

# STS - Zusammenfassung

## Stochastik & Statistik · Zusammenfassung

### 1 Grundbegriffe

#### Häufig gebrauchte Variablen:

$x$ : Merkmal / Beobachtung  
 $a_i$ : Merkmalsausprägung (Wertklasse)  
 $h_i$ : absolute Häufigkeit von  $a_i$   
 $f_i$ : relative Häufigkeit von  $a_i$   
 $H_i$ : kumulierte absolute Häufigkeit  
 $F_i$ : kumulierte relative Häufigkeit  
 $\bar{x}$ : arithmetisches Mittel  
 $x_{\text{med}}$ : Median (2. Quantil)  
 $x_{\text{mod}}$ : Modalwert (häufigster Wert)  
 $s^2$ : Varianz,  $s_x$ : Standardabweichung  
 $s_{\text{korr}}$ : korrigierte Standardabweichung  
 $f(x)$ : PMF/PDF,  $F(x)$ : CMF/CDF

**PMF:**  $f(x)$  Relative Häufigkeit (Stabdiagramm)

**CMF:**  $F(x)$  Kumulative relative Häufigkeit (Treppendiagramm)

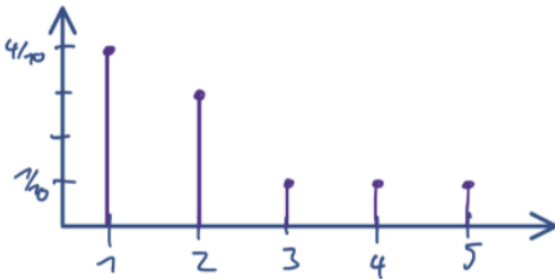
**PDF:**  $f(x)$  Höhe Balken Histogramm

**CDF:**  $F(x)$  Kumulative Fläche Balken Histogramm

### 2 Diagrammtypen

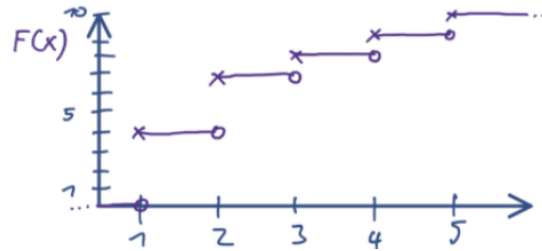
#### 2.1 PMF

$f(x)$  Relative Häufigkeit (Stabdiagramm)



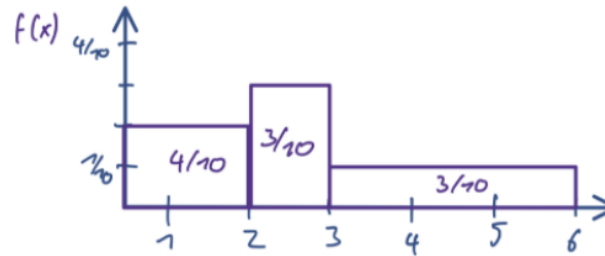
#### 2.2 CMF

$F(x)$  Kumulative relative Häufigkeit (Treppendiagramm)



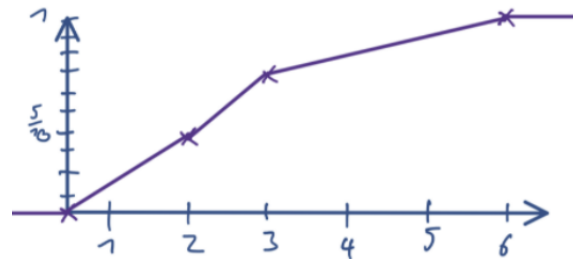
#### 2.3 PDF

$f(x)$  Höhe Balken Histogramm



#### 2.4 CDF

$F(x)$  Kumulative Fläche Balken Histogramm



### 3 Häufigkeiten und Verteilungen

**Neue Variablen in diesem Abschnitt:**

$n$ : Stichprobengröße (Anzahl Beobachtungen)

1. Urliste sortieren
2. Verschiedene Werte  $a_i$  bestimmen
3. Absolute Häufigkeit:  $h_i$  = Anzahl von  $a_i$
4. Relative Häufigkeit:  $f_i = \frac{h_i}{n}$
5. Kumulative Häufigkeit:
  - $H_i = \sum_{k \leq i} h_k$
  - $F_i = \sum_{k \leq i} f_k$

#### Würfel (20 Würfe):

$a_i$	1	2	3	4	5	6
$h_i$	4	3	4	0	6	3
$f_i$	0.2	0.15	0.2	0	0.3	0.15
$F_i$	0.2	0.35	0.55	0.55	0.85	1.0
$H_i$	4	7	11	11	17	20

PMF:  $\frac{h_i}{20}$ , PDF:  $\frac{h_i}{20}$ , CDF:  $F_i$ , CMF:  $H_i$

### 4 Median und Quantile

#### Neue Variablen:

$q$ : Quantilsniveau (z.B. 0.25, 0.5, 0.75)

$Q_q$ : zugehöriges  $q$ -Quantil

$p$ : Positionsindex in der sortierten Liste

#### q-Quantil $Q_q$ :

Position:  $p = n \cdot q$

Falls  $p$  ganzzahlig:

$$Q_q = \frac{1}{2}(x_{(p)} + x_{(p+1)})$$

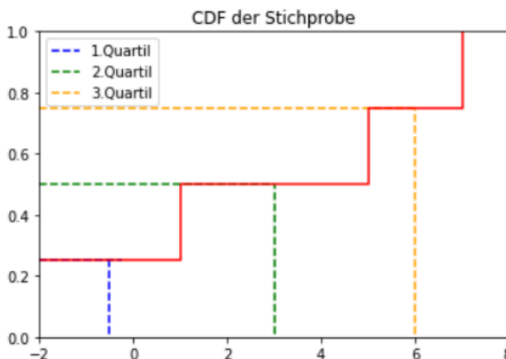
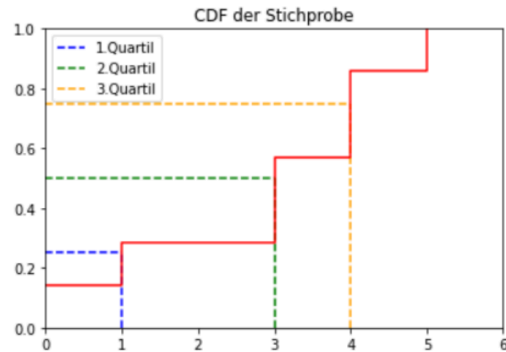
Falls  $p$  nicht ganzzahlig:

$$Q_q = x_{(\lceil p \rceil)}$$

**Spezialfälle:**

- Median:  $q = 0.5$
- 1. Quartil (Q1):  $q = 0.25$
- 3. Quartil (Q3):  $q = 0.75$

