Zentrale Schwankungsintervalle (1)

- Mit den Quantilen lassen sich die Wahrscheinlichkeiten für sogenannte **zentrale Schwankungsintervalle** der Form

$$\mu - c \leq X \leq \mu + c$$

angeben.

- Entweder gibt man c vor und bestimmt dazu α bzw. $1 - \alpha$ so, dass

$$P(\mu - c \leq X \leq \mu + c) = 1 - \alpha$$

gilt.

- Oder man gibt die Wahrscheinlichkeit $1 - \alpha$ vor und bestimmt mittels Quantil das zugehörige c.

Zentrale Schwankungsintervalle (2)

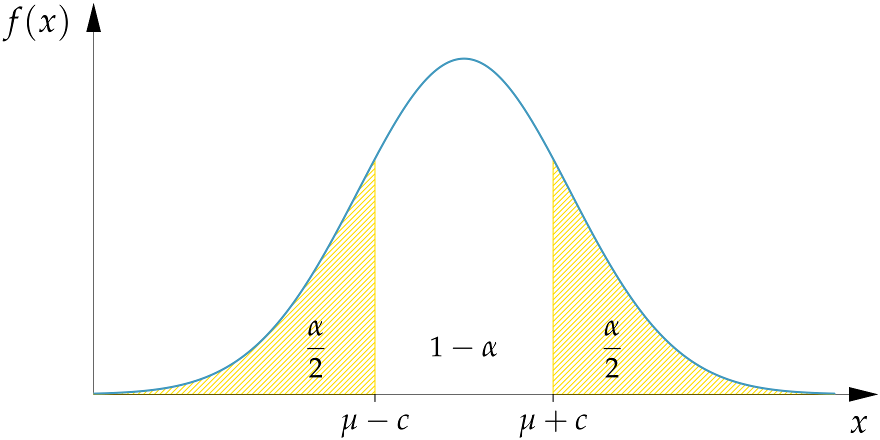


Abb.: Darstellung eines zentralen Schwankungsintervalls der Form $\mu - c \leq X \leq \mu + c$ mit zugehöriger Wahrscheinlichkeit $1 - \alpha$

Quantile und Ws der $N(0,1)$-Verteilung

- Für die Zufallsvariable $Z \sim N(0,1)$ gilt

$$P(-1.64 \leq Z \leq 1.64) = 0.90$$
$$P(-1.96 \leq Z \leq 1.96) = 0.95$$
$$P(-2.58 \leq Z \leq 2.58) = 0.99$$

bzw. allgemein

$$P\left(-z_{1-\alpha/2} \leq Z \leq z_{1-\alpha/2}\right) = 1 - \alpha\,,$$

wobei $z_{1-\alpha/2}$ das $(1-\alpha/2)$-Quantil der Standardnormalverteilung bezeichnet.

Bsp. Lage der mittleren 95 %

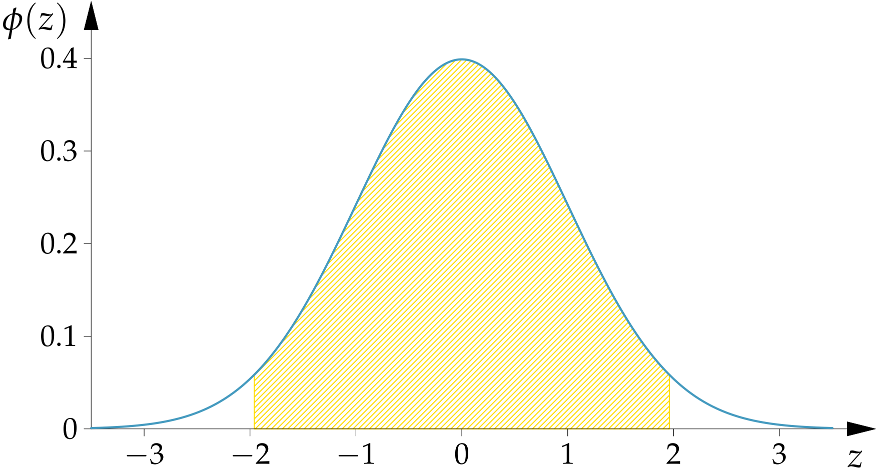


Abb.: Fläche unter der Dichte der Standardnormalverteilung im Intervall $[-1.96, 1.96]$

