# Stochastik und Statistik

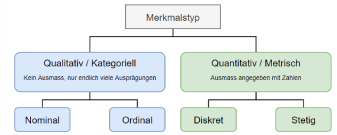
Jil Zerndt, Lucien Perret
January 2025

## Intro

### Begriffe

#### Grundlegende Begriffe

- $\Omega$ = Grundgesamtheit
- $n$ = Anzahl Objekte
- $X$ = Stichprobenwerte
- $a$ = Ausprägungen
- $h$ = Absolute Häufigkeit
- $f$ = Relative Häufigkeit
- $H$ = Kumulative Absolute Häufigkeit
- $F$ = Kumulative Relative Häufigkeit




#### Statistische Grundbegriffe

- **Merkmalsträger/Statistische Einheiten**: Objekte, an denen interessierende Grössen beobachtet werden
- **Grundgesamtheit**: Alle statistischen Einheiten, über die Aussagen gewonnen werden sollen
- **Vollerhebung**: Eigenschaften werden bei jedem Individuum in der Grundgesamtheit erhoben
- **Stichprobe**: Untersuchte Teilmenge der Grundgesamtheit (repräsentativ)
- **Merkmal**: Interessierende Grösse, die an den Einheiten beobachtet wird

#### Merkmalstypen

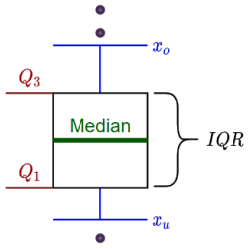
- **Qualitativ/Kategoriell**: Ausprägung und kein Ausmass (endlich viele Ausprägungen)
  - **Nominal**: Reine Kategorisierung (z.B. Parteien bei Wahlen)
  - **Ordinal**: Ordnung vorhanden (z.B. Schulnoten)
- **Quantitativ/Metrisch**: Ausmass wird mit Zahlen angegeben
  - **Diskret**: Abzählbar viele Ausprägungen (z.B. Würfelwurf)
  - **Stetig**: Alle Ausprägungen in einem reellen Intervall (z.B. Länge)

#### Merkmalstypen - Praktische Beispiele

- **Nominal**: Geschlecht, Automarke, Blutgruppe
- **Ordinal**: Bildungsabschluss, Zufriedenheit (1-5), Kaufkraft (tief/mittel/hoch)
- **Diskret metrisch**: Anzahl Kinder, Würfelaugen, Stockwerke
- **Stetig metrisch**: Temperatur, Gewicht, Länge

#### Boxplot

- $Q_1, Q_2 = x_{\text{med}}, Q_3$ (Quartile)
- $IQR = Q_3 - Q_1$ (Interquartilsabstand)
- Untere Antenne $x_u$ : $u = \min[Q_1 - 1.5 \cdot IQR, Q_1]$
- Obere Antenne $x_0$ : $o = \max[Q_3 + 1.5 \cdot IQR, Q_3]$
- Ausreisser: $x_i < x_u \vee x_i > x_0$




#### Erstellen eines Boxplots

1. Berechne die Quartile $Q_1$, $Q_2$ (Median) und $Q_3$
2. Bestimme den Interquartilsabstand IQR $= Q_3 - Q_1$
3. Berechne die Grenzen für Ausreisser:
   - Untere Grenze: $Q_1 - 1.5 \cdot IQR$
   - Obere Grenze: $Q_3 + 1.5 \cdot IQR$
4. Zeichne Box mit:
   - Unterer Rand bei $Q_1$
   - Mittellinie bei $Q_2$
   - Oberer Rand bei $Q_3$
5. Zeichne Antennen bis zum:
   - Kleinsten Wert $\geqslant$ untere Grenze
   - Grössten Wert $\leqslant$ obere Grenze
6. Markiere alle Werte ausserhalb als Ausreisser

**Boxplot - Praktisches Beispiel** Gegeben sind folgende Messwerte: 2, 3, 5, 6, 7, 8, 9, 15, 50

1. Sortiere Werte: 2, 3, 5, 6, 7, 8, 9, 15, 50
2. Bestimme Quartile:
   - $Q_1 = 4$ (25%-Quantil)
   - $Q_2 = 7$ (Median)
   - $Q_3 = 12$ (75%-Quantil)
3. IQR = 12 - 4 = 8
4. Ausreisser-Grenzen:
   - Untere: 4 - 1.5 $\cdot$ 8 = -8
   - Obere: 12 + 1.5 $\cdot$ 8 = 24
5. 50 ist ein Ausreisser ($> 24$)

### Häufigkeiten

#### Häufigkeiten bei nicht-klassierten Daten

- **Absolute Häufigkeit** $h_i$: Anzahl Vorkommen des Wertes $a_i$
- **Relative Häufigkeit** $f_i$: $f_i = \frac{h_i}{n}$
- **Kumulative absolute Häufigkeit** $H_i$: $H_i = \sum_{j=1}^{i} h_j$
- **Kumulative relative Häufigkeit** $F_i$: $F_i = \sum_{j=1}^{i} f_j = \frac{H_i}{n}$

#### Häufigkeiten bei klassierten Daten

Bei grossen Stichproben metrisch stetiger Merkmale werden die Werte in Klassen eingeteilt:

- Klassen sind aneinandergrenzende Intervalle
- Obere Intervallgrenzen gehören zum nächsten Intervall
- Relative Häufigkeit einer Klasse = Anzahl Werte in Klasse / Stichprobengrösse
- Relative Häufigkeitsdichte = Relative Häufigkeit / Klassenbreite

#### Klasseneinteilung

**Faustregeln:**

- Klassen sollten gleich breit gewählt werden
- Anzahl Klassen zwischen 5 und 20
- Anzahl Klassen sollte $\sqrt{n}$ nicht überschreiten
- Klassenbreite = $\frac{\text{Max - Min}}{\text{Anzahl Klassen}}$

**Häufigkeitsverteilung** Noten einer Klasse: 3.5, 4.0, 4.0, 4.5, 4.5, 4.5, 5.0, 5.0, 5.5, 6.0

- $n = 10$ (Stichprobengrösse)
- Absolute Häufigkeiten: $h_{3.5} = 1$, $h_{4.0} = 2$, $h_{4.5} = 3$, $h_{5.0} = 2$, $h_{5.5} = 1$, $h_{6.0} = 1$
- Relative Häufigkeiten: $f_{3.5} = 0.1$, $f_{4.0} = 0.2$, $f_{4.5} = 0.3$, $f_{5.0} = 0.2$, $f_{5.5} = 0.1$, $f_{6.0} = 0.1$
- Kumulative absolute Häufigkeiten: $H_{3.5} = 1$, $H_{4.0} = 3$, $H_{4.5} = 6$, $H_{5.0} = 8$, $H_{5.5} = 9$, $H_{6.0} = 10$

## Deskriptive Statistik

#### Bivariate Daten (Merkmale)

Bivariate Daten beschreiben zwei Merkmale desselben Merkmalsträgers. Die Darstellung hängt von den Merkmalstypen ab:

- 2x kategoriell $\rightarrow$ Kontingenztabelle + Mosaikplot
- 1x kategoriell + 1x metrisch $\rightarrow$ Boxplot oder Stripchart
- 2x metrisch $\rightarrow$ Streudiagramm

#### Erstellen einer Kontingenztabelle

1. Identifiziere die Ausprägungen beider kategorieller Merkmale
2. Erstelle eine Tabelle mit:
   - Zeilen für Ausprägungen des ersten Merkmals
   - Spalten für Ausprägungen des zweiten Merkmals
3. Zähle die Häufigkeiten für jede Kombination
4. Füge Randsummen für Zeilen und Spalten hinzu
5. Optional: Berechne relative Häufigkeiten

**Kontingenztabelle** Studierende nach Studiengang und Geschlecht:

| | Männlich | Weiblich | Total |
|---|---|---|---|
| Informatik | 120 | 30 | 150 |
| Wirtschaft | 80 | 70 | 150 |
| Total | 200 | 100 | 300 |

#### Absolute Häufigkeiten

$$H = \sum_{i=1}^{n} h_i$$

$H$: Absolute Häufigkeit,
$h_i$: Einzelhäufigkeit der $i$-ten Beobachtung,
$n$: Anzahl der Beobachtungen.

#### Relative Häufigkeiten

$$F = \sum_{i=1}^{m} f_i, \quad F(x) = \frac{H(x)}{n}$$

$F$: Relative Häufigkeit,
$f_i$: Einzelrelative Häufigkeit der $i$-ten Beobachtung,
$H(x)$: Absolute Häufigkeit eines Wertes $x$,
$n$: Anzahl der Beobachtungen.

## Kennwerte (Lagemasse)

### Berechnung von Lagekennwerten

1. Sortiere die Daten aufsteigend
2. Berechne den Mittelwert:
   - Summe aller Werte / Anzahl Werte
3. Bestimme den Median:
   - Bei ungerader Anzahl: mittlerer Wert
   - Bei gerader Anzahl: Mittelwert der beiden mittleren Werte
4. Finde den Modus (häufigster Wert)
5. Berechne die Quartile:
   - Q1: 25%-Quantil
   - Q2: Median (50%-Quantil)
   - Q3: 75%-Quantil

### Quantil

$$i = \lceil n \cdot q \rceil, \quad Q = x_i = x_{\lceil n \cdot q \rceil}$$

$i$: Position des Quantils,
$n$: Anzahl der Beobachtungen,
$q$: Quantilswert (z. B. 0.25 für das erste Quartil),
$x_i$: Beobachtung an Position $i$.

### Interquartilsabstand

$$IQR = Q_3 - Q_1$$

$IQR$: Interquartilsabstand,
$Q_3$: Oberes Quartil (75. Perzentil),
$Q_1$: Unteres Quartil (25. Perzentil).

### Berechnung von Quantilen

Gegeben sei die Datenreihe: 2, 4, 4, 5, 7, 8, 9, 10
$n = 8$ Beobachtungen

**Berechnung Q1 (25%-Quantil):**
- $i = \lceil 8 \cdot 0.25 \rceil = \lceil 2 \rceil = 2$
- $Q1 = x_2 = 4$

**Berechnung Q2 (Median):**
- $n$ gerade → Mittelwert von Position 4 und 5
- $Q2 = (5 + 7)/2 = 6$

**Berechnung Q3 (75%-Quantil):**
- $i = \lceil 8 \cdot 0.75 \rceil = \lceil 6 \rceil = 6$
- $Q3 = x_6 = 8$

**Interquartilsabstand:**
- IQR = Q3 - Q1 = 8 - 4 = 4

### Modus

$$x_{\text{mod}} = \text{Häufigste Wert}$$

### Arithmetisches Mittel

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \sum_{i=1}^{m} a_i \cdot f_i$$

$\bar{x}$: Arithmetisches Mittel,
$n$: Anzahl der Beobachtungen,
$x_i$: Einzelbeobachtung,
$a_i$: Klassenmitte,
$f_i$: Relative Häufigkeit der Klasse $i$.

### Median

$$\begin{cases} x_{\left\lceil \frac{n+1}{2} \right\rceil} & n \text{ ungerade} \\ 0.5 \cdot \left( x_{\left\lceil \frac{n}{2} \right\rceil} + x_{\left\lceil \frac{n}{2}+1 \right\rceil} \right) & n \text{ gerade} \end{cases}$$

$n$: Anzahl der Beobachtungen,
$x_{[k]}$: Beobachtung an der $k$-ten Position.

### Vergleich der Lageparameter

Gegeben seien folgende Datensätze:
- A: 2, 2, 3, 4, 4, 5, 8, 12
- B: 2, 4, 4, 4, 4, 4, 6, 8

**Datensatz A:**
- Mittelwert: $\bar{x}_A = 5$
- Median: $x_{med_A} = 4$
- Modus: $x_{mod_A} = 2, 4$ (bimodal)

**Datensatz B:**
- Mittelwert: $\bar{x}_B = 4.5$
- Median: $x_{med_B} = 4$
- Modus: $x_{mod_B} = 4$

Vergleich zeigt:
- Mittelwert reagiert empfindlich auf Ausreißer (A)
- Median ist robuster gegen Ausreißer
- Modus zeigt Häufungen, kann mehrfach auftreten

### Stichprobenvarianz $s^2$ (Streumasse)

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

$$(s_{\text{kor}})^2 = \frac{n}{n-1} \cdot s^2$$

$s^2$: Stichprobenvarianz,
$s^2_{\text{kor}}$: Korrigierte Stichprobenvarianz,
$x_i$: Einzelbeobachtung,
$\bar{x}$: Arithmetisches Mittel,
$n$: Anzahl der Beobachtungen.

### Berechnung der Stichprobenvarianz

1. Berechne den Mittelwert $\bar{x}$
2. Für jeden Wert $x_i$:
   2.1 Berechne Abweichung vom Mittelwert $(x_i - \bar{x})$
   2.2 Quadriere die Abweichung $(x_i - \bar{x})^2$
3. Summiere alle quadrierten Abweichungen
4. Teile durch $(n - 1)$ für korrigierte Varianz
5. Alternative Berechnung:
   5.1 Berechne $\overline{x^2}$ (Mittelwert der quadrierten Werte)
   5.2 Berechne $(\bar{x})^2$ (Quadrat des Mittelwerts)
   5.3 Varianz $= \overline{x^2} - (\bar{x})^2$

### Standardabweichung $s$ (Streumasse)

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2} = \sqrt{\overline{x^2} - \bar{x}^2}, \quad s_{\text{kor}} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$s$: Standardabweichung,
$s_{\text{kor}}$: Korrigierte Standardabweichung,
$x_i$: Einzelbeobachtung,
$\bar{x}$: Arithmetisches Mittel,
$n$: Anzahl der Beobachtungen.

### Berechnung der Standardabweichung

Gegeben sei die Datenreihe: 2, 4, 4, 6, 9

**Schritt 1:** Mittelwert berechnen

$$\bar{x} = \frac{2 + 4 + 4 + 6 + 9}{5} = 5$$

**Schritt 2:** Abweichungen quadrieren

| $x_i$ | $(x_i - \bar{x})$ | $(x_i - \bar{x})^2$ |
|---|---|---|
| 2 | -3 | 9 |
| 4 | -1 | 1 |
| 4 | -1 | 1 |
| 6 | 1 | 1 |
| 9 | 4 | 16 |

**Schritt 3:** Varianz berechnen

$$s^2_{\text{kor}} = \frac{9 + 1 + 1 + 1 + 16}{5 - 1} = \frac{28}{4} = 7$$

**Schritt 4:** Standardabweichung berechnen

$$s_{\text{kor}} = \sqrt{7} \approx 2.65$$

**Alternative Berechnung:**
- $\overline{x^2} = \frac{4+16+16+36+81}{5} = 30.6$
- $(\bar{x})^2 = 5^2 = 25$
- $s^2 = 30.6 - 25 = 5.6$
- $s^2_{\text{kor}} = \frac{5}{4} \cdot 5.6 = 7$

## PDF + CDF

### Nicht klassierte Daten (PMF und CDF)

Die absolute Häufigkeit kann als Funktion $h : \mathbb{R} \to \mathbb{R}$ bezeichnet werden.

$$h_i$$

$h_i$: Absolute Häufigkeit der $i$-ten Beobachtung.

Die relative Häufigkeit kann als Funktion $f : \mathbb{R} \to \mathbb{R}$ bezeichnet werden.

$$f_i = \frac{h_i}{n}$$

$f_i$: Relative Häufigkeit der $i$-ten Beobachtung,
$h_i$: Absolute Häufigkeit der $i$-ten Beobachtung,
$n$: Anzahl der Beobachtungen.

