Verteilung Stichprobenmittelwerte	
$\bar{x}$	$P(\bar{x})$
160	10%
163.34	10%
166.67	20%
170	20%
173.34	20%
176.67	10%
180	10%

7 unterschiedliche Stichprobenmittelwerte

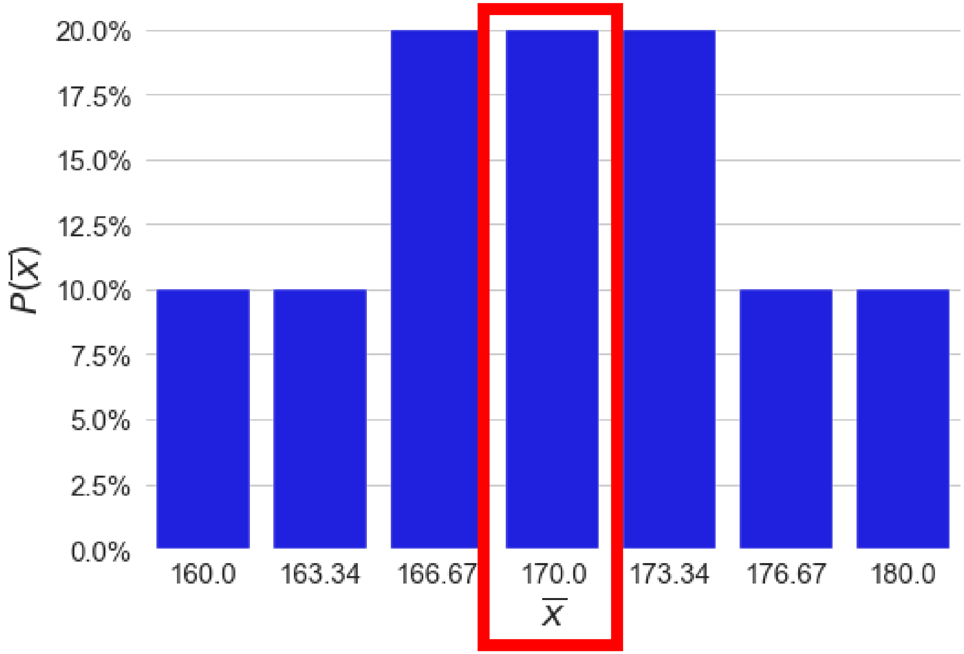
VERTEILUNG DER SCHÄTZWERTE: SCHÄTZER MITTELWERT

Verteilung Stichprobenmittelwerte	
$\overline{x}$	$P(\overline{x})$
160	10%
163.34	10%
166.67	20%
170	20%
173.34	20%
176.67	10%
180	10%



VERTEILUNG DER SCHÄTZWERTE: SCHÄTZER MITTELWERT

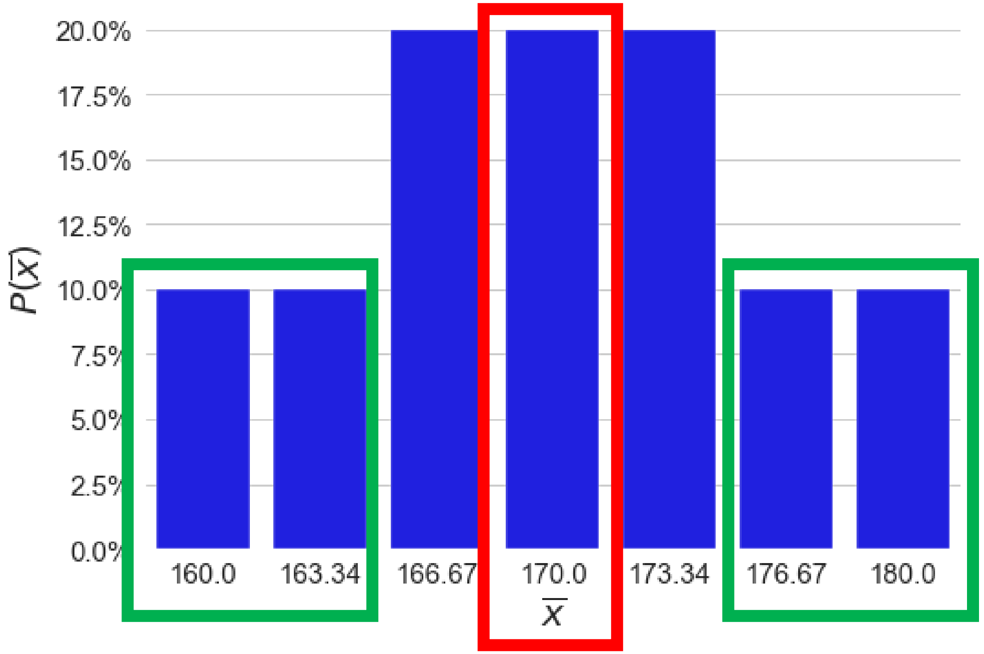
HOCHSCHULE HEILBRONN



20% Wahrscheinlichkeit dass $Schätzwert\ \bar{x} = Populationsmittelwert\ \mu = 170$

VERTEILUNG DER SCHÄTZWERTE: SCHÄTZER MITTELWERT

HOCHSCHULE HEILBRONN

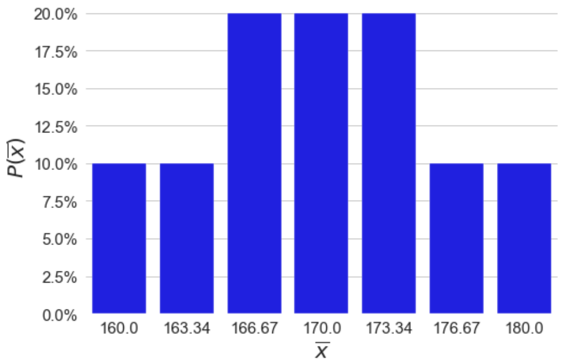


> Wahrscheinlichkeit, dass der Schätzfehler mindestens 5cm ist?

