

# Wahrscheinlichkeit und Statistik

Tobias Forster

*tobias.forster@fhv.at*

WS 2024/2025

# Inhalt

Organisatorisches

Einführung in die Statistik

Deskriptive Statistik

Wahrscheinlichkeitsrechnung

Induktive Statistik

# Vorstellung Vortragender

- ▶ Tobias Forster
- ▶ Externer Lehrender an der FHV
- ▶ Studium der Technische Mathematik sowie Betriebswirtschaftslehre
- ▶ Energiehändler und Datenanalyst bei illwerke vkw AG
- ▶ Foliensatz wurde in Anlehnung an Angewandte Statistik (Wirtschaftsinformatik) von Dr. Kathrin Plankensteiner erstellt

# Lehrveranstaltung inkl. Benotung

- ▶ Integrierte Lehrveranstaltung
- ▶ 3 ECTS / 2 SWS / 75 Stunden
- ▶ Anwesenheit Lehrveranstaltung: 21 Stunden
- ▶ Vorbereitung Abschlussprüfung: 30 Stunden
- ▶ Hausübungen DataCamp: 15 Stunden (laut DataCamp 11 Stunden)
  - ▶ Introduction to Statistics in Python
  - ▶ Introduction to Regression with statsmodels in Python
  - ▶ Foundations of Probability in Python
- ▶ Übungsbeispiele (5 Zettel mit je 4 Aufgaben): 9 Stunden
- ▶ Benotung:
  - ▶ DataCamp (20 %)
  - ▶ Übungsbeispiele (10 %)
  - ▶ Klausur (70 %)
  - ▶ Negative Klausur impliziert negative Endnote

# Prüfung und Notenschlüssel für die Lehrveranstaltung

Prüfung am 11.01.2025:

- ▶ Die Prüfung muss für eine positive Endnote positiv sein.
- ▶ Die Prüfung findet am Zettel statt.
- ▶ 70 % der Endnote.
- ▶ Theorie wird abgefragt.
- ▶ Beispiele (Code in Python interpretieren / verstehen / ergänzen oder auch Beispiele wie auf den Übungsblättern).

Notenschlüssel:

- ▶  $0 \leq x < 50$  %: Nicht Genügend
- ▶  $50 \leq x < 62.5$  %: Genügend
- ▶  $62.5 \leq x < 75$  %: Befriedigend
- ▶  $75 \leq x < 87.5$  %: Gut
- ▶  $87.5 \leq x \leq 100$  %: Sehr Gut

# Übungsablauf

- ▶ 5 Übungszettel mit je 4 Beispielen:
  - ▶ 1. Übungszettel bis zum 21.09.
  - ▶ 2. Übungszettel bis zum 28.09.
  - ▶ 3. Übungszettel bis zum 19.10.
  - ▶ 4. Übungszettel bis zum 09.11.
  - ▶ 5. Übungszettel bis zum 30.11.
- ▶ Beispiele werden präsentiert.
- ▶ Jede bzw. jeder präsentiert mindestens ein Beispiel.
- ▶ 10 DataCamp Kurse, die in der Übung begonnen werden
  - ▶ Computerraum
  - ▶ Kopfhörer mitbringen

# Programmiersprache

- ▶ Python wird als Programmiersprache in diesem Kurs verwendet.
  - ▶ <https://www.python.org>
- ▶ Laut TIOBE Index vom September 2024 ist Python die beliebteste Programmiersprache der Welt:
  - ▶ <https://www.tiobe.com/tiobe-index/>
- ▶ Python ist eine relativ einfache Programmiersprache und hat umfangreiche Standardbibliotheken wie zum Beispiel Pandas.
- ▶ DataCamp hilft beim Erlernen von Statistik und Python und wird in diesem Kurs verwendet
  - ▶ <https://www.datacamp.com>
- ▶ Jupyter Notebooks können verwendet werden.
  - ▶ Installiere Anaconda (siehe <https://www.anaconda.com/download>)
  - ▶ Anaconda bietet verschiedene Tools an.
  - ▶ Jupyter Notebooks <https://jupyter.org> können für interaktives Coden verwendet werden.

# Termine

13.09	15:45 - 17:20	Einführung in die Statistik
14.09	08:00 - 09:35	Deskriptive Statistik
14.09	09:45 - 11:20	Introduction to Statistics in Python 1 and 4
21.09	08:00 - 09:35	Bivariate Statistik / Regression
21.09	09:45 - 11:20	Regression with statsmodels in Python 1
28.09	08:00 - 09:35	Wahrscheinlichkeitsrechnung
28.09	09:45 - 11:20	Regression with statsmodels in Python 2
19.10	14:50 - 16:30	Introduction to Statistics in Python 2
09.11	08:00 - 09:35	Induktive Statistik
09.11	09:45 - 11:20	Introduction to Statistics in Python 3
30.11	14:00 - 14:45	Induktive Statistik
30.11	14:50 - 16:30	Foundations of Probability in Python 1 and 2
14.12	10:35 - 13:05	Foundations of Probability in Python 3 and 4
11.01	08:00 - 09:35	Klausur



# Einführung in die Statistik

# Wann kommen wir mit Statistik in Verbindung?

Diagramme, Tabellen oder statistische Kennzahlen werden in unterschiedlichen Bereichen des täglichen Lebens verwendet:

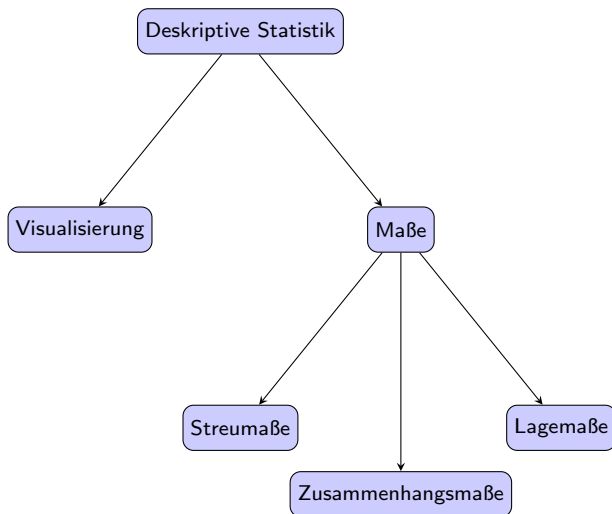
- ▶ Qualitätskontrollen
- ▶ Wahlausgänge
- ▶ Volkszählungen
- ▶ Sportsendungen
- ▶ etc.

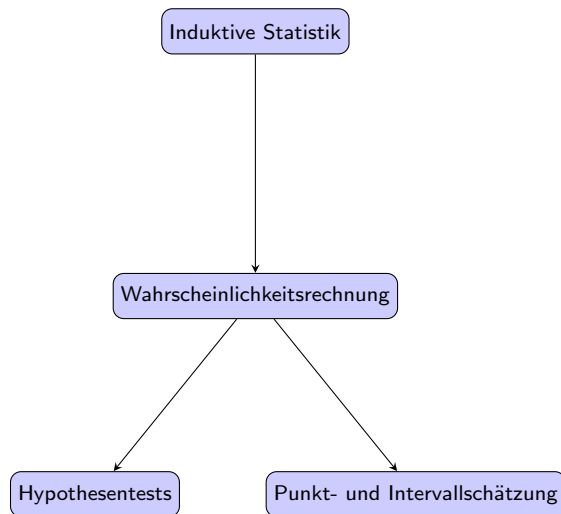
# Einsatzgebiete im Bereich Mechatronik

- ▶ Qualitätskontrolle und -sicherung (z.B. Überwachung von Produktionsprozessen)
- ▶ Prognosemodelle (z.B.: Lebensdauerprognose von Bauteilen)
- ▶ Systemdiagnose und Fehlererkennung (z.B.: frühzeitige Erkennung von Fehlern in Systemen)
- ▶ Produktentwicklung und -design (z.B.: Kundenzufriedenheit)
- ▶ etc.

# Vorstellung und Erwartung

- ▶ Wer bin ich?
- ▶ Wer ist mein Arbeitgeber?
- ▶ Wann bin ich bei meinem Arbeitgeber mit Statistik in Berührung gekommen?
- ▶ Was sind meine Erwartungen an die Lehrveranstaltung?





# Deskriptive Statistik

# Übersicht und Grundbegriffe

## Deskriptive (beschreibende) Statistik

Die deskriptive Statistik verfolgt das Ziel, Daten mithilfe von Tabellen, Grafiken oder auch Kennzahlen darzustellen und zu beschreiben.

## Grundgesamtheit / Population

Die Grundgesamtheit - auch Population genannt - beschreibt die Menge aller Objekte, über die wir eine Aussage treffen wollen. Die Bestimmung von gewissen Eigenschaften der Grundgesamtheit ist das Ziel der schließenden Statistik.

## Stichprobe

Die Stichprobe ist die Menge der beobachteten Objekten. In anderen Worten ist die Stichprobe eine (zufällige) Teilmenge der Grundgesamtheit. Diese Teilmenge wird in der deskriptiven Statistik untersucht.



## Merkmalsträger

Die Objekte dessen Eigenschaften wir beobachten.

## Merkmale

Eine gewisse Eigenschaft, die an den Merkmalsträgern beobachtet wird.

## Merkmalsausprägung / Ausprägung eines Merkmals

Merkmalsausprägungen sind verschiedene Werte, die ein gewisses Merkmal annehmen kann.

# Kategorische Daten (Merkmale)

## Nominale Daten

Die Ausprägungen eines Merkmals können unterschieden werden.

z.B.: Geschlecht, Autokennzeichen, Blutgruppe, Studiengänge an einer Hochschule.

## Ordinale Daten

Ordinalen Daten können zusätzlich zu der Eigenschaft der nominalen Daten sortiert werden. Dabei gilt entweder  $a < b$ ,  $a = b$  oder  $a > b$ . Zudem muss gelten: Wenn  $a > b$  und  $b > c$ , dann muss  $a > c$  sein. Somit können kumulierte Häufigkeiten berechnet werden.

z.B.: Klausurleistung, Kundenzufriedenheit, Lebensdauer eines Bauteils.

# Numerische Daten (Merkmale)

## Diskrete numerische Daten

Die Ausprägungen eines Merkmals können nur bestimmte, abzählbare Werte annehmen. Oft handelt es sich dabei um ganze Zahlen.

z.B.: Anzahl an Bauteilen, Anzahl an Studenten im Kurs.

## Kontinuierliche numerische Daten

Die Ausprägungen eines Merkmals können jeden beliebigen Wert innerhalb eines gegebenen Bereichs annehmen.

z.B.: Verbrauch in MWh pro Einwohner, Lufttemperatur in °C.

## Absolute Häufigkeiten

Die absolute Häufigkeit  $h_i$  beschreibt wie oft die  $i$ -te Ausprägung eines Merkmals vorkommt.

## Relative Häufigkeiten

Die relative Häufigkeit  $r_i$  berechnet sich wie folgt

$$r_i = \frac{h_i}{n},$$

wobei  $n = \sum_i h_i$  die Anzahl an Beobachtungen beschreibt. Zudem gilt  $\sum_i r_i = 1$ .

**Beispiel:**

d	Kleidung
0	T-Shirt
1	Hemd
2	Hemd
3	T-Shirt
4	Hemd
5	Pullover
6	T-Shirt

Kleidung	$h_i$	$r_i$
T-Shirt	3	$3/7$
Hemd	3	$3/7$
Pullover	1	$1/7$

Zudem ist ersichtlich, dass  $n = 7$  und  $\sum_i r_i = 1$  ist.

# Kumulierte Häufigkeiten

## Absolute kumulierte Häufigkeit

Gegeben sind  $n$  Beobachtungen eines ordinalen Merkmals mit  $m$  verschiedenen Ausprägungen  $x_i$  mit der Ordnung  $x_1 < x_2 < \dots < x_m$ . Die absolute kumulierte Häufigkeit wird berechnet als

$$F_{abs}(x_k) = \sum_{i=1}^k h_i \text{ mit } 1 \leq k \leq m,$$

wobei  $h_i$  die absolute Häufigkeit der Ausprägung  $x_i$  beschreibt.

Die absolute kumulierte Häufigkeit ist demnach die Summe der absoluten Häufigkeiten der Ausprägungen von der kleinsten bis zur jeweils gegebenen Schranke.

Beispiel: Anzahl der Studierenden mit einer Note nicht schlechter als eine 3 im Fach Wahrscheinlichkeit und Statistik.

# Kumulierte Häufigkeiten

## Relative kumulierte Häufigkeit

Gegeben sind  $n$  Beobachtungen eines ordinalen Merkmals mit  $m$  verschiedenen Ausprägungen  $x_i$  mit der Ordnung  $x_1 < x_2 < \dots < x_m$ . Die relative kumulierte Häufigkeit wird berechnet als

$$F_{rel}(x_k) = \sum_{i=1}^k r_i \text{ mit } 1 \leq k \leq m,$$

wobei  $r_i$  die relative Häufigkeit der Ausprägung  $x_i$  beschreibt. Die Funktion  $F_{rel}$  wird auch empirische Verteilungsfunktion genannt.