### Erstellen einer Häufigkeitsverteilung

1. Sammle alle verschiedenen Werte
2. Zähle absolute Häufigkeiten:
   - Wie oft kommt jeder Wert vor?
3. Berechne relative Häufigkeiten:
   - Teile jede absolute Häufigkeit durch $n$
4. Berechne kumulative Häufigkeiten:
   - Absolute: Summiere $h_i$ von links nach rechts
   - Relative: Summiere $f_i$ von links nach rechts

# PMF und CDF für diskrete und stetige Daten

## Unterschied zwischen PMF und PDF

- **PMF (Probability Mass Function)**:
  - Für diskrete Daten
  - Wahrscheinlichkeit für exakte Werte
  - Summe aller Wahrscheinlichkeiten = 1
- **PDF (Probability Density Function)**:
  - Für stetige Daten
  - Fläche unter Kurve gibt Wahrscheinlichkeit
  - Integral über gesamten Bereich = 1

## Diskrete Verteilungsfunktionen

Die absolute Häufigkeit kann als Funktion $h : \mathbb{R} \to \mathbb{R}$ bezeichnet werden:

$$h_i$$

Die relative Häufigkeit kann als Funktion $f : \mathbb{R} \to \mathbb{R}$ bezeichnet werden:

$$f_i = \frac{h_i}{n}$$

### Diskrete Häufigkeitsverteilung

| $a_i$ | 397 | 398 | 399 | 400 | Total |
|---|---|---|---|---|---|
| $h_i$ | 1 | 3 | 7 | 5 | 16 |
| $f_i$ | $\frac{1}{16}$ | $\frac{3}{16}$ | $\frac{7}{16}$ | $\frac{5}{16}$ | 1 |
| $H_i$ | 1 | 4 | 11 | 16 | |
| $F_i$ | $\frac{1}{16}$ | $\frac{4}{16}$ | $\frac{11}{16}$ | $\frac{16}{16}$ | |

## Klassenbildung für stetige Daten

1. Bestimme Spannweite (Max - Min)
2. Wähle Anzahl Klassen $k$:
   - $5 \leq k \leq 20$
   - $k \leq \sqrt{n}$
3. Berechne Klassenbreite:
   - $d = \frac{\text{Spannweite}}{k}$
   - Runde auf praktische Zahl
4. Bestimme Klassengrenzen:
   - Start bei Min oder praktischem Wert darunter
   - Ende bei Max oder praktischem Wert darüber
5. Zähle Häufigkeiten in jeder Klasse

## Klassenbildung (Faustregeln)

- Die Klassen sollten gleich breit gewählt werden
- Die Anzahl der Klassen sollte zwischen 5 und 20 liegen, jedoch $\sqrt{n}$ nicht überschreiben
- Klassengrenzen sollten 'runde' Zahlen sein
- Werte auf Klassengrenzen kommen in die obere Klasse

## Stetige Verteilungsfunktionen

Die absolute Häufigkeitsdichtefunktion erhält man, indem der Wert der absoluten Häufigkeit $h_i$ durch die Klassenbreite (Säulenbreite) $d_i$ geteilt wird:

$$h(x) = \frac{h_i}{d_i}$$

Die relative Häufigkeitsdichtefunktion (PDF) $f : \mathbb{R} \to [0, 1]$ erhält man aus der absoluten Häufigkeitsdichtefunktion, indem man den Wert durch die Stichprobengrösse $n$ teilt:

$$\text{PDF} = f(x) = \frac{h(x)}{n}$$

### Stetige Häufigkeitsverteilung

| Klassen | [100,200) | [200,500) | [500,800) | [800,1000) | Total |
|---|---|---|---|---|---|
| $h_i$ | 35 | 182 | 317 | 84 | 618 |
| $f_i$ | $\frac{35}{618}$ | $\frac{182}{618}$ | $\frac{317}{618}$ | $\frac{84}{618}$ | Area = |
| $d_i$ | 100 | 300 | 300 | 200 | |
| $h(x)$ | $\frac{35}{100}$ | $\frac{182}{300}$ | $\frac{317}{300}$ | $\frac{84}{200}$ | |
| $f(x)$ | $\frac{35}{100 \cdot 618}$ | $\frac{182}{300 \cdot 618}$ | $\frac{317}{300 \cdot 618}$ | $\frac{84}{200 \cdot 618}$ | |

## Berechnung von PDF und CDF für klassierte Daten

1. PDF Berechnung:
   1.1 Bestimme für jede Klasse:
   - Absolute Häufigkeit $h_i$
   - Klassenbreite $d_i$

   1.2 Berechne Häufigkeitsdichte:
   - $h(x) = \frac{h_i}{d_i}$

   1.3 Berechne relative Häufigkeitsdichte:
   - $f(x) = \frac{h(x)}{n}$
2. CDF Berechnung:
   2.1 Bestimme kumulative Häufigkeiten $H_i$
   2.2 Teile durch Stichprobengröße:
   - $F(x) = \frac{H(x)}{n}$

## Varianz und Kovarianz

**Varianz** $s_x^2, s_y^2$:

$$(s_x)^2 = \overline{x^2} - \bar{x}^2, \quad (s_y)^2 = \overline{y^2} - \bar{y}^2$$

**Kovarianz** $s_{xy}$:

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

## Berechnung der Kovarianz

1. Methode 1 (direkte Formel):
   1.1 Berechne Mittelwerte $\bar{x}$ und $\bar{y}$
   1.2 Für jedes Paar $(x_i, y_i)$:
   - Berechne $(x_i - \bar{x})(y_i - \bar{y})$

   1.3 Summiere alle Produkte
   1.4 Teile durch $n$
2. Methode 2 (schnellere Berechnung):
   2.1 Berechne $\overline{xy}$ (Mittelwert der Produkte)
   2.2 Berechne $\bar{x} \cdot \bar{y}$
   2.3 Kovarianz = $\overline{xy} - \bar{x} \cdot \bar{y}$

### Berechnung von Kovarianz und Korrelation

Gegeben seien die Wertepaare:

$$(1, 2), (2, 4), (3, 5), (4, 8)$$

**Schritt 1:** Mittelwerte berechnen

$$\bar{x} = \frac{1+2+3+4}{4} = 2.5, \quad \bar{y} = \frac{2+4+5+8}{4} = 4.75$$

**Schritt 2:** Kovarianz berechnen

- $\overline{xy} = \frac{2+8+15+32}{4} = 14.25$
- $\bar{x} \cdot \bar{y} = 2.5 \cdot 4.75 = 11.875$
- $s_{xy} = 14.25 - 11.875 = 2.375$

**Schritt 3:** Korrelationskoeffizient berechnen

- $s_x^2 = \frac{1+4+9+16}{4} - 2.5^2 = 1.25$
- $s_y^2 = \frac{4+16+25+64}{4} - 4.75^2 = 5.6875$
- $r_{xy} = \frac{2.375}{\sqrt{1.25} \cdot \sqrt{5.6875}} = 0.894$

## Abkürzungen

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \text{(Mittelwert der x-Werte)}$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \quad \text{(Mittelwert der y-Werte)}$$

$$\overline{xy} = \frac{1}{n} \sum_{i=1}^{n} x_i \cdot y_i \quad \text{(Mittelwert der Produkte)}$$

## Rang-Varianz und Kovarianz

**Varianz (Ränge)** $(s_{rg(x)})^2, (s_{rg(y)})^2$:

$$(s_{rg(x)})^2 = \overline{rg(x)^2} - (\overline{rg(x)})^2, \quad (s_{rg(y)})^2 = \overline{rg(y)^2} - (\overline{rg(y)})^2$$

**Kovarianz (Ränge)** $s_{rg(xy)}$:

$$s_{rg(xy)} = \overline{rg(xy)} - \overline{rg(x)} \cdot \overline{rg(y)} = \overline{rg(xy)} - \frac{(n+1)^2}{4}$$

### Rangberechnung und Bindungen

1. Sortiere die Werte aufsteigend
2. Weise Ränge zu:
   - Kleinster Wert: Rang 1
   - Zweitkleinster: Rang 2
   - usw.
3. Bei Bindungen (gleiche Werte):
   3.1 Identifiziere gleiche Werte
   3.2 Berechne Durchschnittsrang:
   - Durchschnittsrang $= \frac{\text{Summe der Rangplätze}}{\text{Anzahl gebundener Werte}}$

   3.3 Weise allen gleichen Werten diesen Rang zu

**Rangberechnung mit Bindungen** Gegeben sei die Datenreihe: 3, 7, 7, 4, 9, 7, 2

**Schritt 1:** Sortieren

$$2, 3, 4, 7, 7, 7, 9$$

**Schritt 2:** Ränge zuweisen

- 2: Rang 1
- 3: Rang 2
- 4: Rang 3
- 7: Durchschnittsrang $\frac{4+5+6}{3} = 5$
- 9: Rang 7

**Schritt 3:** Finale Rangzuordnung

| Wert | 3 | 7 | 7 | 4 | 9 | 7 | 2 |
|---|---|---|---|---|---|---|---|
| Rang | 2 | 5 | 5 | 3 | 7 | 5 | 1 |

### Der Korrelationskoeffizient (Pearson) $r_{xy}$

$$r_{xy} = \frac{s_{xy}}{s_x \cdot s_y} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\overline{x^2} - \bar{x}^2} \cdot \sqrt{\overline{y^2} - \bar{y}^2}}$$

Ist der Korrelationskoeffizient $r_{xy}$:

- $r_{xy} \approx 1 \rightarrow$ starker positiver linearer Zusammenhang
- $r_{xy} \approx -1 \rightarrow$ starker negativer linearer Zusammenhang
- $r_{xy} \approx 0 \rightarrow$ keine lineare Korrelation

**Interpretation des Korrelationskoeffizienten** Verschiedene Datensätze mit jeweils 20 $(x, y)$-Paaren:

**Fall A:** $r_{xy} = 0.95$

- Starker positiver linearer Zusammenhang
- $y$ steigt fast proportional mit $x$
- Nur geringe Streuung um die Regressionsgerade

**Fall B:** $r_{xy} = -0.82$

- Starker negativer linearer Zusammenhang
- $y$ sinkt mit steigendem $x$
- Moderate Streuung vorhanden

**Fall C:** $r_{xy} = 0.12$

- Kaum linearer Zusammenhang
- Starke Streuung der Punkte
- Möglicherweise nichtlinearer Zusammenhang

### Prüfung auf Scheinkorrelation

1. Betrachte die Datenpunkte im Streudiagramm:
   - Gibt es Ausreißer?
   - Ist der Zusammenhang wirklich linear?
2. Überlege fachlich:
   - Gibt es plausible Kausalität?
   - Könnte ein drittes Merkmal beide beeinflussen?
3. Prüfe Teilstichproben:
   - Bleibt Korrelation in Untergruppen bestehen?
   - Ändert sich die Stärke deutlich?
4. Bei Zweifeln:
   - Spearman-Korrelation prüfen
   - Weitere Merkmale einbeziehen
   - Fachexperten konsultieren

Bemerkungen
Auch wenn zwischen zwei Grössen eine Korrelation besteht, so muss das noch lange nicht einen **kausalen Zusammenhang** bedeuten. Man spricht von **Scheinkorrelation** wenn:

- Ein drittes Merkmal beide beeinflusst
- Der Zusammenhang zufällig ist
- Ausreißer das Ergebnis verzerren
- Ein nichtlinearer Zusammenhang vorliegt

## Deskriptive Statistik Multivarianz

### Graphische Darstellung

- **Form** linear / gekrümmt
  - Linear: Punkte streuen um Gerade
  - Gekrümmt: Systematische Abweichung von Gerade
- **Richtung** positiver / negativer Zusammenhang
  - Positiv: $y$ steigt mit $x$
  - Negativ: $y$ fällt mit steigendem $x$
- **Stärke** starke / schwache Streuung
  - Stark: Punkte nahe an Linie/Kurve
  - Schwach: Große Streuung um Trend

### Korrelationskoeffizient (Spearman) $r_{sp}$

$$r_{sp} = \frac{s_{rg(xy)}}{s_{rg(x)} \cdot s_{rg(y)}} = \frac{\overline{rg(xy)} - \overline{rg(x)} \cdot \overline{rg(y)}}{\sqrt{\overline{rg(x)^2} - (\overline{rg(x)})^2} \cdot \sqrt{\overline{rg(y)^2} - (\overline{rg(y)})^2}}$$

Vereinfachte Formel, sofern **alle Ränge unterschiedlich** sind:

$$r_{sp} = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n \cdot (n^2 - 1)}, \quad \text{mit } d_i = rg(x_i) - rg(y_i)$$

**Ränge**
Der Rang $rg(x_i)$ des Stichprobenwertes $x_i$ ist definiert als der Index von $x_i$ in der nach der Grösse geordneten Stichprobe.

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $x_i$ | 23 | 27 | 35 | 35 | 42 | 59 |
| $rg(x_i)$ | 1 | 2 | 3.5 | 3.5 | 5 | 6 |

### Berechnung des Spearman-Korrelationskoeffizienten

1. Weise beiden Merkmalen Ränge zu:
   - Sortiere x-Werte, vergebe Ränge
   - Sortiere y-Werte, vergebe Ränge
   - Bei Bindungen: Durchschnittsränge
2. Falls keine Bindungen vorhanden:
   2.1 Berechne Rangdifferenzen $d_i$
   2.2 Quadriere Differenzen $d_i^2$
   2.3 Summiere quadrierte Differenzen
   2.4 Verwende Formel: $r_{sp} = 1 - \frac{6 \sum d_i^2}{n(n^2-1)}$
3. Bei Bindungen:
   3.1 Berechne Rangmittelwerte
   3.2 Berechne Rangvarianzen und -kovarianz
   3.3 Verwende allgemeine Formel

**Vergleich Pearson und Spearman** Gegeben seien die Wertepaare:

$$(1, 1), (2, 4), (3, 9), (4, 16), (5, 25)$$

**Pearson-Korrelation:**

- Zeigt starken linearen Zusammenhang
- $r_{xy} = 0.975$

**Spearman-Korrelation:**

- Perfekter monotoner Zusammenhang
- $r_{sp} = 1.000$

**Vergleich:**

- Pearson erfasst nur linearen Zusammenhang
- Spearman erfasst jeden monotonen Zusammenhang
- Hier: Quadratischer Zusammenhang
- Spearman robuster gegen Ausreißer

Wahl des Korrelationskoeffizienten

- **Pearson verwenden wenn:**
  - Linearer Zusammenhang vermutet
  - Keine/wenige Ausreißer
  - Metrische Daten
- **Spearman verwenden wenn:**
  - Nichtlinearer monotoner Zusammenhang
  - Ausreißer vorhanden
  - Ordinale Daten
  - Robustheit wichtig

## Kombinatorik

### Grundlegende Methoden

**Fakultät** Die Fakultät $n!$ ist für eine natürliche Zahl $n$ rekursiv definiert:

$$n! = n \cdot (n-1)! \text{ mit } 0! = 1$$

$$n! = 1 \cdot 2 \cdot \ldots \cdot n = \prod_{k=1}^{n} k$$

$n$ = Die positive ganze Zahl, für die die Fakultät berechnet wird
$k$ = Laufvariable in der Produktnotation
$\prod$ = Produkt aller Terme von $k = 1$ bis $n$

**Binomialkoeffizient** Der Binomialkoeffizient $\binom{n}{k}$ ist für natürliche Zahlen $0 \leq k \leq n$ definiert als:

$$\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!}$$

Wie viele Möglichkeiten gibt es $k$ Objekte aus einer Gesamtheit von $n$ Objekten auszuwählen.

$$\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!}$$

$n$ = Gesamtanzahl der Objekte in der Menge
$k$ = Anzahl der auszuwählenden Objekte
$n!$ = Fakultät von $n$
$(n-k)!$ = Fakultät von $(n-k)$
$k!$ = Fakultät von $k$

**Fakultät** Die Fakultät einer natürlichen Zahl $n$ ist definiert als das Produkt aller positiven ganzen Zahlen bis zu dieser Zahl:

$$n! = 1 \cdot 2 \cdot \ldots \cdot n = \prod_{k=1}^{n} k$$

mit $0! = 1$ als Definitionsvereinbarung
**Parameter:**
- $n$ = Die positive ganze Zahl, für die die Fakultät berechnet wird
- $k$ = Laufvariable in der Produktnotation
- $\prod$ = Produkt aller Terme von $k = 1$ bis $n$

**Berechnung von Fakultäten** 1. Prüfe Spezialfälle:
- $0! = 1$ (Definition)
- $1! = 1$

2. Für $n > 1$:
- Schreibe alle Zahlen von 1 bis $n$ auf
- Multipliziere der Reihe nach
- Alternative: Nutze rekursive Definition $n! = n \cdot (n-1)!$

**Binomialkoeffizient** Der Binomialkoeffizient $\binom{n}{k}$ gibt an, wie viele Möglichkeiten es gibt, $k$ Objekte aus einer Gesamtheit von $n$ Objekten auszuwählen:

$$\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!}$$

**Parameter:**
- $n$ = Gesamtanzahl der Objekte in der Menge
- $k$ = Anzahl der auszuwählenden Objekte $(0 \leq k \leq n)$
- $n!$ = Fakultät von $n$
- $(n-k)!$ = Fakultät von $(n-k)$
- $k!$ = Fakultät von $k$

**Wichtige Eigenschaften:**
- $\binom{n}{0} = \binom{n}{n} = 1$
- $\binom{n}{k} = \binom{n}{n-k}$ (Symmetrie)
- $\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$ (Pascal'sche Rekursion)

## Systematik

### Grundbegriffe der Kombinatorik

| Variation (mit Reihenfolge) | | Kombination (ohne Reihenfolge) | |
|---|---|---|---|
| Mit Wiederholung | Ohne Wiederholung | Mit Wiederholung | Ohne Wiederholung |
| $n^k$ | $\frac{n!}{(n-k)!}$ | $\binom{n+k-1}{k}$ | $\binom{n}{k}$ |
| Zahlenschloss | Schwimmwettkampf | Zahnarzt | Lotto |

**Entscheidungsweg für kombinatorische Probleme** 1. **Bestimme die relevanten Parameter**
- $n$: Wie viele Objekte gibt es insgesamt?
- $k$: Wie viele Objekte sollen ausgewählt werden?