40%

VERTEILUNG DER SCHÄTZWERTE: MITTELWERT UND STANDARDABWEICHUNG

Verteilung der Stichprobenmittelwerte:



- Mittelwert:

$$\mu_{\bar{X}} = \text{Populationsmittelwert } \mu$$

- Standardabweichung = Standardfehler:

$$se_{\bar{X}} = \frac{Populationsstandardabweichung}{\sqrt{Stichprobengröße}} = \frac{\sigma}{\sqrt{n}}$$

Je höher der Standardfehler, desto wahrscheinlicher ist ein großer Schätzfehler

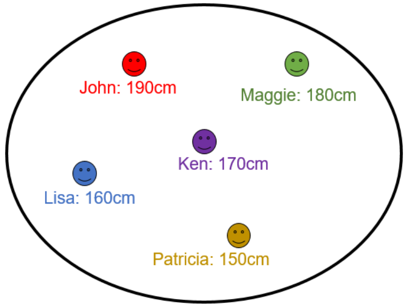
PUNKTSCHÄTZER UND STANDARDFEHLER IHRER VERTEILUNG

HOCHSCHULE HEILBRONN

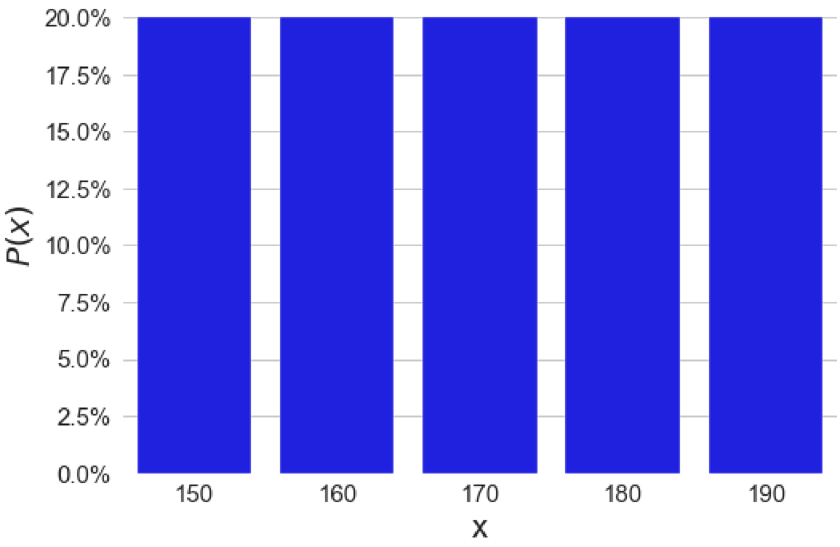
Parameter	Punktschätzer	Standardfehler
Mittelwert μ	Stichprobenmittelwert $\bar{x}$	$se_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$
Differenz Mittelwerte $\mu_X - \mu_Y$	Differenz Stichprobenmittelwerte $\bar{x} - \bar{y}$	$se_{\bar{X}-\bar{Y}=}\sqrt{se_{\bar{X}}^2 + se_{\bar{Y}}^2}$
Varianz σ	Bessel-korrigierte Stichprobenvarianz: $\hat{\sigma}^2 = Stichprobenvarianz \cdot \frac{n}{n-1}$	$se_{\hat{\sigma}^2} = \sqrt{\frac{2}{n-1}} \times \hat{\sigma}^2$
Korrelationskoeffizient ρ_{XY}	Fisher-Transformation: $\hat{\rho}_{XY}^f = \text{arctanh}(r_{xy}) = \frac{1}{2}\ln\left(\frac{1+r_{xy}}{1-r_{xy}}\right)$ > r_{xy} = Stichprobenkorrelationskoeffizient	$se_f = \sqrt{\frac{1}{n-3}}$

ZENTRALER GRENZWERTSATZ

Beispiel 1:

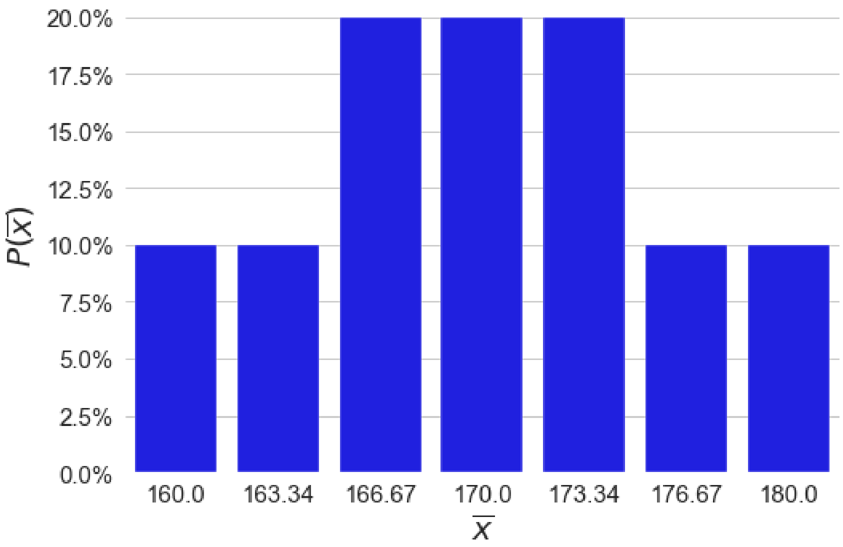


Population:
Gleichverteilung



Stichproben
$(n = 3)$

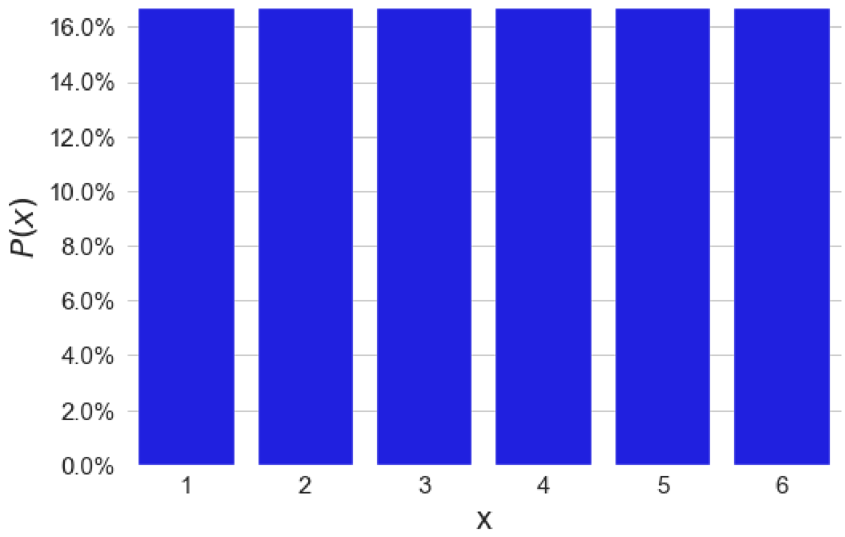
Verteilung Stichprobenmittelwerte:
Keine Gleichverteilung