Die relative kumulierte Häufigkeit ist die Summe der relativen Häufigkeiten der Ausprägungen von der kleinsten bis zur jeweils gegebenen Schranke.

## Klasseneinteilung

Die Klasseneinteilung beschreibt die Einteilung in Abhängigkeit der Ausprägung eines Merkmals in Klassen.

Wann ist eine Klasseneinteilung sinnvoll?

- ▶ Bei vielen Ausprägungen eines Merkmals
- ▶ Falls fast kein Unterschied zwischen den Ausprägungen vorhanden ist
- ▶ Falls einzelne Ausprägungen selten oder gar nicht vorkommen



## Was ist bei der Klasseneinteilung zu beachten?

- ▶ Jede Ausprägung wird genau einer Klasse zugeordnet.
- ▶ Die Klassen sind disjunkt und zudem entweder links offen und rechts abgeschlossen oder vice versa.
- ▶ Die erste Klasse kann bis  $-\infty$  und die letzte Klasse kann bis  $+\infty$  gehen.
- ▶ Die Reduktion der Daten geht mit einem Verlust an (wesentlichen) Informationen einher.
- ▶ Faustregeln für die Bestimmung der Anzahl an Klassen sind:

	Anzahl Beobachtungen	Klassen
	$< 50$	5 bis 7
i)	$50 \leq x < 100$	6 bis 10
	$100 \leq x < 250$	7 bis 12
	$\geq 250$	10 bis 20

- ii)  $\text{Anzahl an Klassen} \leq \sqrt{\text{Anzahl an Ausprägungen}}$ .

## Lagemaß

Ein Lagemaß ist eine statistische Kennzahl, die durch Zusammenfassen von Informationen zu einem repräsentativen Wert ermittelt werden kann. Dieser repräsentative Wert liefert Positionsinformationen der zugrunde liegenden Daten.

Beispiele für gängige Lagemaße:

- ▶ Arithmetischer Mittelwert
- ▶ Median und Modus
- ▶ Quantile

## Arithmetisches Mittel

Gegeben sind  $n$  Beobachtungen mit Ausprägungen  $x_i$ . Das arithmetische Mittel - auch Mittelwert, Mittel oder Durchschnitt genannt - wird wie folgt berechnet

$$\bar{x} = \frac{x_1 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Eigenschaften vom arithmetischen Mittel sind:

- ▶ Die Abweichungen  $\epsilon_i$  der Ausprägungen  $x_i$  vom Mittelwert  $\bar{x}$  berechnen sich wie folgt:  $\epsilon_i = x_i - \bar{x}$  für  $i = 1, \dots, n$ .
- ▶ Falls die Summe der Abweichungsquadrate minimiert werden soll ( $\min_z \sum_{i=1}^n (x_i - z)^2$ ), liefert  $z = \bar{x}$  den minimierenden Wert.
- ▶ Das arithmetische Mittel ist eindeutig.

## Median

Gegeben sind  $n$  Beobachtungen eines ordinalen Merkmals mit geordneten Ausprägungen  $x_i$ , d.h.  $x_1 \leq \dots \leq x_n$ . Der Median - auch Zentralwert genannt - wird wie folgt berechnet

$$x_{median} = \begin{cases} x_{m+1} & \text{für ungerades } n = 2m + 1 \\ \frac{1}{2}(x_m + x_{m+1}) & \text{für gerades } n = 2m. \end{cases}$$

- ▶ Der Median ist im Vergleich mit dem arithmetischen Mittel robust gegenüber Ausreißern.
- ▶ Der Median ist eindeutig.

## Modus

Der Modus - auch als Modalwert bekannt - beschreibt jene Ausprägungen eines Merkmals, die am häufigsten in der Stichprobe vorkommen.

- ▶ Der Modus ist nicht immer eindeutig, kann aber immer berechnet werden.
- ▶ Um den Modus berechnen zu können, müssen zumindest nominale Daten vorliegen.
- ▶ Verteilungen werden
  - ▶ mit einem Modus unimodale Verteilung,
  - ▶ mit zwei Modi bimodale Verteilung und
  - ▶ mit zwei oder mehr Modi multimodale Verteilunggenannt.

## Quantil

Gegeben sind  $n$  Beobachtungen eines ordinalen Merkmals mit geordneten Ausprägungen  $x_i$ , d.h.  $x_1 \leq \dots \leq x_n$ . Sei  $p \in (0, 1)$ , dann berechnet sich das empirische  $p$ -Quantil  $x_p$  wie folgt

$$x_p = \begin{cases} \frac{1}{2}(x_{n \cdot p} + x_{n \cdot p + 1}) & \text{falls } n \cdot p \in \mathbb{N} \\ x_{\lfloor n \cdot p + 1 \rfloor} & \text{falls } n \cdot p \notin \mathbb{N}. \end{cases}$$

Hier bezeichnet  $\lfloor \cdot \rfloor$  die Abrundungsfunktion (auch Gaußklammer genannt).

- ▶ Median: Der Median entspricht dem 0.5-Quantil.
- ▶ Terzil: Das untere Terzil entspricht dem  $\frac{1}{3}$ -Quantil und das obere Terzil dem  $\frac{2}{3}$ -Quantil.
- ▶ Quartil: Das erste Quartil ( $Q_1$ ) beschreibt das 0.25-Quantil, das zweite Quartil ( $Q_2$ ) den Median und das dritte Quartil ( $Q_3$ ) das 0.75-Quantil.
- ▶ Dezil:  $p \in \{0.1, 0.2, \dots, 0.9\}$ .
- ▶ Perzentil:  $p \in \{0.01, 0.02, \dots, 0.99\}$ .

## Streumaß

Ein Streumaß - auch Dispersionsmaß genannt - ist eine statistische Kennzahl, die die Streuung der Beobachtungen angibt. Um ein Streumaß definieren zu können, wird ein Abstandsmaß benötigt. Ein kleines (großes) Streumaß bedeutet, dass die Beobachtungen konzentriert (zerstreut) sind.

Streumaße können wie folgt unterteilt werden:

- ▶ Abstand zwischen Lagemaße ( $R$ ,  $IQA$ )
- ▶ Streuung um ein Lagemaß
- ▶ Streuung relativ zu einem Lagemaß

# Abstand zwischen Lagemaße

## Spannweite

Die Spannweite berechnet sich aus der Differenz zwischen der maximalen ( $x_{max}$ ) und der minimalen ( $x_{min}$ ) Ausprägung eines Merkmals in der Stichprobe

$$R = x_{max} - x_{min}.$$

## Interquartilsabstand

Der Interquartilsabstand berechnet sich aus der Differenz zwischen dem dritten Quartil ( $Q_3$ ) und dem ersten Quartil ( $Q_1$ )

$$IQA = Q_3 - Q_1.$$

- ▶  $R$  wird im Vergleich mit  $IQA$  stark durch Ausreißer beeinflusst.
- ▶  $R$  und  $IQA$  existieren immer und sind eindeutig.
- ▶ Länge der Box im Boxplot.



## Ausreißer bestimmen mit IQA

Ein Ausreißer ist eine Beobachtung, die von den meisten anderen Beobachtungen weit entfernt liegt. Eine Möglichkeit Ausreißer zu identifizieren bietet der IQA. Demnach ist eine Beobachtung  $x$  ein Ausreißer, falls:

$$x < Q_1 - 1.5 \cdot IQA \text{ oder}$$

$$x > Q_3 + 1.5 \cdot IQA.$$

# Streuung um ein Lagemaß

## Mittlere absolute Abweichung

Gegeben sind  $n$  Beobachtungen  $x_1, \dots, x_n$ . Die mittlere absolute Abweichung berechnet sich wie folgt

$$MAD_{mean} = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

$$MAD_{median} = \frac{1}{n} \sum_{i=1}^n |x_i - x_{median}|.$$

- ▶ Hier wird als Abstandsmaß die Betragsfunktion  $|\cdot|$  verwendet.
- ▶  $MAD_{mean}$  und  $MAD_{median}$  sind nicht robust gegenüber Ausreißer.

## Empirische Varianz

Gegeben sind  $n$  Beobachtungen  $x_1, \dots, x_n$ . Die empirische Varianz  $s^2$  berechnet sich wie folgt

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

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

Hier wird als Abstandsmaß der quadratische Abstand  $(.)^2$  verwendet.

- ▶ Falls die Grundgesamtheit gegeben ist, wird Gleichung 1 verwendet.
- ▶ Falls eine Stichprobe gegeben ist, wird Gleichung 2 verwendet.
- ▶ Das  $s^2$  in Gleichung 2 ist ein erwartungstreuer Schätzer für die Varianz der Grundgesamtheit.
- ▶ Die empirische Standardabweichung  $s$  ist gegeben durch  $s = \sqrt{s^2}$ .

# Streuung relativ zu einem Lagemaß

Um unterschiedliche Datenreihen miteinander vergleichen zu können, wird eine dimensionslose Größe benötigt.

## Variationskoeffizient

Der Variationskoeffizient - auch Abweichungskoeffizient genannt - ist ein relatives Streumaß, das nicht von der Maßeinheit der Beobachtungen abhängt. Dieses Streumaß berechnet sich wie folgt

$$V = \frac{s}{\bar{x}}.$$

- ▶  $V$  ist nur sinnvoll für Datenreihen mit ausschließlich positiven (oder ausschließlich negativen) Werten.
- ▶ Dieses Streumaß beschreibt die normierte Standardabweichung.
- ▶ Der Variationskoeffizient ermöglicht wegen seiner Dimensionslosigkeit einen Vergleich der Streuung unterschiedlicher Datenreihen.
- ▶  $V$  ist nicht robust gegenüber Ausreißer.

# Streuung relativ zu einem Lagemaß

## Quartilsdispersionskoeffizient

Der Quartilsdispersionskoeffizient ist ein relatives Streumaß, das nicht von der Maßeinheit der Beobachtungen abhängt. Dieses Streumaß berechnet sich wie folgt

$$V_r = \frac{IQA}{x_{median}}.$$

- ▶  $V_r$  ist eine robuste Version von  $V$ .
- ▶ Dieses Streumaß beschreibt den normierten Interquartilsabstand.

## Übersicht

- ▶ Nominale sowie ordinale Daten:
  - ▶ Zusammenhangsmaß: Es wird der Chi-Quadrat Koeffizient und daraus der Kontingenzkoeffizient bestimmt.
- ▶ Metrische Daten: Die Ausprägung eines Merkmals besteht aus einer Zahl, einer Dimension und einem Nullpunkt. Zum Beispiel: Einkommen, Alter, Geschwindigkeit.
  - ▶ Zusammenhangsmaß: Die Kovarianz und der Korrelationskoeffizient werden bestimmt.

# Randhäufigkeiten

Seien  $x_1, \dots, x_m$  Merkmalsausprägungen der Variable  $X$  und  $y_1, \dots, y_k$  Merkmalsausprägungen der Variable  $Y$ . Wie oben bereits definiert, beschreibt  $h_{ij}$  die absolute Häufigkeit der Merkmalsausprägung  $x_i$  mit  $y_j$ .

	$y_1$	$\dots$	$y_k$	<b>Randhäufigkeiten von <math>X</math></b>
$x_1$	$h_{11}$	$\dots$	$h_{1k}$	$h_{1\bullet}$
$x_2$	$h_{21}$	$\dots$	$h_{2k}$	$h_{2\bullet}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$x_m$	$h_{m1}$	$\dots$	$h_{mk}$	$h_{m\bullet}$
<b>Randhäufigkeiten von <math>Y</math></b>	$h_{\bullet 1}$	$\dots$	$h_{\bullet k}$	$h_{\bullet \bullet}$

Die Randhäufigkeiten werden wie folgt berechnet:  $h_{\bullet j} = \sum_i h_{ij}$  oder  $h_{i\bullet} = \sum_j h_{ij}$ .  
Die Summe der Randhäufigkeiten ergibt sich aus  $n = h_{\bullet \bullet} = \sum_j \sum_i h_{ij}$ .