2. **Prüfe die Reihenfolge**
- Spielt die Reihenfolge eine Rolle? → Variation
- Ist nur die Auswahl wichtig? → Kombination

3. **Prüfe Wiederholungen**
- Können Objekte mehrfach gewählt werden? → Mit Wiederholung
- Darf jedes Objekt nur einmal vorkommen? → Ohne Wiederholung

4. **Wähle die passende Formel**
- Variation mit Wiederholung: $n^k$
- Variation ohne Wiederholung: $\frac{n!}{(n-k)!}$
- Kombination mit Wiederholung: $\binom{n+k-1}{k}$
- Kombination ohne Wiederholung: $\binom{n}{k}$

**Systematik der Kombinatorik**
- **Variation mit Wiederholung**: $n^k$
- **Variation ohne Wiederholung**: $\frac{n!}{(n-k)!}$
- **Kombination mit Wiederholung**: $\binom{n+k-1}{k}$
- **Kombination ohne Wiederholung**: $\binom{n}{k}$

**Entscheidungsbaum für Kombinatorik** 1. **Spielt die Reihenfolge eine Rolle?** - Ja: Variation - Nein: Kombination 2. **Sind Wiederholungen erlaubt?** - Ja: Mit Wiederholung - Nein: Ohne Wiederholung 3. **Formel auswählen:**
- Variation mit Wiederholung: $n^k$
- Variation ohne Wiederholung: $\frac{n!}{(n-k)!}$
- Kombination mit Wiederholung: $\binom{n+k-1}{k}$
- Kombination ohne Wiederholung: $\binom{n}{k}$

Zahlenschloss Ein Zahlenschloss hat 6 Stellen, jede mit Ziffern 0-9.
- Reihenfolge wichtig? Ja (Variation)
- Wiederholungen erlaubt? Ja
- Formel: $n^k = 10^6$ mögliche Kombinationen

Lotto 6 aus 49 ziehen:
- Reihenfolge wichtig? Nein (Kombination)
- Wiederholungen erlaubt? Nein
- Formel: $\binom{49}{6} = \frac{49!}{6!(49-6)!}$ mögliche Kombinationen

Zahnarztproblem 3 Spielzeuge aus 5 Töpfen ziehen:
- Reihenfolge wichtig? Nein (Kombination)
- Wiederholungen erlaubt? Ja
- Formel: $\binom{5+3-1}{3} = \binom{7}{3}$ mögliche Kombinationen

**Lösungsstrategie für Kombinatorik-Aufgaben** 1. **Analysiere die Aufgabe** - Was wird gezählt? - Welche Einschränkungen gibt es? 2. **Identifiziere die Art** - Reihenfolge wichtig? - Wiederholungen erlaubt? 3. **Wähle die passende Formel** 4. **Setze die Zahlen ein** 5. **Berechne das Ergebnis**

Komplexeres Beispiel: Mannschaftsauswahl In einer Klasse von 20 Studierenden sollen:
- Eine 11er Fußballmannschaft gebildet werden
- Mit genau 6 Frauen und 5 Männern
- Es gibt 8 Frauen und 12 Männer in der Klasse

**Lösung:** 1. Wähle 6 aus 8 Frauen: $\binom{8}{6}$ 2. Wähle 5 aus 12 Männern: $\binom{12}{5}$ 3. Multipliziere: $\binom{8}{6} \cdot \binom{12}{5} = 22{,}176$ Möglichkeiten

**Schritte zur Berechnung von Binomialkoeffizienten** 1. **Prüfe Grundfälle** - $\binom{n}{0} = 1$ - $\binom{n}{n} = 1$ - $\binom{n}{1} = n$ 2. **Nutze Symmetrie** - $\binom{n}{k} = \binom{n}{n-k}$ 3. **Pascal's Dreieck** - $\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$ 4. **Direkte Berechnung** - $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

## Kombinatorik

Variation mit Wiederholung (Zahlenschloss) **Aufgabe:** Wie viele Möglichkeiten gibt es bei einem Zahlenschloss (0 – 9) mit 6 Zahlenkränzen?
**Lösung:**
- $n = 10$ (Ziffern 0-9)
- $k = 6$ (Stellen)
- Reihenfolge wichtig: Ja ($123456 \neq 654321$)
- Wiederholungen erlaubt: Ja (11111 ist möglich)
- Formel: $n^k = 10^6 = 1\,000\,000$ mögliche Kombinationen

Variation ohne Wiederholung (Schwimmwettkampf) **Aufgabe:** Bei einem Schwimmwettkampf starten 10 Teilnehmer. Wie viele mögliche Platzierungen der ersten drei Plätze (Podest) gibt es?
**Lösung:**
- $n = 10$ (Teilnehmer)
- $k = 3$ (Podestplätze)
- Reihenfolge wichtig: Ja (1., 2., 3. Platz unterschiedlich)
- Wiederholungen erlaubt: Nein (niemand kann mehrere Plätze belegen)
- Formel: $\frac{n!}{(n-k)!} = \frac{10!}{7!} = 720$ mögliche Platzierungen

Kombination mit Wiederholung (Zahnarzt) **Aufgabe:** 3 Spielzeuge werden aus 5 Töpfen gezogen. Jeder Topf ist mit einer (unterschiedlichen) Art von Spielzeug befüllt. Wie viele Möglichkeiten hat das Kind?
**Lösung:**
- $n = 5$ (Arten von Spielzeug)
- $k = 3$ (zu wählende Spielzeuge)
- Reihenfolge wichtig: Nein (nur Anzahl pro Art relevant)
- Wiederholungen erlaubt: Ja (mehrere Spielzeuge gleicher Art möglich)
- Formel: $\binom{n+k-1}{k} = \binom{7}{3} = 35$ Möglichkeiten

Kombination ohne Wiederholung (Lotto) **Aufgabe:** Wie gross sind die Chancen beim Lotto 6 aus 49 Zahlen richtig zu ziehen?
**Lösung:**
- $n = 49$ (Zahlen insgesamt)
- $k = 6$ (zu wählende Zahlen)
- Reihenfolge wichtig: Nein (nur Auswahl relevant)
- Wiederholungen erlaubt: Nein (jede Zahl nur einmal)
- Formel: $\binom{49}{6} = 13\,983\,816$ Möglichkeiten
- Gewinnwahrscheinlichkeit: $\frac{1}{13\,983\,816} \approx 0.0000000715$

Komplexeres Beispiel: Passwörter **Aufgabe:** Ein Passwort muss bestehen aus:
- Genau 8 Zeichen
- Mindestens ein Großbuchstabe (26 mögliche)
- Mindestens eine Ziffer (10 mögliche)
- Kleine Buchstaben erlaubt (26 mögliche)

**Lösung:** 1. Gesamtzahl aller möglichen 8-stelligen Passwörter mit den Zeichen:
- $n = 26 + 26 + 10 = 62$ Zeichen
- Variation mit Wiederholung: $62^8$

2. Abziehen der ungültigen Kombinationen:
- Ohne Großbuchstaben: $(36)^8$
- Ohne Ziffern: $(52)^8$
- Ohne beide: $(26)^8$

3. Nach dem Inklusions-Exclusions-Prinzip:

$$\text{Gültige Passwörter} = 62^8 - 36^8 - 52^8 + 26^8$$

**Lösen komplexer kombinatorischer Probleme** 1. **Problem zerlegen**
- Teile das Problem in unabhängige Teilprobleme
- Identifiziere abhängige Entscheidungen

2. **Für jedes Teilproblem**
- Bestimme $n$ und $k$
- Prüfe Reihenfolge und Wiederholung
- Wähle passende Formel

3. **Kombiniere Teillösungen**
- Unabhängige Ereignisse: Multipliziere
- Sich ausschließende Ereignisse: Addiere
- Prüfe Überlappungen (Inklusions-Exclusions)

Variation mit Wiederholung (Zahlenschloss)
Wie viele Möglichkeiten gibt es bei einem Zahlenschloss (0 – 9) mit 6 Zahlenkränzen?

$$n = 10, \quad k = 6$$
$$n^k = 10^6$$

Variation ohne Wiederholung (Schimmwettkampf)
Bei einem Schwimmwettkampf starten 10 Teilnehmer. Wie viele mögliche Platzierungen der ersten drei Plätze (Podest) gibt es?

$$n = 10, \quad k = 3$$
$$\frac{n!}{(n-k)!} = \frac{10!}{(10-3)!} = \frac{10!}{7!}$$

Kombination mit Wiederholung (Zahnarzt)
3 Spielzeuge werden aus 5 Töpfen gezogen. Jeder Topf ist mit einer (unterschiedlichen) Art von Spielzeug befüllt.
Wie viele Möglichkeiten hat das Kind?

$$n = 5, \quad k = 3$$
$$\binom{n+k-1}{k} = \binom{5+3-1}{3} = \binom{7}{3}$$

Kombination ohne Wiederholung (Lotto)
Wie gross sind die Chancen beim Lotto 6 aus 49 Zahlen richtig zu ziehen? Jede Zahl ist nur einmal vorhanden und die Zahlen werden nicht zurückgelegt. Die Reihenfolge in der gezogen wird spielt keine Rolle.

$$n = 49, \quad k = 6$$
$$\binom{n}{k} = \binom{49}{6}$$

# Wahrscheinlichkeitsrechnung

## Grundlagen

**Grundlegende Strategien**
- **Aufteilung in Kombinationen:** Komplexe Probleme in einfachere Teilprobleme zerlegen
- **Berechnung über Inverse:** Manchmal ist es einfacher, die Gegenwahrscheinlichkeit zu berechnen
- **Prozentrechnung:** Wahrscheinlichkeit / Gesamt-Wahrscheinlichkeit · 100%
- **Vierfeldertafel:** Zur Übersicht bei zwei binären Merkmalen

**Grundschritte der Wahrscheinlichkeitsberechnung** 1. **Ergebnisraum identifizieren**
- Alle möglichen Ergebnisse auflisten
- Prüfen, ob es sich um einen Laplace-Raum handelt

2. **Ereignis präzisieren**
- Exakte mathematische Beschreibung des gesuchten Ereignisses
- Zerlegung in Teilmengen falls nötig

3. **Berechnungsstrategie wählen**
- Direkte Berechnung: $P(A) = \frac{|A|}{|\Omega|}$
- Über Gegenereignis: $P(A) = 1 - P(\bar{A})$
- Über bedingte Wahrscheinlichkeit falls abhängig

4. **Berechnung durchführen**
- Kombinatorische Formeln anwenden
- Zwischenergebnisse notieren
- Probe durch Plausibilitätskontrolle

Wahrscheinlichkeit bei Rommé
Beim Rommé spielt man mit **110 Karten: sechs** davon sind **Joker**. Zu Beginn eines Spiels erhält jeder Spieler genau **12 Karten**.
In wieviel Prozent aller möglichen Fälle sind darunter **genau zwei** Joker?

$$\frac{\binom{6}{2} \cdot \binom{104}{10}}{\binom{110}{12}}$$

In wieviel Prozent aller möglichen Fälle ist darunter **mindestens ein** Joker?

$$1 - \frac{\binom{104}{12}}{\binom{110}{12}}$$

Geschwister und Geburtsmonat
Sind in mehr als 60% aller Fälle von vier (nicht gleichaltrigen) Geschwistern mindestens zwei im gleichen Monat geboren?

$$1 - \frac{12 \cdot 11 \cdot 10 \cdot 9}{12^4}$$

Anordnung von Büchern
Auf wie viele Arten lassen sich 10 Bücher in ein Regal reihen?

$$n = 10, \quad k = 10$$
$$\frac{n!}{(n-k)!} = 10!$$

Glühbirnen auswählen
Von **100 Glühbirnen** sind genau **drei defekt**. Es werden nun **6 Glühbirnen** zufällig ausgewählt.
Wie viele Möglichkeiten gibt es, wenn sich **mindestens eine defekte** Glühbirne in der Auswahl befinden soll?

$$\binom{100}{6} - \binom{97}{6} = 203'880'032$$

Mit wie viel Prozent Chancen ist bei einer Auswahl von 6 Glühbirnen **keine defekt**?

$$\frac{\binom{97}{6}}{\binom{100}{6}}$$

Buchstabenkombinationen
Wie viele Worte lassen sich aus den Buchstaben des Wortes ABRAKADABRA bilden? (Nur Worte in denen alle Buchstaben vorkommen!)
$A = 5x, \quad B = 2x, \quad R = 2x, \quad D = 1x, \quad K = 1x$

$$\binom{11}{5} \cdot \binom{6}{2} \cdot \binom{4}{2} \cdot \binom{2}{1} \cdot \binom{1}{1} = 83160$$

## Wahrscheinlichkeitstheorie

**Ergebnisraum und Laplace-Raum** Der **Ergebnisraum** $\Omega$ ist die Menge aller möglichen Ergebnisse des Zufallsexperiments. Die **Zähldichte** $\rho : \Omega \to [0, 1]$ ordnet jedem Ereignis seine Wahrscheinlichkeit zu.
Für einen **Laplace-Raum** $(\Omega, P)$ gilt:

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

**Parameter:**
- $\Omega$ = Ergebnisraum (Menge aller möglichen Ergebnisse)
- $P(M)$ = Wahrscheinlichkeit des Ereignisses $M$
- $|M|$ = Anzahl der für $M$ günstigen Ergebnisse
- $|\Omega|$ = Anzahl aller möglichen Ergebnisse

**Wahrschinlichkeits Ausdrücke**
- $P(A)$ = Wahrscheinlichkeit von Ereignis $A$
- $P(B)$ = Wahrscheinlichkeit von Ereignis $B$
- $P(\bar{A})$ = Wahrscheinlichkeit des Gegenereignisses von $A$
- $P(B|A)$ = Wahrscheinlichkeit von $B$ unter der Bedingung dass $A$ eingetreten ist
- $P(B|\bar{A})$ = Wahrscheinlichkeit von $B$ unter der Bedingung dass $A$ nicht eingetreten ist
- $P(A \cap B)$ = Wahrscheinlichkeit dass beide Ereignisse eintreten
- $P(A \cup B)$ = Wahrscheinlichkeit dass mindestens eines der Ereignisse eintritt

### Stochastische Unabhängigkeit

Zwei Ereignisse $A$ und $B$ heissen stochastisch unabhängig, falls:

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

Zwei Zufallsvariablen $X : \Omega \to \mathbb{R}$ und $Y : \Omega \to \mathbb{R}$ heissen stochastisch unabhängig, falls:

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y), \quad \text{für alle } x, y \in \mathbb{R}$$

$P(X = x, Y = y)$ = Wahrscheinlichkeit dass $X$ den Wert $x$ und $Y$ den Wert $y$ annimmt
$P(X = x)$ = Wahrscheinlichkeit dass $X$ den Wert $x$ annimmt
$P(Y = y)$ = Wahrscheinlichkeit dass $Y$ den Wert $y$ annimmt

### Wahrscheinlichkeitsregeln

- **Additionssatz:** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- **Komplementärregel:** $P(\bar{A}) = 1 - P(A)$
- **Multiplikationssatz:** $P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$
- **Totale Wahrscheinlichkeit:** $P(B) = P(A) \cdot P(B|A) + P(\bar{A}) \cdot P(B|\bar{A})$
- **Satz von Bayes:** $P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B)}$
- **Bedingte Wahrscheinlichkeit:** $P(B|A) = \frac{P(B \cap A)}{P(A)}$
- **Stochastische Unabhängigkeit:** $P(A \cap B) = P(A) \cdot P(B)$
- **Stochastische Unabhängigkeit (Zufallsvariablen):** $P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$

Kartenspiel: Rommé **Aufgabe:** Beim Rommé spielt man mit 110 Karten, davon sind 6 Joker. Jeder Spieler erhält 12 Karten.
**Teil 1:** Berechne die Wahrscheinlichkeit für genau zwei Joker.

- **Ergebnisraum:** Alle möglichen 12-Karten-Hände: $|\Omega| = \binom{110}{12}$
- **Günstige Ereignisse:**
  - 2 Joker aus 6: $\binom{6}{2}$
  - 10 Nicht-Joker aus 104: $\binom{104}{10}$
- **Berechnung:** $P(2 \text{ Joker}) = \frac{\binom{6}{2} \cdot \binom{104}{10}}{\binom{110}{12}}$

**Teil 2:** Berechne die Wahrscheinlichkeit für mindestens einen Joker.

- **Strategie:** Berechnung über Gegenereignis (kein Joker)
- **Berechnung:** $P(\text{mind. 1 Joker}) = 1 - \frac{\binom{104}{12}}{\binom{110}{12}}$

Glühbirnen-Problem **Aufgabe:** Von 100 Glühbirnen sind 3 defekt. Es werden 6 zufällig ausgewählt.
**Teil 1:** Anzahl Möglichkeiten mit mindestens einer defekten Glühbirne.

- **Gesamtmöglichkeiten:** $\binom{100}{6}$
- **Gegenereignis:** Keine defekte = $\binom{97}{6}$
- **Lösung:** $\binom{100}{6} - \binom{97}{6} = 203'880'032$

**Teil 2:** Wahrscheinlichkeit für keine defekte Glühbirne.

$$P(\text{keine defekt}) = \frac{\binom{97}{6}}{\binom{100}{6}}$$

### Problemlösung mit Gegenereignis

1. **Prüfe, ob Gegenereignis einfacher ist**
   - Original: "Mindestens eine...öder "Mehr als..."
   - Gegenereignis: "Keine...öder "Höchstens..."
2. **Berechne Wahrscheinlichkeit des Gegenereignis**
   - Oft einfacher zu zählen
   - Weniger Fälle zu berücksichtigen
3. **Wende Komplementärregel an**
   - $P(A) = 1 - P(\bar{A})$
   - Überprüfe Plausibilität des Ergebnisses

## Bedingte Wahrscheinlichkeit

### Bedingte Wahrscheinlichkeit

Die bedingte Wahrscheinlichkeit von $B$ unter der Bedingung $A$ ist:

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

**Parameter:**

- $P(B \cap A)$ = Wahrscheinlichkeit des Durchschnitts ??