ZENTRALER GRENZWERTSATZ

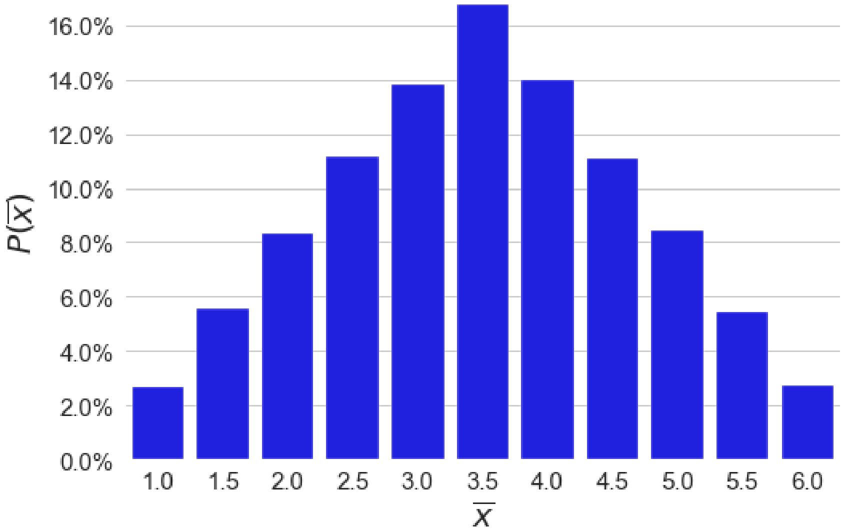
Beispiel 2. Zweimaliger Würfelwurf: Stichprobenmittelwert = Durchschnitt der Augenzahlen.

Population:
Gleichverteilung



100'000
Stichproben

Verteilung Stichprobenmittelwerte:
Keine Gleichverteilung



ZENTRALER GRENZWERTSATZ

Zentraler Grenzwertsatz
Summe und Mittelwert von n unabhängig und identisch verteilten Zufallsvariablen sind approximativ normalverteilt, wenn n hinreichend groß ist.

Implikationen:

- Normalverteilung Population → Stichprobenverteilung normal (n egal)
- Keine Normalverteilung Population → Stichprobenverteilung normal (wenn n groß)

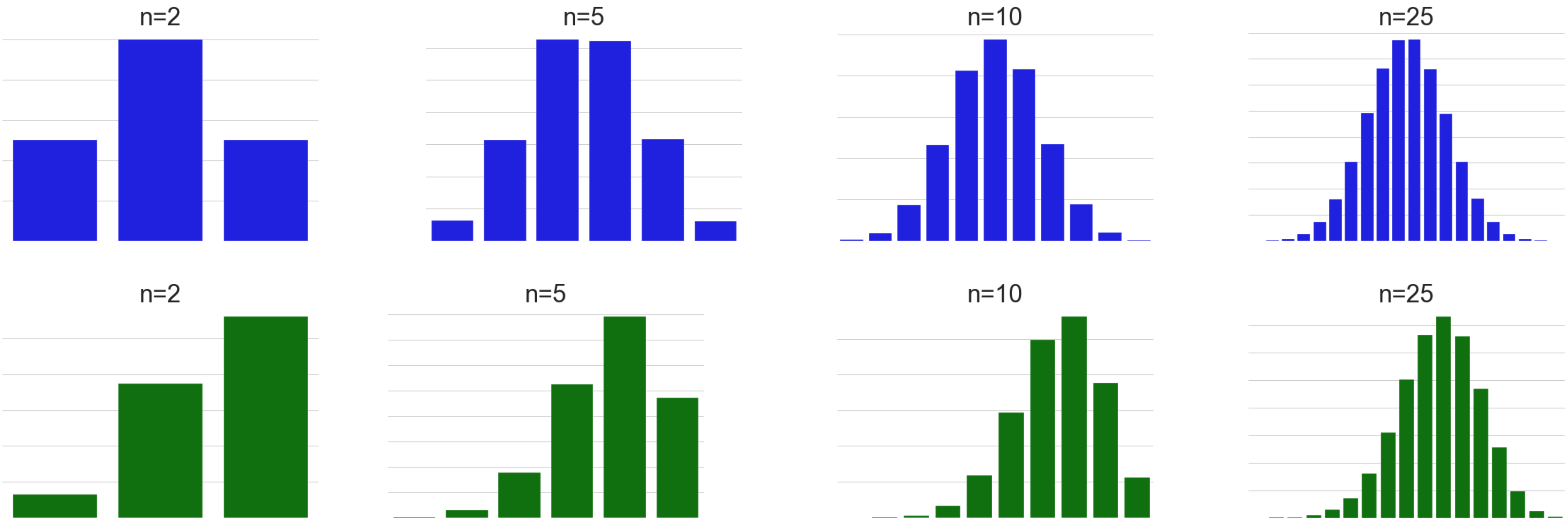
Daumenregel: Die Stichprobenverteilung ist annähernd normalverteilt, wenn

$$Stichprobengröße \geq 30$$

ZENTRALER GRENZWERTSATZ

Illustration. Populationen: Bernoulli-verteilt mit $\pi = 50\%$ und $\pi = 75\%$

> Relative Häufigkeitsverteilung des Stichprobenmittels bei 100'000 Stichproben:



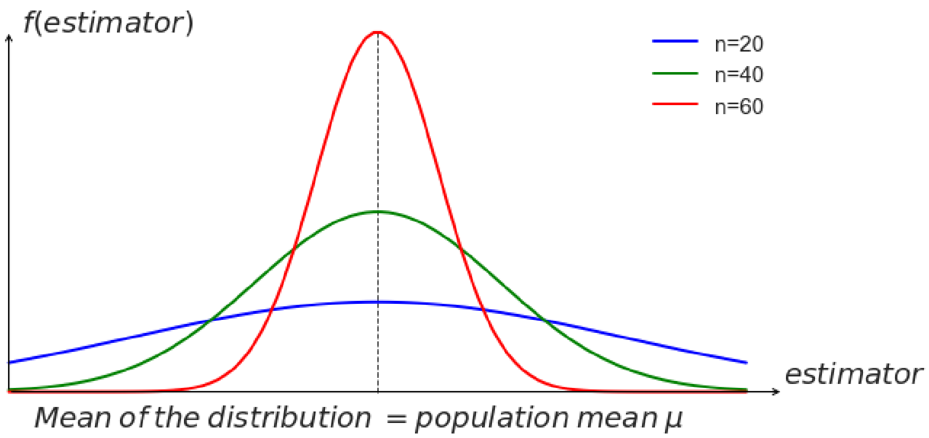
ZENTRALER GRENZWERTSATZ: SCHÄTZFEHLER

$$Schätzfehler \;=\; Schätzwert - Wahrer\ Parameter$$

> Mittelwert:

$$Schätzfehler = Schätzer - \mu = \text{Stichprobenmittelwert } (\bar{x}) - \mu$$

> Welche der folgenden Verteilungen ergibt im Schnitt den kleinsten Schätzfehler?



ZENTRALER GRENZWERTSATZ: SCHÄTZFEHLER

Beispiel. Körpergröße ist verteilt mit Mittelwert 167cm (Standardabweichung = 10cm).