## Von absoluten zu relativen Häufigkeiten

Wir können die Tabelle von der vorherigen Folie auch in relative Häufigkeiten umwandeln. Dafür erinnern wir uns, dass  $r_i = \frac{h_i}{n}$  die relative Häufigkeit der  $i$ -ten Ausprägung und  $n$  die Anzahl an Beobachtungen beschreibt. Demnach gilt:

	$y_1$	$\dots$	$y_k$	<b>Randhäufigkeiten von X</b>
$x_1$	$u_{11}$	$\dots$	$u_{1k}$	$u_{1\bullet}$
$x_2$	$u_{21}$	$\dots$	$u_{2k}$	$u_{2\bullet}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$x_m$	$u_{m1}$	$\dots$	$u_{mk}$	$u_{m\bullet}$
<b>Randhäufigkeiten von Y</b>	$u_{\bullet 1}$	$\dots$	$u_{\bullet k}$	$u_{\bullet\bullet}$

Auch hier gilt:  $u_{\bullet j} = \sum_i u_{ij}$ ,  $u_{i\bullet} = \sum_j u_{ij}$  und  $u_{\bullet\bullet} = \sum_j \sum_i u_{ij} = 1$ .



## Beispiel 1: Kontingenztafel

	BB	VZ	<b>Summe (Zeile)</b>
Männlich	12	20	32
Weiblich	3	18	21
<b>Summe (Spalte)</b>	15	38	n = 53

- ▶ Wie hoch war der Anteil der Studenten im Studiendesign BB insgesamt?
- ▶ Wie hoch war der Anteil der weiblichen Studenten im Studiendesign VZ?
- ▶ Wie hoch war der Anteil vom Studiendesign VZ an den männlichen Studierenden?

# Statistische Abhängigkeit und Unabhängigkeit

Unabhängig ist die Tabelle, wenn das Geschlecht nicht mit dem Studiendesign zusammenhängt. Dafür muss folgende Gleichung erfüllt sein:

$$\tilde{h}_{ij} * n = h_{i\bullet} \cdot h_{\bullet j}.$$

$\tilde{h}_{ij}$  beschreibt den erwarteten Wert der absoluten Häufigkeit der Merkmalskombination in der Zeile  $i$  und der Spalte  $j$ .

Falls die Gleichung für eine Zelle nicht erfüllt ist, gibt es eine Abhängigkeit der Merkmale.

Ergebnis:

$$12 \cdot 53 \neq 15 \cdot 32$$

$$3 \cdot 53 \neq 15 \cdot 21$$

$$20 \cdot 53 \neq 38 \cdot 32$$

$$18 \cdot 53 \neq 38 \cdot 21$$

In diesem Fall liegt eine Abhängigkeit der Merkmale vor.

## Beispiel 2 Kontingenztafel

	$y_1$	$y_2$	<b>Summe (Zeile)</b>
$x_1$	20	20	40
$x_2$	6	6	12
<b>Summe (Spalte)</b>	26	26	$n = 52$

In diesem Fall liegt eine Unabhängigkeit der Merkmale vor.

## Chi-Quadrat Koeffizient

Chi-Quadrat wird wie folgt berechnet

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

- ▶  $\chi^2$  beschreibt Chi-Quadrat.
- ▶  $m$  beschreibt die Anzahl an Zeilen.
- ▶  $k$  beschreibt die Anzahl an Spalten.
- ▶  $h_{ij}$  beschreibt die absolute Häufigkeit der Merkmalskombination in der Zeile  $i$  und der Spalte  $j$ .
- ▶  $\tilde{h}_{ij}$  beschreibt den erwarteten Wert der absoluten Häufigkeit der Merkmalskombination in der Zeile  $i$  und der Spalte  $j$ .

# Ergebnis für Beispiel 1

- ▶ Schritt 1: Berechnung  $\tilde{h}_{ij}$ .
  - ▶  $\tilde{h}_{11} : \frac{15 \cdot 32}{53} \approx 9$
  - ▶  $\tilde{h}_{21} : \frac{15 \cdot 21}{53} \approx 6$
  - ▶  $\tilde{h}_{12} : \frac{38 \cdot 32}{53} \approx 23$
  - ▶  $\tilde{h}_{22} : \frac{38 \cdot 21}{53} \approx 15$
- ▶ Schritt 2: Berechnung  $(h_{ij} - \tilde{h}_{ij})^2$ .
  - ▶  $(12 - 9)^2 = 9$
  - ▶  $(3 - 6)^2 = 9$
  - ▶  $(20 - 23)^2 = 9$
  - ▶  $(18 - 15)^2 = 9$
- ▶ Schritt 3: Berechnung  $\frac{(h_{ij} - \tilde{h}_{ij})^2}{\tilde{h}_{ij}}$ .
  - ▶  $\frac{9}{9} = 1$
  - ▶  $\frac{9}{6} = 1.5$
  - ▶  $\frac{9}{23} \approx 0.39$
  - ▶  $\frac{9}{15} = 0.6$
- ▶ Schritt 4:  $X^2 = 1 + 1.5 + 0.39 + 0.6 = 3.49$

# Aussagekraft von $\chi^2$

- ▶ Aussagekraft des Koeffizienten ist gering.
- ▶  $\chi^2 \in [0, n \cdot (\min\{m, k\} - 1)]$ .
- ▶  $\chi^2 = 0$ : Völlige Unabhängigkeit der Merkmale.
- ▶ Wert bei vollkommener Abhängigkeit hängt ab von:
  - ▶ Anzahl an Ausprägungen.
  - ▶ Untersuchter Gesamtheit  $n$ .
- ▶  $\chi^2$  ist demnach nicht standardisiert und daher nur begrenzt vergleichbar.

Aus diesem Grund können wir den Chi-Quadrat Koeffizienten in den Kontingenzkoeffizient umwandeln.

## Korrigierter Kontingenzkoeffizient

Der korrigierte Kontingenzkoeffizient wird wie folgt berechnet

$$K^P = \sqrt{\frac{X^2}{n + X^2} \cdot \frac{\min\{m, k\}}{\min\{m, k\} - 1}}.$$

- ▶  $K^P$  beschreibt den korrigierten Kontingenzkoeffizient.
- ▶  $X^2$  beschreibt den Chi-Quadrat Koeffizient.
- ▶  $n$  beschreibt die Größe der Stichprobe.
- ▶  $\min\{m, k\}$  nimmt das Minimum der Anzahl an Zeilen und Spalten.

## Beispiel und Interpretation

- ▶  $K^P = \sqrt{\frac{3.49}{53+3.49} \cdot \frac{2}{1}} \approx 0.35$
- ▶ Der Kontingenzkoeffizient nimmt Werte zwischen 0 und 1 an.
- ▶ Ein Wert von Null bedeutet, dass kein Zusammenhang zwischen den Merkmalen vorliegt.
- ▶ Ein Wert von Eins bedeutet, dass ein vollständiger Zusammenhang zwischen den Merkmalen existiert.
- ▶ In unserem Beispiel liegt ein schwacher statistischer Zusammenhang zwischen den Merkmalen Geschlecht und Studiendesign vor.
- ▶ Der korrigierte Kontingenzkoeffizient ermöglicht den Vergleich von Kontingenztabellen unterschiedlicher Größen.



## Korrigierte empirische Kovarianz

Die empirische korrigierte Kovarianz einer Stichprobe mit  $n$  Beobachtungen berechnet sich wie folgt:

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}).$$

- ▶ Die korrigierte empirische Kovarianz ist ein erwartungstreuer Schätzer der Kovarianz der gesamten Population mithilfe einer Stichprobe.
- ▶ Die empirische Kovarianz beschreibt den linearen Zusammenhang zweier Merkmale.
- ▶ Die Kovarianz ist keine dimensionslose Größe.
- ▶ Falls  $s_{xy} > 0$  ist, liegt ein positiver Zusammenhang vor. Vice versa für  $s_{xy} < 0$ .

## Empirischer Korrelationskoeffizient

Der empirische Korrelationskoeffizient für eine zweidimensionale Stichprobe  $(x_i, y_i)_{i=1}^n$  berechnet sich wie folgt:

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

- ▶ Dieses Zusammenhangsmaß beschreibt den Grad des linearen Zusammenhangs zwischen zwei Merkmalen und ist eine dimensionslose Größe.
- ▶ Die Korrelation bedeutet nicht, dass die Änderung einer Variable eine Änderung der anderen Variable zur Folge hat.
- ▶ Falls  $r_{xy} \approx 0$  kann trotzdem ein nicht-linearer Zusammenhang vorliegen.

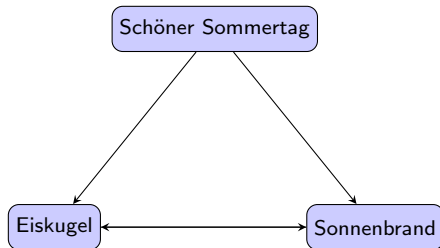
# Interpretation

- ▶  $|r_{xy}| = 1$ : Die zwei Merkmale liegen im 2-dimensionalen Koordinatensystem auf einer Linie.
- ▶ Ab  $|r_{xy}| > 0.5$  sprechen wir davon, dass der lineare Zusammenhang groß ist.
- ▶ Bei  $|r_{xy}| \approx 0.3$  sprechen wir von einer moderaten Korrelation.
- ▶ Bei  $|r_{xy}| \approx 0.1$  sprechen wir von einer geringen Korrelation.
- ▶ Sprechweise bei z.B.  $r_{xy} = 0.8$ : Ein hohes  $x$  geht mit einem hohen  $y$  einher.

# Kausalität

- ▶ Unter Kausalität wird verstanden, dass die Änderung einer Variable  $x$  eine Änderung der Variable  $y$  zur Folge hat.
- ▶ Korrelation impliziert keine Kausalität.

## Beispiel Korrelation vs. Kausalität



Es liegt eine positive Korrelation zwischen Eiskugel und Sonnenbrand vor. Kausalität: Ein schöner Sommertag hat zur Folge, dass mehr Eis konsumiert wird und die Individuen einen Sonnenbrand bekommen können.

Falls ein drittes Merkmal, das nicht berücksichtigt wurde, die Ursache des Zusammenhangs von zwei Merkmalen ist, so nennt man dies eine Scheinkorrelation.

- ▶ Störche und Geburten
  - ▶ Urbanisierung und Industrialisierung: Den Störchen wird ihr Lebensraum entzogen.
  - ▶ Urbanisierung und Industrialisierung: Zunahme der Erwerbstätigkeit von Frauen und somit sinkende Geburtenrate.
- ▶ Schuhgröße und Einkommen
  - ▶ Drittes Merkmal Geschlecht entscheidend.
  - ▶ Frauen haben im Allgemeinen kleinere Schuhnummern und Einkommen.
  - ▶ Vice Versa für Männer.

## Exkurs: Einfache lineare Regression

- ▶ Die lineare Regression ist eine Methodik zur Bestimmung des Einflusses einer oder mehrerer Variablen auf eine Zielgröße.
- ▶ Wir sprechen von einfacher linearer Regression, falls es genau eine unabhängige Variable gibt.
- ▶ In der einfachen linearen Regression wird versucht die beobachteten Werte durch eine Gerade abzubilden.
- ▶ Falls die einfache lineare Regression als Prognosemodell dient, liegen die Prognosen auf dieser Geraden.
- ▶ Beispiele:
  - ▶ Größe und Gewicht
  - ▶ Lernzeit und Punkte im Abschlusstest



## Einfache lineare Regression

Sei  $X$  die Einflussgröße (auch unabhängige Variable genannt) und  $Y$  die gesuchte Zielgröße (auch abhängige Variable genannt). Mithilfe der einfachen linearen Regression wird eine Gerade durch die Punktwolke  $\{(x_1, y_1), \dots, (x_n, y_n)\}$  gelegt, sodass der lineare Zusammenhang zwischen  $X$  und  $Y$  möglichst genau beschrieben wird.

Die Gleichung der einfachen linearen Regression ist gegeben durch:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i \in \{1, \dots, n\}.$$

$\epsilon_i$  beschreibt den Fehler zwischen dem wahren Wert  $y$  und dem Schätzwert  $\hat{y}$ ,  $\beta_0$  die Regressionskonstante (auch intercept genannt) und  $\beta_1$  das Gewicht der unabhängigen Variable  $X$ .

Falls alle Punkte auf einer Geraden liegen würden, könnte die einfache lineare Regression die abhängige Variable perfekt erklären und es würden keine Fehler existieren. Dies ist in den seltensten Fällen der Fall.

## Wie wird die Gerade durch die Punktwolke gelegt?

Das Ziel ist die Fehler zwischen  $y_i$  und  $\hat{y}_i$  zu minimieren. Da sich positive und negative Fehler nicht aufheben sollen, können beispielsweise die absoluten oder quadrierten Fehler betrachtet werden.