### Multiplikationssatz

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

**Anwendung:**

- Berechnung von Schnittwahrscheinlichkeiten
- Prüfung auf stochastische Unabhängigkeit
- Zerlegung von mehrstufigen Experimenten

### Erstellen einer Vierfeldertafel

1. **Aufbau der Tabelle**
   - Zeilen: Erstes Merkmal (A und nicht A)
   - Spalten: Zweites Merkmal (B und nicht B)
   - Randwahrscheinlichkeiten notieren
2. **Eintragen der Wahrscheinlichkeiten**
   - Schnittwahrscheinlichkeiten in die Felder
   - Zeilensummen = P(A) bzw. P(nicht A)
   - Spaltensummen = P(B) bzw. P(nicht B)
3. **Berechnung bedingter Wahrscheinlichkeiten**
   - $P(B|A) = \frac{P(A \cap B)}{P(A)}$
   - $P(A|B) = \frac{P(A \cap B)}{P(B)}$

Medizinischer Test **Aufgabe:** Ein Test auf eine Krankheit hat folgende Eigenschaften:

- 1% der Bevölkerung hat die Krankheit
- Test ist bei Kranken zu 98% positiv
- Test ist bei Gesunden zu 95% negativ

**Lösung mit Vierfeldertafel:**

| | Test + | Test - | Summe |
|---|---|---|---|
| Krank | 0.0098 | 0.0002 | 0.01 |
| Gesund | 0.0495 | 0.9405 | 0.99 |
| Summe | 0.0593 | 0.9407 | 1 |

**Berechnung:** Wahrscheinlichkeit krank bei positivem Test:

$$P(\text{krank}|\text{positiv}) = \frac{0.0098}{0.0593} \approx 0.165 = 16.5\%$$

## Spezielle Sätze

### Satz von der Totalen Wahrscheinlichkeit

$$P(B) = P(A) \cdot P(B|A) + P(\bar{A}) \cdot P(B|\bar{A})$$

**Anwendung:**

- Berechnung von P(B) durch Fallunterscheidung
- Basis für den Satz von Bayes
- Wichtig bei Entscheidungsbäumen

### Anwendung des Satzes der totalen Wahrscheinlichkeit

### Satz von Bayes

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

**Anwendung:**

- Umkehrung bedingter Wahrscheinlichkeiten
- Aktualisierung von Wahrscheinlichkeiten
- Diagnostische Tests

### Anwendung des Satzes von Bayes

1. **Identifiziere die bekannten Größen**
   - A priori Wahrscheinlichkeit P(A)
   - Bedingte Wahrscheinlichkeit P(B|A)
   - Totale Wahrscheinlichkeit P(B)
2. **Berechne P(B) falls nötig**
   - Nutze Satz der totalen Wahrscheinlichkeit
   - $P(B) = P(A) \cdot P(B|A) + P(\bar{A}) \cdot P(B|\bar{A})$
3. **Berechne P(A|B)**
   - Setze in Bayes-Formel ein
   - Interpretiere das Ergebnis

Qualitätskontrolle **Aufgabe:** Eine Maschine produziert Teile.

- 95% der Teile sind fehlerfrei
- Ein Test erkennt fehlerhafte Teile zu 98%
- Der Test klassifiziert 3% der guten Teile falsch

**Gesucht:** Wahrscheinlichkeit für tatsächlich fehlerhaftes Teil bei positivem Test
**Lösung:**

- $P(F) = 0.05$ (fehlerhaft)
- $P(T|F) = 0.98$ (Test positiv wenn fehlerhaft)
- $P(T|\neg F) = 0.03$ (Test positiv wenn gut)
- $P(T) = 0.05 \cdot 0.98 + 0.95 \cdot 0.03 = 0.0775$
- $P(F|T) = \frac{0.05 \cdot 0.98}{0.0775} \approx 0.632 = 63.2\%$

## Erwartungswert und Varianz

### Kenngrössen von Zufallsvariablen

**Wichtige Eigenschaften:**

- **Erwartungswert:** $E(X+Y) = E(X) + E(Y)$, $E(\alpha X) = \alpha E(X)$
- **Varianz:** $V(X) = E(X^2) - E(X)^2$
- **Standardabweichung:** $S(X) = \sqrt{V(X)}$
- **Lineare Transformation:** $V(\alpha X + \beta) = \alpha^2 \cdot V(X)$

## Kenngrössen (Varianz und Erwartungswert)

$$E(X+Y) = E(X) + E(Y), \quad E(\alpha X) = \alpha E(X)$$

$$V(X) = E(X^2) - E(X)^2 = \left[\sum_{x\in\mathbb{R}} P(X=x)\cdot x^2\right] - E(X)^2$$

$$V(\alpha X+\beta) = \alpha^2 \cdot V(X), \quad S(X) = \sqrt{V(X)}$$

$E(X)$ = Erwartungswert der Zufallsvariable $X$
$V(X)$ = Varianz der Zufallsvariable $X$
$S(X)$ = Standardabweichung der Zufallsvariable $X$
$\alpha, \beta$ = Konstanten
$P(X=x)$ = Wahrscheinlichkeit, dass $X$ den Wert $x$ annimmt
$\sum_{x\in\mathbb{R}}$ = Summe über alle möglichen Werte von $x$ in den reellen Zahlen

## Verteilungen und Erwartungswerte

Für diskrete Verteilungen:

$$E(X) = \sum_{x\in\mathbb{R}} f(x)\cdot x$$

$$V(X) = \sum_{x\in\mathbb{R}} f(x)\cdot (x-E(X))^2$$

$E(X)$ = Erwartungswert der Zufallsvariable $X$
$V(X)$ = Varianz der Zufallsvariable $X$
$f(x)$ = Wahrscheinlichkeitsfunktion
$x$ = Mögliche Werte der Zufallsvariable

Für stetige Verteilungen:

$$E(X) = \int_{-\infty}^{\infty} f(x)\cdot x dx$$

$$V(X) = \int_{-\infty}^{\infty} f(x)\cdot (x-E(X))^2 dx$$

$E(X)$ = Erwartungswert der Zufallsvariable $X$
$V(X)$ = Varianz der Zufallsvariable $X$
$f(x)$ = Dichtefunktion
$x$ = Mögliche Werte der Zufallsvariable

## Berechnung von Erwartungswert und Varianz

1. **Erwartungswert bestimmen**
   - Diskret: $E(X) = \sum_x x \cdot P(X=x)$
   - Stetig: $E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$
2. **Varianz berechnen (2 Methoden)**
   - Direkte Methode: $V(X) = \sum_x (x-E(X))^2 \cdot P(X=x)$
   - Verschiebungssatz: $V(X) = E(X^2) - (E(X))^2$
3. **Bei Standardabweichung**
   - Wurzel aus Varianz ziehen
   - Einheit beachten (gleich wie Ursprungsdaten)

Erwartungswert bei Würfelspiel **Aufgabe:** Bei einem Würfelspiel gewinnt man:
- Bei 6: 5€
- Bei 5: 2€
- Bei 1-4: verliert man 1€

**Lösung:**
1. **Wahrscheinlichkeiten und Werte aufstellen:**
   - P(X = 5€) = 1/6
   - P(X = 2€) = 1/6
   - P(X = -1€) = 4/6
2. **Erwartungswert berechnen:**

$$E(X) = 5\cdot\frac{1}{6} + 2\cdot\frac{1}{6} + (-1)\cdot\frac{4}{6} = \frac{5+2-4}{6} = \frac{3}{6} = 0.5$$

3. **Varianz berechnen:**

$$E(X^2) = 25\cdot\frac{1}{6} + 4\cdot\frac{1}{6} + 1\cdot\frac{4}{6} = \frac{25+4+4}{6} = \frac{33}{6}$$

$$V(X) = E(X^2) - (E(X))^2 = \frac{33}{6} - (\frac{1}{2})^2 = \frac{33}{6} - \frac{1}{4} \approx 5.25$$

**Interpretation:**
- Positiver Erwartungswert: Spiel ist langfristig profitabel
- Hohe Varianz: Große Schwankungen möglich

Lotterie mit bedingten Gewinnen **Aufgabe:** Bei einer Lotterie gewinnt man zunächst mit p = 0.1 einen Bonus-Los. Mit diesem Los kann man dann mit p = 0.2 den Hauptpreis von 1000€ gewinnen. Berechne den Erwartungswert.

**Lösung:**
1. **Ereignisbaum erstellen:**
   - P(Bonus) = 0.1
   - P(Hauptgewinn|Bonus) = 0.2
2. **Mögliche Ausgänge:**
   - 1000€: P = 0.1 · 0.2 = 0.02
   - 0€: P = 0.98
3. **Erwartungswert:**

$$E(X) = 1000\cdot 0.02 + 0\cdot 0.98 = 20$$

## Interpretation von Erwartungswert und Varianz

1. **Erwartungswert**
   - Langfristiger Durchschnitt
   - Schwerpunkt der Verteilung
   - Nicht unbedingt ein möglicher Wert
2. **Varianz**
   - Maß für die Streuung
   - Quadratische Einheit beachten
   - Je größer, desto unsicherer die Vorhersage
3. **Standardabweichung**
   - Gleiche Einheit wie Daten
   - Typische Abweichung vom Mittelwert
   - Oft für Konfidenzintervalle verwendet

Aktienportfolio **Aufgabe:** Ein Portfolio besteht aus:
- Aktie A: 60% Anteil, E(A) = 8%, V(A) = 25
- Aktie B: 40% Anteil, E(B) = 12%, V(B) = 36

**Lösung:**
1. **Erwartungswert des Portfolios:**

$$\begin{aligned} E(P) &= 0.6\cdot E(A) + 0.4\cdot E(B) \\ &= 0.6\cdot 8\% + 0.4\cdot 12\% \\ &= 4.8\% + 4.8\% = 9.6\% \end{aligned}$$

2. **Varianz des Portfolios** (bei Unabhängigkeit):

$$\begin{aligned} V(P) &= (0.6)^2\cdot V(A) + (0.4)^2\cdot V(B) \\ &= 0.36\cdot 25 + 0.16\cdot 36 \\ &= 9 + 5.76 = 14.76 \end{aligned}$$

3. **Standardabweichung:**

$$S(P) = \sqrt{14.76} \approx 3.84\%$$

## Kovarianz und Korrelation

Die **Kovarianz** zweier Zufallsvariablen ist:

$$Cov(X,Y) = E((X-E(X))(Y-E(Y))) = E(XY) - E(X)E(Y)$$

Der **Korrelationskoeffizient** ist:

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}$$

**Eigenschaften:**
- $-1 \le \rho_{XY} \le 1$
- $\rho_{XY} = \pm 1$: perfekter linearer Zusammenhang
- $\rho_{XY} = 0$: unkorreliert

## Anwendung von Kovarianz und Korrelation

1. **Kovarianz berechnen**
   - Direkter Weg: $Cov(X,Y) = E(XY) - E(X)E(Y)$
   - Alternativ: $\frac{1}{n}\sum(x_i-\bar{x})(y_i-\bar{y})$
2. **Korrelation bestimmen**
   - Kovarianz durch Produkt der Standardabweichungen
   - Normierung auf [-1,1]
3. **Interpretation**
   - Vorzeichen: Richtung des Zusammenhangs
   - Betrag: Stärke des Zusammenhangs
   - Unabhängig von Maßeinheiten

Portfoliorisiko mit Korrelation **Aufgabe:** Zwei Aktien mit:
- A: E(A) = 10%, S(A) = 5%
- B: E(B) = 8%, S(B) = 4%
- Korrelation: $\rho_{AB} = 0.3$

**Portfolio:** 60% A, 40% B

**Lösung:**
1. **Erwartungswert:**

$$E(P) = 0.6\cdot 10\% + 0.4\cdot 8\% = 9.2\%$$

2. **Varianz mit Korrelation:**

$$\begin{aligned} V(P) &= (0.6)^2 V(A) + (0.4)^2 V(B) + 2(0.6)(0.4)\rho_{AB} S(A) S(B) \\ &= 0.36\cdot (5\%)^2 + 0.16\cdot (4\%)^2 + 2(0.24)(0.3)(5\%)(4\%) \\ &= 0.09\% + 0.0256\% + 0.0288\% = 0.1444\% \end{aligned}$$

3. **Standardabweichung:**

$$S(P) = \sqrt{0.1444\%} \approx 3.8\%$$

## Stochastische Unabhängigkeit

**Stochastische Unabhängigkeit** Zwei Ereignisse $A$ und $B$ heißen stochastisch unabhängig, falls:

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

Zwei Zufallsvariablen $X$ und $Y$ heißen stochastisch unabhängig, falls für alle $x, y \in \mathbb{R}$:

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$$

**Eigenschaften:**
- Für unabhängige Ereignisse: $P(A|B) = P(A)$
- Für unabhängige Zufallsvariablen: $E(XY) = E(X)E(Y)$
- Varianz der Summe: $V(X + Y) = V(X) + V(Y)$

**Prüfung auf stochastische Unabhängigkeit** 1. **Für Ereignisse**
- Berechne $P(A \cap B)$
- Berechne $P(A) \cdot P(B)$
- Vergleiche die Werte

2. **Für Zufallsvariablen**
- Stelle Verbundverteilung auf
- Prüfe für alle Wertepaare
- Alternative: Prüfe Kovarianz = 0

3. **Praktische Überlegungen**
- Physikalische/logische Abhängigkeit?
- Kausaler Zusammenhang?
- Gemeinsame Einflussfaktoren?

Würfelwurf und Münzwurf **Aufgabe:** Ein Würfel wird geworfen und eine Münze geworfen. Ereignisse:
- A: "Würfel zeigt eine 6"
- B: "Münze zeigt Kopf"

**Lösung:**
1. **Einzelwahrscheinlichkeiten:**
   - $P(A) = \frac{1}{6}$
   - $P(B) = \frac{1}{2}$
2. **Schnittwahrscheinlichkeit:** $P(A \cap B) = \frac{1}{12} = \frac{1}{6} \cdot \frac{1}{2} = P(A) \cdot P(B)$
3. **Schlussfolgerung:** Die Ereignisse sind stochastisch unabhängig

Kartenziehen ohne Zurücklegen **Aufgabe:** Aus einem Kartenspiel werden nacheinander zwei Karten gezogen. Ereignisse:
- A: Ërste Karte ist Herz"
- B: SZweite Karte ist Herz"

**Lösung:**
1. **Wahrscheinlichkeiten:**
   - $P(A) = \frac{13}{52} = \frac{1}{4}$
   - $P(B|A) = \frac{12}{51}$
   - $P(B|\bar{A}) = \frac{13}{51}$
2. **Prüfung:**

$$P(B) = \frac{13}{52} \neq P(B|A)$$

3. **Schlussfolgerung:** Die Ereignisse sind stochastisch abhängig

# Spezielle Verteilungen

## Diskrete und Stetige Zufallsvariablen

**Verteilungen und Erwartungswerte** **Für diskrete Verteilungen:**

$$E(X) = \sum_{x \in \mathbb{R}} f(x) \cdot x$$

$$V(X) = \sum_{x \in \mathbb{R}} f(x) \cdot (x - E(X))^2$$

**Für stetige Verteilungen:**

$$E(X) = \int_{-\infty}^{\infty} f(x) \cdot x dx$$

$$V(X) = \int_{-\infty}^{\infty} f(x) \cdot (x - E(X))^2 dx$$

**Parameter:**
- $E(X)$ = Erwartungswert der Zufallsvariable $X$
- $V(X)$ = Varianz der Zufallsvariable $X$
- $f(x)$ = Wahrscheinlichkeitsfunktion (diskret) oder Dichtefunktion (stetig)
- $x$ = Mögliche Werte der Zufallsvariable

**Berechnung von Erwartungswert und Varianz** 1. **Diskrete Verteilung**
- Liste alle möglichen Werte $x_i$ auf
- Bestimme zugehörige Wahrscheinlichkeiten $P(X = x_i)$
- $E(X) = \sum x_i \cdot P(X = x_i)$
- $V(X) = \sum (x_i - E(X))^2 \cdot P(X = x_i)$

2. **Stetige Verteilung**
- Identifiziere Dichtefunktion $f(x)$
- Berechne $E(X) = \int x \cdot f(x) dx$
- Berechne $V(X) = \int (x - E(X))^2 \cdot f(x) dx$

## Diskrete Verteilungen

**Bernoulli-Verteilung** Experiment mit genau zwei möglichen Ausgängen (Erfolg/Misserfolg).

$$P(X = 1) = p, \quad P(X = 0) = 1 - p = q$$

**Kenngrößen:**
- $E(X) = p$
- $V(X) = p(1 - p)$

**Anwendung der Bernoulli-Verteilung** 1. **Prüfe Voraussetzungen**
- Genau zwei mögliche Ausgänge
- Unabhängige Wiederholungen
- Konstante Erfolgswahrscheinlichkeit

2. **Parameter identifizieren**
- p = Erfolgswahrscheinlichkeit
- q = 1-p = Misserfolgswahrscheinlichkeit

3. **Berechnung**
- E(X) = p
- V(X) = pq

**Bernoulliverteilung** Bernoulli-Experimente sind Zufallsexperimente mit nur zwei möglichen Ergebnissen (1 und 0):

$$P(X = 1) = p, \quad P(X = 0) = 1 - p = q$$

Es gilt:
1. $E(X) = E(X^2) = p$
2. $V(X) = p \cdot (1 - p)$

$E(X)$ = Erwartungswert
$V(X)$ = Varianz
$P(X = 1)$ = Wahrscheinlichkeit für Erfolg
$p$ = Erfolgswahrscheinlichkeit
$q$ = Gegenwahrscheinlichkeit $(1 - p)$

**Bernoulliverteilung** Experiment mit genau zwei möglichen Ausgängen:

$$P(X = 1) = p, \quad P(X = 0) = 1 - p = q$$

**Parameter:**
- $p$ = Erfolgswahrscheinlichkeit
- $q = 1 - p$ = Gegenwahrscheinlichkeit

**Kenngrößen:**
1. $E(X) = E(X^2) = p$
2. $V(X) = p \cdot (1 - p) = pq$

Münzwurf **Aufgabe:** Faire Münze wird geworfen. X = 1 bei Kopf, X = 0 bei Zahl.

**Lösung:**
- p = 0.5 (faire Münze)
- $E(X) = 0.5$
- $V(X) = 0.5 \cdot 0.5 = 0.25$
- $P(X = 1) = 0.5$
- $P(X = 0) = 0.5$

**Binomialverteilung** n-malige unabhängige Wiederholung eines Bernoulli-Experiments.

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

**Kenngrößen:**
- $E(X) = np$
- $V(X) = np(1 - p)$

**Notation:** $X \sim B(n, p)$

Qualitätskontrolle mit Binomialverteilung **Aufgabe:** Eine Maschine produziert Teile mit Ausschussquote 5%. In einer Stichprobe von 100 Teilen:
- a) Wie viele defekte Teile sind zu erwarten?
- b) Wie groß ist die Wahrscheinlichkeit für genau 3 defekte Teile?
- c) Wie groß ist die Wahrscheinlichkeit für höchstens 2 defekte Teile?