> Stichprobe mit n = 36 Personen

> Wahrscheinlichkeit, dass der Mittelwert um mindestens 3cm überschätzt wird?

ZENTRALER GRENZWERTSATZ: SCHÄTZFEHLER

Beispiel. Monatliches Einkommen von Männern (M) und Frauen (W).

> Mittelwerte: $\mu_M = 3'201€$ und $\mu_W = 3'001€$

> Standardabweichungen: $\sigma_M = 420€$ und $\sigma_W = 305€$

> Stichprobe mit 36 Frauen und 49 Männer

> Wahrscheinlichkeit, dass Durchschnittseinkommen Frauen $\geq$ Männer?

INTERVALLSCHÄTZER: MARGIN OF ERROR (MOE)

HOCHSCHULE HEILBRONN

- Schätzfehler Mittelwert = Stichprobenmittelwert – Populationsmittelwert (unbekannt)
- Z-Wert:

Zufallsvariable des Schätzfehlers

$$Z(\bar{X}) = \frac{\bar{X} - \mu}{Standardfehler\ (se)}$$

- Also:

$$Schätzfehler\ (moe) = Z \cdot se$$

- Standardfehler ist bekannt, Z muss festgelegt werden

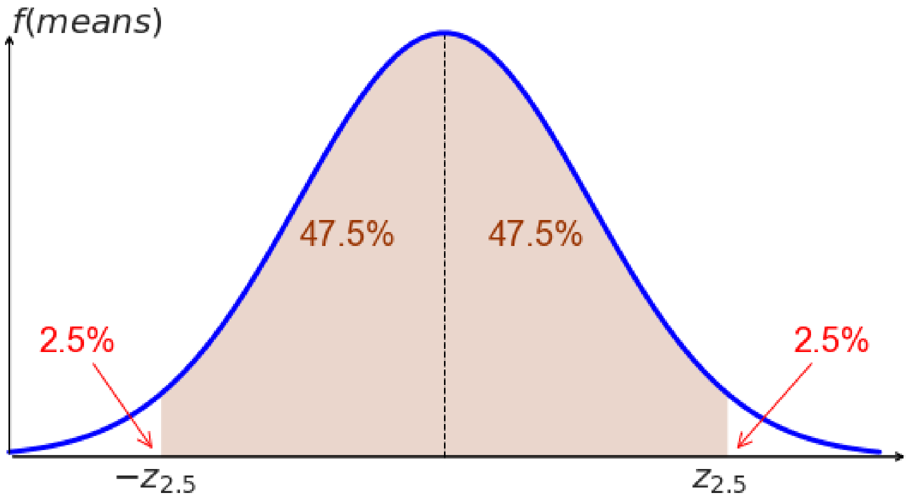
INTERVALLSCHÄTZER: KONFIDENZNIVEAU z FESTLEGEN

HOCHSCHULE HEILBRONN

> Wir können ein Konfidenzniveau $(1-\alpha)$ festlegen

> $(1-\alpha)$ = 95% bedeutet z.B., dass der Schätzfehler zu 95% in dem Bereich liegt

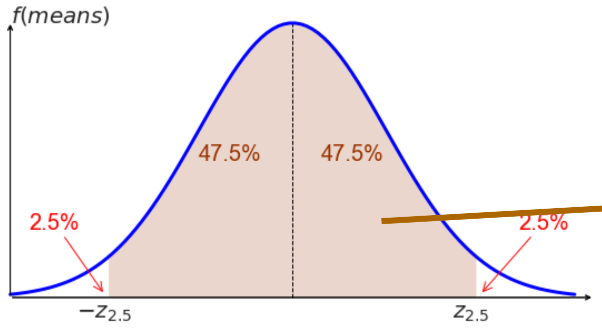
> Wir lassen eine Fehlerwahrscheinlichkeit von $\alpha/2 = 2.5\%$ auf jeder Seite zu:



BEREICH $z_{-\alpha/2}$ BIS $z_{\alpha/2}$

Beispiel. $\alpha = 5\% \rightarrow z_{2.5}$?

> Nun: Z-Wert für Fläche finden!



> Also:

$$z_{2.5} = 1.96$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.10	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.20	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.30	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.40	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.50	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.60	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.70	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.80	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.90	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.00	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.10	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.20	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.30	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.40	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.50	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.60	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.70	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.80	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.90	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.00	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.10	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.20	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.30	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.40	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.50	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.60	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.70	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.80	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.90	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.00	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

INTERVALLSCHÄTZER: KONFIDENZINTERVALLE

HOCHSCHULE HEILBRONN

1. Fixiere ein Konfidenzniveau $(1 - \alpha)$ und bestimme hierfür den Wert $z_{\alpha/2}$

2. Bestimme margin of error:

$$moe = z_{\alpha/2} \cdot Standardfehler$$

3. $(1 - \alpha)$-Konfidenzintervall:

$$Punktschätzer \pm moe$$

INTERVALLSCHÄTZER: KONFIDENZINTERVALLE

HOCHSCHULE HEILBRONN

Beispiel. Zahlungsbereitschaft von Kunden für eine Übernachtung im Hotel bestimmen.

> In einer Stichprobe von n = 39 Befragten liegt die Zahlungsbereitschaft bei 86 €

> Standardabweichung = 26 €

a) Bestimmen Sie das 99%-Konfidenzintervall

Hausaufgabe.

b) Bestimmen Sie das 90%-Konfidenzintervall

Lösung: 90%-Konfidenzintervall = $[79, 93]$

INTERVALLSCHÄTZER: KONFIDENZINTERVALLE

HOCHSCHULE HEILBRONN

Beispiel. Wir möchten den durchschnittlichen IQ von Studierenden der Hochschule Heilbronn schätzen.

> In einer Stichprobe mit 47 Studierenden ist der durchschnittliche IQ 102
> Standardabweichung = 13

a) Bestimmen Sie das 90%-Konfidenzintervall

Hausaufgabe.

b) Bestimmen Sie das 95%-Konfidenzintervall

Lösung: 95%-Konfidenzintervall = $[98.28, 105.72]$

INTERVALLSCHÄTZER: DESIDERATA

Optimales Konfidenzintervall: So **schmal wie möglich** und **Konfidenzniveau nahe 1**

- Je größer das Konfidenzniveau, desto sicherer ist wahrer Parameter enthalten
- Je enger das Konfidenzintervall, desto exakter ist die Schätzung
- Je größer die Stichprobe, desto enger ist das Intervall

Aber:

- Je höher das Konfidenzniveau, desto breiter das Intervall

INTERVALLSCHÄTZER: OPTIMALE STICHPROBENGRÖẞE

HOCHSCHULE HEILBRONN

Frage wenn Daten erhoben werden → Welche Stichprobengröße?

> Frage:

Wie akkurat sollen die Schätzungen sein?