#### 4.1 Quartil aus CDF ablesen



1. Daten sortieren:  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$
2. Position berechnen:  $p = n \cdot q$
3. Falls  $p$  ganzzahlig  $\rightarrow Q_q = \frac{1}{2}(x_{(p)} + x_{(p+1)})$
4. Falls nicht  $\rightarrow k = \lceil p \rceil$ , dann  $Q_q = x_{(k)}$

**Stichprobe:** 4, 4, 0, 3, 5, 3, 1

**Sortiert:** 0, 1, 3, 3, 4, 4, 5 ( $n = 7$ )

**Q1:**  $p = 7 \cdot 0.25 = 1.75 \rightarrow k = 2 \rightarrow Q_1 = 1$

**Median:**  $p = 7 \cdot 0.5 = 3.5 \rightarrow k = 4 \rightarrow Q_2 = 3$

**Q3:**  $p = 7 \cdot 0.75 = 5.25 \rightarrow k = 6 \rightarrow Q_3 = 4$

## 5 Klassierte Daten

### 5.1 Grundbegriffe

3, 7, 12, 18, 19, 20, 25, 27, 28, 29, 31, 32, 34, 37, 38, 40, 41, 45, 47.

Gruppieren Sie die Stichprobenwerte in geeignete Klassen und bestimmen Sie die absoluten und die relativen Häufigkeiten der Klassen. Fertigen Sie Plots der *PDF* und *CDF* an.

Klasse $c_i$ Von ... bis weniger als ...	[0,10[	[10,20[	[20,30[	[30,40[	[40,50[	Total
Absolute Häufigkeit $h_i$ des Intervalls	2	3	6	5	4	20
Relative Häufigkeit $f_i$ des Intervalls	$\frac{2}{20}$	$\frac{3}{20}$	$\frac{6}{20}$	$\frac{5}{20}$	$\frac{4}{20}$	1
Säulenhöhe im Histogramm ( <i>PDF</i> ) $f(x)$	$\frac{2}{20 \cdot 10}$	$\frac{3}{20 \cdot 10}$	$\frac{6}{20 \cdot 10}$	$\frac{5}{20 \cdot 10}$	$\frac{4}{20 \cdot 10}$	
Werte der kumu- lativen Verteilungs- funktion ( <i>CDF</i> ) $F(x)$ an der rechten Intervallgrenze	$2/20=0.1$	$5/20=0.25$	$11/20=0.55$	$16/20=0.8$	1	

#### • Mittelwert:

$$\bar{x} = \left(\frac{1}{n}\right) \sum_{i=1}^k m_i \cdot h_i$$

$m_i$ : Klassenmitte der i-ten Klasse (Mittelwert der Klassengrenzen)

$h_i$ : absolute Häufigkeit der i-ten Klasse

#### • Varianz:

$$s^2 = \left(\frac{1}{n}\right) \sum_{i=1}^k (h_i \cdot m_i^2) - \bar{x}^2$$

Klasse $a_i$ Von ... bis weniger als ...	[100,200[	[200,500[	[500,800[	[800,1000[	[1000,2000[	Total
Absolute Häufigkeit $h_i$ des Intervalls	35	182	317	84	132	750
Relative Häufigkeit $f_i$ des Intervalls	$35/750$	$182/750$	$317/750$	$84/750$	$132/750$	1
Werte der kumu- lativen Verteilungs- funktion ( <i>CDF</i> ) $F(x)$ an der rechten Intervallgrenze	$35/750 \approx 0.047$	$217/750 \approx 0.29$	$534/750 \approx 0.712$	$618/750 \approx 0.824$	1	
Klassenmitten	150	350	650	900	1500	

Stichprobenmittelwert:

$$\bar{x} = \frac{1}{750} [150 \cdot 35 + 350 \cdot 182 + 650 \cdot 317 + 900 \cdot 84 + 1500 \cdot 132] = 10972/15 \approx 731.5 \text{ [CHF]}$$

Stichprobenvarianz:

$$s^2 = \frac{1}{750} [150^2 \cdot 35 + 350^2 \cdot 182 + 650^2 \cdot 317 + 900^2 \cdot 84 + 1500^2 \cdot 132] - \bar{x}^2 \approx 161030 \text{ [CHF}^2\text{]}$$

Standardabweichung:

$$s = \sqrt{s^2} \approx 401.29 \text{ [CHF]}$$

Korrigierte Stichprobenvarianz:

$$s^2 = \frac{750}{749} s^2 \approx 161245 \text{ [Franken}^2\text{]}$$

Median:

Gesucht ist der Wert  $R_{0.5} = F^{-1}(0.5)$  für die klassierten Daten. Dazu nehmen wir die Klasse 500-800 und lösen die Gleichung:

$$\frac{F(800) - F(500)}{800 - 500} = \frac{R_{0.5} - 500}{R_{0.5} - 500} \Rightarrow R_{0.5} = 500 + \frac{300 \cdot (0.5 - F(500))}{F(800) - F(500)} = 500 + \frac{300 \cdot (0.5 - 217/750)}{317/750} \approx 649.53$$

### 5.2 Quantile bei klassierten Daten

#### Neue Variablen:

$a_{k-1}, a_k$ : Klassenunter-/obergrenze

$F(a_k)$ : CDF an rechter Klassengrenze

Gegeben: Klasse  $[a_{k-1}, a_k]$  mit  $F(a_{k-1}) < q \leq F(a_k)$

#### Lineare Interpolation:

$$Q_q = a_{k-1} + \frac{a_k - a_{k-1}}{F(a_k) - F(a_{k-1})} \cdot (q - F(a_{k-1}))$$

1. Kumulierte Verteilung  $F(a_k)$  aus Tabelle ablesen
2. Klasse finden, in der  $F(a_{k-1}) < q \leq F(a_k)$
3. Formel anwenden (lineare Interpolation)

#### Mieten (0.7351-Quantil):

Klasse 1000-1500:  $F(1000) = 0.47, F(1500) = 0.775$

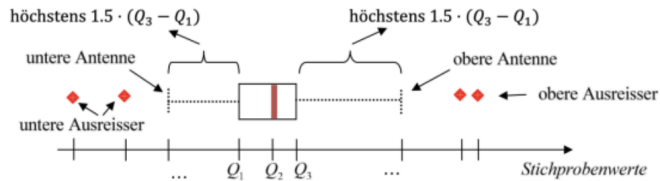
$$\begin{aligned} Q_{0.7351} &= 1000 + \frac{1500 - 1000}{0.775 - 0.47} \cdot (0.7351 - 0.47) \\ &= 1000 + \frac{500}{0.305} \cdot 0.2651 \\ &\approx 1434.78 \end{aligned}$$

## 6 Boxplot-Kennwerte

#### Neue Variablen:

IQR: Interquartilsabstand

Untere/obere Antenne: Bereich ohne Ausreißer



**Interquartilsabstand (IQR):**

$$\text{IQR} = Q_3 - Q_1$$

**Antennen (Whisker):**

- Untere: Minimum in  $[Q_1 - 1.5 \cdot \text{IQR}, Q_1]$
- Obere: Maximum in  $[Q_3, Q_3 + 1.5 \cdot \text{IQR}]$