### Zielfunktion

Aus diesem Grund minimiert die einfache lineare Regression die quadrierten Abweichungen zwischen  $y_i$  und  $\hat{y}_i$ . Die Zielfunktion lautet:

$$\min \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \min \sum_{i=1}^n \epsilon_i^2.$$

Diese Methode wird Methode der kleinsten Quadrate sowie OLS (engl. *ordinary least squares*) genannt. Demnach werden diejenigen  $\beta_0$  sowie  $\beta_1$  gesucht, die folgendes Problem lösen:

$$\min Q = \min \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2.$$

Um dieses Minimierungsproblem zu lösen, leiten wir nach  $\beta_0$  und  $\beta_1$  ab und setzen gleich Null. Wir erhalten

$$\frac{\partial Q}{\partial \beta_1} \stackrel{!}{=} 0$$

$$-2 \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i)) x_i = 0$$

$$\sum_{i=1}^n x_i y_i = \beta_0 \sum_{i=1}^n x_i + \beta_1 \sum_{i=1}^n x_i^2$$

sowie

$$\frac{\partial Q}{\partial \beta_0} \stackrel{!}{=} 0$$

$$-2 \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i)) = 0$$

$$\sum_{i=1}^n y_i = \beta_0 n + \beta_1 \sum_{i=1}^n x_i$$

Dies können wir wie folgt darstellen:

$$\begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

Wir wissen, dass eine Matrix wie folgt invertiert wird:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Daraus ergibt sich:

$$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \frac{1}{n \sum x_i^2 - (\sum x_i)^2} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix} \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}.$$

Nun wäre noch zu zeigen, dass es ein Minimum ist. Dies lässt sich über die zweite Ableitung zeigen.

## Interpretation von $\beta_0$

Die Regressionskonstante gibt den Schnittpunkt der Regressionsgeraden mit der y-Achse bei  $x = 0$  an. Falls  $\beta_0 = 0$  ist, geht die Regressionsgerade durch den Ursprung.

## Interpretation von $\beta_1$

Der Vorhersagewert des Modells erhöht sich um  $\beta_1$  Einheiten, falls die unabhängige Variable sich um eine Einheit erhöht. Das Vorzeichen von  $\beta_1$  gibt demnach die Richtung des Effekts an.

## Multiple lineare Regression

Falls eine beobachtete abhängige Variable durch mehrere unabhängige Variablen erklärt wird, sprechen wir von multipler linearer Regression. Dies ist eine Verallgemeinerung der einfachen linearen Regression.

## DataCamp

Im DataCamp Kurs *Introduction to Regression with statsmodels in Python* lernen wir die lineare Regression besser kennen und wenden diese an gewissen Datensätzen an.

## Exkurs: Mengenlehre

# Definitionen I

## Menge

Eine Menge beschreibt eine Zusammenfassung von bestimmten wohlunterscheidbaren Objekten zu einem Ganzen.

## Teilmenge

$A$  heißt Teilmenge von  $B$ , falls folgendes gilt:

$$A \subseteq B :\Leftrightarrow \forall x \in A : x \in B.$$

## Leere Menge

Die leere Menge enthält kein Element. Schreibweise:  $\emptyset$ .



## Definitionen II

### Vereinigung zweier Mengen

$$A \cup B :\Leftrightarrow \{x : x \in A \vee x \in B\}$$

### Schnitt zweier Menge

$$A \cap B :\Leftrightarrow \{x : x \in A \wedge x \in B\}$$

### Differenz zweier Mengen

$$A \setminus B :\Leftrightarrow \{x : x \in A \wedge x \notin B\}$$

# Definitionen III

## Komplement

$$A^C :\Leftrightarrow \{x : x \notin A\}$$

## Symmetrische Differenz von Mengen

$$A \triangle B :\Leftrightarrow \{x : (x \in A \wedge x \notin B) \vee (x \notin A \wedge x \in B)\}$$

# Gesetzmäßigkeiten I

Sei  $A, B, C \subseteq \Omega$ , dann gilt:

Assoziativgesetz:

$$(A \cap B) \cap C = A \cap (B \cap C)$$

$$(A \cup B) \cup C = A \cup (B \cup C)$$

Distributivgesetz:

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$

$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$

Kommutativ:

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

# Gesetzmäßigkeiten II

## De Morgansche Gesetze

$$(A \cup B)^C = A^C \cap B^C$$

$$(A \cap B)^C = A^C \cup B^C$$

## Absorptionsgesetz

$$(A \cup B) \cap A = A$$

$$(A \cap B) \cup A = A$$

# Wahrscheinlichkeitsrechnung

## Grundraum oder Ereignisraum $\Omega$

Die Menge  $\Omega$  der möglichen Ereignisse eines Experiments nennen wir Ereignisraum.

Beispiele:

- ▶ Würfeln mit einem Würfel:  $\Omega = \{1, 2, 3, 4, 5, 6\}$ .
- ▶ Würfeln mit zwei Würfeln:  
 $\Omega = \{(1, 1), \dots, (1, 6), (2, 1), \dots, (2, 6), (3, 1), \dots, (3, 6), \dots, (6, 1), \dots, (6, 6)\}$ .
- ▶ Messung der Inflationsrate (in Prozent und ohne Nachkommastellen):  
 $\Omega = \{0, 1, 2, 100\}$ .
- ▶ Noten im Fach Wahrscheinlichkeit und Statistik:  $\Omega = \{1, 2, 3, 4, 5\}$ .

## Ereignisse $A \subseteq \Omega$

Ereignisse sind Teilmengen von  $\Omega$ .

Beispiele für Ereignisse beim Würfeln mit einem Würfel:

- ▶ Ungerade Zahlen:  $\{1, 3, 5\}$ .
- ▶ Primzahlen:  $\{2, 3, 5\}$ .
- ▶ Unmögliches Ereignis:  $\emptyset$ .
- ▶ Sicheres Ereignis:  $\Omega$ .
- ▶ Eins:  $\{1\}$ .

## Elementarereignis $\omega$

Ein Elementarereignis  $\omega$  ist eine einelementige Teilmenge von  $\Omega$ .

## Zufallsvariable

Eine Zufallsvariable ist eine Funktion, welche jedem möglichen Elementarereignis eines Zufallsexperiments einen Wert zuordnet.

# Axiomatische Definition (Axiome von Kolmogorow)

Eine Funktion  $P$  heißt Wahrscheinlichkeitsverteilung (oder auch Wahrscheinlichkeitsfunktion), wenn drei Bedingungen erfüllt sind:

- ▶ Nichtnegativität für  $A \subseteq \Omega$ :  $P(A) \geq 0$ .
- ▶ Normiertheit:  $P(\Omega) = 1$ .
- ▶ Additivität für  $A \cap B = \emptyset$ :  $P(A \cup B) = P(A) + P(B)$ .



## Wahrscheinlichkeitsverteilung (engl.: probability distribution)

Eine Wahrscheinlichkeitsverteilung  $P$  einer Zufallsvariable ist eine Funktion, welche jedem Ereignis  $A$  eine Wahrscheinlichkeit  $P(A)$  zuweist.

## Gleichverteilung (engl.: uniform distribution)

Eine Wahrscheinlichkeitsverteilung  $P$  wird Gleichverteilung genannt, falls es nur endlich viele Ereignisse gibt und alle gleich wahrscheinlich sind. Für alle Ereignisse  $A$  gilt:

$$P(A) = \frac{|A|}{|\Omega|}$$

.

Beispiele:

- Wahrscheinlichkeit einer ungeraden Zahl beim Würfeln mit einem Würfel:

$$P(\{1, 3, 5\}) = \frac{|\{1, 3, 5\}|}{|\Omega|} = \frac{3}{6}.$$

- Zweimaliges Würfeln mit einem Würfel ergibt die Summe 5:

$$P(\{(1, 4), (2, 3), (3, 2), (4, 1)\}) = \frac{4}{36}.$$

## Additionstheorem

Das Additionstheorem besagt für  $A, B \subseteq \Omega$ :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Überprüfe dieses Theorem beim Würfeln mit einem Würfel mit:

- ▶  $P(\{1, 3, 4\} \cup \{2, 6\})$ .
- ▶  $P(\{1, 3, 4\} \cup \{2, 3, 4\})$ .

## Unmögliches und sicheres Ereignis

Das unmögliche Ereignis tritt nie ein und das sichere Ereignis tritt immer ein. Somit gilt:

$$P(\emptyset) = 0$$

$$P(\Omega) = 1.$$

Für ein Ereignis  $A$  gilt mithilfe des Additionstheorems:

$$P(A) + P(A^C) = P(A \cup A^C) = P(\Omega) = 1.$$

Und daher

$$P(A^C) = 1 - P(A).$$

**Annahme:** Wir wissen, dass eine gerade Zahl gewürfelt wird und daher sind 1, 3 und 5 als Ausgänge nicht weiter möglich. 2, 4 und 6 sind weiterhin gleichverteilt. Also gilt für  $B = \{2, 4, 6\}$ :

- ▶  $P(\{1\}|B) = P(\{3\}|B) = P(\{5\}|B) = 0$ ,
- ▶  $P(\{2\}|B) = P(\{4\}|B) = P(\{6\}|B) = \frac{1}{3}$ .

$B^C$  kann aus allen möglichen Ergebnissen entfernt werden, indem wir mit  $B$  schneiden. Zum Beispiel erhalten wir für  $A = \{3, 4, 5, 6\}$ :

- ▶  $A \cap B = \{4, 6\}$ ,
- ▶  $\Omega \cap B = \{2, 4, 6\}$ .

Die bedingte Wahrscheinlichkeit von  $A$  gegeben  $B$  kann demnach wie folgt berechnet werden:

$$P(A|B) = \frac{\# \text{ der noch möglichen Elemente in } A}{\# \text{ der noch möglichen Elemente in } \Omega} = \frac{|A \cap B|}{|\Omega \cap B|} = \frac{2}{3}.$$

## Bedingte Wahrscheinlichkeit im Fall einer Gleichverteilung

Im Fall einer Gleichverteilung wird die bedingte Wahrscheinlichkeit von A gegeben B wie folgt berechnet:

$$P(A|B) = \frac{\# \text{ der noch möglichen Elemente in } A}{\# \text{ der noch möglichen Elemente in } \Omega} = \frac{|A \cap B|}{|\Omega \cap B|}.$$

Diese Formel kann auch wie folgt umformuliert werden:

$$P(A|B) = \frac{|A \cap B|}{|\Omega \cap B|} = \frac{\frac{|A \cap B|}{|\Omega|}}{\frac{|\Omega \cap B|}{|\Omega|}} = \frac{P(A \cap B)}{P(\Omega \cap B)} = \frac{P(A \cap B)}{P(B)}.$$

## Bedingte Wahrscheinlichkeit (engl.: *conditional probability*)

Die bedingte Wahrscheinlichkeit von A gegeben B wird im allgemeinen Fall wie folgt berechnet:

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

## Unabhängigkeit von Ereignissen

Ereignis  $A$  und  $B$  sind unabhängig, falls das Folgende gilt:

$$P(A \cap B) = P(A) \cdot P(B).$$

Falls  $A$  und  $B$  unabhängig sind und  $P(B) \neq 0$  gilt:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \cdot P(B)}{P(B)} = P(A).$$

Demnach entspricht die bedingte Wahrscheinlichkeit gleich der unbedingten Wahrscheinlichkeit.

### Beispiel 1:

- ▶ Wir würfeln mit einem Würfel. Sind die Ereignisse  $A = \{2, 4, 6\}$  und  $B = \{4, 5, 6\}$  unabhängig?
- ▶  $P(A \cap B) = P(\{4, 6\}) = \frac{2}{6} = \frac{1}{3}$ .
- ▶  $P(A) \cdot P(B) = \frac{3}{6} \cdot \frac{3}{6} = \frac{9}{36} = \frac{1}{4}$ .
- ▶ **Lösung:** Die Ereignisse sind abhängig.

### Beispiel 2:

- ▶ Wir würfeln zweimal mit einem Würfel. Sind die Ereignisse  $A = \{(1, 1), \dots, (1, 5), (1, 6)\}$  (Eins bei 1. Wurf) und  $B = \{(1, 1), \dots, (5, 1), (6, 1)\}$  (Eins bei 2. Wurf) unabhängig?
- ▶  $P(A \cap B) = P(\{(1, 1)\}) = \frac{1}{36}$ .
- ▶  $P(A) \cdot P(B) = \frac{6}{36} \cdot \frac{6}{36} = \frac{1}{36}$ .
- ▶ **Lösung:** Die Ereignisse sind unabhängig.