> Wir können den moe benutzen, um diese Frage zu beantworten

INTERVALLSCHÄTZER: OPTIMALE STICHPROBENGRÖßE

HOCHSCHULE HEILBRONN

Beispiel. Sie besitzen eine große Brauerei und möchten Ihre Tagesproduktion schätzen.

> Ihre durchschnittliche Tagesproduktion der letzten 30 Tage ist 871 Kubikmeter Bier

> Standardabweichung: 21

Wie viele Tage müssen Sie die Tagesproduktion beobachten, um die langfristige Tagesproduktion mit 95% zu schätzen bei einem moe von 4 Kubikmetern?

HYPOTHESEN-TESTS

WARUM HYPOTHESENTESTS?

In einer Stichprobe mit 35 Frauen und 44 Männer ist das Durchschnittseinkommen der Frauen um 200 € niedriger als das der Männer

> Ist dies ein Beleg dafür, dass Frauen allgemein im Schnitt weniger verdienen?

Der Umsatz einer Firma war im letzten Jahr 300'000 € höher als der Durchschnitt ihres langfristigen jährlichen Umsatzes

> Ist dies ein Beleg dafür, dass sich die Nachfrage für die Produkte erhöht hat?

Derartige Fragen werden anhand von Hypothesentests beantwortet

HYPOTHESENTEST: KOMPONENTEN

HOCHSCHULE HEILBRONN

1. Einem Hypothesenpaar:
   - $Alternativhypothese\ H_1$ = die Hypothese, welche man testen möchte
   - $Nullhypothese\ H_0$ = Gegenteil Alternativhypothese
2. Einer Teststatistik
   - Zahl, die auf Basis einer Stichprobe berechnet wird
3. Einer Entscheidung
   - $Ablehnung\ Nullhypothese$ = Beleg für Alternativhypothese
   - $Keine\ Ablehnung\ Nullhypothese$ = Kein Beleg für Alternativhypothese

HYPOTHESENTEST: NULL- UND ALTERNATIVHYPOTHESE

HOCHSCHULE HEILBRONN

Beispiel. Wir möchten untersuchen, ob Frauen weniger verdienen als Männer.

- Alternativhypothese: $Durchschnittseinkommen_{Frauen} < Durchschnittseinkommen_{Männer}$

Wir untersuchen nun wie wahrscheinlich das Gegenteil ist:

- Nullhypothese: $Durchschnittseinkommen_{Frauen} \geq Durchschnittseinkommen_{Männer}$

HYPOTHESENTEST: NULL- UND ALTERNATIVHYPOTHESE

HOCHSCHULE HEILBRONN

Hypothesen

Alternativhypothese H_1	Nullhypothese H_0
> Hypothese, die getestet werden soll > Postuliert üblicherweise ein Resultat	> Exaktes Gegenteil der Alternativhypothese > Verneint das Resultat

HYPOTHESENTEST: NULL- UND ALTERNATIVHYPOTHESE

HOCHSCHULE HEILBRONN

Beispiele.

Wir möchten wissen…	Alternativhypothese (H_1)	Nullhypothese (H_0)
ob Absolventen der HS Heilbronn im Schnitt (μ_{Alter}) älter sind als 26 (μ_0)	$\mu_{Alter} > 26$	$\mu_{Alter} \leq 26$
ob sich der Durchschnitts-IQ (μ_{IQ}) von Studierenden verändert hat (langfristiger Durchschnitt: $\mu_0 = 101$)	$\mu_{IQ} \neq 101$	$\mu_{IQ} = 101$
ob die Nachfrage (μ_{Nach}) für das Produkt einer Firma abgenommen hat (langfristiger Durchschnitt: $\mu_0 = 41'000$)	$\mu_{Nach} < 41'000$	$\mu_{Nach} \geq 41'000$

HYPOTHESENTEST: NULL- UND ALTERNATIVHYPOTHESE

HOCHSCHULE HEILBRONN

Beispiel. In einer Stichprobe mit 35 Frauen und 44 Männer ist das monatliche Durchschnittseinkommen der Frauen $2.9\,T€$ $(Varianz = 0.7)$ und das der Männer $3.1\,T€$ $(Varianz = 0.4)$.

Ist dies ein Beleg dafür ist, dass Frauen im Schnitt weniger verdienen?

a) Schreiben Sie die Null- und die Alternativhypothese zu dieser Fragestellung auf.

> Alternativhypothese:

$$Durchschnittseinkommen\ Pop.Frauen\ \mu_f < Durchschnittseinkommen\ Pop.Männer\ \mu_m$$

> Nullhypothese:

$$Durchschnittseinkommen\ Pop.Frauen\ \mu_f \geq Durchschnittseinkommen\ Pop.Männer\ \mu_m$$

HYPOTHESENTEST: NULL- UND ALTERNATIVHYPOTHESE

HOCHSCHULE HEILBRONN

Null- und Alternativhypothese repräsentieren **zwei unterschiedliche Welten:**

Welt 0 (Nullhypothese H_0):

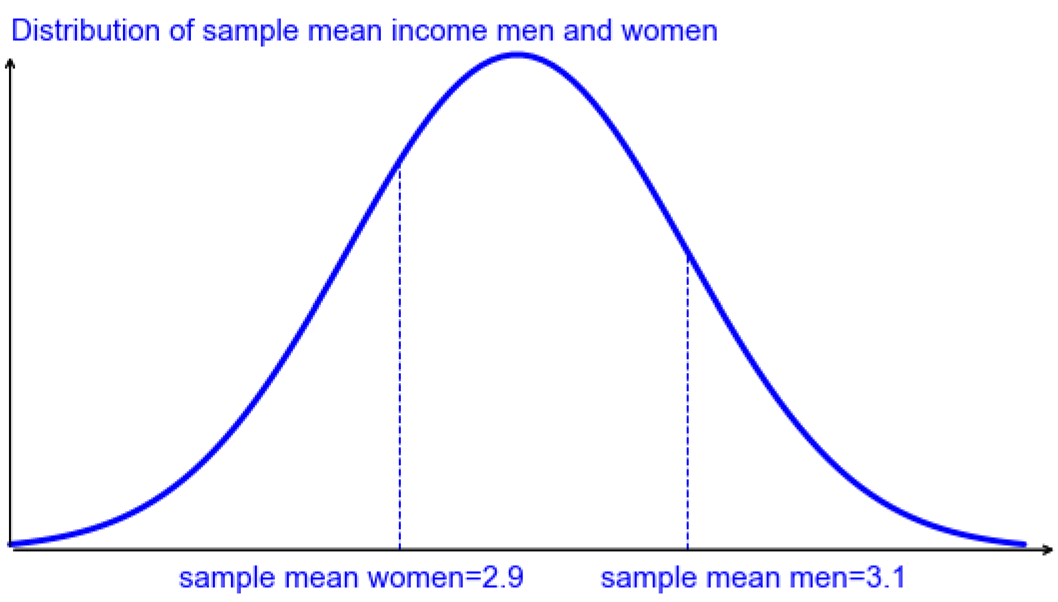
Frauen verdienen nicht weniger.