**Ausreißer:** Werte außerhalb  $[Q_1 - 1.5 \cdot \text{IQR}, Q_3 + 1.5 \cdot \text{IQR}]$

**Daten:** 5, 7, 8, 9, 10, 12, 15, 20, 22

$$Q_1 = 8, \quad \text{Median} = 10, \quad Q_3 = 15$$

$$\text{IQR} = 15 - 8 = 7$$

**Whisker:**

- Untere:  $Q_1 - 1.5 \cdot \text{IQR} = 8 - 10.5 = -2.5 \rightarrow \text{Min}(\text{Daten}) = 5$
- Obere:  $Q_3 + 1.5 \cdot \text{IQR} = 15 + 10.5 = 25.5 \rightarrow \text{Max}(\text{Daten}) = 22$

**Ergebnis:** Alle Werte in  $[5, 22] \rightarrow$  keine Ausreißer

## 7 Mittelwert und Varianz

**Neue Variablen:**

$s$ : Standardabweichung

$s_{\text{korrt}}$ : korrigierte Standardabweichung

**Arithmetischer Mittelwert:**

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

**Empirische Varianz:**

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

**Verschiebungssatz:**

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

**Standardabweichung:**

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

**Korrigierte Varianz:**

$$s_{\text{korrt}}^2 = \frac{n}{n-1} \cdot s^2$$

1. Mittelwert  $\bar{x}$  berechnen
2. Varianz entweder direkt oder mit Verschiebungssatz
3. Standardabweichung:  $s = \sqrt{s^2}$
4. Für Schätzungen: korrigierte Varianz verwenden

**Gewinnspiel (100 Spiele):**

Gewinn	-1	0	4	8	10	20
$h_i$	74	13	3	5	4	1

$$\begin{aligned}\bar{x} &= \frac{1}{100}(-1 \cdot 74 + 0 \cdot 13 + 4 \cdot 3 + 8 \cdot 5 + 10 \cdot 4 + 20 \cdot 1) \\ &= -0.38 \text{ CHF}\end{aligned}$$

$$\begin{aligned}s^2 &= \frac{1}{100}(1 \cdot 74 + 0 + 16 \cdot 3 + 64 \cdot 5 + 100 \cdot 4 + 400 \cdot 1) - 0.38^2 \\ &= 12.28 \text{ CHF}^2\end{aligned}$$

$$s = \sqrt{12.28} \approx 3.5 \text{ CHF}$$

## 8 Lineare Regression (KQM)

**Neue Variablen:**

$a$ : Achsenabschnitt

$b$ : Steigung der Regressionsgeraden

$R^2$ : Bestimmtheitsmaß

**Regressionsgerade:**  $y = a + bx$

**Steigung:**

$$b = \frac{s_{xy}}{s_x^2} = r \cdot \frac{s_y}{s_x}$$

**Achsenabschnitt:**

$$a = \bar{y} - b \cdot \bar{x}$$

**Residualvarianz:**

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

**Bestimmtheitsmaß:**

$$R^2 = r^2 = \frac{s_y^2 - s_{\text{Res}}^2}{s_y^2}$$

1. Korrelationskoeffizient  $r$  berechnen
2. Steigung:  $b = s_{xy}/s_x^2$
3. Achsenabschnitt:  $a = \bar{y} - b\bar{x}$
4. Prüfung: Residuenplot (sollte unsystematisch streuen)
5. Güte:  $R^2$  berechnen (Anteil erklärter Varianz)

**Größe-Gewicht (15 Personen):**

$$\bar{x} = 173 \text{ cm}, s_x = 6.047 \text{ cm}$$

$$\bar{y} = 72.267 \text{ kg}, s_y = 7.474 \text{ kg}$$

$$s_{xy} = 41.071, r = 0.898$$

$$b = \frac{41.071}{6.047^2} \approx 1.123$$

$$a = 72.267 - 1.123 \cdot 173 \approx -122.02$$

**Regressionsgerade:**

$$y = -122.02 + 1.123x$$

$$R^2 = 0.898^2 \approx 0.806$$

$\rightarrow$  80.6% der Varianz erklärt

## 9 Bivariate Daten & Kovarianz

**Neue Variablen:**

$(x_i, y_i)$ : Wertepaare zweier Merkmale  
 $s_{xy}$ : empirische Kovarianz  
 $\overline{xy}$ : Mittelwert der Produkte

#### Empirische Kovarianz:

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

#### Verschiebungssatz:

$$s_{xy} = \overline{xy} - \bar{x} \cdot \bar{y}$$

wobei:

$$\overline{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i$$

#### Korrigierte Kovarianz:

$$s_{xy, \text{korr}} = \frac{n}{n-1} \cdot s_{xy}$$

#### Interpretation:

- $s_{xy} > 0$ : positiver Zusammenhang
- $s_{xy} \approx 0$ : kein linearer Zusammenhang
- $s_{xy} < 0$ : negativer Zusammenhang

1. Mittelwerte  $\bar{x}, \bar{y}$  berechnen
2. Produkte  $x_i y_i$  bilden
3. Verschiebungssatz anwenden:  $s_{xy} = \overline{xy} - \bar{x} \bar{y}$
4. Interpretation: Vorzeichen gibt Richtung des Zusammenhangs

**Stichprobe:**  $x = [1, 2, 3], y = [4, -1, 2]$

$$\bar{x} = 2, \quad \bar{y} = \frac{5}{3}$$

$$\overline{xy} = \frac{1}{3} (1 \cdot 4 + 2 \cdot (-1) + 3 \cdot 2) = \frac{8}{3}$$

$$s_{xy} = \frac{8}{3} - 2 \cdot \frac{5}{3} = -\frac{2}{3}$$

Negativer Wert  $\rightarrow$  negativer Zusammenhang

## 10 Spearman-Rangkorrelation

#### Neue Variablen:

$\text{rg}(x_i)$ : Rang von  $x_i$  in sortierter Liste  
 $r_{\text{Sp}}$ : Spearman-Korrelationskoeffizient  
 $d_i$ : Differenz der Ränge

#### Definition (allgemein):

$$r_{\text{Sp}} = \frac{\sum_{i=1}^n (\text{rg}(x_i) - \overline{\text{rg}(x)}) (\text{rg}(y_i) - \overline{\text{rg}(y)})}{\sqrt{\sum_{i=1}^n (\text{rg}(x_i) - \overline{\text{rg}(x)})^2} \cdot \sqrt{\sum_{i=1}^n (\text{rg}(y_i) - \overline{\text{rg}(y)})^2}}$$

#### Vereinfachte Formel (keine gleichen Ränge):

$$r_{\text{Sp}} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

wobei  $d_i = \text{rg}(x_i) - \text{rg}(y_i)$

#### Interpretation:

- Misst monotonen (nicht nur linearen) Zusammenhang
- Robust gegen Ausreißer
- Werte:  $-1 \leq r_{\text{Sp}} \leq 1$