**Lösung:**

1. **Parameter:**
   - n = 100 (Stichprobenumfang)
   - p = 0.05 (Ausschusswahrscheinlichkeit)
   - X ∼ B(100, 0.05)
2. **Erwartungswert:**

$$E(X) = np = 100 \cdot 0.05 = 5$$

3. **Genau 3 defekte:**

$$P(X = 3) = \binom{100}{3}(0.05)^3(0.95)^{97} \approx 0.1404$$

4. **Höchstens 2 defekte:**

$$P(X \leq 2) = \sum_{k=0}^{2} \binom{100}{k}(0.05)^k(0.95)^{100-k} \approx 0.0861$$

## Binomialverteilung (Mit zurücklegen)

- $n$ = Anzahl Wiederholungen
- $p$ = Wahrscheinlichkeit für ein Ergebnis 1
- $q = 1 - p$

$$P(X = x) = \binom{n}{x} \cdot p^x \cdot q^{n-x}$$

Schreibweise: $X \sim B(n; p)$
1. $\mu = E(X) = np$ 2. $\sigma^2 = V(X) = npq$ 3. $\sigma = S(X) = \sqrt{npq}$

**Binomialverteilung** n-malige unabhängige Wiederholung eines Bernoulli-Experiments:

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}$$

**Parameter:**
- $n$ = Anzahl Wiederholungen
- $p$ = Wahrscheinlichkeit für Erfolg
- $q = 1 - p$ = Gegenwahrscheinlichkeit

**Kenngrößen:**
1. $E(X) = np$
2. $V(X) = npq$
3. $\sigma = \sqrt{npq}$

**Notation:** $X \sim B(n; p)$

Binomialverteilung in der Qualitätskontrolle **Aufgabe:** Ein Produktionsprozess hat eine Fehlerquote von 5%. In einer Stichprobe von 100 Teilen:

**Parameter:**
- n = 100 (Stichprobenumfang)
- p = 0.05 (Fehlerwahrscheinlichkeit)
- X ∼ B(100, 0.05)

**Berechnung:**
- $E(X) = 100 \cdot 0.05 = 5$ defekte Teile erwartet
- $V(X) = 100 \cdot 0.05 \cdot 0.95 = 4.75$
- $P(X = 3) = \binom{100}{3} \cdot 0.05^3 \cdot 0.95^{97} \approx 0.1754$
- $P(X \leq 2) = \sum_{k=0}^{2} \binom{100}{k} \cdot 0.05^k \cdot 0.95^{100-k} \approx 0.1247$

## Hypergeometrische Verteilung (Ohne zurücklegen)

- $N$ = Objekte gesamthaft
- $M$ = Objekte einer bestimmten Sorte
- $n$ = Stichprobengrösse
- $x$ = Merkmalsträger

$$P(X = x) = \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}}$$

Schreibweise: $X \sim H(N, M, n)$
1. $\mu = E(X) = n \cdot \frac{M}{N}$ 2. $\sigma^2 = V(X) = n \cdot \frac{M}{N} \cdot (1 - \frac{M}{N}) \cdot \frac{N-n}{N-1}$ 3. $\sigma = S(X) = \sqrt{V(X)}$

**Hypergeometrische Verteilung** Ziehen ohne Zurücklegen aus einer endlichen Grundgesamtheit.

**Parameter:**
- N: Grundgesamtheit
- M: Anzahl der Merkmalsträger
- n: Stichprobenumfang

$$P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

**Kenngrößen:**
- $E(X) = n\frac{M}{N}$
- $V(X) = n\frac{M}{N}(1 - \frac{M}{N})\frac{N-n}{N-1}$

**Notation:** $X \sim H(N, M, n)$

Ziehung ohne Zurücklegen **Aufgabe:** In einer Urne sind 20 Kugeln, davon 8 rot. Es werden 5 Kugeln ohne Zurücklegen gezogen.

**Lösung:**

1. **Parameter:**
   - N = 20 (Gesamtanzahl)
   - M = 8 (rote Kugeln)
   - n = 5 (Ziehungen)
2. **Erwartungswert:**

$$E(X) = 5 \cdot \frac{8}{20} = 2$$

3. **Varianz:**

$$V(X) = 5 \cdot \frac{8}{20} \cdot \frac{12}{20} \cdot \frac{15}{19} \approx 1.184$$

4. **P(genau 2 rote):**

$$P(X = 2) = \frac{\binom{8}{2}\binom{12}{3}}{\binom{20}{5}} \approx 0.3682$$

**Hypergeometrische Verteilung** Ziehen ohne Zurücklegen:

$$P(X = k) = \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}$$

**Parameter:**
- $N$ = Grundgesamtheit
- $M$ = Anzahl Merkmalsträger
- $n$ = Stichprobengröße
- $k$ = Erfolge in Stichprobe

**Kenngrößen:**
1. $E(X) = n \cdot \frac{M}{N}$
2. $V(X) = n \cdot \frac{M}{N} \cdot (1 - \frac{M}{N}) \cdot \frac{N-n}{N-1}$
3. $\sigma = \sqrt{V(X)}$

**Notation:** $X \sim H(N, M, n)$

Lotterie **Aufgabe:** In einer Urne sind 100 Lose, davon 10 Gewinnerlose. Ein Spieler zieht 5 Lose.

**Parameter:**
- N = 100 (Gesamtlose)
- M = 10 (Gewinnerlose)
- n = 5 (gezogene Lose)

**Berechnung:**
- $E(X) = 5 \cdot \frac{10}{100} = 0.5$ Gewinne erwartet
- $V(X) = 5 \cdot \frac{10}{100} \cdot \frac{90}{100} \cdot \frac{95}{99} \approx 0.432$
- $P(X = 1) = \frac{\binom{10}{1} \cdot \binom{90}{4}}{\binom{100}{5}} \approx 0.3726$

## Poisson Verteilung

- $\lambda$ = Rate

$$P(X = x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}, \quad \lambda > 0$$

Schreibweise: $X \sim Poi(\lambda)$
1. $\mu = E(X) = \lambda$ 2. $\sigma^2 = V(X) = \lambda$ 3. $\sigma = S(X) = \sqrt{\lambda}$

**Poisson-Verteilung** Modelliert seltene Ereignisse in festem Zeit- oder Raumintervall.

$$P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}$$

**Parameter:**
- $\lambda$: Erwartungswert pro Intervall

**Kenngrößen:**
- $E(X) = \lambda$
- $V(X) = \lambda$

**Notation:** $X \sim Poi(\lambda)$

**Poisson-Verteilung** Modelliert seltene Ereignisse:

$$P(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}, \quad \lambda > 0$$

**Parameter:**
- $\lambda$ = Erwartungswert/Rate

**Kenngrößen:**
1. $E(X) = \lambda$
2. $V(X) = \lambda$
3. $\sigma = \sqrt{\lambda}$

**Notation:** $X \sim Poi(\lambda)$

Anrufe in Call-Center **Aufgabe:** Ein Call-Center erhält durchschnittlich 4 Anrufe pro Stunde.

**Parameter:**
- $\lambda = 4$ (Anrufe pro Stunde)
- X $\sim$ Poi(4)

**Berechnung:**
- $E(X) = 4$ Anrufe erwartet
- $V(X) = 4$
- $P(X = 3) = \frac{4^3}{3!} \cdot e^{-4} \approx 0.1954$
- $P(X \leq 2) = e^{-4} \cdot (1 + 4 + \frac{16}{2}) \approx 0.2381$

Poisson-Verteilung in der Praxis **Aufgabe:** Ein Callcenter erhält durchschnittlich 3 Anrufe pro 10 Minuten.
- a) Wahrscheinlichkeit für genau 2 Anrufe in 10 Minuten?
- b) Wahrscheinlichkeit für mehr als 4 Anrufe?

**Lösung:**
1. **Parameter:**
   - $\lambda = 3$ (Erwartungswert)
   - X $\sim$ Poi(3)
2. **Genau 2 Anrufe:**

$$P(X = 2) = \frac{3^2}{2!} e^{-3} \approx 0.2240$$

3. **Mehr als 4 Anrufe:**

$$P(X > 4) = 1 - \sum_{k=0}^{4} \frac{3^k}{k!} e^{-3} \approx 0.1847$$

**Wahl der richtigen Verteilung** 1. **Prüfe Ziehungsart**
- Mit Zurücklegen → Binomialverteilung
- Ohne Zurücklegen → Hypergeometrische Verteilung
- Seltene Ereignisse → Poisson-Verteilung

2. **Prüfe Grundgesamtheit**
- Endlich, klein → Hypergeometrische Verteilung
- Sehr groß/unendlich → Binomialverteilung
- Zeitlich/räumlich kontinuierlich → Poisson-Verteilung

3. **Beachte Approximationen**
- Binomial → Poisson für $n \to \infty$, $p \to 0$, $np = \lambda$
- Hypergeometrisch → Binomial für $\frac{n}{N} \leq 0.05$

**Wahl der richtigen diskreten Verteilung** 1. **Bernoulli-Verteilung**
- Genau zwei mögliche Ausgänge
- Ein einzelner Versuch
- Konstante Erfolgswahrscheinlichkeit

2. **Binomial-Verteilung**
- Feste Anzahl unabhängiger Versuche
- Mit Zurücklegen/große Grundgesamtheit
- Konstante Erfolgswahrscheinlichkeit

3. **Hypergeometrische Verteilung**
- Ziehen ohne Zurücklegen
- Endliche Grundgesamtheit
- Veränderliche Wahrscheinlichkeiten

4. **Poisson-Verteilung**
- Seltene Ereignisse
- Festes Zeitintervall/Raumbereich
- Rate $\lambda$ bekannt

## Normalverteilung

**Gauss-Verteilung** Die Dichtefunktion der Normalverteilung ist:

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

**Standardnormalverteilung** ($\mu = 0$, $\sigma = 1$):

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2}$$

**Parameter:**
- $\mu$ = Erwartungswert
- $\sigma$ = Standardabweichung
- $\varphi_{\mu,\sigma}(x)$ = Dichtefunktion
- $\phi_{\mu,\sigma}(x)$ = Verteilungsfunktion

**Notation:** $X \sim N(\mu; \sigma)$

**Gauss-Verteilung** Die stetige Zufallsvariable $X$ folgt der Normalverteilung mit den Parametern $\mu, \sigma \in \mathbb{R}, \sigma > 0$:

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

Standardnormalverteilung ($\mu = 0$ und $\sigma = 1$):

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2}$$

$\varphi_{\mu,\sigma}(x)$ = Dichtefunktion der Normalverteilung
$\varphi(x)$ = Dichtefunktion der Standardnormalverteilung

$\mu$ = Erwartungswert
$\sigma$ = Standardabweichung
$e$ = Eulersche Zahl
$\pi$ = Kreiszahl Pi

**Die Verteilungsfunktion der Normalverteilung**
Die kumulative Verteilungsfunktion (CDF) von $\varphi_{\mu,\sigma}(x)$ wird mit $\phi_{\mu,\sigma}(x)$ bezeichnet. Sie ist definiert durch:

$$\phi_{\mu,\sigma}(x) = P(X \leq x) = \int_{-\infty}^{x} \varphi_{\mu,\sigma}(t)dt = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \int_{-\infty}^{x} e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2} dt$$

$\phi_{\mu,\sigma}(x)$ = Verteilungsfunktion der Normalverteilung
$\varphi_{\mu,\sigma}(x)$ = Dichtefunktion der Normalverteilung
$P(X \leq x)$ = Wahrscheinlichkeit dass $X$ kleiner oder gleich $x$ ist
$\mu$ = Erwartungswert
$\sigma$ = Standardabweichung
$\pi$ = Kreiszahl Pi
$e$ = Eulersche Zahl

**Approximation durch die Normalverteilung**
- Binomialverteilung: $\mu = np, \sigma^2 = npq$
- Poissonverteilung: $\mu = \lambda, \sigma^2 = \lambda$

$$P(a \leq X \leq b) = \sum_{x=a}^{b} P(X = x) \approx \phi_{\mu,\sigma}(b + \frac{1}{2}) - \phi_{\mu,\sigma}(a - \frac{1}{2})$$

$P(a \leq X \leq b)$ = Wahrscheinlichkeit dass $X$ zwischen $a$ und $b$ liegt
$\phi_{\mu,\sigma}$ = Verteilungsfunktion der Normalverteilung
$a, b$ = Untere und obere Grenze

**Standardisierung der Normalverteilung**
Bei einer stetigen Zufallsvariable $X$ lässt sich die Verteilungsfunktion als Integral einer Funktion $f$ darstellen:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u) \cdot du$$

Liegt eine beliebige Normalverteilung $N(\mu, \sigma)$ vor, muss standardisiert werden. Statt ursprünglichen Zufallsvariablen $X$ betrachtet man die Zufallsvariable:

$$U = \frac{X - \mu}{\sigma}$$

$F(x)$ = Verteilungsfunktion
$P(X \leq x)$ = Wahrscheinlichkeit dass $X$ kleiner oder gleich $x$ ist
$f(u)$ = Dichtefunktion
$U$ = Standardisierte Zufallsvariable
$X$ = Ursprüngliche Zufallsvariable
$\mu$ = Erwartungswert
$\sigma$ = Standardabweichung

**Erwartungswert und Varianz der Normalverteilung**
Für eine Zufallsvariable $X \sim N(\mu; \sigma)$ gilt:

$$E(X) = \mu, \quad V(X) = \sigma^2$$

$E(X)$ = Erwartungswert der Zufallsvariable $X$
$V(X)$ = Varianz der Zufallsvariable $X$
$\mu$ = Erwartungsparameter
$\sigma^2$ = Varianzparameter

**Arbeiten mit der Normalverteilung** 1. **Standardisierung**
- $Z = \frac{X-\mu}{\sigma}$ transformiert zu N(0,1)
- Benutze Tabelle der Standardnormalverteilung
- Beachte: $\phi(z) = 1 - \phi(-z)$

2. **Stetigkeitskorrektur**
- Bei Approximation diskreter Verteilungen
- Untere Grenze: $a - 0.5$
- Obere Grenze: $b + 0.5$

3. **Faustregel für Intervalle**
- $\mu \pm \sigma$: ca. 68
- $\mu \pm 2\sigma$: ca. 95
- $\mu \pm 3\sigma$: ca. 99.7

Körpergrößen **Aufgabe:** Körpergrößen in einer Population sind normalverteilt mit $\mu = 175$ cm und $\sigma = 10$ cm.
**Berechnung:**
- $P(X \leq 185) = \phi(\frac{185-175}{10}) = \phi(1) \approx 0.8413$
- $P(165 \leq X \leq 185) = \phi(1) - \phi(-1) \approx 0.6826$
- $P(X > 195) = 1 - \phi(2) \approx 0.0228$

## Zentraler Grenzwertsatz und Approximationen

**Zentraler Grenzwertsatz** Für eine Folge von Zufallsvariablen $X_1, X_2, \ldots, X_n$ mit gleichem Erwartungswert $\mu$ und gleicher Varianz $\sigma^2$ gilt für die Summe $S_n$ und das arithmetische Mittel $\bar{X}_n$:

$$E(S_n) = n \cdot \mu, \quad V(S_n) = n \cdot \sigma^2$$

$$E(\bar{X}_n) = \mu, \quad V(\bar{X}_n) = \frac{\sigma^2}{n} = \frac{1}{n^2} \cdot V(S_n)$$

Die standardisierte Zufallsvariable ist:

$$U_n = \frac{(X_1 + X_2 + \cdots + X_n) - n\mu}{\sqrt{n} \cdot \sigma} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Für $n \to \infty$ konvergiert die Verteilungsfunktion $F_n(u)$ gegen die Standardnormalverteilung:

$$\lim_{n\to\infty} F_n(u) = \phi(u) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{u} e^{-\frac{1}{2}t^2} dt$$

**Anwendung des Zentralen Grenzwertsatzes** 1. **Prüfe Voraussetzungen**
- Unabhängige Zufallsvariablen
- Identische Verteilung
- Endliche Varianz
- Genügend große Stichprobe (n $\geq$ 30)

2. **Berechne Parameter**
- $\mu_{S_n} = n\mu$
- $\sigma_{S_n} = \sqrt{n}\sigma$
- $\mu_{\bar{X}} = \mu$
- $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

3. **Standardisiere**
- Transformiere zu $Z = \frac{X-\mu}{\sigma}$
- Verwende Tabelle der Standardnormalverteilung

**Zentraler Grenzwertsatz**
Für eine Folge von Zufallsvariablen $X_1, X_2, \ldots, X_n$ mit gleichem Erwartungswert $\mu$ und gleicher Varianz $\sigma^2$ gilt:

$$E(S_n) = n \cdot \mu, \quad V(S_n) = n \cdot \sigma^2, \quad E(\bar{X}_n) = \mu, \quad V(\bar{X}_n) = \frac{\sigma^2}{n} = \frac{1}{n^2} \cdot V(S$$

$S_n$ = Summe der Zufallsvariablen
$\bar{X}_n$ = Arithmetisches Mittel der Zufallsvariablen
$n$ = Anzahl der Zufallsvariablen
$\mu$ = Erwartungswert der einzelnen Zufallsvariablen
$\sigma^2$ = Varianz der einzelnen Zufallsvariablen

Die standardisierte Zufallsvariable:

$$U_n = \frac{((X_1 + X_2 + \cdots + X_n) - n\mu)}{\sqrt{n} \cdot \sigma} = \frac{(\bar{X} - \mu)}{\frac{\sigma}{\sqrt{n}}}$$

Sind die Zufallsvariablen alle identisch $N(\mu, \sigma)$ verteilt, so sind die Summe $S_n$ und das arithmetische Mittel $\bar{X}_n$ wieder normalverteilt mit:
- $S_n$ : $N(n \cdot \mu, \sqrt{n} \cdot \sigma)$
- $\bar{X}_n$ : $N(\mu, \frac{\sigma}{\sqrt{n}})$

Verteilungsfunktion $F_n(u)$ konvergiert für $n \to \infty$ gegen die Verteilungsfunktion $\phi(u)$ der Standardnormalverteilung:

$$\lim_{n\to\infty} F_n(u) = \phi(u) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{u} e^{-\frac{1}{2}t^2} dt$$

**Faustregeln für Approximationen**
- Die Approximation (Binomialverteilung) kann verwendet werden, wenn $npq > 9$
- Für grosses $n (n \geq 50)$ und kleines $p (p \leq 0.1)$ kann die Binomial- durch die Poisson-Verteilung approximiert werden:

$$B(n, p) \approx \text{Poi}(n \cdot p)$$

$B(n, p)$ = Binomialverteilung
$\text{Poi}(\lambda)$ = Poissonverteilung mit Parameter $\lambda = n \cdot p$

- Eine Hypergeometrische Verteilung kann durch eine Binomialverteilung angenähert werden, wenn $n \leq \frac{N}{20}$:

$$H(N, M, n) \approx B(n, \frac{M}{N})$$

$H(N, M, n)$ = Hypergeometrische Verteilung
$B(n, p)$ = Binomialverteilung
$N$ = Grundgesamtheit
$M$ = Anzahl der Erfolge in der Grundgesamtheit
$n$ = Stichprobengröße

**Faustregeln für Approximationen** **Binomialverteilung durch Normalverteilung:**
- Bedingung: $npq > 9$
- Parameter: $\mu = np$, $\sigma^2 = npq$
- Stetigkeitskorrektur beachten!