Bsp. Lage der äußeren 5%

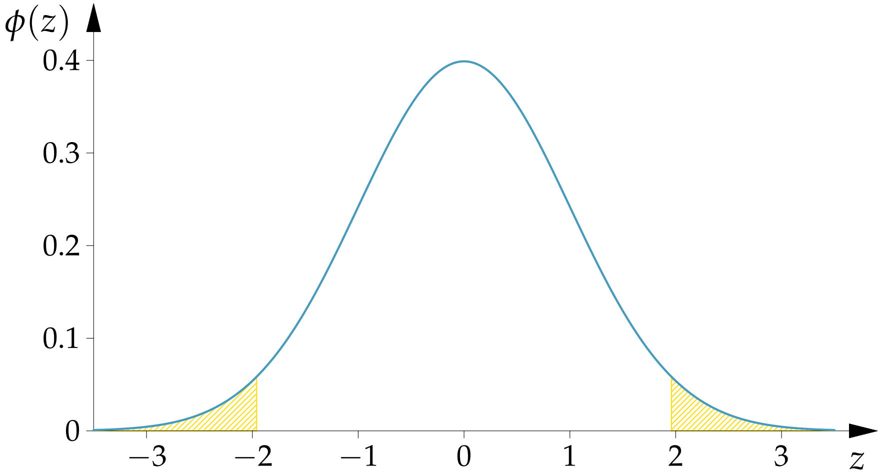


Abb.: Fläche unter der Dichte der $N(0,1)$-Verteilung unterhalb von -1.96 und oberhalb von 1.96

Quantile und Ws der $N(\mu, \sigma)$-Verteilung

- Für die Zufallsvariable $X \sim N(\mu, \sigma)$ gilt

$$P(\mu - 1.64 \cdot \sigma \leq X \leq \mu + 1.64 \cdot \sigma) = 0.90$$
$$P(\mu - 1.96 \cdot \sigma \leq X \leq \mu + 1.96 \cdot \sigma) = 0.95$$
$$P(\mu - 2.58 \cdot \sigma \leq X \leq \mu + 2.58 \cdot \sigma) = 0.99$$

bzw. allgemein

$$P\left(\mu - z_{1-\alpha/2} \cdot \sigma \leq X \leq \mu + z_{1-\alpha/2} \cdot \sigma\right) = 1 - \alpha\,,$$

wobei $z_{1-\alpha/2}$ das $(1 - \alpha/2)$-Quantil der Standardnormalverteilung bezeichnet.

Zentrale Schwankungsintervalle (3)

- Für $X \sim N(\mu, \sigma)$ und $z_{1-\alpha/2} = k$ erhält man die $k\sigma$-Bereiche:

$$\begin{aligned}
k = 1: &\quad P(\mu - \sigma \leq X \leq \mu + \sigma) = 0.6827 \\
k = 2: &\quad P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.9545 \\
k = 3: &\quad P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.9973
\end{aligned}$$

Zentrale Schwankungsintervalle (4)

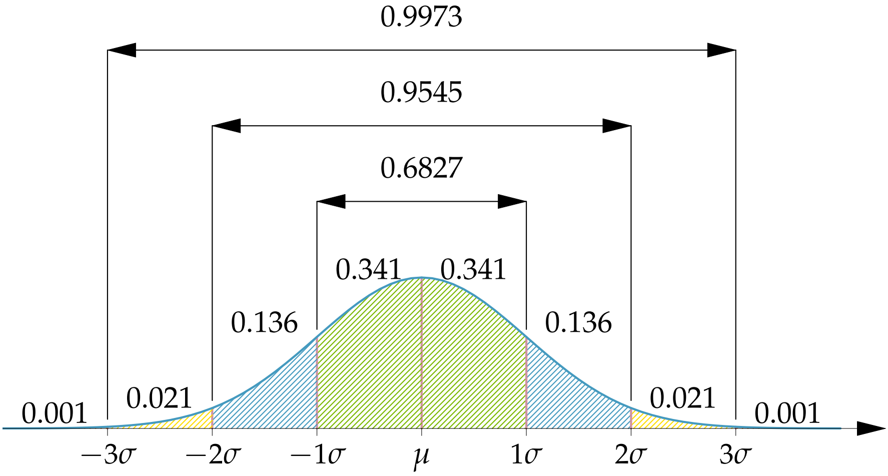


Abb.: Die 68-95-99.7-Regel für Normalverteilungen

Bsp. Zentrale Schwankungsintervalle

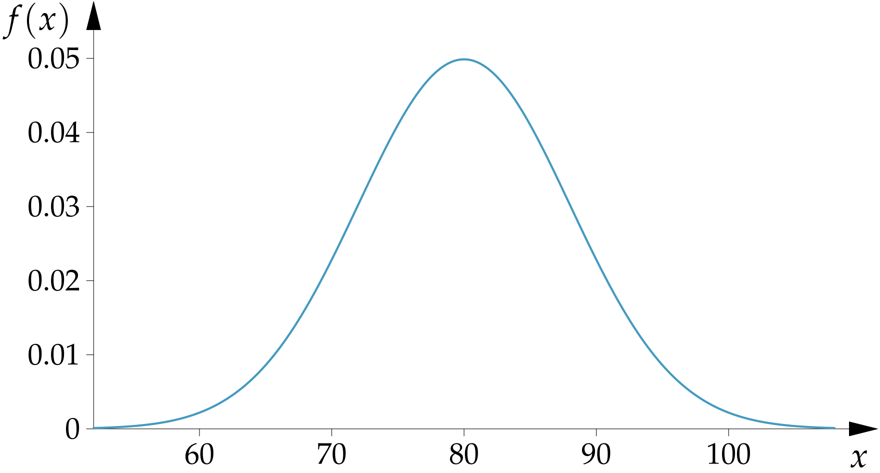


Abb.: Dichte der Normalverteilung mit $\mu = 80$ und $\sigma = 8$

Prüfverteilungen

- Prüfverteilungen
  - dienen dazu, die Verteilung von statistischen Kenngrößen zu beschreiben, z. B.
    - des Stichprobenmittelwerts,
    - der Stichprobenvarianz.
  - werden für Schätz- und Testverfahren der induktiven Statistik benötigt.
  - werden auch als Testverteilungen bezeichnet.