1. Ränge für  $x$  und  $y$  separat vergeben (1 = kleinster Wert)
2. Bei gleichen Werten: Durchschnittsrang vergeben
3. Rangdifferenzen  $d_i$  berechnen
4. Formel anwenden (vereinfacht falls keine Bindungen)
5. Interpretation wie bei Pearson

#### Alter vs. Laufzeit (6 Personen):

$i$	$x_i$	$y_i$	$\text{rg}(x_i)$	$\text{rg}(y_i)$
1	59	14.6	6	6
2	35	11.8	3	2
3	43	14.3	5	5
4	23	13.0	1	3
5	42	14.2	4	4
6	27	11.0	2	1

$$\sum d_i^2 = 0 + 1 + 0 + 4 + 0 + 1 = 6$$

$$r_{\text{Sp}} = 1 - \frac{6 \cdot 6}{6(36 - 1)} = 1 - \frac{36}{210} \approx 0.83$$

Starker positiver monotoner Zusammenhang

## 11 Korrelation nach Bravais-Pearson

#### Neue Variablen:

$y$ : zweite Variable  
 $s_x, s_y$ : Standardabweichungen von  $x, y$   
 $s_{xy}$ : Kovarianz von  $x, y$   
 $r$ : Korrelationskoeffizient

#### Korrelationskoeffizient:

$$r = \frac{s_{xy}}{s_x \cdot s_y}$$

wobei:

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

#### Interpretation:

- $r \approx 1$ : starker positiver Zusammenhang
- $r \approx 0$ : kein linearer Zusammenhang
- $r \approx -1$ : starker negativer Zusammenhang

**Bestimmtheitsmaß:**  $r^2$  = Anteil erklärter Varianz

1. Mittelwerte  $\bar{x}, \bar{y}$  berechnen
2. Standardabweichungen  $s_x, s_y$  berechnen
3. Kovarianz  $s_{xy}$  berechnen
4. Korrelationskoeffizient:  $r = s_{xy} / (s_x s_y)$

#### Beispiel: Berechnung des Bravais-Pearson Korrelationskoeffizienten

Gegeben sind folgende Wertepaare  $(x_i, y_i)$ :

$i$	1	2	3	4	5	6	7	8
$x_i$	5	10	20	8	4	6	12	15
$y_i$	27	46	73	40	30	28	47	59

(1) Mittelwerte berechnen:  $\bar{x} = 10, \bar{y} = 43.75$

$x^2$	25	100	400	64	16	36	144	225
$y^2$	729	2116	5329	1600	900	784	2209	3481
$x_i y_i$	135	460	1460	320	120	168	564	885

$$\sum x_i^2 = 1010, \sum y_i^2 = 17148, \sum x_i y_i = 4112$$

(2) **Standardabweichungen berechnen:**

$$s_x = \sqrt{\frac{1}{n} \sum x_i^2 - (\bar{x})^2} = \sqrt{126.25 - 100} = \sqrt{26.25} = 5.12$$

$$s_y = \sqrt{\frac{1}{n} \sum y_i^2 - (\bar{y})^2} = \sqrt{2143.5 - 1914.06} = 15.15$$

(3) **Kovarianz berechnen:**

$$s_{xy} = \frac{1}{8} \sum x_i y_i - \bar{x} \cdot \bar{y} = 514 - 10 \cdot 43.75 = 76.5$$

(4) **Korrelationskoeffizient berechnen:**

$$r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{76.5}{5.12 \cdot 15.15} \approx 0.986$$

→ **Starker positiver linearer Zusammenhang** zwischen  $x$  und  $y$

## 11.1 Kombinatorik

**Neue Variablen:**

$n$ : Größe der Grundmenge

$k$ : Größe der Teilmenge

1. Art der Auswahl bestimmen (Variation/Kombination, mit/ohne Wiederholung)
2. Passende Formel aus Übersichtstabelle auswählen
3. Werte für  $n$ ,  $k$  einsetzen und berechnen

Art	Mit Wiederholung	Ohne Wiederholung
<b>Variation</b> (mit Reihenfolge)	$n^k$	$\frac{n!}{(n-k)!}$
<b>Kombination</b> (ohne Reihenfolge)	$\binom{n+k-1}{k}$	$\binom{n}{k}$

**Permutation:** (alle Elemente, keine Wiederholung)

$$N = n!$$

→ Es werden **alle  $n$  Elemente** in **verschiedener Reihenfolge** angeordnet.

**Permutation mit Wiederholungen**

$$N = \frac{n!}{n_1! n_2! \dots n_r!}$$

$n_i$ : Anzahl gleiche Elemente der  $i$ -ten Sorte (3 für T in TATTOO)

→ Es gibt **mehrere gleiche Elemente**, z. B. bei Wörtern wie „TATTOO“

**Beispiel: TATTOO**

Das Wort „TATTOO“ hat 6 Buchstaben:

- T kommt 3-mal vor
- A kommt 1-mal vor
- O kommt 2-mal vor

$$N = \frac{6!}{3! \times 1! \times 2!} = \frac{720}{6 \times 1 \times 2} = \frac{720}{12} = 60$$

Es gibt also **60 verschiedene Anordnungen** des Wortes „TATTOO“.

### 11.1.1 Wahrscheinlichkeit berechnen

Obige Kombinatorikformeln können zur Berechnung von Wahrscheinlichkeiten verwendet werden.

$$P = \frac{N_{\text{günstig}}}{N_{\text{möglich}}}$$

## 11.2 hypergeometrische Verteilung Anwendung

Formel siehe unten bei Kontingenztabellen.

## 12 Aufgabe

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

a) genau zwei bzw. b) mindestens ein Joker?

### 12.1 Lösung

a)

$$\frac{\binom{6}{2} * \binom{104}{10}}{\binom{110}{12}} \approx 11.13\%$$

b)

$$1 - \frac{\binom{104}{12}}{\binom{110}{12}} \approx 50.85\%$$

## 13 Kontingenztabellen

**Neue Variablen:**

$n_{ij}$ : absolute Häufigkeit in Zelle  $(i, j)$

$f_{ij}$ : relative Häufigkeit in Zelle  $(i, j)$

Randsummen: Summen über Zeilen/Spalten

**Kontingenztafel** für Merkmale  $A, B$ :

	$B_1$	$B_2$	Summe
$A_1$	$n_{11}$	$n_{12}$	$n_{1.}$
$A_2$	$n_{21}$	$n_{22}$	$n_{2.}$
Summe	$n_{.1}$	$n_{.2}$	$n$