**Binomialverteilung durch Poissonverteilung:**
- Bedingung: $n \geq 50$ und $p \leq 0.1$
- $B(n, p) \approx Poi(n \cdot p)$

**Hypergeometrische durch Binomialverteilung:**
- Bedingung: $n \leq \frac{N}{20}$
- $H(N, M, n) \approx B(n, \frac{M}{N})$

Approximation der Binomialverteilung **Aufgabe:** Eine Produktionsanlage produziert mit Ausschusswahrscheinlichkeit 5%. In einer Charge von 200 Teilen:
- Wie groß ist die Wahrscheinlichkeit für 15 oder mehr defekte Teile?

**Lösung:**
1. **Prüfung Approximationsbedingung:**
   - $npq = 200 \cdot 0.05 \cdot 0.95 = 9.5 > 9$
   - Normalapproximation ist zulässig
2. **Parameter der Normalverteilung:**
   - $\mu = np = 200 \cdot 0.05 = 10$
   - $\sigma = \sqrt{npq} = \sqrt{9.5} \approx 3.08$
3. **Berechnung mit Stetigkeitskorrektur:**

$$\begin{aligned} P(X \geq 15) &= 1 - P(X \leq 14) \\ &= 1 - P(X \leq 14.5) \\ &= 1 - \phi(\frac{14.5 - 10}{3.08}) \\ &= 1 - \phi(1.46) \\ &\approx 0.0721 \end{aligned}$$

Approximation durch Poissonverteilung **Aufgabe:** Ein seltener Gendefekt tritt mit Wahrscheinlichkeit p = 0.001 auf. In einer Gruppe von 2000 Menschen:
- Wie groß ist die Wahrscheinlichkeit für genau 3 Betroffene?

**Lösung:**
1. **Prüfung Approximationsbedingung:**
   - $n = 2000 \geq 50$ und $p = 0.001 \leq 0.1$
   - Poissonapproximation ist zulässig
2. **Parameter:**
   - $\lambda = np = 2000 \cdot 0.001 = 2$
3. **Berechnung:**

$$P(X = 3) = \frac{2^3}{3!} \cdot e^{-2} \approx 0.180$$

4. **Vergleich mit Binomialverteilung:**

$$P_{Bin}(X = 3) = \binom{2000}{3} \cdot 0.001^3 \cdot 0.999^{1997} \approx 0.180$$

### Entscheidung über Approximationen

1. **Prüfe Stichprobenumfang**
- Klein (n < 30): Exakte Verteilung
- Mittel (30 ≤ n < 50): Je nach p
- Groß (n ≥ 50): Approximation möglich

2. **Prüfe Wahrscheinlichkeit**
- p ≤ 0.1: Poisson möglich
- 0.1 < p < 0.9: Normal möglich
- npq > 9: Normal empfohlen

3. **Wähle Approximation**
- Binomial → Normal: Große Stichproben, mittleres p
- Binomial → Poisson: Große n, kleines p
- Hypergeometrisch → Binomial: Kleine Stichprobe relativ zur Grundgesamtheit

4. **Beachte**
- Stetigkeitskorrektur bei Normal
- Rundungsregeln bei Grenzen
- Vergleich mit exakter Lösung wenn möglich

## Methode der kleinsten Quadrate

### Lineare Regression

Gegeben sind Datenpunkte $(x_i; y_i)$ mit $1 \leq i \leq n$. Die Residuen / Fehler $\epsilon_i = g(x_i) - y_i$ dieser Datenpunkte sind Abstände in $y$-Richtung zwischen $y_i$ und der Geraden $g$. Die Ausgleichs- oder Regressiongerade ist diejenige Gerade, für die die Summe der quadrierten Residuen $\sum_{i=1}^{n} \epsilon_i^2$ am kleinsten ist.
$(x_i, y_i)$ = Datenpunkte
$\epsilon_i$ = Residuum (Abweichung) des $i$-ten Datenpunkts
$g(x_i)$ = Wert der Regressionsgerade an der Stelle $x_i$
$n$ = Anzahl der Datenpunkte

### Lineare Regression berechnen

1. Berechne arithmetische Mittel $\bar{x}$ und $\bar{y}$ 2. Berechne Kovarianzen und Varianzen:
- $s_{xy} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$
- $s_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$
- $s_y^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2$

3. Berechne Steigung $m$ und y-Achsenabschnitt $d$:
- $m = \frac{s_{xy}}{s_x^2}$
- $d = \bar{y} - m\bar{x}$

4. Regressionsgerade: $g(x) = mx + d$

**Lineare Regression** Gegeben sind die Datenpunkte:

| $x_i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $y_i$ | 2.1 | 4.0 | 6.3 | 7.8 | 9.9 |

1. $\bar{x} = 3$, $\bar{y} = 6.02$
2. Kovarianzen und Varianzen:
- $s_{xy} = 3.945$
- $s_x^2 = 2$
- $s_y^2 = 8.4916$

3. Parameter:
- $m = \frac{3.945}{2} = 1.9725$
- $d = 6.02 - 1.9725 \cdot 3 = 0.1025$

4. Regressionsgerade: $g(x) = 1.9725x + 0.1025$

### Regressionsgerade

Die Regressionsgerade $g(x) = mx + d$ mit den Parametern $m$ und $d$ ist die Gerade, für welche die Residualvarianz $s_\epsilon^2$ minimal ist.

Steigung: $m = \frac{s_{xy}}{s_x^2}$, y-Achsenabschnitt: $d = \bar{y} - m\bar{x}$, $s_\epsilon^2 = s_y^2 - \frac{s_{xy}^2}{s_x^2}$

$m$ = Steigung der Regressionsgerade
$d$ = y-Achsenabschnitt
$s_{xy}$ = Kovarianz von $x$ und $y$
$s_x^2$ = Varianz der $x$-Werte
$s_y^2$ = Varianz der $y$-Werte
$\bar{x}$ = Arithmetisches Mittel der $x$-Werte
$\bar{y}$ = Arithmetisches Mittel der $y$-Werte
$s_\epsilon^2$ = Residualvarianz

### Residuen und Residuenplot analysieren

1. Berechne die Residuen für jeden Datenpunkt:
- $\epsilon_i = y_i - (mx_i + d)$

2. Erstelle Residuenplot:
- x-Achse: Prognostizierte Werte $\hat{y}_i = mx_i + d$
- y-Achse: Residuen $\epsilon_i$

3. Prüfe Eigenschaften:
- Residuen sollten zufällig um Null streuen
- Keine systematischen Muster erkennbar
- Gleiche Streubreite über alle $\hat{y}_i$

## Bestimmtheitsmass

### Varianzaufspaltung

Die Totale Varianz setzt sich zusammen aus der Residualvarianz und der Varianz der prognostizierten Werte:
- $s_y^2$ Totale Varianz
- $s_{\hat{y}}^2$ prognostizierte (erklärte) Varianz
- $s_\epsilon^2$ Residualvarianz

$$s_y^2 = s_\epsilon^2 + s_{\hat{y}}^2$$

$s_y^2$ = Totale Varianz der beobachteten $y$-Werte
$s_\epsilon^2$ = Varianz der Residuen
$s_{\hat{y}}^2$ = Varianz der durch die Regression geschätzten Werte

### Bestimmtheitsmass berechnen

1. Berechne die totale Varianz $s_y^2$ 2. Berechne die Residualvarianz $s_\epsilon^2$ 3. Berechne die erklärte Varianz $s_{\hat{y}}^2$ 4. Berechne das Bestimmtheitsmass:

$$R^2 = \frac{s_{\hat{y}}^2}{s_y^2} = 1 - \frac{s_\epsilon^2}{s_y^2}$$

5. Interpretation:
- $R^2 \approx 1$: Sehr gute Anpassung
- $R^2 \approx 0$: Schlechte Anpassung

### Bestimmtheitsmass

Das Bestimmtheitsmass $R^2$ beurteilt die globale Anpassungsgüte einer Regression über den Anteil der prognostizierten Varianz $s_{\hat{y}}^2$ an der totalen Varianz $s_y^2$:

$$R^2 = \frac{s_{\hat{y}}^2}{s_y^2}$$

$R^2$ = Bestimmtheitsmass (zwischen 0 und 1)
$s_{\hat{y}}^2$ = Varianz der prognostizierten Werte
$s_y^2$ = Totale Varianz

Das Bestimmtheitsmass $R^2$ entspricht dem Quadrat des Korrelationskoeffizienten:

$$R^2 = \frac{s_{xy}^2}{s_x^2 \cdot s_y^2} = (r_{xy})^2$$

$s_{xy}$ = Kovarianz von $x$ und $y$
$s_x^2$ = Varianz der $x$-Werte
$s_y^2$ = Varianz der $y$-Werte
$r_{xy}$ = Korrelationskoeffizient

## Linearisierungsfunktionen

### Transformationen

| Ausgangsfunktion | Transformation |
|---|---|
| $y = q \cdot x^m$ | $\log(y) = \log(q) + m \cdot \log(x)$ |
| $y = q \cdot m^x$ | $\log(y) = \log(q) + \log(m) \cdot x$ |
| $y = q \cdot e^{m \cdot x}$ | $\ln(y) = \ln(q) + m \cdot x$ |
| $y = \frac{1}{q + m \cdot x}$ | $V = q + m \cdot x; V = \frac{1}{y}$ |
| $y = q + m \cdot \ln(x)$ | $y = q + m \cdot U; u = \ln(x)$ |
| $y = \frac{1}{q \cdot m^x}$ | $\log(\frac{1}{y}) = \log(q) + \log(m) \cdot x$ |

$y$ = Abhängige Variable
$x$ = Unabhängige Variable
$q, m$ = Parameter der Funktion
$e$ = Eulersche Zahl
ln = Natürlicher Logarithmus
log = Logarithmus zur Basis 10

### Nichtlineare Regression durch Linearisierung

1. Bestimme passende Transformation aus Tabelle 2. Führe Transformation durch 3. Wende lineare Regression auf transformierte Daten an 4. Transformiere Parameter zurück

Beispiel für exponentielles Wachstum $y = q \cdot e^{mx}$:

1. Transformation: $\ln(y) = \ln(q) + mx$
2. Setze $Y = \ln(y)$, $b = \ln(q)$
3. Lineare Regression für $Y = mx + b$
4. Rücktransformation: $q = e^b$

**Exponentielles Wachstum** Gegeben sind die Messwerte:

| $x$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $y$ | 2.1 | 4.2 | 8.1 | 15.9 |

1. Transformation $Y = \ln(y)$:

| $x$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $Y$ | 0.742 | 1.435 | 2.092 | 2.766 |

2. Lineare Regression ergibt: $Y = 0.674x + 0.071$
3. Rücktransformation:
   - $m = 0.674$
   - $q = e^{0.071} = 1.074$
4. Ergebnis: $y = 1.074 \cdot e^{0.674x}$

## Methode der kleinsten Quadrate

### Matrix-Darstellung

Die Parameter $m$ und $q$ der Regressionsgeraden werden mit der Matrix $A$ berechnet:

$$A = \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}, \quad A^T \cdot A \cdot \begin{pmatrix} m \\ q \end{pmatrix} = A^T \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

### Matrix-Methode für lineare Regression

1. Erstelle Design-Matrix $A$:

$$A = \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}$$

2. Berechne $A^T \cdot A$ 3. Berechne $(A^T \cdot A)^{-1}$ 4. Berechne Parameter:

$$\begin{pmatrix} m \\ q \end{pmatrix} = (A^T \cdot A)^{-1} \cdot A^T \cdot \vec{y}$$

### Residuenberechnung

Die Residuen $\epsilon_i$ ergeben sich als:

$$\epsilon_i = y_i - \hat{y}_i = y_i - (mx_i + q)$$

Die Summe der quadrierten Residuen wird minimiert:

$$\sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - (mx_i + q))^2 \to \min$$

### Mehrfachregression

1. Aufstellen der Designmatrix:

$$A = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1(k-1)} & 1 \\ x_{21} & x_{22} & \cdots & x_{2(k-1)} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{n(k-1)} & 1 \end{pmatrix}$$

2. Berechnung der Parameter:

$$\vec{p} = (A^T A)^{-1} A^T \vec{y}$$

3. Residuen berechnen:

$$\vec{\epsilon} = \vec{y} - A\vec{p}$$

4. Bestimmtheitsmass:

$$R^2 = 1 - \frac{\sum \epsilon_i^2}{\sum (y_i - \bar{y})^2}$$

**Mehrfachregression** Ein Gebrauchtwagenhändler möchte den Preis (P) seiner Autos basierend auf Alter (A) und Kilometerstand (K) berechnen. Gegeben sind folgende Daten:

| Auto | Alter (Jahre) | km (10000) | Preis (1000 CHF) |
|---|---|---|---|
| 1 | 2 | 3 | 25 |
| 2 | 3 | 4 | 20 |
| 3 | 4 | 6 | 15 |
| 4 | 5 | 7 | 12 |

1. Designmatrix aufstellen:

$$A = \begin{pmatrix} 2 & 3 & 1 \\ 3 & 4 & 1 \\ 4 & 6 & 1 \\ 5 & 7 & 1 \end{pmatrix}$$

2. Parameter berechnen:

$$\vec{p} = \begin{pmatrix} -3 \\ -1.5 \\ 35 \end{pmatrix}$$

3. Resultierende Funktion:

$$P = -3A - 1.5K + 35$$

### Polynomiale Regression

Für Regression mit Polynomen höheren Grades:

1. Erweitere Designmatrix:

$$A = \begin{pmatrix} x_1^n & x_1^{n-1} & \cdots & x_1 & 1 \\ x_2^n & x_2^{n-1} & \cdots & x_2 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_m^n & x_m^{n-1} & \cdots & x_m & 1 \end{pmatrix}$$

2. Löse wie bei linearer Regression:

$$\vec{p} = (A^T A)^{-1} A^T \vec{y}$$

3. Polynom aufstellen:

$$y = p_1 x^n + p_2 x^{n-1} + \ldots + p_n x + p_{n+1}$$

**Quadratische Regression** Gegeben sind Messwerte:

| $x$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $y$ | 1 | 2.1 | 5.2 | 10.1 | 17.2 |

1. Designmatrix für quadratisches Polynom:

$$A = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \\ 16 & 4 & 1 \end{pmatrix}$$

2. Parameter berechnen:

$$\vec{p} = \begin{pmatrix} 1 \\ 0.1 \\ 1 \end{pmatrix}$$

3. Quadratische Funktion:

$$y = x^2 + 0.1x + 1$$

### Gütekriterien für Regression

1. Bestimmtheitsmass $R^2$:
   - $R^2 > 0.9$: Sehr gute Anpassung
   - $0.7 < R^2 < 0.9$: Gute Anpassung
   - $0.5 < R^2 < 0.7$: Mittelmässige Anpassung
   - $R^2 < 0.5$: Schlechte Anpassung
2. Residuenanalyse:
   - Residuen sollten zufällig um 0 schwanken
   - Keine systematischen Muster erkennbar
   - Residuen sollten normalverteilt sein
3. Prognosegüte:
   - Mittlerer quadratischer Fehler (MSE)
   - Wurzel des mittleren quadratischen Fehlers (RMSE)
   - Mittlerer absoluter Fehler (MAE)

**Modellwahl durch Residuenanalyse** Für einen Datensatz wurden drei Modelle getestet:

- Linear: $y = 2x + 1$
- Quadratisch: $y = x^2 + x + 1$
- Exponentiell: $y = 2e^{0.5x}$

Bestimmtheitsmasse:

- Linear: $R^2 = 0.85$
- Quadratisch: $R^2 = 0.98$
- Exponentiell: $R^2 = 0.92$

Residuenanalyse zeigt:

- Linear: Systematische Krümmung in Residuen
- Quadratisch: Zufällige Verteilung der Residuen
- Exponentiell: Leichte Systematik in Residuen

Schlussfolgerung: Das quadratische Modell ist am besten geeignet.