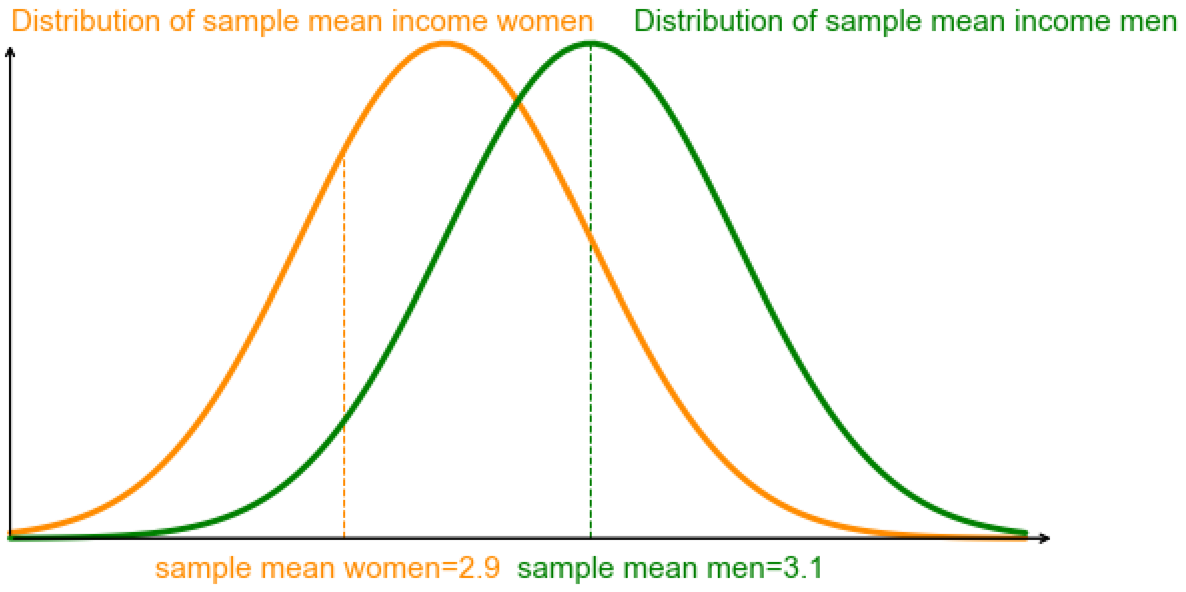
> Gleiche Einkommensverteilung

Welt 1 (Alternativhypothese H_1):

Frauen verdienen weniger.

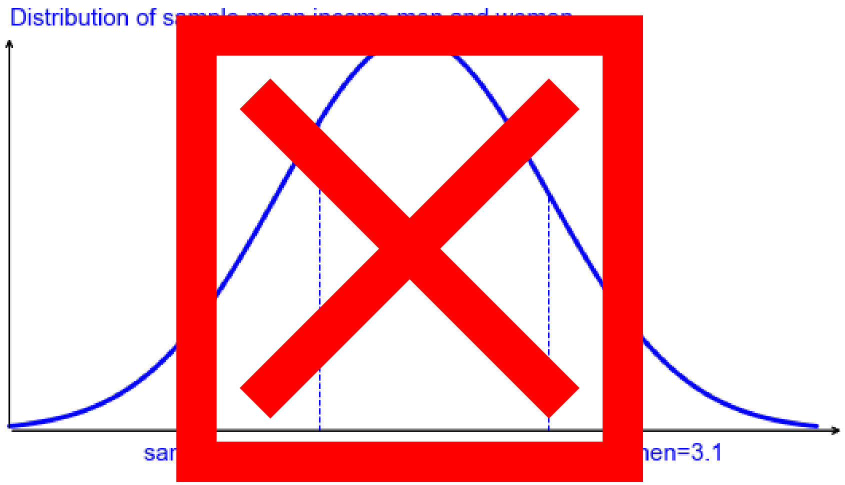
> Unterschiedliche Einkommensverteilungen.





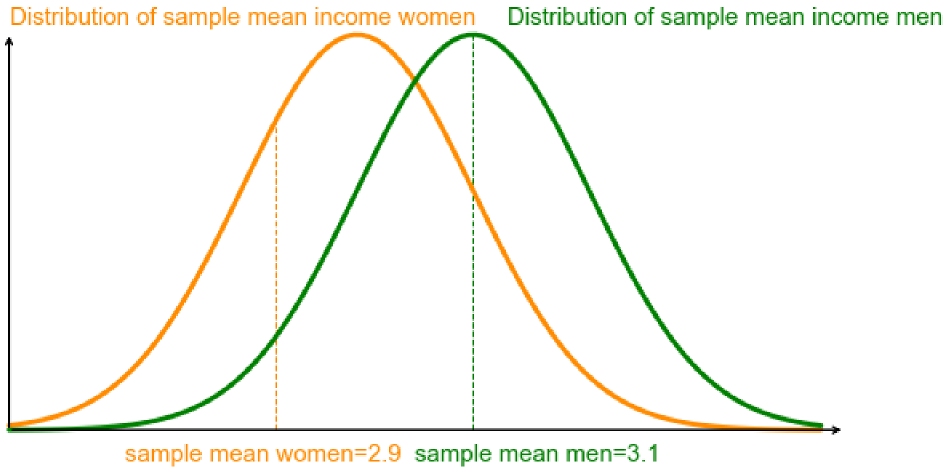
HYPOTHESENTEST: NULL- UND ALTERNATIVHYPOTHESE

HOCHSCHULE HEILBRONN



Welt 0 (Nullhypothese H_0)

In welcher Welt leben wir?



Welt 1 (Alternativhypothese H_a)

> Ein Hypothesentest sagt uns, ob wir die Nullhypothese ablehnen können

> Wenn wir die Nullhypothese ablehnen können, dann ist der Test erfolgreich

HYPOTHESENTEST: TESTSTATISTIK

HOCHSCHULE HEILBRONN

Welches Kriterium bestimmt, ob wir die Nullhypothese (Welt 0) ablehnen können?

Schritt 1. Berechne Teststatistik:

$$Teststatistik = \frac{Punktschätzer\ - Populationsmittelwert\ Welt\ 0}{Standardfehler}$$

HYPOTHESENTEST: TESTSTATISTIK

Beispiel (Fortsetzung). In einer Stichprobe mit 35 Frauen und 44 Männer ist das monatliche Durchschnittseinkommen der Frauen $2.9\,T€$ $(Varianz = 0.7)$ und das der Männer $3.1\,T€$ $(Varianz = 0.4)$.