**Randhäufigkeiten:**

- Zeilen:  $n_{i.} = \sum_j n_{ij}$
- Spalten:  $n_{.j} = \sum_i n_{ij}$

**Relative Häufigkeiten:**

$$f_{ij} = \frac{n_{ij}}{n}$$

**Bedingte Häufigkeit** ( $A_i$  gegeben  $B_j$ ):

$$f(A_i | B_j) = \frac{n_{ij}}{n_{.j}}$$

1. Tabelle mit Zeilen (Merkmal  $A$ ) und Spalten (Merkmal  $B$ ) aufstellen
2. Absolute Häufigkeiten  $n_{ij}$  eintragen
3. Randsummen berechnen
4. Relative/bedingte Häufigkeiten bei Bedarf berechnen
5. Visualisierung: Mosaikplot

**Zivilstand vs. Geschlecht (100 Personen):**

	Männer	Frauen	Summe
Ledig	25	20	45

Verheiratet	30	25	55
Summe	55	45	100

$$f(\text{Ledig} \mid \text{Mann}) = \frac{25}{55} \approx 0.45$$

$$f(\text{Frau} \mid \text{Verheiratet}) = \frac{25}{55} \approx 0.45$$

## 14 Scatterplot & Streudiagramm

evaluieren ob irgendwo gebraucht

**Verwendung:** Visualisierung zweier metrischer Merkmale

**Interpretation:**

- Form: linear, gekrümmt, mehrere Cluster
- Richtung: positiv (steigend), negativ (fallend)
- Stärke: eng um Kurve  $\rightarrow$  stark, weit gestreut  $\rightarrow$  schwach

**Warnung:** Korrelation  $\neq$  Kausalität (Scheinkorrelation möglich)

- Wertepaare  $(x_i, y_i)$  als Punkte in Koordinatensystem eintragen
- Visuelle Inspektion: Form, Richtung, Stärke erkennen
- Ausreißer identifizieren
- Korrelationskoeffizient berechnen (Pearson oder Spearman)
- Immer Scatterplot + Korrelationskoeffizient zusammen angeben!

## 15 Übersichtstabelle Verteilungen

Verteilung	PMF/PDF	E(X)	Var(X)
Bernoulli	$p^x(1-p)^{1-x}$	$p$	$p(1-p)$
Binomial	$\binom{n}{k}p^k(1-p)^{n-k}$	$np$	$np(1-p)$
Poisson	$\frac{\lambda^k e^{-\lambda}}{k!}$	$\lambda$	$\lambda$
Hypergeom.	$\frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$	$n \frac{M}{N}$	...
Normal	$\frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$	$\mu$	$\sigma^2$

## 16 Wichtige Quantile

**Neue Begriffe:**

$z_p$ : p-Quantil der Standardnormalverteilung

$t_{df,p}$ : p-Quantil der t-Verteilung mit df Freiheitsgraden

Niveau	z-Wert	Verwendung
--------	--------	------------

90%	1.645	Normalvert.
95%	1.960	Normalvert.
99%	2.576	Normalvert.
95% (n=7)	2.365	t-Vert. (df=7)
95% (n=10)	2.262	t-Vert. (df=9)

## 17 Bedingte Wahrscheinlichkeit

**Neue Variablen:**

$A, B$ : Ereignisse

$A_i$ : Zerlegung des Ergebnisraums in Teilereignisse

$P(\cdot)$ : Wahrscheinlichkeit eines Ereignisses

**Definition:**

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0$$

**Multiplikationssatz:**

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

**Satz von Bayes:**

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

**Totale Wahrscheinlichkeit:**

$$P(B) = \sum_i P(A_i) \cdot P(B \mid A_i)$$

- Vierfeldertafel erstellen (Randsummen!)
- Gegebene Wahrscheinlichkeiten eintragen
- Formel  $P(A|B) = \frac{P(A \cap B)}{P(B)}$  anwenden
- Bei Bayes: Totale Wahrscheinlichkeit im Nenner

**Raucher/Nichtraucher (1400 Personen):**

	Raucher	Nichtraucher	Summe
Frauen	100	200	300
Männer	400	700	1100

Summe	500	900	1400
-------	-----	-----	------

$$P(\text{Raucherin} \mid \text{Frau}) = \frac{100}{300} = \frac{1}{3}$$

$$P(\text{Mann} \mid \text{Raucher}) = \frac{400}{500} = \frac{4}{5}$$

## 18 Ereignisbäume & Satz von Bayes

**Neue Begriffe:**

**Pfadwahrscheinlichkeit:** Produkt der Kantenwahrscheinlichkeiten

- Ereignisbaum zeichnen mit allen Verzweigungen
- Pfadwahrscheinlichkeit = Produkt entlang Pfad
- Totale Wahrscheinlichkeit = Summe aller Pfade zum Ereignis
- Bayes:  $P(A|B) = \frac{\text{Pfad zu A und B}}{\text{Alle Pfade zu B}}$

**Steckdosen (defekt):**

Fabrik 1:  $P(F_1) = 0.8, P(D|F_1) = 0.05$

Fabrik 2:  $P(F_2) = 0.2, P(D|F_2) = 0.1$

**Totale Wahrscheinlichkeit (defekt):**

$$P(D) = 0.8 \cdot 0.05 + 0.2 \cdot 0.1 = 0.06$$

**Bayes (F1 gegeben defekt):**

$$P(F_1 \mid D) = \frac{0.8 \cdot 0.05}{0.06} = \frac{2}{3}$$

## 19 Binomialverteilung

**Neue Variablen:**

$n$ : Anzahl Versuche

$p$ : Erfolgswahrscheinlichkeit pro Versuch

$k$ : Anzahl Erfolge

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

**Dichtefunktion (PMF):**

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

**Erwartungswert:**

$$E(X) = n \cdot p$$

**Varianz:**

$$\text{Var}(X) = n \cdot p \cdot (1 - p)$$

**Standardabweichung:**

$$\sigma = \sqrt{n \cdot p \cdot (1 - p)}$$

1. Prüfen:  $n$  Wiederholungen, konstantes  $p$ , unabhängig?
2. Parameter  $n, p$  festlegen
3.  $P(X = k)$  mit PMF-Formel berechnen
4. Für Intervalle: CDF nutzen oder summieren

**12× Würfeln, X = Anzahl Sechsen**

$$n = 12, p = \frac{1}{6}$$

$$P(X = 3) = \binom{12}{3} \cdot \left(\frac{1}{6}\right)^3 \cdot \left(\frac{5}{6}\right)^9 \approx 0.1974$$

$$P(X \leq 3) = \sum_{k=0}^3 P(X = k) \approx 0.8748$$

$$E(X) = 12 \cdot \frac{1}{6} = 2$$

$$\text{Var}(X) = 12 \cdot \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{3}$$

## 20 Poissonverteilung

**Neue Variablen:**

$\lambda$ : mittlere Ereignisrate pro Intervall

**Notation:**  $X \sim \text{Pois}(\lambda)$

**Dichtefunktion:**

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

**Erwartungswert & Varianz:**

$$E(X) = \lambda, \quad \text{Var}(X) = \lambda$$

**Anwendung:** Seltene Ereignisse in festem Intervall

1. Modell: Ereignisse pro Zeit/Fläche/Volumen
2. Parameter  $\lambda$  = erwartete Anzahl
3.  $P(X = k)$  mit PMF berechnen
4. Für Summen: einzeln addieren oder Tabelle nutzen