### Prüfungsaufgaben lösen

1. Aufgabentyp identifizieren:
   - Einfache lineare Regression
   - Mehrfachregression
   - Nichtlineare Regression mit Transformation
   - Polynomiale Regression
2. Vorgehen wählen:
   - Linear: Direkte Berechnung mit Formeln
   - Nichtlinear: Transformation und lineare Regression
   - Polynomial: Erweiterte Designmatrix
   - Mehrfach: Matrix-Methode
3. Berechnungen durchführen:
   - Parameter bestimmen
   - Bestimmtheitsmass berechnen
   - Residuen analysieren
4. Ergebnisse interpretieren:
   - Modellgüte bewerten
   - Residuen beurteilen
   - Prognosen erstellen

**Klausuraufgabe - Linearisierung** Gegeben sind Messwerte für ein exponentielles Wachstum:

| $t$ (h) | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $N$ | 100 | 150 | 225 | 340 |

Finden Sie eine Funktion der Form $N(t) = N_0 e^{kt}$

1. Transformation:

$$\ln(N) = \ln(N_0) + kt$$

2. Neue Wertetabelle:

| $t$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $\ln(N)$ | 4.61 | 5.01 | 5.42 | 5.83 |

3. Lineare Regression:

$$\ln(N) = 0.405t + 4.61$$

4. Rücktransformation:

$$N(t) = 100.4e^{0.405t}$$

5. Bestimmtheitsmass: $R^2 = 0.999$

## Parameter- und Intervallschätzung

### Grundlagen der Schätztheorie

Die Schätztheorie befasst sich mit zwei Hauptproblemen:

- Punktschätzung: Bestimmung eines einzelnen Schätzwerts
- Intervallschätzung: Bestimmung eines Vertrauensbereichs

Wichtige Begriffe:

- $\theta$: Unbekannter Parameter der Grundgesamtheit
- $\Theta$: Schätzfunktion (Zufallsvariable)
- $\hat{\theta}$: Schätzwert (konkreter Wert)
- $n$: Stichprobenumfang

### Erwartungstreue Schätzfunktion

Eine Schätzfunktion $\Theta$ eines Parameters $\theta$ heisst erwartungstreu, wenn:

$$E(\Theta) = \theta$$

$E(\Theta)$: Erwartungswert der Schätzfunktion
$\theta$: Wahrer Parameter der Grundgesamtheit

### Effizienz einer Schätzfunktion

Gegeben sind zwei erwartungstreue Schätzfunktionen $\Theta_1$ und $\Theta_2$ desselben Parameters $\theta$. Man nennt $\Theta_1$ effizienter als $\Theta_2$, falls:

$$V(\Theta_1) < V(\Theta_2)$$

$V(\Theta_1), V(\Theta_2)$: Varianzen der Schätzfunktionen

### Konsistenz einer Schätzfunktion

Eine Schätzfunktion $\Theta$ heisst konsistent, wenn:

$$E(\Theta) \to \theta \text{ und } V(\Theta) \to 0 \text{ für } n \to \infty$$

$n$: Stichprobenumfang

### Prüfen von Schätzfunktionen

1. Erwartungstreue:
   - Erwartungswert $E(\Theta)$ berechnen
   - Mit Parameter $\theta$ vergleichen
   - Erwartungstreu, wenn $E(\Theta) = \theta$
2. Effizienz:
   - Varianzen $V(\Theta_1)$ und $V(\Theta_2)$ berechnen
   - Varianzen vergleichen
   - Kleinere Varianz = effizienter
3. Konsistenz:
   - Grenzwert für $n \to \infty$ betrachten
   - $E(\Theta) \to \theta$?
   - $V(\Theta) \to 0$?

**Prüfung auf Erwartungstreue**
Gegeben sei die Schätzfunktion $\Theta_1 = \frac{1}{3}(2X_1 + X_2)$ für den Erwartungswert $\mu$.
1. Erwartungswert berechnen:

$$E(\Theta_1) = E(\frac{1}{3}(2X_1 + X_2)) = \frac{1}{3}(2E(X_1) + E(X_2))$$

2. Einsetzen der Erwartungswerte:

$$E(\Theta_1) = \frac{1}{3}(2\mu + \mu) = \frac{3\mu}{3} = \mu$$

3. Da $E(\Theta_1) = \mu$, ist die Schätzfunktion erwartungstreu.
Effizienzberechnung:

$$\begin{aligned} V(\Theta_1) &= V(\frac{1}{3}(2X_1 + X_2)) \\ &= \frac{1}{9}V(2X_1 + X_2) \\ &= \frac{1}{9}(4V(X_1) + V(X_2)) \\ &= \frac{1}{9}(4\sigma^2 + \sigma^2) \\ &= \frac{5\sigma^2}{9} \end{aligned}$$

**Erwartungstreue Schätzfunktion**
Grundgesamtheit mit Erwartungswert $\mu$, Varianz $\sigma^2$ und Zufallsstichprobe $X_1, X_2, X_3$. Die folgende Schätzfunktion ist gegeben:

$$\Theta_1 = \frac{1}{3} \cdot (2X_1 + X_2)$$

$\Theta_1$ = Schätzfunktion
$X_1, X_2$ = Zufallsvariablen aus der Stichprobe

Ist diese Schätzfunktion erwartungstreu (Parameter: $\mu$)?

$$E(\Theta_1) = E(\frac{1}{3} \cdot (2X_1 + X_2)) = \frac{1}{3} \cdot (2E(X_1) + E(X_2))$$

$$E(\Theta_1) = \frac{1}{3} \cdot (2\mu + \mu) = \frac{3\mu}{3} = \mu$$

$E(\Theta_1)$ = Erwartungswert der Schätzfunktion
$E(X_1), E(X_2)$ = Erwartungswerte der einzelnen Zufallsvariablen
$\mu$ = Wahrer Parameterwert

Da $E(\Theta_1) = \mu$ ist die Funktion erwartungstreu.

### Effizienz einer Schätzfunktion

Grundgesamtheit mit Erwartungswert $\mu$, Varianz $\sigma^2$ und Zufallsstichprobe $X_1, X_2, X_3$. Gegeben ist die Schätzfunktion:

$$\Theta_1 = \frac{1}{3} \cdot (2X_1 + X_2)$$

**Berechnung der Effizienz:**

$$\begin{aligned} V(\Theta_1) &= V(\frac{1}{3} \cdot (2X_1 + X_2)) \\ &= \frac{1}{9} \cdot V(2X_1 + X_2) \\ &= \frac{1}{9} \cdot (V(2X_1) + V(X_2)) \\ &= \frac{1}{9} \cdot (4V(X_1) + V(X_2)) \\ &= \frac{1}{9} \cdot (4\sigma^2 + \sigma^2) \\ &= \frac{5\sigma^2}{9} \end{aligned}$$

$V(\Theta_1)$ = Varianz der Schätzfunktion
$V(X_1), V(X_2)$ = Varianzen der einzelnen Zufallsvariablen
$\sigma^2$ = Varianz der Grundgesamtheit

Die Effizienz der Schätzfunktion ist also $\frac{5\sigma^2}{9}$.

## Maximum-Likelihood-Schätzung

### Erwartungswert und Varianz (Funktion und Wert)

**Erwartungswert:**

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i, \quad \hat{\mu} = \bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

$\bar{X}$ = Arithmetischer Mittelwert (Zufallsvariable)
$\hat{\mu} = \bar{x}$ = Arithmetischer Mittelwert (Stichprobenwert)
$n$ = Stichprobenumfang
$X_i$ = $i$-te Zufallsvariable
$x_i$ = $i$-ter Stichprobenwert

**Varianz:**

$$S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \bar{X})^2, \quad \hat{\sigma}^2 = s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$S^2$ = Stichprobenvarianz (Zufallsvariable)
$\hat{\sigma}^2 = s^2$ = Stichprobenvarianz (Stichprobenwert)
$\bar{X}$ = Arithmetischer Mittelwert (Zufallsvariable)
$\bar{x}$ = Arithmetischer Mittelwert (Stichprobenwert)

### Likelyhood-Funktion

Wir betrachten eine Zufallsvariable $X$ und ihre Dichte (PDF) $f_x(x|\theta)$, welche von $x$ und einem oder mehreren Parametern $\theta$ abhängig sind.
Für eine Stichprobe vom Umfang $n$ mit $x_1, \ldots, x_n$ nennen wir die vom Parameter $\theta$ abhängige Funktion die Likelyhood-Funktion der Stichprobe:

$$L(\theta) = f_x(x_1|\theta) \cdot f_x(x_2|\theta) \cdot \ldots \cdot f_x(x_n|\theta)$$

### Vorgehen bei Maximum-Likelihood-Schätzung

1. Likelyhood-Funktion bestimmen
2. Maximalstelle der Funktion bestimmen:
   - (Partielle) Ableitung $L'(\theta) = 0$

### Maximum-Likelihood-Schätzung

Die Likelihood-Funktion für eine Stichprobe $x_1, \ldots, x_n$ ist:

$$L(\theta) = \prod_{i=1}^{n} f_X(x_i|\theta)$$

$f_X(x|\theta)$: Wahrscheinlichkeitsdichte
$\theta$: zu schätzender Parameter
$n$: Stichprobenumfang
Der Maximum-Likelihood-Schätzer maximiert $L(\theta)$ bzw. $\ln(L(\theta))$.

### Maximum-Likelihood-Schätzung

1. Likelihood-Funktion aufstellen:

$$L(\theta) = \prod_{i=1}^{n} f_X(x_i|\theta)$$

2. Logarithmieren:

$$\ln(L(\theta)) = \sum_{i=1}^{n} \ln(f_X(x_i|\theta))$$

3. Ableitung Null setzen:

$$\frac{d}{d\theta} \ln(L(\theta)) = 0$$

4. Nach $\theta$ auflösen:

$$\hat{\theta}_{ML} = \text{Lösung}$$

5. Maximum prüfen:

$$\frac{d^2}{d\theta^2} \ln(L(\theta)) < 0$$

### Maximum-Likelihood Normalverteilung

Gegeben sei eine Stichprobe aus einer Normalverteilung $N(\mu, \sigma^2)$.
1. Likelihood-Funktion:

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

2. Log-Likelihood:

$$\ln(L) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

3. Ableitungen:

$$\frac{\partial \ln(L)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0$$

$$\frac{\partial \ln(L)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} (x_i - \mu)^2 = 0$$

4. ML-Schätzer:

$$\hat{\mu}_{ML} = \bar{x}$$

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

## Vertrauensintervalle

### Vertrauensintervall

Ein Vertrauensintervall zum Niveau $\gamma$ ist ein zufälliges Intervall $[\Theta_u, \Theta_o]$ mit:

$$P(\Theta_u \leq \theta \leq \Theta_o) = \gamma$$

$\gamma$: Vertrauensniveau (stat. Sicherheit)
$\alpha = 1 - \gamma$: Irrtumswahrscheinlichkeit
$\Theta_u, \Theta_o$: Unter- und Obergrenze

### Intervallschätzung

Verteilungstypen und zugehörige Quantile:

| Verteilung | Parameter | Quantile |
|---|---|---|
| Normalverteilung ($\sigma^2$ bekannt) | $\mu$ | $c = u_p, p = \frac{1+\gamma}{2}$ |
| t-Verteilung ($\sigma^2$ unbekannt) | $\mu$ | $c = t_{(p;f=n-1)}, p = \frac{1+\gamma}{2}$ |
| Chi-Quadrat-Verteilung | $\sigma^2$ | $c_1 = \chi^2_{(\frac{1-\gamma}{2};n-1)}, c_2 = \chi^2_{(\frac{1+\gamma}{2};n-1)}$ |

Berechnung eines Vertrauensintervalls
Geben Sie das Vertrauensintervall für $\mu$ an ($\sigma^2$ unbekannt). Gegeben sind:

$$n = 10, \quad \bar{x} = 102, \quad s^2 = 16, \quad \gamma = 0.99$$

1. Verteilungstyp mit Param $\mu$ und $\sigma^2$ unbekannt $\rightarrow$ T-Verteilung
2. $f = n - 1 = 9$, $p = \frac{1+\gamma}{2} = 0.995$, $c = t_{(p;f)} = t_{(0.995;9)} = 3.25$
3. $e = c \cdot \frac{S}{\sqrt{n}} = 4.111$, $\Theta_u = \bar{X} - e = 97.89$, $\Theta_o = \bar{X} + e = 106.11$

### Vertrauensintervalle berechnen

1. Verteilungstyp bestimmen:
- Normalverteilung ($\sigma^2$ bekannt)
- t-Verteilung ($\sigma^2$ unbekannt)
- Chi-Quadrat (für Varianz)

2. Quantile bestimmen:
- Normalverteilung: $c = u_p$ mit $p = \frac{1+\gamma}{2}$
- t-Verteilung: $c = t_{(p;f)}$ mit $f = n - 1$
- Chi-Quadrat: $c_1 = \chi^2_{(p_1;f)}$, $c_2 = \chi^2_{(p_2;f)}$

3. Intervallgrenzen berechnen:
- Mittelwert: $[\bar{x} \pm c \cdot \frac{s}{\sqrt{n}}]$
- Varianz: $[\frac{(n-1)s^2}{c_2}, \frac{(n-1)s^2}{c_1}]$

#### Vertrauensintervall für Mittelwert

Gegeben: $n = 25$, $\bar{x} = 102$, $s = 4$, $\gamma = 0.95$

1. Verteilung: t-Verteilung mit $f = 24$ ($\sigma^2$ unbekannt)
2. Quantil:
- $p = \frac{1+0.95}{2} = 0.975$
- $c = t_{(0.975;24)} = 2.064$

3. Intervallgrenzen:
- $e = 2.064 \cdot \frac{4}{\sqrt{25}} = 1.652$
- $[102 - 1.652; 102 + 1.652] = [100.348; 103.652]$

### Minimum Stichprobenumfang bestimmen

1. Voraussetzungen:
- Gewünschte Genauigkeit $d$
- Vertrauensniveau $\gamma$
- Standardabweichung $\sigma$ (wenn bekannt)

2. Kritischen Wert bestimmen:
- $c = u_p$ oder $t_{(p;f)}$
- $p = \frac{1+\gamma}{2}$

3. Stichprobenumfang berechnen:
- $n \geq (\frac{2c\sigma}{d})^2$ für Mittelwert
- Auf nächste ganze Zahl aufrunden

#### Stichprobenumfang

Ein Parameter soll mit einer Genauigkeit von $d = 0.2$ bei $\gamma = 0.99$ geschätzt werden. Die Standardabweichung ist $\sigma = 0.5$ bekannt.

1. Kritischer Wert:
- $p = \frac{1+0.99}{2} = 0.995$
- $c = u_{0.995} = 2.576$

2. Mindestumfang:
- $n \geq (\frac{2 \cdot 2.576 \cdot 0.5}{0.2})^2 = 41.47$
- $n = 42$ (aufgerundet)

### Übersicht Vertrauensintervalle

| Parameter | Verteilung | Test-Statistik | Intervall |
|---|---|---|---|
| $\mu$ ($\sigma^2$ bekannt) | Normal | $U = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ | $[\bar{x} \pm u_p \frac{\sigma}{\sqrt{n}}]$ |
| $\mu$ ($\sigma^2$ unbek.) | t | $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ | $[\bar{x} \pm t_p \frac{s}{\sqrt{n}}]$ |
| $\sigma^2$ | $\chi^2$ | $Z = \frac{(n-1)S^2}{\sigma^2}$ | $[\frac{(n-1)s^2}{c_2}, \frac{(n-1)s^2}{c_1}]$ |

## Detaillierte Vertrauensintervalle

### Verteilungstypen und Quantile

| Verteilung | Parameter | Standardisierung | Quantil |
|---|---|---|---|
| Normalverteilung ($\sigma^2$ bekannt) | $\mu$ | $U = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ | $c = u_p, p =$ |
| t-Verteilung ($\sigma^2$ unbekannt) | $\mu$ | $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ | $c = t_{(p;f=n-1)}$, |
| Chi-Quadrat | $\sigma^2$ | $Z = (n-1)\frac{s^2}{\sigma^2}$ | $c_1 = \chi^2_{(\frac{1-\gamma}{2}}$ $c_2 = \chi^2_{(\frac{1+\gamma}{2}}$ |

### Konfidenzintervalle

Für verschiedene Verteilungen ergeben sich folgende Intervallgrenzen:

1. Normalverteilung ($\sigma^2$ bekannt):

$$\Theta_u = \bar{X} - c\frac{\sigma}{\sqrt{n}}, \quad \Theta_o = \bar{X} + c\frac{\sigma}{\sqrt{n}}$$

2. t-Verteilung ($\sigma^2$ unbekannt):

$$\Theta_u = \bar{X} - c\frac{S}{\sqrt{n}}, \quad \Theta_o = \bar{X} + c\frac{S}{\sqrt{n}}$$

3. Chi-Quadrat-Verteilung:

$$\Theta_u = \frac{(n-1)s^2}{c_2}, \quad \Theta_o = \frac{(n-1)s^2}{c_1}$$

## Typische Prüfungsaufgaben

### Typische Prüfungsaufgaben

1. Maximum-Likelihood:
- Likelihood-Funktion aufstellen
- Logarithmieren und ableiten
- ML-Schätzer bestimmen

2. Vertrauensintervalle:
- Verteilungstyp bestimmen
- Quantile nachschlagen
- Intervall berechnen

3. Schätzer prüfen:
- Erwartungstreue nachweisen
- Effizienz vergleichen
- Konsistenz zeigen

4. Stichprobenumfang:
- Genauigkeit berücksichtigen
- Vertrauensniveau einbeziehen
- Mindestumfang bestimmen

## Hypothesentests

### Hypothesentest

Ein statistischer Test zur Überprüfung einer Behauptung bzw. Hypothese über einen oder mehrere Parameter einer Grundgesamtheit.
- $H_0$: Nullhypothese (zu überprüfende Behauptung)
- $H_A$: Alternativhypothese (Gegenhypothese)
- $\alpha$: Signifikanzniveau (Irrtumswahrscheinlichkeit)

### Ablauf eines Hypothesentests

1. Hypothesen formulieren:
- $H_0$ aufstellen (punktförmig)
- $H_A$ aufstellen (ein- oder zweiseitig)

2. Signifikanzniveau $\alpha$ festlegen:
- Meist $\alpha = 0.05$ oder $\alpha = 0.01$

3. Testvariable und Verteilung bestimmen:
- Passende Zeile aus Tabelle 8.2.1 wählen
- Standardisierte Testvariable notieren

4. Kritische Werte bestimmen:
- Einseitig: ein Wert $c$
- Zweiseitig: zwei Werte $c_u$ und $c_o$

5. Testwert berechnen:
- Stichprobenwerte einsetzen
- Standardisierung durchführen

6. Entscheidung treffen:
- $H_0$ annehmen oder ablehnen
- Ergebnis interpretieren

#### Mittelwerttest (bekannte Varianz)

Ein Automobilhersteller gibt den mittleren Verbrauch mit $\mu_0 = 8.2$ l/100km an. In einer Stichprobe von $n = 25$ Fahrzeugen wurde ein Mittelwert von $\bar{x} = 9.1$ l/100km bei bekannter Standardabweichung $\sigma = 2.1$ l/100km gemessen.