## Multiplikationssatz

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

$$P(A \cap B \cap C) = P(A \cap B) \frac{P((A \cap B) \cap C)}{P(A \cap B)} = P(A)P(B|A)P(C|(A \cap B)).$$

Mithilfe des Distributivgesetzes kann das Folgende hergeleitet werden:

## Gesetz der totalen Wahrscheinlichkeit

$$\begin{aligned} P(A) &= P(A \cap \Omega) = P(A \cap (B \cup B^C)) \\ &= P((A \cap B) \cup (A \cap B^C)) \\ &= P(A \cap B) + P(A \cap B^C) \\ &= P(B)P(A|B) + P(B^C)P(A|B^C). \end{aligned}$$



# Satz von Bayes

Der Satz von Bayes für  $B \cup B^C = \Omega$  lautet:

$$\begin{aligned} P(B|A) &= \frac{P(B \cap A)}{P(A)} = \frac{P(B \cap A)}{P(B)P(A|B) + P(B^C)P(A|B^C)} \\ &= \frac{P(B)P(A|B)}{P(B)P(A|B) + P(B^C)P(A|B^C)}. \end{aligned}$$

Falls  $B_1, \dots, B_n$  paarweise disjunkt und  $\cup_i B_i = \Omega$ , dann gilt für  $i \in \{1, \dots, n\}$ :

$$\begin{aligned} P(B_i|A) &= \frac{P(B_i \cap A)}{P(A)} = \frac{P(B_i \cap A)}{P(B_1)P(A|B_1) + \dots + P(B_n)P(A|B_n)} \\ &= \frac{P(B_i)P(A|B_i)}{P(B_1)P(A|B_1) + \dots + P(B_n)P(A|B_n)}. \end{aligned}$$

## Diskrete Zufallsvariable

Eine diskrete Zufallsvariable  $X$  nimmt nur endlich viele oder abzählbar unendlich viele Werte  $x_1, x_2, \dots, x_n$  an.

Anmerkung: Eine Menge wird abzählbar unendlich genannt, falls sie zur Menge  $\mathbb{N}$  gleichmächtig ist. Primzahlen sind zum Beispiel abzählbar unendlich, da sie eine Teilmenge von  $\mathbb{N}$  sind.  $\mathbb{R}$  hingegen ist überabzählbar.

## Stetige Zufallsvariable

Eine stetige Zufallsvariable nimmt unendlich viele, nicht abzählbare Werte an.

## Wahrscheinlichkeitsverteilung (pdf) einer diskreten Zufallsvariable

Die diskrete Zufallsvariable  $X$  kann Werte  $x_1, \dots, x_n$  annehmen. Der Wert  $x_i$ , zu welchem das Ereignis  $X = x_i$  gehört, tritt mit Wahrscheinlichkeit

$$p_i = P(X = x_i)$$

ein. Die möglichen Realisierungen  $x_i$  gemeinsam mit den Wahrscheinlichkeiten  $p_i$  nennen wir **Wahrscheinlichkeitsverteilung** (pdf) oder **Verteilung** der Zufallsvariable  $X$ .

### Beispiele:

- ▶ Einmaliges Werfen eines fairen Würfels:  $(x_1, p_1) = (1, \frac{1}{6})$ , ...,  $(x_6, p_6) = (6, \frac{1}{6})$ .
- ▶ Einmaliges Werfen einer fairen Münze:  $(x_1, p_1) = (\text{Kopf}, \frac{1}{2})$ ,  $(x_2, p_2) = (\text{Zahl}, \frac{1}{2})$ .

## Kumulative Verteilungsfunktion (cdf) einer diskreten Zufallsvariable

Die diskrete Zufallsvariable  $X$  kann Werte  $x_1, \dots, x_n$  annehmen. Wir nennen für  $x \in \mathbb{R}$  die Funktion

$$F(x) = P(X \leq x) = \sum_{\{i: x_i \leq x\}} p_i$$

**kumulative Verteilungsfunktion** (cdf) der Zufallsvariable  $X$ .

## Zusammenhang pdf und cdf am Beispiel eines fairen Würfels

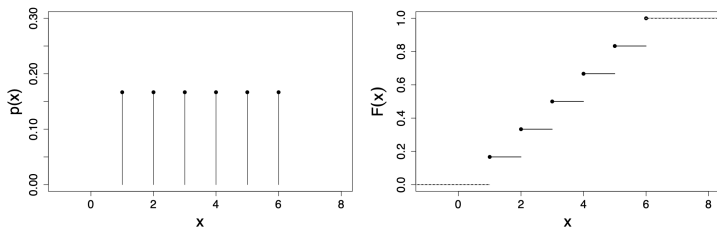


Abb.: Links die pdf und rechts die cdf.

**Anmerkung:** Die kumulative Verteilungsfunktion hat für eine diskrete Zufallsvariable  $X$  immer die Form einer Treppenfunktion.

## Stetige Zufallsvariable

Wir nennen eine Zufallsvariable  $X$  stetig, wenn ihre kumulative Verteilungsfunktion  $F(x) = P(X \leq x)$  stetig ist.

## Dichtefunktion (pdf) einer stetigen Zufallsvariable $X$

Die stetige Zufallsvariable  $X$  besitzt eine (stückweise) differenzierbare Verteilungsfunktion  $F(x)$ . Wir nennen

$$f(x) = F'(x)$$

die Dichtefunktion. Zudem gilt:

$$F(x) = \int_{-\infty}^x f(s) ds.$$

## Zusammenhang pdf und cdf einer stetigen Zufallsvariable

Siehe <https://demonstrations.wolfram.com/ConnectingTheCDFAndThePDF/>.

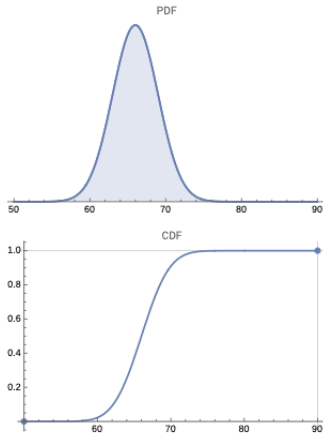


Abb.: Oben die pdf und unten die cdf.

## Anmerkungen zur pdf einer stetigen Zufallsvariable

Sei  $X$  eine stetige Zufallsvariable und  $f(x)$  ihre Dichtefunktion.

- ▶ Die Dichtefunktion ist normiert und es gilt  $\int_{-\infty}^{+\infty} f(s) ds = 1$ .
- ▶  $f(x)$  ist auf ganz  $\mathbb{R}$  definiert und nimmt nur nichtnegative Werte an.
- ▶ Wir nennen  $\{x \in \mathbb{R} : f(x) > 0\}$  den Träger der Dichtefunktion  $f$ .
- ▶ Es gilt für stetige Zufallsvariablen:  $P(X = x_i) = 0$ . Die Punktwahrscheinlichkeit ist demnach stets 0.



## Anmerkungen zur cdf einer stetigen Zufallsvariable

Sei  $X$  eine stetige Zufallsvariable und  $F(x)$  ihre Verteilungsfunktion.

- ▶ Die cdf ist monoton steigend und es gilt:
  - ▶  $\lim_{x \rightarrow -\infty} F(x) = 0$ .
  - ▶  $\lim_{x \rightarrow \infty} F(x) = 1$ .
  - ▶ Für  $x < y$ :  $F(x) \leq F(y)$ .
- ▶ Falls die cdf an einer Stelle springt, entspricht die Höhe des Sprungs genau der Wahrscheinlichkeit, mit welcher das Ereignis an jener Stelle eintritt.
- ▶ Die Verteilungsfunktion  $F(x)$  ist rechtsseitig stetig.

## Berechnung von Quantilen

Sei  $X$  eine stetige oder diskrete Zufallsvariable. Wir nennen den Wert  $x_p \in \mathbb{R}$  mit  $F(x_p) = p$  das  $p$ -Quantil von  $X$ .

## Intervallswahrscheinlichkeit einer stetigen Zufallsvariable

Sei  $X$  eine stetige Zufallsvariable und die Wahrscheinlichkeit eines Ereignisses im Intervall  $[s_1, s_2]$  ist

$$P(s_1 \leq x \leq s_2) = F(s_2) - F(s_1) = \int_{s_1}^{s_2} f(s) ds.$$

**Anmerkung:** Analoges gilt für  $[s_1, s_2)$ ,  $(s_1, s_2]$  oder  $(s_1, s_2)$ . Die rechte Seite von oben bleibt dieselbe, da die Wahrscheinlichkeit  $P(X = s_i)$  für  $i \in \{1, 2\}$  einer stetigen Zufallsvariable Null ist.

## Erwartungswert von Zufallsvariablen

Sei  $X$  eine diskrete oder stetige Zufallsvariable. Der Erwartungswert von  $X$  wird wie folgt berechnet:

$$\mu = E(X) = \begin{cases} \sum_{i=1}^n x_i \cdot p_i & , \text{ für } X \text{ diskret,} \\ \int_{-\infty}^{+\infty} x \cdot f(x) dx & , \text{ für } X \text{ stetig.} \end{cases}$$

Zudem gilt für  $a, b \in \mathbb{R}$  und  $X$  und  $Y$  Zufallsvariablen:

$$E(a \cdot X + b \cdot Y) = a \cdot E(X) + b \cdot E(Y).$$

Falls die Zufallsvariablen  $X$  und  $Y$  unabhängig sind, gilt

$$E(X \cdot Y) = E(X) \cdot E(Y).$$

Für eine Zufallsvariable  $X$ , die eine symmetrische Dichtefunktion um einen Wert  $m$  hat, gilt:  $E(X) = m$ .

## Erwartungswert einer Funktion $g(\cdot)$

Sei  $X$  eine diskrete oder stetige Zufallsvariable und  $g(\cdot)$  eine beliebige reelle Funktion. Es gilt:

$$E(g(X)) = \begin{cases} \sum_{i=1}^n g(x_i) \cdot p_i & , \text{ für } X \text{ diskret,} \\ \int_{-\infty}^{+\infty} g(x) \cdot f(x) dx & , \text{ für } X \text{ stetig.} \end{cases}$$

**Beispiel 1:** Ein Zufallsgenerator erzeugt eine reelle Zahl zwischen 0 und 4. Die Dichtefunktion der Zufallsvariable ist:

$$f(x) = \begin{cases} 0 & \text{für } x < 0 \\ \frac{1}{4} & \text{für } 0 \leq x \leq 4 \\ 0 & \text{für } x > 4. \end{cases}$$

Berechne den Erwartungswert.

### Lösung Beispiel 1:

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x \cdot f(x) dx = \int_{-\infty}^0 x \cdot 0 dx + \int_0^4 x \cdot \frac{1}{4} dx + \int_4^{+\infty} x \cdot 0 dx \\ &= \left[ \frac{1}{4} \cdot \frac{x^2}{2} \right]_0^4 \\ &= \frac{16}{8} - 0 \\ &= 2. \end{aligned}$$

**Beispiel 2:** Pro Jahr treten folgende Anzahl an Hitzetage mit entsprechenden Wahrscheinlichkeiten auf.

Anzahl	0	1	2
$p_i$	0.1	0.5	0.4

Dies verursacht folgende Kosten in tausend Euro bei einem Bauunternehmer:

$$K(x) = 300 - \frac{100}{5 + 2 \cdot x}.$$

Was sind die zu erwartenden Kosten pro Jahr?

## Lösung Beispiel 2:

$$\begin{aligned} E(K(X)) &= \sum_{i=0}^2 K(x_i) \cdot p_i \\ &= 0.1 \cdot \left(300 - \frac{100}{5 + 2 \cdot 0}\right) + 0.5 \cdot \left(300 - \frac{100}{5 + 2 \cdot 1}\right) + 0.4 \cdot \left(300 - \frac{100}{5 + 2 \cdot 2}\right) \\ &= 286.41 \text{ tsd Euro} \end{aligned}$$



## Varianz von Zufallsvariablen

Sei  $X$  eine diskrete oder stetige Zufallsvariable mit  $E(X) = \mu$ . Die Varianz von  $X$  berechnet sich wie folgt:

$$\sigma^2 = \text{Var}(X) = \begin{cases} \sum_{i=1}^n (x_i - \mu)^2 \cdot p_i & , \text{ für } X \text{ diskret,} \\ \int_{-\infty}^{+\infty} (x - \mu)^2 \cdot f(x) dx & , \text{ für } X \text{ stetig.} \end{cases}$$

Zudem gilt für  $a, b \in \mathbb{R}$  das Folgende:

$$\text{Var}(a \cdot X + b) = a^2 \cdot \text{Var}(X).$$

Außerdem gilt:

- ▶  $\text{Var}(X) = E((X - \mu)^2)$
- ▶  $E((X - \mu)^2) = E(X^2) - \mu^2.$

## Kovarianz von zwei Zufallsvariablen

Seien  $X$  und  $Y$  Zufallsvariablen mit Erwartungswerten  $E(X) = \mu_X$  und  $E(Y) = \mu_Y$ . Die Kovarianz ist wie folgt definiert:

$$\text{Cov}(X, Y) = E((X - \mu_X)(Y - \mu_Y)).$$

Für  $a, b \in \mathbb{R}$  gilt:

$$\text{Var}(X + Y) = \text{Var}(X) + 2 \cdot \text{Cov}(X, Y) + \text{Var}(Y).$$

Zwei Zufallsvariablen  $X$  und  $Y$  heißen unkorreliert, wenn  $\text{Cov}(X, Y) = 0$  ist.