**Anrufe (120/Stunde = 2/Minute):**

$$\lambda = 2$$

$$P(X = 0) = \frac{2^0 \cdot e^{-2}}{0!} \approx 0.135$$

$$P(X = 1) = \frac{2^1 \cdot e^{-2}}{1!} \approx 0.271$$

$$P(X \leq 3) = \sum_{k=0}^3 P(X = k) \approx 0.857$$

## 21 Hypergeometrische Verteilung

**Neue Variablen:**

$N$ : Gesamtzahl Objekte

$M$ : Anzahl Merkmalsträger in Grundgesamtheit

$n$ : Stichprobengröße

$k$ : Merkmalsträger in Stichprobe

**Ziehen ohne Zurücklegen:**

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

**Erwartungswert:**

$$E(X) = n \cdot \frac{M}{N}$$

**Varianz:**

$$\text{Var}(X) = n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) \cdot \frac{N-n}{N-1}$$

1. Modell: Ziehen ohne Zurücklegen aus endlicher Grundgesamtheit

2. Parameter  $N, M, n$  bestimmen
3.  $P(X = k)$  mit Binomialkoeffizienten berechnen
4. Bei großem  $N$ : Approximation durch Binomialverteilung

**Lotto 6 aus 49:**

$$N = 49, M = 6, n = 6$$

$$P(X \geq 4) = \sum_{k=4}^6 \frac{\binom{6}{k} \cdot \binom{43}{6-k}}{\binom{49}{6}} \approx 0.00099$$

Etwa 1 von 1000 Tipps hat  $\geq 4$  Richtige.

**Approximation durch Binomial:**

$$\text{Faustregel: } \frac{n}{N} \leq 0.05 \rightarrow B\left(n, \frac{M}{N}\right)$$

## 22 Normalverteilung

**Neue Variablen:**

$\mu$ : Erwartungswert der Normalverteilung

$\sigma$ : Standardabweichung

$\Phi$ : Verteilungsfunktion der Standardnormalverteilung

**Notation:**  $X \sim N(\mu, \sigma^2)$

**Dichtefunktion (PDF):**

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

**Standardisierung:**

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

**Intervallwahrscheinlichkeit:**

$$P(a \leq X \leq b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

**68-95-99.7-Regel:**

- ca. 68% in  $[\mu - \sigma, \mu + \sigma]$
- ca. 95% in  $[\mu - 2\sigma, \mu + 2\sigma]$
- ca. 99.7% in  $[\mu - 3\sigma, \mu + 3\sigma]$

1. Gegeben:  $X \sim N(\mu, \sigma^2)$



2. Grenzen  $a, b$  standardisieren:  $z = \frac{x-\mu}{\sigma}$
3. Tabellenwerte  $\Phi(z_a), \Phi(z_b)$  ablesen
4.  $P(a \leq X \leq b) = \Phi(z_b) - \Phi(z_a)$

**Gegeben:**  $X \sim N(3, 4) \rightarrow \mu = 3, \sigma = 2$

**Gesucht:**  $P(1.26 \leq X \leq 5.12)$

$$z_1 = \frac{1.26 - 3}{2} = -0.87$$

$$z_2 = \frac{5.12 - 3}{2} = 1.06$$

$$\begin{aligned} P(1.26 \leq X \leq 5.12) &= \Phi(1.06) - \Phi(-0.87) \\ &= 0.8554 - (1 - 0.8078) \\ &= 0.6632 \end{aligned}$$

## 23 Normalapproximation der Binomialverteilung

**Neue Begriffe:**

$Y$ : approximierende Normalverteilung zu  $X$   
Stetigkeitskorrektur: Anpassung um 0.5

**Faustregel:**  $n \cdot p \cdot (1 - p) \geq 9$

**Approximation:**

$$B(n, p) \approx N(np, np(1 - p))$$

**Stetigkeitskorrektur:**

$$P(a \leq X \leq b) \approx P(a - 0.5 \leq Y \leq b + 0.5)$$

wobei  $Y \sim N(np, np(1 - p))$

1. Faustregel prüfen:  $np(1 - p) \geq 9$
2. Parameter berechnen:  $\mu = np, \sigma^2 = np(1 - p)$
3. Stetigkeitskorrektur anwenden ( $\pm 0.5$ )
4. Mit Standardnormalverteilung rechnen

**200 Geräte,  $p = 0.06, P(10 \leq X \leq 15)$**

Prüfung:  $200 \cdot 0.06 \cdot 0.94 = 11.28 \geq 9 \checkmark$

$$\mu = 12, \quad \sigma = \sqrt{11.28} \approx 3.36$$

Mit Stetigkeitskorrektur:

$$\begin{aligned} P(9.5 \leq Y \leq 15.5) &= \Phi\left(\frac{15.5 - 12}{3.36}\right) - \Phi\left(\frac{9.5 - 12}{3.36}\right) \\ &= \Phi(1.04) - \Phi(-0.74) \\ &\approx 0.85 - 0.23 = 0.62 \end{aligned}$$

## 24 Zentraler Grenzwertsatz

**Neue Variablen:**

$X_i$ : i-te Zufallsvariable in Stichprobe

$\bar{X}$ : Stichprobenmittel

$n$ : Stichprobengröße (hier als Zufallsvariablenzahl)

**Aussage:** Summe/Mittelwert von i.i.d. Zufallsvariablen ist asymptotisch normalverteilt.

Gegeben:  $X_1, X_2, \dots, X_n$  i.i.d. mit  $E(X_i) = \mu, \text{Var}(X_i) = \sigma^2$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

**Für großes  $n$ :**

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

**Standardisiert:**

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

**Produktion (6 Arbeitsschritte):**

Jeder Schritt: gleichverteilt in  $[1, 2]$  Stunden

$$\mu = 1.5, \quad \sigma^2 = \frac{1}{12}$$

Gesamtdauer  $T = \sum_{i=1}^6 X_i$ :

$$E(T) = 6 \cdot 1.5 = 9$$

$$\text{Var}(T) = 6 \cdot \frac{1}{12} = \frac{1}{2}$$

$$\sigma_T = \frac{1}{\sqrt{2}} \approx 0.707$$

$$P(T \leq 10) \approx \Phi\left(\frac{10 - 9}{0.707}\right) \approx \Phi(1.41) \approx 0.92$$

## 25 Punktschätzungen

**Neue Variablen:**

$\hat{\theta}$ : Schätzer für Parameter  $\theta$

$\hat{\mu}$ : Schätzer für Mittelwert

$\hat{\sigma}^2$ : Schätzer für Varianz

**Erwartungswert:**

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

**Varianz:**

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

**Eigenschaften:**

- **Erwartungstreu:**  $E(\hat{\theta}) = \theta$
- **Konsistent:**  $\hat{\theta} \rightarrow \theta$  für  $n \rightarrow \infty$
- **Effizient:** minimale Varianz