Ist dies ein Beleg dafür ist, dass Frauen im Schnitt weniger verdienen?

a) Schreiben Sie die Null- und die Alternativhypothese zu dieser Fragestellung auf.

b) Führen Sie einen Test durch, um zu überprüfen, ob der Unterschied signifikant ist.

HYPOTHESENTEST: P-WERT

Welches Kriterium bestimmt, ob wir die Nullhypothese (Welt 0) ablehnen können?

Schritt 1. Berechne Teststatistik

Schritt 2. Berechne p-Wert

$$p - Wert = P(Teststatistik = wie\ beobachtet\ oder\ extremer \mid Nullhypothese\ wahr)$$

HYPOTHESENTEST: P-WERT EINSEITIG UND ZWEISEITIG

HOCHSCHULE HEILBRONN

- $Einseitige\ Tests$ haben eine Richtung: mehr/weniger, höher/niedriger, …
- $Zweiseitige\ Tests$ haben keine Richtung: Unterschied ja/nein

Beispiel. Überprüfen Sie, ob die folgenden Tests einseitig oder zweiseitig sind:

Wir möchten wissen…	Alternativhypothese (H_1)	Nullhypothese (H_0)
ob Absolventen der HS Heilbronn im Schnitt (μ_{Alter}) älter sind als 26 (μ_0)	$\mu_{Alter} > 26$	$\mu_{Alter} \leq 26$
ob sich der Durchschnitts-IQ (μ_{IQ}) von Studierenden verändert hat (langfristiger Durchschnitt: $\mu_0 = 101$)	$\mu_{IQ} \neq 101$	$\mu_{IQ} = 101$
ob die Nachfrage (μ_{Nach}) für das Produkt einer Firma abgenommen hat (langfristiger Durchschnitt: $\mu_0 = 41'000$)	$\mu_{Nach} < 41'000$	$\mu_{Nach} \geq 41'000$

HYPOTHESENTEST: P-WERT EINSEITIG UND ZWEISEITIG

Sei $P_0 = P(X > Betrag(Teststatistik) \mid H_0\ wahr)$, dann

$$p - Wert\ einseitiger\ Test = P_0$$

$$p - Wert\ zweiseitiger\ Test = 2 \times P_0$$

HYPOTHESENTEST: P-WERT

Beispiel (Fortsetzung).

b) Führen Sie einen Test durch, um zu überprüfen, ob der Unterschied signifikant ist.

Schritt 1. Teststatistik = -1.17 ✓

Schritt 2. Berechne p-Wert.

$$p - Wert = P(X > 1.17) = 0.5 - 0.379 = 12.1\%$$

HYPOTHESENTEST: ENTSCHEIDUNG

Was sagt uns p-Wert = 12.1%?

> Wenn p-Wert $\leq$ konventionelle Obergrenzen → Ablehnung Nullhypothese

p-Wert $\leq$ Obergrenze α (*)	Interpretation
5%	Statistisch signifikant
1%	Stark statistisch signifikant

(*) Die Obergrenze α heißt *Signifikanzniveau*

> p-Wert = 12.1% > 5% → der Einkommensunterschied ist nicht statistisch signifikant

> Nullhypothese kann nicht abgelehnt werden → keine Evidenz für einen Unterschied

HYPOTHESENTEST: EFFEKTSTÄRKE

> Signifikanz (p-Wert) eines Tests → Kann der beobachtete Effekt verallgemeinert werden

> $Effektstärke$ → Wie stark ist der Effekt

HYPOTHESENTEST: EFFEKTSTÄRKE

Test	Berechnung Effektstärke	Interpretation
Einstichproben-Z	$Cohens\ d = \frac{MW_{Stichprobe} - MW_{Pop\ Nullwelt}}{\hat{\sigma}}$	> Schwach: $\lvert d \rvert$ von 0,2 bis 0,5 > Mittel: $\lvert d \rvert$ von 0,5 bis 0,8 > Stark: $\lvert d \rvert > 0{,}8$
Zweistichproben-Z	$Cohens\ d = \frac{MW_{Stichprobe1} - MW_{Stichprobe2}}{\sqrt{\frac{(n_1 - 1) \cdot \hat{\sigma}_1^2 + (n_2 - 1) \cdot \hat{\sigma}_2^2}{n_1 + n_2 - 1}}}$	
Korrelationstest	Korrelationskoeffizient r	> Schwach: 0,1 bis 0,29 > Mittel: 0,3 bis 0,49 > Stark: > 0,49
Mann-Whitney-U	$CLES = \frac{U}{n1 \times n2}$	> 0,5 (kein Unterschied) > < > 0,5 (X < >Y)
Chi2-Unabhängigkeit	$Cramer's\ V = \sqrt{\frac{\chi^2}{n \cdot \min(n_{Kategorien1} - 1, n_{Kategorien_2} - 1)}}$	wie Korrelation

PARAMETRISCHE TESTS: EINSTICHPROBEN Z-TEST

HOCHSCHULE HEILBRONN

$Einstichproben\ Test$ = Vergleich Mittelwert **einer** Population mit einer Zahl

Beispiele. Fragen:

- Hat sich der durchschnittliche IQ (bisher: 101 = μ_0) von Studierenden verändert?
- Verbringen Studierende mehr als 4h / Tag (= μ_0) in sozialen Netzwerken?

EINSTICHPROBEN Z-TEST: BEISPIEL 1

Das Einstiegsgehalt von Absolventen der Hochschule Heilbronn betrug im langfristigen Durchschnitt ca. 31 T€/Jahr (Standardabweichung: 21 T€). Das durchschnittliche Einstiegsgehalt in einer Stichprobe mit 44 Absolventen ist 26 T€/Jahr.

Ist das ein Beleg, dass sich das Einstiegsgehalt signifikant verschlechtert hat?

a) Handelt es sich um einen einseitigen oder zweiseitigen Test?

b) Schreiben Sie die Null- und Alternativhypothese auf.

c) Führen Sie einen Test durch, um die Frage zu beantworten.

d) Berechnen und interpretieren Sie die Effektstärke.

EINSTICHPROBEN Z-TEST: BEISPIEL 2

Bei einem Online-Anbieter waren 54 der letzten 87 Kunden unter 30.

Ist dies Evidenz, dass der Anteil der unter 30-Jährigen Käufern von 50% abweicht?

a) Handelt es sich um einen einseitigen oder zweiseitigen Test?

b) Schreiben Sie die Null- und Alternativhypothese auf.

c) Führen Sie einen Test durch, um die Frage zu beantworten.

d) Berechnen und interpretieren Sie die Effektstärke.