Falls  $X$  und  $Y$  unabhängig sind, gilt ebenfalls

$$\text{Cov}(X, Y) = 0.$$

Demnach sind unabhängige Zufallsvariablen auch unkorreliert. Aus der Unkorreliertheit folgt aber nicht zwingend, dass die Zufallsvariablen unabhängig sind.

**Anmerkung:** Die Kovarianz ist ein Maß für die gegenseitige Abhängigkeit von  $X$  und  $Y$ . Sie gibt die Richtung aber nicht die Stärke des Zusammenhangs an.

## Korrelation von zwei Zufallsvariablen

Um die Stärke des Zusammenhangs zu bestimmen, wird die Korrelation verwendet.

Für Zufallsvariablen  $X$  und  $Y$  mit Erwartungswert  $E(X) = \mu_X$  und  $E(Y) = \mu_Y$  sowie Varianz  $\text{Var}(X) = \sigma_X^2$  und  $\text{Var}(Y) = \sigma_Y^2$  wird die Korrelation wie folgt berechnet:

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

**Anmerkung:** Die Standardabweichung  $\sigma$  ergibt sich als Wurzel aus der Varianz  $\sigma^2$ .

## Schiefe

Die Schiefe gibt an, ob und wie stark die Verteilung sich nach rechts oder nach links neigt.

Sei  $X$  eine Zufallsvariable mit Erwartungswert  $\mu$  und Standardabweichung  $\sigma$ . Die Schiefe (auch drittes zentrales Moment genannt) ist wie folgt definiert:

$$\gamma = E\left(\left(\frac{X - \mu}{\sigma}\right)^3\right).$$

Für eine symmetrische Verteilung gilt:  $\gamma = 0$ . Falls  $\gamma < 0$  liegt eine linksschiefe Verteilung vor. Bei  $\gamma > 0$  liegt eine rechtsschiefe Verteilung vor.

Es gilt:

- ▶ **Rechtsschief** ist identisch mit dem Begriff **linkssteil**.
- ▶ **Linksschief** ist identisch mit dem Begriff **rechtssteil**.

## Wölbung

Die Wölbung (auch Kurtosis genannt) gibt die Steilheit einer Verteilung an. Sie wird für eine Zufallsvariable  $X$  mit Erwartungswert  $\mu$  und Standardabweichung  $\sigma$  wie folgt berechnet:

$$\kappa = E\left(\left(\frac{X - \mu}{\sigma}\right)^4\right).$$

**Anmerkung:** Für eine Normalverteilung gilt  $\kappa = 3$ . Falls  $\kappa > 3$  ist die vorliegende Verteilung spitzer als die Normalverteilung. Für  $\kappa < 3$  ist sie flachgipfliger.

## Wichtige Wahrscheinlichkeitsverteilungen

Die wichtigsten Verteilungen lernen wir im **DataCamp Kurs** kennen. Diese sind unter anderem:

- ▶ Diskrete Gleichverteilung
- ▶ Geometrische Verteilung
- ▶ Normalverteilung
- ▶ Poisson-Verteilung
- ▶ t-Verteilung
- ▶ Exponentialverteilung
- ▶ ...

## Identisch verteilte Zufallsvariablen

Seien  $X$  und  $Y$  Zufallsvariablen. Sie heißen **identisch verteilt**, wenn beide dieselbe Verteilungsfunktion (cdf) haben. Demnach muss  $\forall x \in \mathbb{R}$  gelten:

$$P(X \leq x) = P(Y \leq x).$$

Für identisch verteilte Zufallsvariablen  $X_1, \dots, X_n$  mit

$$\mu = E(X_1) = \dots = E(X_n)$$

ist der Erwartungswert des arithmetischen Mittels  $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$  gegeben durch

$$\begin{aligned} E(\bar{X}) &= \frac{1}{n}(E(X_1) + \dots + E(X_n)) \\ &= \frac{1}{n}(\mu + \dots + \mu) \\ &= \mu. \end{aligned}$$

### i.i.d.

$X_1, \dots, X_n$  werden **i.i.d.** (engl: independent and identically distributed) genannt, falls sie unabhängig und identisch verteilt sind.

## Das Gesetz der großen Zahlen

Seien  $X_1, \dots, X_n$  i.i.d. Zufallsvariablen mit  $E(X_i) = \mu$  und  $\text{Var}(X_i) = \sigma^2$ . Das arithmetische Mittel der Zufallsvariablen berechnet sich wie folgt:  $\bar{X} = \frac{X_1 + \dots + X_n}{n}$  und ist ebenfalls eine Zufallsvariable. Für ein beliebiges  $\epsilon > 0$  gilt:

$$\lim_{n \rightarrow \infty} P(|\bar{X} - \mu| < \epsilon) = 1.$$

### Beispiel:

Anzahl Würfe faire Münze	10	100	1000	10000
Relative Häufigkeit Zahl	0.35	0.42	0.52	0.49

**Achtung:** Es gibt kein Gesetz des Ausgleichs. Falls zum Beispiel beim Roulette häufig die Farbe schwarz der Ausgang war, sagt das Gesetz der großen Zahlen nicht, dass die Farbe rot ihren Rückstand aufholt. Es handelt sich in jeder Runde um unabhängige Experimente.



## Der zentrale Grenzwertsatz

Sei  $X_1, X_2, \dots$  eine Folge von i.i.d. Zufallsvariablen mit  $E(X_i) = \mu$  und  $\text{Var}(X_i) = \sigma^2$ . Die  $n$ -te Teilsumme dieser Folge ist definiert als  $S_n = X_1 + X_2 + \dots + X_n$  und es gilt:

$$\begin{aligned} E(S_n) &= n \cdot \mu, \\ \text{Var}(S_n) &= n \cdot \sigma^2. \end{aligned}$$

Dann ist die standardisierte  $n$ -te Teilsumme definiert als

$$Z_n = \frac{S_n - n \cdot \mu}{\sqrt{n \cdot \sigma^2}},$$

mit  $E(Z_n) = 0$  und  $\text{Var}(Z_n) = 1$ . Ist  $\Phi$  die Verteilungsfunktion der Standardnormalverteilung  $N(0, 1)$ , dann gilt für beliebiges  $x \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} P(Z_n \leq x) = \Phi(x).$$

**Anmerkung:** Unerheblich welche Verteilung  $X_1, X_2, \dots$  hat, die  $n$ -te Teilsumme ist für hinreichend großes  $n$  standardnormalverteilt.

# Induktive Statistik

## Induktive Statistik

Induktive Statistik, auch schließende Statistik genannt, ist ein Teilbereich der Statistik, der darauf abzielt, von einer Stichprobe auf die Grundgesamtheit zu schließen. Die induktive Statistik wird verwendet, um auf Basis von Daten einer Stichprobe Aussagen über die Population zu machen, ohne alle Daten der Population zu kennen.

## Stichprobe

Eine Stichprobe ist eine Teilmenge von Objekten, die aus der Grundgesamtheit zufällig ausgewählt werden. Stichproben werden in der Statistik verwendet, um Daten zu erheben, ohne die Grundgesamtheit untersuchen zu müssen.

Eine zufällige Stichprobe vom Umfang  $n$  ist eine Folge  $X_1, X_2, \dots, X_n$  von unabhängig, identisch verteilten Zufallsvariablen, wobei  $X_i$  die Ausprägung des Merkmals des  $i$ -ten Objekts beschreibt. Wir nennen  $X_i$  eine Stichprobenvariable.

## Schätzfunktion

Eine Funktion  $T(X_1, \dots, X_n)$  nennen wir Stichprobenfunktion und auch diese Funktion ist ebenfalls eine Zufallsvariable. Falls wir sie für die Schätzung eines Parameters  $\psi$  der Grundgesamtheit verwenden, nennen wir sie Schätzfunktion bzw. Schätzer für den Parameter  $\psi$ .

# Punktschätzer

## Punktschätzer

Aus Stichproben können ein oder auch mehrere Werte errechnet werden (z.B. Median und Varianz), die Schätzwerte für die Grundgesamtheit darstellen. Wir sprechen in diesem Fall von Punktschätzern, weil sie jeweils genau einen Wert schätzen.

## Eigenschaften von Schätzern

- ▶ Erwartungstreue:  $E(T) = \psi$ . In Worten: Der Schätzer trifft im Durchschnitt den wahren Wert  $\psi$ . Das bedeutet, dass der Schätzer im Durchschnitt den wahren Wert liefert, auch wenn er bei einzelnen Schätzungen vom wahren Wert abweicht.
- ▶ Effizienz: Je kleiner die Varianz des Schätzers, desto näher wird ein Schätzwert am wahren Wert  $\psi$  liegen. Ein effizienter Schätzer ist derjenige, der die geringstmögliche Varianz aufweist.
- ▶ Konsistenz: Mit wachsender Stichprobe werden die Abweichungen vom wahren Wert kleiner.

## Bias (Verzerrung)

Ein Schätzer, der nicht erwartungstreu ist, ist verzerrt. Die Verzerrung (auch Bias genannt) eines Schätzers ist die Differenz zwischen dem Erwartungswert des Schätzers und dem wahren Wert:  $Bias(T) = E(T) - \psi$ . Wir nennen einen Schätzer unverzerrt, falls  $Bias(T) = 0$  ist.

## Beispiele für erwartungstreue Schätzer

- ▶ Das arithmetische Mittel  $\bar{x}$  ist ein erwartungstreuer Schätzer des Erwartungswertes der Grundgesamtheit.
- ▶ Die empirische Varianz  $s^2$  ist ein erwartungstreuer Schätzer für die Varianz der Grundgesamtheit.
- ▶ Die korrigierte empirische Kovarianz  $s_{xy}$  ist ein erwartungstreuer Schätzer der Kovarianz der Grundgesamtheit.

## Erwartungstreue des arithmetischen Mittels

Seien  $X_1, \dots, X_n$  i.i.d. mit  $E(X_1) = \dots = E(X_n) = \mu$ . Das arithmetische Mittel der Zufallsvariablen berechnet sich wie folgt  $\bar{X} = \frac{X_1 + \dots + X_n}{n}$  und ist ein erwartungstreuer Schätzer des Erwartungswerts der Grundgesamtheit, da das Folgende gilt:

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{X_1 + \dots + X_n}{n}\right) \\ &= \frac{1}{n} E(X_1 + \dots + X_n) \\ &= \frac{1}{n} (E(X_1) + \dots + E(X_n)) \\ &= \frac{1}{n} \cdot n \cdot \mu \\ &= \mu. \end{aligned}$$

## Erwartungstreue der empirischen Varianz $s^2$

Seien  $X_1, \dots, X_n$  i.i.d. mit  $E(X_1) = \dots = E(X_n) = \mu$  und  $\text{Var}(X_1) = \dots = \text{Var}(X_n) = \sigma^2$ .  
 $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  ist ein erwartungstreuer Schätzer für die Varianz, da:

$$\begin{aligned} E(s^2) &= \frac{1}{n-1} E\left(\sum_{i=1}^n (X_i - \bar{X})^2\right) \\ &= \frac{1}{n-1} E\left(\sum_{i=1}^n ((X_i - \mu) - (\bar{X} - \mu))^2\right) \\ &= \frac{1}{n-1} E\left(\sum_{i=1}^n ((X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2)\right) \\ &= \frac{1}{n-1} E\left(\left(\sum_{i=1}^n (X_i - \mu)^2\right) - 2(\bar{X} - \mu)\left(\sum_{i=1}^n (X_i - \mu)\right) + n(\bar{X} - \mu)^2\right) \\ &= \frac{1}{n-1} E\left(\left(\sum_{i=1}^n (X_i - \mu)^2\right) - 2(\bar{X} - \mu)n\frac{1}{n}\left(\sum_{i=1}^n (X_i - \mu)\right) + n(\bar{X} - \mu)^2\right) \\ &= \frac{1}{n-1} E\left(\left(\sum_{i=1}^n (X_i - \mu)^2\right) - 2n(\bar{X} - \mu)^2 + n(\bar{X} - \mu)^2\right) \\ &= \frac{1}{n-1} E\left(\left(\sum_{i=1}^n (X_i - \mu)^2\right) - n(\bar{X} - \mu)^2\right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n E((X_i - \mu)^2) - nE((\bar{X} - \mu)^2)\right) = \frac{1}{n-1} (n\sigma^2 - n\frac{\sigma^2}{n}) = \sigma^2. \end{aligned}$$



# Intervallschätzung

Warum könnte es Sinn machen Punktschätzer durch Intervallschätzer zu ersetzen?

Betrachten wir eine Stichprobe  $X_1, \dots, X_n$  aus einer  $N(\mu, 1)$  verteilten Population.