1. Stichprobe  $x_1, \dots, x_n$  gegeben
2. Schätzer wählen (z.B.  $\bar{x}$  für  $\mu$ )
3. Schätzwert berechnen
4. Eigenschaften prüfen (erwartungstreu, konsistent)



(1) Verteilung Grundgesamtheit	(2) Param.	(3) Schätzfunktionen	(4) standardisierte Zufallsvariable	(5) Verteilung, benötigte Quantile	(6) Intervallgrenzen
Normalverteilung (Varianz $\sigma^2$ bekannt)	$\mu$	$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$	$U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$	<b>Standardnormalverteilung</b> (Tabelle 2) $c = u_p$ mit $p = \frac{1+\gamma}{2}$	$\theta_u = \bar{X} - c \cdot \frac{\sigma}{\sqrt{n}}$ $\theta_o = \bar{X} + c \cdot \frac{\sigma}{\sqrt{n}}$
Normalverteilung (Varianz $\sigma^2$ unbekannt und $n \leq 30$ ; sonst Fall 1 mit $s$ als Schätzwert für $\sigma$ )	$\mu$	$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ $S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (X_i - \bar{X})^2$	$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$	<b>t-Verteilung</b> (Tabelle 4) mit $f = n - 1$ $c = t_{(p;f)}$ mit $p = \frac{1+\gamma}{2}$	$\theta_u = \bar{X} - c \cdot \frac{S}{\sqrt{n}}$ $\theta_o = \bar{X} + c \cdot \frac{S}{\sqrt{n}}$
Normalverteilung	$\sigma^2$	$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ $S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (X_i - \bar{X})^2$	$Z = (n-1) \frac{S^2}{\sigma^2}$	<b>Chi-Quadrat-Verteilung</b> (Tabelle 3) mit $f = n - 1$ $c_1 = z_{(p_1;f)}$ mit $p_1 = \frac{1-\gamma}{2}$ $c_2 = z_{(p_2;f)}$ mit $p_2 = \frac{1+\gamma}{2}$	$\theta_u = \frac{(n-1) \cdot S^2}{c_1}$ $\theta_o = \frac{(n-1) \cdot S^2}{c_2}$
Bernoulli-Verteilung Anteilsschätzung (mit $n\hat{p}(1-\hat{p}) > 9$ )	$p$	$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ $X_i$ 0/1-wertig mit $P(X_i = 1) = p$	$U = \frac{\bar{X} - p}{\sqrt{p(1-p)/n}}$	<b>Standardnormalverteilung</b> (näherungsweise), Tabelle 2 $c = u_q$ mit $q = \frac{1+\gamma}{2}$	$\theta_u = \bar{X} - c \cdot \sqrt{\frac{\bar{X} \cdot (1-\bar{X})}{n}}$ $\theta_o = \bar{X} + c \cdot \sqrt{\frac{\bar{X} \cdot (1-\bar{X})}{n}}$
beliebig mit $n > 30$	$\mu, \sigma^2$	wie im Fall 1 (gegebenenfalls mit $s$ als Schätzwert für $\sigma$ ) bzw. im Fall 3			

#### Vertrauensintervall für $\mu$ ( $\sigma$ bekannt)

**Gewichte** ( $\sigma = 2$  kg,  $n = 100$ ,  $\bar{x} = 75$  kg, 95%-Niveau):

1. **Vertrauensniveau festlegen:** 95%  $\rightarrow \alpha = 0.05$
2. **Quantil aus Tabelle 2:**  $z_{1-\alpha/2} = z_{0.975} = 1.960$

3. **Halbbreite berechnen:**

$$h = z \cdot \frac{\sigma}{\sqrt{n}} = 1.960 \cdot \frac{2}{\sqrt{100}} = 0.392$$

4. **Intervallgrenzen:**

$$\theta_u = \bar{x} - h = 75 - 0.392 = 74.608$$

$$\theta_o = \bar{x} + h = 75 + 0.392 = 75.392$$

**Vertrauensintervall:** [74.608, 75.392]

#### Vertrauensintervall für $\mu$ ( $\sigma$ unbekannt)

**Marroni** ( $n = 8$ ,  $\bar{x} = 18$ ,  $s = 2.39$ , 95%-Niveau):

1. **Vertrauensniveau:** 95%  $\rightarrow \alpha = 0.05$
2. **Freiheitsgrade:**  $f = n - 1 = 7$
3. **t-Quantil aus Tabelle 4:**  $t_{7;0.975} = 2.365$
4. **Halbbreite berechnen:**

$$h = t \cdot \frac{s}{\sqrt{n}} = 2.365 \cdot \frac{2.39}{\sqrt{8}} \approx 2.0$$

5. **Intervallgrenzen:**

$$\theta_u = 18 - 2.0 = 16.0$$

$$\theta_o = 18 + 2.0 = 20.0$$

**Vertrauensintervall:** [16.0, 20.0]

#### Vertrauensintervall für $\sigma^2$

**Marroni** ( $n = 8$ ,  $s^2 = 5.71$ , 95%-Niveau):

1. **Vertrauensniveau:** 95%  $\rightarrow \alpha = 0.05$
2. **Freiheitsgrade:**  $f = n - 1 = 7$
3.  **$\chi^2$ -Quantile aus Tabelle 3:**
  - $\chi_{7;0.975}^2 = 16.01$  (für untere Grenze)
  - $\chi_{7;0.025}^2 = 1.69$  (für obere Grenze)
4. **Intervallgrenzen:**

$$\theta_u = \frac{(n-1) \cdot s^2}{\chi_{0.975}^2} = \frac{7 \cdot 5.71}{16.01} = 2.50$$

$$\theta_o = \frac{(n-1) \cdot s^2}{\chi_{0.025}^2} = \frac{7 \cdot 5.71}{1.69} = 23.66$$

**Vertrauensintervall:** [2.50, 23.66]

#### Vertrauensintervall für Anteilswert $p$

**Umfrage** ( $n = 1200$ ,  $k = 473$  Ja, 99%-Niveau):

1. **Anteil berechnen:**

$$\hat{p} = \frac{k}{n} = \frac{473}{1200} = 0.394$$

2. **Faustregel prüfen:**

$$n\hat{p}(1-\hat{p}) = 1200 \cdot 0.394 \cdot 0.606 = 286.5 \geq 9$$

3. **Quantil (Tabelle 2):**  $z_{0.995} = 2.576$

4. **Halbbreite:**

$$h = z \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 2.576 \sqrt{\frac{0.394 \cdot 0.606}{1200}} \approx 0.036$$