PARAMETRISCHE TESTS: ZWEISTICHPROBEN Z-TEST

HOCHSCHULE HEILBRONN

$Zweistichproben\ Test$ = Vergleich Mittelwerte **zweier** Populationen

Beispiele. Fragen:

> Gibt es einen Unterschied im Einkommen von Frauen und Männern?

> Sind Nutzer von sozialen Netzwerken glücklicher?

ZWEISTICHPROBEN Z-TEST
BEISPIEL 1

In einer Stichprobe mit 32 Mitarbeitern der Firma A und 35 Mitarbeitern der Firma B werden folgende Durchschnittsjahresgehälter gemessen: Firma A: 43 T€ (Varianz: 450.11), Firma B: 29.8 T€ (Varianz: 294.05).

Unterscheiden sich die Durchschnittsgehälter von Firma A und B (1%-Niveau)?

a) Handelt es sich um einen einseitigen oder zweiseitigen Test?

b) Schreiben Sie die Null- und Alternativhypothese auf.

c) Führen Sie einen Test durch, um die Frage zu beantworten.

d) Berechnen und interpretieren Sie die Effektstärke.

ZWEISTICHPROBEN Z-TEST
BEISPIEL 2

In einer Stichprobe sind 63 von 90 Frauen und 27 von 52 Männer umweltbewusst. Die Stichprobenvarianz entspricht jeweils ungefähr der Populationsvarianz.

Ist der Anteil der umweltbewussten Frauen höher als der der Männer?

a) Handelt es sich um einen einseitigen oder zweiseitigen Test?

b) Schreiben Sie die Null- und Alternativhypothese auf.

c) Führen Sie einen Test durch, um die Frage zu beantworten.

PEARSON KORRELATIONSTEST: BEISPIEL

In einer Stichprobe mit 54 Probanden beträgt der Korrelationskoeffizient zwischen Selbstbewusstsein und Lebenszufriedenheit $r_{sl} = 0{,}32$.

Ist dies ein Beleg für eine positive Korrelation zwischen Lebenszufriedenheit und Selbstbewusstsein?

a) Handelt es sich um einen einseitigen oder zweiseitigen Test?

b) Schreiben Sie die Null- und Alternativhypothese auf.

c) Führen Sie einen Test durch, um die Frage zu beantworten.

PARAMETRISCHE TESTS: VORAUSSETZUNGEN

Die Variablen müssen

1. numerisch,
2. normalverteilt und
3. unabhängig

sein.

NICHT-PARAMETRISCHE TESTS: MANN-WHITNEY U TEST

HOCHSCHULE HEILBRONN

Mann-Whitney U Test = Vergleich von Gruppen

- mit ordinalen kategorialen und/oder
- nicht normalverteilten Variablen

Der Test überprüft die relative Wahrscheinlichkeit von Werten beider Populationen.

NICHT-PARAMETRISCHE TESTS: MANN-WHITNEY U TEST

HOCHSCHULE HEILBRONN

Punktschätzer U:

$$U = n_{max} \cdot n_{min} + \frac{n_{max}(n_{max} + 1)}{2} - R_{max},$$

wobei

- n_{max}, n_{min} = Stichprobengröße der Gruppe mit der größeren / kleineren Rangsumme
- R_{max} = Größere Rangsumme

NICHT-PARAMETRISCHE TESTS: MANN-WHITNEY U TEST

HOCHSCHULE HEILBRONN

Z-transformierte Teststatistik U:

$$z_U = \frac{U - \mu_U}{\sigma_U},$$

wobei

$$\mu_U = Mittelwert\ U = \frac{n_{max} \cdot n_{min}}{2}$$

und (bei nicht-verbundenen Rängen)

$$\sigma_U = Standardfehler\ U = \sqrt{\frac{n_{max} \cdot n_{min}(n_{max} + n_{min} + 1)}{12}}$$

MANN-WHITNEY U TEST: BEISPIEL

HOCHSCHULE HEILBRONN

Belegt folgende Stichprobe, dass Frauen mehr Wert auf Mode legen als Männer? Schreiben Sie zunächst die Hypothesen auf und führen dann den Test durch. Berechnen Sie außerdem den CLES und interpretieren diesen.

Geschlecht	**Mode Wichtig (1 bis 10)**
Frau	10
Mann	9
Mann	4
Frau	8
Mann	7
Mann	5
Frau	3

NICHT-PARAMETRISCHE TESTS: CHI2- / χ^2- TEST

Chi2-Tests:

- $Chi2 - Verteilungstest$:

  Stimmt eine beobachtete Verteilung mit einer gegebenen Verteilung überein?

- $Chi2 - Unabhängigkeitstest$:

  Sind zwei Variablen X und Y unabhängig?

NICHT-PARAMETRISCHE TESTS: CHI2- / χ^2- TEST

Chi2-Wert:

> basiert auf Differenz beobachtete (B) und erwartete (E) Häufigkeiten:

$$\chi^2 = \sum \frac{(B - E)^2}{E}$$

NICHT-PARAMETRISCHE TESTS: CHI2- / χ^2- TEST

Chi2-z-Teststatistik:

$$z_{\chi^2} = \frac{\chi^2 - dof}{\sqrt{2 \times dof}},$$

wobei dof = Freiheitsgrade

- Chi2-Verteilungstest:

$$dof = Anzahl\ Kategorien - 1$$

- Chi2-Unabhängigkeitstest:

$$dof = (Anzahl\ Kategorien\ Variable\ 1 - 1) \times (Anzahl\ Kategorien\ Variable\ 2 - 1)$$

CHI2-VERTEILUNGSTEST: BEISPIEL

Bei einem Würfelspiel würfelt ein Spieler Augenzahlen in folgender Häufigkeit:

⚀	⚁	⚂	⚃	⚄	⚅
1	2	2	3	4	17

Ist dies ein Beleg, dass der Würfel nicht fair ist?

a) Schreiben Sie die Null- und die Alternativhypothese auf.

b) Führen Sie einen Test durch, um die Frage zu beantworten.

CHI2-UNABHÄNGIGKEITSTEST: BEISPIEL

HOCHSCHULE HEILBRONN

Stichprobe mit 50 Männern und 52 Frauen. Jeder Teilnehmer gibt an, ob er/sie mehr als 2h/Tag auf Social-Media-Plattformen verbringt oder nicht:

	Männer	**Frauen**	**Gesamt**
$\leq$ 2h	33	27	60
> 2h	17	25	42
Gesamt	50	52	102

Sind Geschlecht und Social-Media-Nutzungsdauer unabhängig (5%)?

a) Schreiben Sie die Null- und die Alternativhypothese auf.

b) Führen Sie einen Test durch, um die Frage zu beantworten.

c) Berechnen und interpretieren Sie die Effektstärke.

THE END!

Please refer any questions to:
Prof. Dr. Florian Kauffeldt
Faculty of International Business
Florian.kauffeldt@hs-heilbronn.de