- ▶ Wir kennen  $\mu$  nicht, können diesen Parameter aber mit  $\bar{X}$  schätzen.
- ▶ Es gilt  $P(\bar{X} = \mu) = 0$ .
- ▶ Beispiel eines Intervallschätzers ist  $[\bar{X} - 1; \bar{X} + 1]$ .
- ▶ Wir verlieren an Genauigkeit der Schätzung.
- ▶ Andererseits erhalten wir mit positiver Wahrscheinlichkeit den korrekten Parameterwert, da  $P(\mu \in [\bar{X} - 1; \bar{X} + 1]) > 0$ .

## Intervallschätzung

- ▶ Die Intervallschätzung verfolgt das Ziel, einen unbekannten Parameter zu schätzen.
- ▶ Sie liefert uns ein Intervall, das den wahren Wert mit einer bestimmten Wahrscheinlichkeit enthält.
- ▶ Einen wahren Wert genau zu treffen ist schwierig und daher wird eine Bandbreite angegeben.

## Gebräuchlichste Verfahren: Konfidenzintervall KI

- ▶ Ein KI ist ein Intervall, das den wahren Parameter mit einer bestimmten Wahrscheinlichkeit enthält.
- ▶ Die Grenzen des Intervalls berechnen sich aus den Werten der Stichprobe.
- ▶ Formal kann das Konfidenzintervall wie folgt dargestellt werden:  
 $[T_u(X_1, \dots, X_n); T_o(X_1, \dots, X_n)]$ .
- ▶ Beispiel: Die Varianz liegt mit 90 % Wahrscheinlichkeit zwischen [50, 60].
- ▶ Beispiele:  
[https://shinyapps.wiwi.hu-berlin.de/mmstat\\_en/confidence\\_mean/](https://shinyapps.wiwi.hu-berlin.de/mmstat_en/confidence_mean/).

## Konfidenzniveau und Irrtumsniveau

- Das **Konfidenzniveau** gibt an, mit welcher Wahrscheinlichkeit der zu schätzende Parameter im Intervall enthalten ist.

Zum Beispiel für das Konfidenzniveau  $1 - \alpha = 0.95$ :

$$P(\psi \in [T_u(X_1, \dots, X_n); T_o(X_1, \dots, X_n)]) = 0.95.$$

- Die **Irrtumswahrscheinlichkeit** gibt die Wahrscheinlichkeit an, dass der zu schätzende Parameter nicht im Intervall liegt.

Zum Beispiel für die Irrtumswahrscheinlichkeit  $\alpha = 0.05$ :

$$P(\psi \notin [T_u(X_1, \dots, X_n); T_o(X_1, \dots, X_n)]) = 0.05.$$

## Zweiseitige Konfidenzintervalle zum Konfidenzniveau $1 - \alpha$ eines normalverteilten Merkmals

**Anmerkung:**  $n$  beschreibt die Stichprobengröße,  $\bar{x}$  das arithmetische Mittel und  $s$  die empirische Standardabweichung.

- KI für den Erwartungswert  $\mu$ , falls  $\sigma^2$  der Population bekannt ist:

$$\left[ \bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}; \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$$

**Anmerkung:**  $z_{1-\frac{\alpha}{2}}$  schneidet die oberen  $\frac{\alpha}{2}$  der Fläche unter der Normalverteilung ab und ist das  $(1 - \frac{\alpha}{2})$ -Quantil der Standardnormalverteilung.

- KI für den Erwartungswert  $\mu$ , falls  $\sigma^2$  der Population unbekannt ist:

$$\left[ \bar{x} - t_{n-1;1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}; \bar{x} + t_{n-1;1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right]$$

**Anmerkung:**  $t_{n-1;1-\frac{\alpha}{2}}$  ist das  $(1 - \frac{\alpha}{2})$ -Quantil der t-Verteilung mit  $n - 1$  Freiheitsgraden.

- KI für die Varianz  $\sigma^2$ :

$$\left[ \frac{(n-1)s^2}{\chi^2_{n-1; \frac{\alpha}{2}}}, \frac{(n-1)s^2}{\chi^2_{n-1; 1-\frac{\alpha}{2}}} \right]$$

**Anmerkung:**  $\chi^2_{n-1}$  beschreibt die Chi-Quadrat-Verteilung mit  $n - 1$  Freiheitsgraden.

## Freiheitsgrad

Ein Freiheitsgrad gibt die Anzahl frei wählbarer Werte für einen Parameter an. Die Anzahl der Freiheitsgrade nimmt mit zunehmender Stichprobengröße zu und fällt mit der Anzahl geschätzter Parameter.

Zum Beispiel haben drei Maschinen folgendes Gewicht: 110 kg, 120 kg und 130 kg. Der Mittelwert ist 120 kg. Basierend auf dem Mittelwert könnten die Maschinen auch folgende Gewichte aufweisen: 100 kg, 120 kg, 140 kg. Dabei sind nur die ersten beiden Gewichte der Maschinen frei wählbar und das dritte Gewicht ergibt sich aus den ersten und dem Mittelwert. In diesem Beispiel ergeben sich somit für die Verteilung zwei Freiheitsgrade.

## Einseitige Konfidenzintervalle zum Konfidenzniveau $1 - \alpha$ eines normalverteilten Merkmals

- KI für den Erwartungswert  $\mu$ , falls  $\sigma^2$  der Population bekannt ist:

$$\left( -\infty; \bar{x} + z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right] \text{ oder } \left[ \bar{x} - z_{1-\alpha} \frac{\sigma}{\sqrt{n}}; +\infty \right)$$

- KI für den Erwartungswert  $\mu$ , falls  $\sigma^2$  der Population unbekannt ist:

$$\left( -\infty; \bar{x} + t_{n-1;1-\alpha} \frac{s}{\sqrt{n}} \right] \text{ oder } \left[ \bar{x} - t_{n-1;1-\alpha} \frac{s}{\sqrt{n}}; +\infty \right)$$

- KI für die Varianz  $\sigma^2$ :

$$\left( 0; \frac{(n-1)s^2}{\chi_{n-1;\alpha}^2} \right] \text{ oder } \left[ \frac{(n-1)s^2}{\chi_{n-1;1-\alpha}^2}; +\infty \right)$$



**Beispiel 1:** Wir betrachten das Merkmal Alter einer Stichprobe von 100 Personen aus Österreich. Der Mittelwert dieser Stichprobe liegt bei 43 Jahren und die Standardabweichung ist 10 Jahre. Zum Konfidenzniveau 0.95 beträgt das zweiseitige Konfidenzintervall:

$$\blacktriangleright T_u(X_1, \dots, X_{100}) = 43 - 1.96 \frac{10}{\sqrt{100}} = 41.04$$

$$\blacktriangleright T_o(X_1, \dots, X_{100}) = 43 + 1.96 \frac{10}{\sqrt{100}} = 44.96$$

Somit können wir schlussfolgern, dass das Durchschnittsalter in Österreich mit 95 % Wahrscheinlichkeit zwischen 41 und 45 Jahren liegt.

**Anmerkung 1:** Die Standardnormalverteilungstabelle ist unter <https://de.wikipedia.org/wiki/Standardnormalverteilungstabelle> zu finden.

**Anmerkung 2:** Die  $t$ -Verteilung ähnelt einer Normalverteilung und mit steigender Anzahl an Freiheitsgraden können wir die  $t$ -Verteilung mithilfe der Normalverteilung approximieren. Eine Faustregel besagt, dass ab 30 Freiheitsgraden die  $t$ -Verteilung durch die Normalverteilung approximiert werden kann.

**Beispiel 2a:** Eine Stichprobe aus 10 Studierenden des Studienfachs Mechatronik hat an einem IQ-Test teilgenommen. Das Ergebnis war: 105, 100, 123, 90, 100, 128, 105, 109, 104, 115. Wir sind am Konfidenzintervall zum Konfidenzniveau  $1 - \alpha = 90\%$  für den Erwartungswert der IQ-Werte für die Population der Studierenden des Studienfachs Mechatronik interessiert.

**Lösung:**

- ▶ Mittelwert der Stichprobe:  $\bar{x} = 107.9$ .
- ▶ Geschätzte Populationsvarianz:  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 128.99$ .
- ▶  $t$ -Wert für  $n - 1 = 9$  Freiheitsgrade und  $1 - \alpha = 0.9$ :  $t_{n-1; 1-\frac{\alpha}{2}} = t_{9; 0.95} = 1.833$ .
- ▶  $T_u(X_1, \dots, X_{10}) = 107.9 - 1.833 \cdot \frac{\sqrt{128.99}}{\sqrt{10}} = 101.32$ .
- ▶  $T_o(X_1, \dots, X_{10}) = 107.9 + 1.833 \cdot \frac{\sqrt{128.99}}{\sqrt{10}} = 114.48$ .

Demnach kann mit 90 % Sicherheit gesagt werden, dass der wahre IQ-Wert zwischen 101.32 und 114.48 liegt.

**Anmerkung:** Die Tabelle einiger  $t$ -Quantile finden wir unter [https://de.wikipedia.org/wiki/Studentsche\\_t-Verteilung](https://de.wikipedia.org/wiki/Studentsche_t-Verteilung).

**Beispiel 2b:** Wir sind am einseitigen Konfidenzintervall  $[a, +\infty)$  zum Konfidenzniveau 90 % interessiert.

**Lösung:**

- ▶ Mittelwert der Stichprobe:  $\bar{x} = 107.9$ .
- ▶ Geschätzte Populationsvarianz:  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 128.99$ .
- ▶  $t$ -Wert für  $n - 1 = 9$  Freiheitsgrade und  $1 - \alpha = 0.9$ :  $t_{n-1;1-\alpha} = t_{9;0.9} = 1.383$ .
- ▶  $T_u(X_1, \dots, X_{10}) = 107.9 - 1.383 \cdot \frac{\sqrt{128.99}}{\sqrt{10}} = 102.93$ .

Demnach kann mit 90 % Sicherheit gesagt werden, dass der wahre IQ-Wert zwischen 102.93 und  $+\infty$  liegt.

**Beispiel 3:** Uns liegt eine zufällige Stichprobe vom Umfang  $n = 20$  Haushalten vor, die ein mittleres Haushaltsnettoeinkommen von  $\bar{x} = 2800$  Euro haben. Als Punktschätzer für die unbekannte Varianz  $\sigma^2$  erhalten wir aus der Stichprobe  $s^2 = 500$ . Wir sind am zweiseitigen Konfidenzintervall für die Varianz  $\sigma^2$  zum Konfidenzniveau 95 % interessiert.

**Lösung:**

- ▶ Anzahl an Freiheitsgraden:  $n - 1 = 20 - 1 = 19$ .
- ▶  $\chi^2_{n-1; 1 - \frac{\alpha}{2}} = \chi^2_{19; 0.975} = 8.907$ .
- ▶  $\chi^2_{n-1; \frac{\alpha}{2}} = \chi^2_{19; 0.025} = 32.852$ .
- ▶ Untere Schranke:  $\frac{19 \cdot 500}{32.852} = 289.18$ .
- ▶ Obere Schranke:  $\frac{19 \cdot 500}{8.907} = 1066.58$ .

Demnach kann mit 95 % Sicherheit gesagt werden, dass die Varianz der Population zwischen 289.18 und 1066.58 liegt.

**Anmerkung:** Die Tabelle einiger Quantile der Chi-Quadrat-Verteilung finden wir unter <https://datatab.de/tutorial/tabelle-chi-quadrat> sowie unter [https://www.uibk.ac.at/econometrics/einf/tab\\_chisq.pdf](https://www.uibk.ac.at/econometrics/einf/tab_chisq.pdf).

# Hypothesentests

**Allgemein:** Ein Hypothesentest verfolgt das Ziel, die aufgestellte Hypothese auf ihre Gültigkeit zu testen. Für diesen Test muss eine Nullhypothese und eine Alternativhypothese aufgestellt werden, welche sich gegenseitig ausschließen. Die Nullhypothese wird nur abgelehnt, wenn ausreichend Evidenz gegen sie vorliegt.

**Parametrischer Test:** Falls wir beim Hypothesentest Aussagen bezüglich einem Parameter  $\psi$  eines Merkmals der Population treffen, sprechen wir von einem parametrischen Test.

- ▶ **Nullhypothese  $H_0$ :** Diese Hypothese wird getestet und stellt in der Regel die Annahme dar, dass es keinen Effekt oder keinen Unterschied gibt. Ein Beispiel für eine Nullhypothese wäre: "Der Mittelwert ist gleich 100."
- ▶ **Alternativhypothese  $H_1$ :** Sie widerspricht der Nullhypothese und besagt, dass es einen Effekt bzw. einen Unterschied gibt. Zum Beispiel: "Der Mittelwert ist ungleich 100."
- ▶ **Signifikanzniveau  $\alpha$ :** Das Signifikanzniveau, oft mit 0.05 festgelegt, beschreibt die maximale Wahrscheinlichkeit, die Nullhypothese fälschlicherweise zu verwerfen. Ein Signifikanzniveau von 0.05 bedeutet, dass wir bereit sind, in maximal 5 % der Fälle eine falsche Entscheidung zu treffen.
- ▶ **p-Wert:** Dieser Wert beschreibt die Wahrscheinlichkeit, unter der Annahme von  $H_0$ , ein Ergebnis zu erhalten, das mindestens so extrem ist, wie das beobachtete Ergebnis. Demnach wird der  $p$ -Wert durch die Stichprobe bestimmt. Ein niedriger  $p$ -Wert (z.B. kleiner als 0.05) spricht dafür, die Nullhypothese abzulehnen. Daher wird die Nullhypothese verworfen, wenn der  $p$ -Wert  $\leq \alpha$  ist.