5. **Intervallgrenzen:**

$$\theta_u = 0.394 - 0.036 = 0.358$$

$$\theta_o = 0.394 + 0.036 = 0.430$$

**Vertrauensintervall:** [0.358, 0.430]

## 26 Lineare Regression (KQM)

### Neue Variablen:

$a$ : Achsenabschnitt

$b$ : Steigung der Regressionsgeraden

$R^2$ : Bestimmtheitsmaß

**Regressionsgerade:**  $y = ax + b$

### Steigung:

$$a = \frac{s_{xy}}{s_x^2} = r \cdot \frac{s_y}{s_x}$$

### Achsenabschnitt:

$$b = \bar{y} - a \cdot \bar{x}$$

### Residualvarianz:

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

### Bestimmtheitsmaß:

$$R^2 = r^2 = \frac{s_y^2 - s_{\text{Res}}^2}{s_y^2}$$

1. Korrelationskoeffizient  $r$  berechnen
2. Steigung:  $b = s_{xy}/s_x^2$
3. Achsenabschnitt:  $a = \bar{y} - b\bar{x}$
4. Prüfung: Residuenplot (sollte unsystematisch streuen)
5. Güte:  $R^2$  berechnen (Anteil erklärter Varianz)

### Größe-Gewicht (15 Personen):

$\bar{x} = 173 \text{ cm}$ ,  $s_x = 6.047 \text{ cm}$

$\bar{y} = 72.267 \text{ kg}$ ,  $s_y = 7.474 \text{ kg}$

$s_{xy} = 41.071$ ,  $r = 0.898$

$$b = \frac{41.071}{6.047^2} \approx 1.123$$

$a = 72.267 - 1.123 \cdot 173 \approx -122.02$

### Regressionsgerade:

$$y = -122.02 + 1.123x$$

$$R^2 = 0.898^2 \approx 0.806$$

→ 80.6% der Varianz erklärt

## 27 Übersichtstabelle Verteilungen

Verteilung	PMF/PDF	E(X)	Var(X)
Bernoulli	$p^x(1-p)^{1-x}$	$p$	$p(1-p)$
Binomial	$\binom{n}{k}p^k(1-p)^{n-k}$	$np$	$np(1-p)$
Poisson	$\frac{\lambda^k e^{-\lambda}}{k!}$	$\lambda$	$\lambda$
Hypergeom.	$\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$	$n \frac{M}{N}$	...
Normal	$\frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$	$\mu$	$\sigma^2$

## 28 Wichtige Quantile

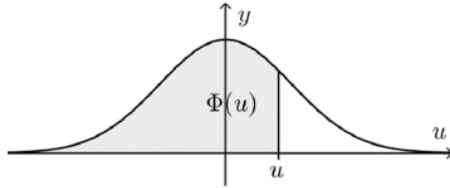
### Neue Begriffe:

$z_p$ : p-Quantil der Standardnormalverteilung

$t_{df,p}$ : p-Quantil der t-Verteilung mit df Freiheitsgraden

Niveau	z-Wert	Verwendung
90%	1.645	Normalvert.
95%	1.960	Normalvert.
99%	2.576	Normalvert.
95% (n=7)	2.365	t-Vert. (df=7)
95% (n=10)	2.262	t-Vert. (df=9)

**Tabelle 1: CDF  $\Phi(u)$  der Standardnormalverteilung**



$$P(U \leq u) = \Phi(u)$$

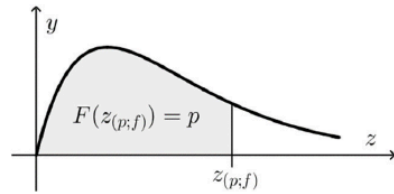
$$P(U \geq u) = 1 - \Phi(u)$$

$$P(-u \leq U \leq u) = 2 \cdot \Phi(u) - 1$$

$$\Phi(-u) = 1 - \Phi(u)$$

$u$	0	1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

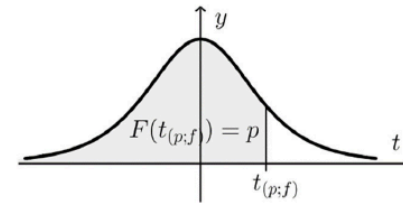
**Tabelle 3: Quantile der Chi-Quadrat-Verteilung**



$p$  : vorgegebene Wahrscheinlichkeit

$z_{(p;f)}$  : zur Wahrscheinlichkeit  $p$  gehöriges Quantil bei  $f$  Freiheitsgraden

**Tabelle 4: Quantile der t-Verteilung von «Student»**



$p$  : vorgegebene Wahrscheinlichkeit

$t_{(p;f)}$  : zur Wahrscheinlichkeit  $p$  gehöriges Quantil bei  $f$  Freiheitsgraden

	$p$									
$f$	0.005	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99	0.995
1	0.000	0.000	0.001	0.004	0.016	2.71	3.84	5.02	6.63	7.88
2	0.010	0.020	0.051	0.103	0.211	4.61	5.99	7.38	9.21	10.60
3	0.072	0.115	0.216	0.352	0.584	6.25	7.81	9.35	11.34	12.84
4	0.207	0.297	0.484	0.711	1.064	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.95
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
22	8.6	9.5	11.0	12.3	14.0	30.8	33.9	36.8	40.3	42.8
24	9.9	10.9	12.4	13.8	15.7	33.2	36.4	39.4	43.0	45.6
26	11.2	12.2	13.8	15.4	17.3	35.6	38.9	41.9	45.6	48.3
28	12.5	13.6	15.3	16.9	18.9	37.9	41.3	44.5	48.3	51.0
30	13.8	15.0	16.8	18.5	20.6	40.3	43.8	47.0	50.9	53.7
40	20.7	22.2	24.4	26.5	29.1	51.8	55.8	59.3	63.7	66.8
50	28.0	29.7	32.4	34.8	37.7	63.2	67.5	71.4	76.2	79.5
60	35.5	37.5	40.5	43.2	46.5	74.4	79.1	83.3	88.4	92.0
70	43.3	45.4	48.8	51.7	55.3	85.5	90.5	95.0	100.4	104.2
80	51.2	53.5	57.2	60.4	64.3	96.6	101.9	106.6	112.3	116.3
90	59.2	61.8	65.6	69.1	73.3	107.6	113.1	118.1	124.1	128.3
100	67.3	70.1	74.2	77.9	82.4	118.5	124.3	129.6	135.8	140.2

	$p$				
$f$	0.9	0.95	0.975	0.99	0.995
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
22	1.321	1.717	2.074	2.508	2.819
24	1.318	1.711	2.064	2.492	2.797
26	1.315	1.706	2.056	2.479	2.779
28	1.313	1.701	2.048	2.467	2.763
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
50	1.299	1.676	2.009	2.403	2.678
60	1.296	1.671	2.000	2.390	2.660
100	1.290	1.660	1.984	2.364	2.626
200	1.286	1.653	1.972	2.345	2.601
500	1.283	1.648	1.965	2.334	2.586
∞	1.282	1.645	1.960	2.326	2.576

$$t_{(1-p;f)} = -t_{(p;f)}$$