1. Hypothesen:
- $H_0 : \mu = 8.2$
- $H_A : \mu \neq 8.2$ (zweiseitig)

2. Signifikanzniveau: $\alpha = 0.05$
3. Testvariable: $U = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ (standardnormalverteilt)
4. Kritische Werte:
- $c_u = -1.96$
- $c_o = 1.96$

5. Testwert:

$$\hat{u} = \frac{9.1 - 8.2}{2.1/\sqrt{25}} = 2.14$$

6. Entscheidung: $\hat{u} > c_o$, also $H_0$ ablehnen

### Unterscheidung abhängiger/unabhängiger Stichproben

1. Abhängige Stichproben:
- Gleicher Stichprobenumfang
- Messungen am gleichen Objekt
- Paarweise Zuordnung möglich

2. Unabhängige Stichproben:
- Unterschiedliche Objekte
- Keine Zuordnung möglich
- Stichprobenumfänge können verschieden sein

3. Auswirkung auf Test:
- Abhängig: Test der Differenzen
- Unabhängig: Test der einzelnen Stichproben

### Abhängige Stichproben (t-Test)

Vergleich zweier Messgeräte an denselben 5 Widerständen:

| i | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Gerät 1 | 100.5 | 102.4 | 104.3 | 101.5 | 98.4 |
| Gerät 2 | 98.2 | 99.1 | 102.4 | 101.1 | 96.2 |

1. Hypothesen:
- $H_0 : \mu_d = 0$
- $H_A : \mu_d \neq 0$

2. $\alpha = 0.01$
3. Differenzen bilden:
- $\bar{d} = 2.02$
- $s_d = 1.047$

4. Testvariable: $T = \frac{\bar{D}}{S_d/\sqrt{n}}$ (t-verteilt mit $f = 4$)
5. Testwert:

$$\hat{t} = \frac{2.02}{1.047/\sqrt{5}} = 4.313$$

6. Kritische Werte: $c_u = -4.604$, $c_o = 4.604$
7. Entscheidung: $|\hat{t}| < c_o$, also $H_0$ annehmen

### Fehlerarten bei Hypothesentests

| | $H_0$ annehmen | $H_0$ ablehnen |
|---|---|---|
| $H_0$ wahr | Richtige Entscheidung | Fehler 1. Art ($\alpha$) |
| $H_0$ falsch | Fehler 2. Art ($\beta$) | Richtige Entscheidung |

- Fehler 1. Art: $\alpha$ (Signifikanzniveau)
- Fehler 2. Art: $\beta$ (abhängig vom wahren Wert)
- Teststärke: $1 - \beta$ (Wahrscheinlichkeit für richtige Ablehnung)

### p-Wert berechnen

1. Testvariable und Verteilung bestimmen:
- Aus Tabelle 8.2.1 auswählen
- Testwert berechnen

2. p-Wert ermitteln:
- Einseitig: $P(T \geq |\hat{t}|)$ oder $P(T \leq |\hat{t}|)$
- Zweiseitig: $2 \cdot P(T \geq |\hat{t}|)$

3. Vergleich mit $\alpha$:
- $p \leq \alpha$: $H_0$ ablehnen
- $p > \alpha$: $H_0$ annehmen

### p-Wert berechnen

Bei einer Qualitätskontrolle wurden in einer Stichprobe von $n = 400$ Teilen 20 Defekte gefunden. Die Nullhypothese lautet $H_0 : p = 0.03$ gegen $H_A : p > 0.03$.

1. Testvariable:

$$U = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

2. Testwert:

$$\hat{u} = \frac{0.05 - 0.03}{\sqrt{\frac{0.03 \cdot 0.97}{400}}} = 2.345$$

3. p-Wert (einseitig):

$$p = P(U \geq 2.345) = 1 - \Phi(2.345) = 0.0095$$

4. Entscheidung: $p < \alpha = 0.05$, also $H_0$ ablehnen

### Typische Prüfungsaufgaben

1. Parametertests:
- Mittelwerttest ($\sigma^2$ bekannt/unbekannt)
- Varianztest
- Anteilstest

2. Vergleich zweier Stichproben:
- Abhängige Stichproben
- Unabhängige Stichproben
- Gleiche/verschiedene Varianzen

3. p-Wert Berechnung:
- Ein-/zweiseitige Tests
- Verschiedene Verteilungen

### Verteilungen der Testvariablen

- Normalverteilung ($\sigma^2$ bekannt):

$$U = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

- t-Verteilung ($\sigma^2$ unbekannt):

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

- Chi-Quadrat (Varianztest):

$$Z = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}$$

- Anteilstest (für großes n):

$$U = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1)$$

### Varianztest

Die Varianz eines Produktionsprozesses soll $\sigma_0^2 = 25$ nicht überschreiten. Eine Stichprobe von $n = 12$ Teilen ergab eine empirische Varianz von $s^2 = 40$.

1. Hypothesen:
- $H_0 : \sigma^2 = 25$
- $H_A : \sigma^2 > 25$

2. $\alpha = 0.05$
3. Testvariable: $Z = \frac{(n-1)S^2}{\sigma_0^2}$ ($\chi^2$-verteilt mit $f = 11$)
4. Testwert:

$$\hat{z} = \frac{11 \cdot 40}{25} = 17.6$$

5. Kritischer Wert: $c = \chi^2_{(0.95;11)} = 19.675$
6. Entscheidung: $\hat{z} < c$, also $H_0$ annehmen

### Einseitige vs. zweiseitige Tests

1. Zweiseitiger Test ($H_A : \theta \neq \theta_0$):
- Kritische Werte: $c_u$ und $c_o$
- Ablehnbereich: $(-\infty, c_u) \cup (c_o, \infty)$
- p-Wert: $2 \cdot P(T \geq |\hat{t}|)$

2. Rechtsseitiger Test ($H_A : \theta > \theta_0$):
- Kritischer Wert: $c_o$
- Ablehnbereich: $(c_o, \infty)$
- p-Wert: $P(T \geq \hat{t})$

3. Linksseitiger Test ($H_A : \theta < \theta_0$):
- Kritischer Wert: $c_u$
- Ablehnbereich: $(-\infty, c_u)$
- p-Wert: $P(T \leq \hat{t})$

### Maximum-Likelihood-Schätzung

1. Likelyhood-Funktion aufstellen:
- Dichte der Verteilung identifizieren
- Produkt der Einzelwahrscheinlichkeiten bilden
- $L(\theta) = \prod_{i=1}^{n} f_X(x_i|\theta)$

2. Logarithmierte Likelyhood-Funktion bilden:
- $\ln(L(\theta))$ berechnen
- Produkte werden zu Summen

3. Ableitung Null setzen:
- $\frac{d}{d\theta} \ln(L(\theta)) = 0$
- Nach $\theta$ auflösen

4. Maximum prüfen:
- Zweite Ableitung muss negativ sein
- $\frac{d^2}{d\theta^2} \ln(L(\theta)) < 0$

### Maximum-Likelihood Exponentialverteilung

Gegeben sei eine Stichprobe $x_1, ..., x_n$ aus einer Exponentialverteilung mit Parameter $\lambda$:

$$f(x|\lambda) = \lambda e^{-\lambda x}, \quad x \geq 0$$

1. Likelihood-Funktion:

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum x_i}$$

2. Log-Likelihood:

$$\ln(L(\lambda)) = n \ln(\lambda) - \lambda \sum x_i$$

3. Ableitung Null setzen:

$$\frac{d}{d\lambda} \ln(L(\lambda)) = \frac{n}{\lambda} - \sum x_i = 0$$

4. ML-Schätzer:

$$\hat{\lambda} = \frac{n}{\sum x_i} = \frac{1}{\bar{x}}$$

### Vertrauensintervalle berechnen

1. Verteilungstyp und Parameter bestimmen:
- Normalverteilung mit $\sigma^2$ bekannt
- Normalverteilung mit $\sigma^2$ unbekannt
- Chi-Quadrat für Varianz

2. Vertrauensniveau $\gamma$ und Freiheitsgrade:
- $\gamma$ aus Aufgabe entnehmen
- $f = n - 1$ bei t- und $\chi^2$-Verteilung

3. Kritische Werte bestimmen:
- Normalverteilung: $c = u_p$ mit $p = \frac{1+\gamma}{2}$
- t-Verteilung: $c = t_{(p;f)}$ mit $p = \frac{1+\gamma}{2}$
- Chi-Quadrat: $c_1 = \chi^2_{(p_1;f)}$ und $c_2 = \chi^2_{(p_2;f)}$

4. Intervallgrenzen berechnen:
- Mittelwert: $[\bar{x} - e; \bar{x} + e]$ mit $e = c \cdot \frac{s}{\sqrt{n}}$
- Varianz: $[\frac{(n-1)s^2}{c_2}; \frac{(n-1)s^2}{c_1}]$

### Vertrauensintervall für Mittelwert

Eine Maschine produziert Schrauben. Bei einer Stichprobe von $n = 16$ Schrauben wurde der Durchmesser gemessen:
- Mittelwert: $\bar{x} = 5.2$ mm
- Standardabweichung: $s = 0.15$ mm
- Vertrauensniveau: $\gamma = 95\%$

1. Verteilungstyp:
- $\sigma^2$ unbekannt $\rightarrow$ t-Verteilung
- $f = n - 1 = 15$ Freiheitsgrade

2. Kritischer Wert:
- $p = \frac{1+0.95}{2} = 0.975$
- $c = t_{(0.975;15)} = 2.131$

3. Intervallgrenzen:
- $e = 2.131 \cdot \frac{0.15}{\sqrt{16}} = 0.080$
- $[\bar{x} - e; \bar{x} + e] = [5.12; 5.28]$

### Vertrauensintervall für Varianz

Für die obigen Schrauben soll ein 95%-Vertrauensintervall für die Varianz berechnet werden.

1. Verteilungstyp:
- Chi-Quadrat-Verteilung
- $f = n - 1 = 15$ Freiheitsgrade

2. Kritische Werte:
- $p_1 = \frac{1-0.95}{2} = 0.025$
- $p_2 = \frac{1+0.95}{2} = 0.975$
- $c_1 = \chi^2_{(0.025;15)} = 6.262$
- $c_2 = \chi^2_{(0.975;15)} = 27.488$

3. Intervallgrenzen:
- $s^2 = 0.15^2 = 0.0225$
- $\theta_u = \frac{15 \cdot 0.0225}{27.488} = 0.0123$
- $\theta_o = \frac{15 \cdot 0.0225}{6.262} = 0.0539$

4. Vertrauensintervall für $\sigma^2$: $[0.0123; 0.0539]$

### Bestimmung des Stichprobenumfangs

1. Gegebene Verteilung und Parameter:
- Normalverteilung mit $\sigma^2$ bekannt
- Vertrauensniveau $\gamma$
- Maximal zulässige Intervallbreite $d$

2. Kritischen Wert bestimmen:
- $p = \frac{1+\gamma}{2}$
- $c = u_p$ für Normalverteilung

3. Stichprobenumfang berechnen:
- $n \geq (\frac{2c\sigma}{d})^2$
- Auf nächste ganze Zahl aufrunden

4. Bei unbekannter Varianz:
- Vorerhebung durchführen
- Varianz schätzen
- t-Verteilung statt Normalverteilung

### Stichprobenumfang bestimmen

Ein Prozess produziert Teile mit bekannter Standardabweichung $\sigma = 0.5$ mm. Der Mittelwert soll mit einer Genauigkeit von $\pm 0.2$ mm bei einem Vertrauensniveau von 99% geschätzt werden.

1. Gesucht:
- Intervallbreite $d = 0.4$ mm
- $\gamma = 0.99$

2. Kritischer Wert:
- $p = \frac{1+0.99}{2} = 0.995$
- $c = u_{0.995} = 2.576$

3. Stichprobenumfang:
- $n \geq (\frac{2 \cdot 2.576 \cdot 0.5}{0.4})^2 = 41.47$
- $n = 42$ (aufgerundet)

### Übersicht statistische Schätzverfahren

1. Punktschätzung:
- Maximum-Likelihood
- Momentenmethode
- Kleinste-Quadrate

2. Intervallschätzung:
- Vertrauensintervalle für Mittelwert
- Vertrauensintervalle für Varianz
- Vertrauensintervalle für Anteilswerte

3. Gütekriterien:
- Erwartungstreue
- Effizienz
- Konsistenz
- Minimale Varianz

### Typische Prüfungsaufgaben

1. Maximum-Likelihood-Schätzung:
- Likelihood-Funktion aufstellen
- Logarithmieren
- Ableitung Null setzen
- ML-Schätzer bestimmen
- Maximum prüfen

2. Vertrauensintervalle:
- Verteilungstyp bestimmen
- Kritische Werte nachschlagen
- Intervallgrenzen berechnen
- Interpretation der Ergebnisse

3. Stichprobenumfang:
- Genauigkeitsanforderungen
- Vertrauensniveau
- Minimal notwendigen Umfang bestimmen

## Beispiele

Erwartungstreue Schätzfunktion Grundgesamtheit mit Erwartungswert $\mu$, Varianz $\sigma^2$ und Zufallsstichprobe $X_1, X_2, X_3$. Die folgende Schätzfunktion ist gegeben:

$$\Theta_1 = \frac{1}{3} \cdot (2X_1 + X_2)$$

$\Theta_1$ = Schätzfunktion
$X_1, X_2$ = Zufallsvariablen aus der Stichprobe

Ist diese Schätzfunktion erwartungstreu (Parameter: $\mu$)?

$$E(\Theta_1) = E(\frac{1}{3} \cdot (2X_1 + X_2)) = \frac{1}{3} \cdot (2E(X_1) + E(X_2))$$

$$E(\Theta_1) = \frac{1}{3} \cdot (2\mu + \mu) = \frac{3\mu}{3} = \mu$$

$E(\Theta_1)$ = Erwartungswert der Schätzfunktion
$E(X_1), E(X_2)$ = Erwartungswerte der einzelnen Zufallsvariablen
$\mu$ = Wahrer Parameterwert

Da $E(\Theta_1) = \mu$ ist die Funktion erwartungstreu.

Intervallschätzung für die Varianz Für die Varianz $\sigma^2$ einer Normalverteilung mit Stichprobenumfang $n = 10$ und Stichprobenvarianz $s^2 = 16$ soll ein 99%-Vertrauensintervall berechnet werden.

1. Verteilungstyp: Chi-Quadrat-Verteilung
2. Freiheitsgrade: $f = n - 1 = 9$
3. Quantile: $c_1 = \chi^2_{(0.005;9)} = 1.735$, $c_2 = \chi^2_{(0.995;9)} = 23.589$
4. Vertrauensintervall:

$$\frac{(n-1)s^2}{c_2} \leq \sigma^2 \leq \frac{(n-1)s^2}{c_1}$$

$n$ = Stichprobenumfang
$s^2$ = Stichprobenvarianz
$c_1, c_2$ = Chi-Quadrat-Quantile
$\sigma^2$ = Wahre Varianz der Grundgesamtheit

$$\frac{9 \cdot 16}{23.589} \leq \sigma^2 \leq \frac{9 \cdot 16}{1.735}$$

$$6.10 \leq \sigma^2 \leq 82.99$$

Bernoulli-Anteilsschätzung Ein Vertrauensintervall für den Parameter $p$ einer Bernoulli-Verteilung soll aus einer Stichprobe mit $n = 100$ und $\bar{x} = 0.42$ bei einem Vertrauensniveau von 95% berechnet werden.

1. Prüfen der Voraussetzung: $n\hat{p}(1-\hat{p}) = 100 \cdot 0.42 \cdot 0.58 = 24.36 > 9$
2. Quantil: $c = u_{0.975} = 1.96$
3. Standardfehler: $\sqrt{\frac{\bar{x}(1-\bar{x})}{n}} = \sqrt{\frac{0.42 \cdot 0.58}{100}} = 0.0494$
4. Vertrauensintervall:

$$0.42 \pm 1.96 \cdot 0.0494 = [0.323; 0.517]$$

$n$ = Stichprobenumfang
$\bar{x}$ = Stichprobenmittelwert (Anteil der Erfolge)
$\hat{p}$ = Geschätzter Parameter der Bernoulli-Verteilung
$u_{0.975} = 97.5$