**Einfache Hypothese:** Wir sprechen von einer einfachen Hypothese, falls die Hypothese für den Parameter  $\psi$  aus einem Wert besteht. Zum Beispiel:  $\psi = \psi_0$ .

**Zusammengesetzte Hypothese:** Wir sprechen von einer zusammengesetzten Hypothese, falls die Hypothese für den Parameter  $\psi$  aus mehreren Werten besteht. Zum Beispiel:  $\psi \leq \psi_0$ ,  $\psi \geq \psi_0$ ,  $\psi \neq \psi_0$ .

Um den Hypothesentest durchführen zu können, brauchen wir eine Stichprobe und eine Teststatistik  $T(X_1, \dots, X_n)$ , welche die Stichprobe als Input nimmt. Die Teststatistik bildet die Grundlage für die Entscheidung. Falls der Wert der Teststatistik für die Alternativhypothese spricht, lehnen wir die Nullhypothese ab.



$\alpha$ -Fehler sowie  $\beta$ -Fehler:

	Test entscheidet sich für $H_0$	Test entscheidet sich für $H_1$
$H_0$ ist wahr	✓	$\alpha$ -Fehler: Fehler 1. Art
$H_1$ ist wahr	$\beta$ -Fehler: Fehler 2. Art	✓

Den  $\alpha$ -Fehler sowie  $\beta$ -Fehler beim Testen können wir mit einem Gerichtsverfahren vergleichen, das sich ebenfalls in vier Fällen darstellen lässt:

	unschuldig	schuldig
Freispruch	✓	Freispruch wegen fehlenden Beweisen
Verurteilung	Fehlverurteilung	✓

Beim Gerichtsverfahren ohne Zeugen kann der Angeklagte seine Unschuld nicht beweisen. Beim Testen von Hypothesen ist es auch nicht möglich zu beweisen, dass die Nullhypothese gültig ist.

### Zweiseitiger Hypothesentest:

$$H_0 : \psi = \psi_0 \quad \text{vs.} \quad H_1 : \psi \neq \psi_0$$

### Einseitiger Hypothesentest:

$$H_0 : \psi \geq \psi_0 \quad \text{vs.} \quad H_1 : \psi < \psi_0$$

$$H_0 : \psi \leq \psi_0 \quad \text{vs.} \quad H_1 : \psi > \psi_0$$

Konfidenzintervalle spielen eine zentrale Rolle bei Hypothesentests. Wenn die berechnete Teststatistik innerhalb des Konfidenzintervalls liegt, wird die Nullhypothese  $H_0$  nicht verworfen. Befindet sich die Teststatistik jedoch außerhalb des Konfidenzintervalls, wird die Nullhypothese abgelehnt.

**Anmerkung 1:** Wir bezeichnen den Bereich, in dem wir  $H_0$  ablehnen als kritischen Bereich (auch Ablehnungsbereich genannt). Sein Komplement ist der Bereich, in welchem wir  $H_0$  beibehalten.

**Anmerkung 2:** Entweder wird  $H_0$  beibehalten, da wir nicht ausreichend Beweise zum Verwerfen haben oder  $H_0$  wird verworfen.  $H_0$  können wir niemals beweisen.

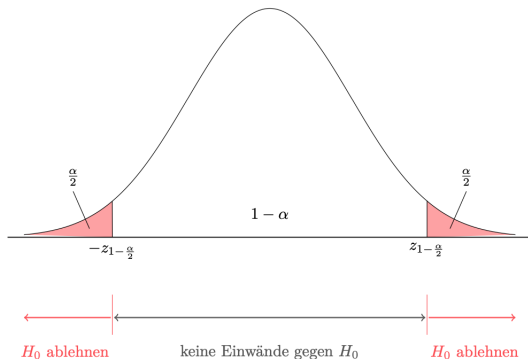


Abb.: Ablehnungsbereich und Nicht-Ablehnungsbereich der Nullhypothese.

Der **Gauß Test** wird zum Testen von Hypothesen über Mittelwerte bei bekannter Standardabweichung verwendet. Voraussetzung ist ein normalverteiltes Merkmal.

Wir wählen ein Signifikanzniveau  $\alpha$  und eine der folgenden Hypothesen (je nach Fragestellung):

1.  $H_0 : \mu = \mu_0$  vs.  $H_1 : \mu \neq \mu_0$
2.  $H_0 : \mu \geq \mu_0$  vs.  $H_1 : \mu < \mu_0$
3.  $H_0 : \mu \leq \mu_0$  vs.  $H_1 : \mu > \mu_0$ .

Die Teststatistik wird wie folgt berechnet:

$$z = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma},$$

wobei  $n$  die Stichprobengröße,  $\sigma$  die Standardabweichung der Grundgesamtheit und  $\mu_0$  der Erwartungswert unter  $H_0$  beschreibt.

Wir lehnen  $H_0$  ab, wenn

1.  $|z| > z_{1-\frac{\alpha}{2}}$
2.  $z < -z_{1-\alpha}$
3.  $z > z_{1-\alpha}$ ,

wobei  $z_\alpha$  das  $\alpha$ -Quantil der Standardnormalverteilung ist.

Der **t-Test** wird zum Testen von Hypothesen über Mittelwerte bei unbekannter Standardabweichung verwendet. Voraussetzung ist ein normalverteiltes Merkmal.

Wir wählen ein Signifikanzniveau  $\alpha$  und eine der folgenden Hypothesen (je nach Fragestellung):

1.  $H_0 : \mu = \mu_0$  vs.  $H_1 : \mu \neq \mu_0$
2.  $H_0 : \mu \geq \mu_0$  vs.  $H_1 : \mu < \mu_0$
3.  $H_0 : \mu \leq \mu_0$  vs.  $H_1 : \mu > \mu_0$ .

Die Teststatistik wird wie folgt berechnet:

$$t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{s},$$

wobei  $n$  die Stichprobengröße,  $s$  die empirische Standardabweichung und  $\mu_0$  der Erwartungswert unter  $H_0$  beschreibt.

Wir lehnen  $H_0$  ab, wenn

1.  $|t| > t_{n-1; 1-\frac{\alpha}{2}}$
2.  $t < -t_{n-1; 1-\alpha}$
3.  $t > t_{n-1; 1-\alpha}$ ,

wobei  $t_{n-1; \alpha}$  das  $\alpha$ -Quantil der  $t$ -Verteilung mit  $n - 1$  Freiheitsgraden ist.

**Beispiel 1:** Wir untersuchen das Abfüllgewicht einer 1000 Gramm Nussmischung und wollen nachweisen, dass das Abfüllgewicht kleiner als 1000 Gramm ist. Die Stichprobe ergibt folgende Werte:

$$\{980, 1000, 1010, 1000, 970, 1010, 960, 980, 1010, 960\}.$$

Wir erhalten:  $\bar{x} = 988$  und  $s = 20.44$ . Wir setzen  $\alpha = 0.05$  fest. Der einseitige Hypothesentest lautet:

$$H_0 : \mu \geq 1000 \quad \text{vs.} \quad H_1 : \mu < 1000$$

Wir erhalten aufgrund der Stichprobe folgende Teststatistik:

$$t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{s} = \frac{\sqrt{10}(988 - 1000)}{20.44} = -1.86.$$

Das 0.95-Quantil der  $t$ -Verteilung mit 9 Freiheitsgraden ist:

$$t_{n-1;1-\alpha} = t_{9;0.95} = 1.833.$$

Siehe [https://de.wikipedia.org/wiki/Studentsche\\_t-Verteilung](https://de.wikipedia.org/wiki/Studentsche_t-Verteilung).

Somit ist der Ablehnungsbereich  $A = \{r \in \mathbb{R} : r < -1.833\}$ . Die Wert der Teststatistik liegt in  $A$  und demnach lehnen wir die Nullhypothese ab und haben einen signifikanten Nachweis, dass  $\mu < 1000g$  ist.

Der  $\chi^2$  **Test** wird zum Testen von Hypothesen für die Varianz verwendet.  
Voraussetzung ist ein normalverteiltes Merkmal.

Wir wählen ein Signifikanzniveau  $\alpha$  und eine der folgenden Hypothesen (je nach Fragestellung):

1.  $H_0 : \sigma^2 = \sigma_0^2$  vs.  $H_1 : \sigma^2 \neq \sigma_0^2$
2.  $H_0 : \sigma^2 \geq \sigma_0^2$  vs.  $H_1 : \sigma^2 < \sigma_0^2$
3.  $H_0 : \sigma^2 \leq \sigma_0^2$  vs.  $H_1 : \sigma^2 > \sigma_0^2$ .

Die Teststatistik wird wie folgt berechnet:

$$c = \frac{(n-1)s^2}{\sigma_0^2},$$

wobei  $n$  die Stichprobengröße und  $s$  die empirische Standardabweichung beschreibt.

Wir lehnen  $H_0$  ab, wenn

1.  $c < \chi_{n-1; \frac{\alpha}{2}}^2$  oder  $c > \chi_{n-1; 1-\frac{\alpha}{2}}^2$
2.  $c < \chi_{n-1; \alpha}^2$
3.  $c > \chi_{n-1; 1-\alpha}^2$ ,

wobei  $\chi_{n-1; \alpha}^2$  das  $(1 - \alpha)$ -Quantil der Chi-Quadrat-Verteilung mit  $n - 1$  Freiheitsgraden ist.

**Beispiel 2:** Die Produktion eines bestimmten Impfstoffes soll 1 Milligramm Phenol enthalten. Eine Stichprobe wurde entnommen und die gemessenen Mengen an Phenol in Milligramm sind:

0.995, 1.003, 1.001, 0.998, 1.002, 0.997, 1.000, 1.004, 0.996, 0.999.

1. Muss der Produktionsprozess neu kalibriert werden oder kann weiterhin  $\mu = 1\text{mg}$  als Sollwert angenommen werden? Verwenden Sie ein Signifikanzniveau von  $\alpha = 0.05$ .
2. Die maximal zulässige Streuung für Phenol darf  $\sigma = 0.015$  nicht übersteigen. Kann dies zu einem Signifikanzniveau von  $\alpha = 0.01$  gewährleistet werden?



**Lösung Frage 2.1:** Wir erhalten  $\bar{x} = 0.9995$  und  $s = 0.00303$ . Der zweiseitige Hypothesentest lautet:

$$H_0 : \mu = 1 \quad \text{vs.} \quad H_1 : \mu \neq 1.$$

Wir erhalten aufgrund der Stichprobe folgende Teststatistik:

$$t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{s} = -0.522.$$

Das 0.975-Quantil der  $t$ -Verteilung mit 9 Freiheitsgraden ist  $t_{9;0.975} = 2.262$ .

Da  $|t| = |-0.522|$  nicht größer als 2.262 ist, wird die Nullhypothese nicht abgelehnt. Somit muss der Produktionsprozess nicht neu kalibriert werden, da kein signifikanter Unterschied festgestellt wurde.

**Lösung Frage 2.2.** Wir wollen nachweisen, dass  $\sigma$  den Wert 0.015 nicht übersteigt und wählen daher die Alternativhypothese entsprechend dieser Aussage. Daher gilt:

$$H_0 : \sigma^2 \geq 0.015^2 \quad \text{vs.} \quad H_1 : \sigma^2 < 0.015^2.$$

Die Teststatistik wird wie folgt berechnet:

$$c = \frac{(n-1)s^2}{\sigma_0^2} = \frac{9 \cdot 0.00303^2}{0.015^2} = 0.3672.$$

Wir lehnen  $H_0$  ab, wenn  $c < \chi_{n-1;\alpha}^2 = \chi_{9;0.01}^2$  gilt.

Laut Tabellen aus <https://datatab.de/tutorial/tabelle-chi-quadrat> sowie [https://www.uibk.ac.at/econometrics/einf/tab\\_chisq.pdf](https://www.uibk.ac.at/econometrics/einf/tab_chisq.pdf) gilt:  
 $\chi_{9;0.01}^2 = 21.666$ . Demnach wird  $H_0$  abgelehnt und  $H_1$  angenommen.

# Vielen Dank!

Tobias Forster  
tobias.forster@fhv.at