χ^2-Verteilung

- Seien $X_1, \ldots, X_n$ unabhängige und identisch $N(0,1)$-verteilte Zufallsvariablen. Dann heißt die Verteilung der Zufallsvariable

$$Z = X_1^2 + \ldots + X_n^2$$

Chi-Quadrat-Verteilung mit n Freiheitsgraden, kurz $\chi^2(n)$-Verteilung, und Z heißt $\chi^2(n)$-verteilt, kurz $Z \sim \chi^2(n)$.

- Es gilt

$$E(Z) = n \,, \quad Var(Z) = 2n \,.$$

Bsp. χ^2-Verteilung

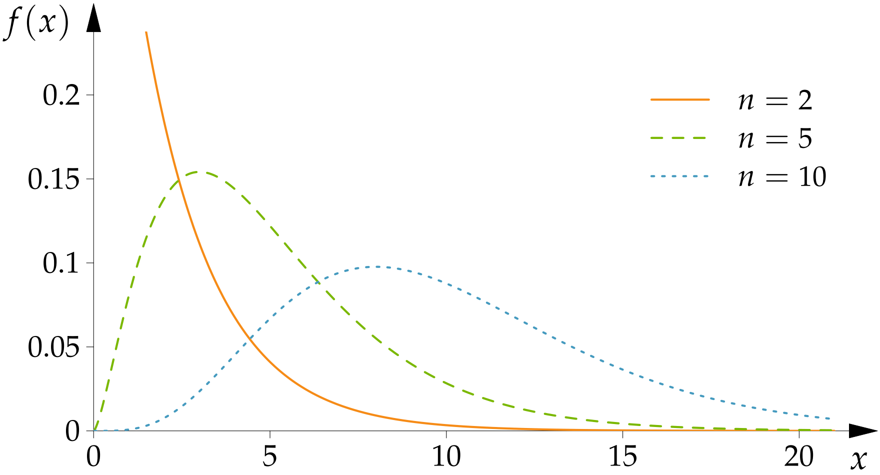


Abb.: Wahrscheinlichkeitsdichten der χ^2-Verteilung für unterschiedliche Freiheitsgrade n

t-Verteilung

- Seien $X \sim N(0,1)$, $Z \sim \chi^2(n)$ sowie X und Z unabhängig. Dann heißt die Verteilung der Zufallsvariable

  $$T = \frac{X}{\sqrt{Z/n}}$$

  ***t*-Verteilung** mit n Freiheitsgraden, kurz $t(n)$-Verteilung. Die Zufallsvariable T heißt $t(n)$-verteilt, kurz $T \sim t(n)$.

- Es gilt

  $$E(T) = 0 \;\; (n \geq 2), \quad Var(T) = \frac{n}{n-2} \;\; (n \geq 3).$$

Bsp. t-Verteilung

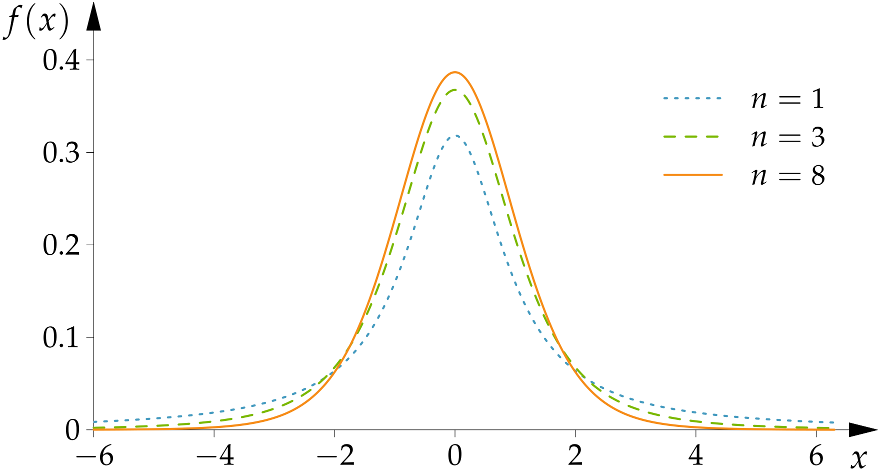


Abb.: Wahrscheinlichkeitsdichten der t-Verteilung für unterschiedliche Freiheitsgrade n

Bsp. t- und Standardnormalverteilung

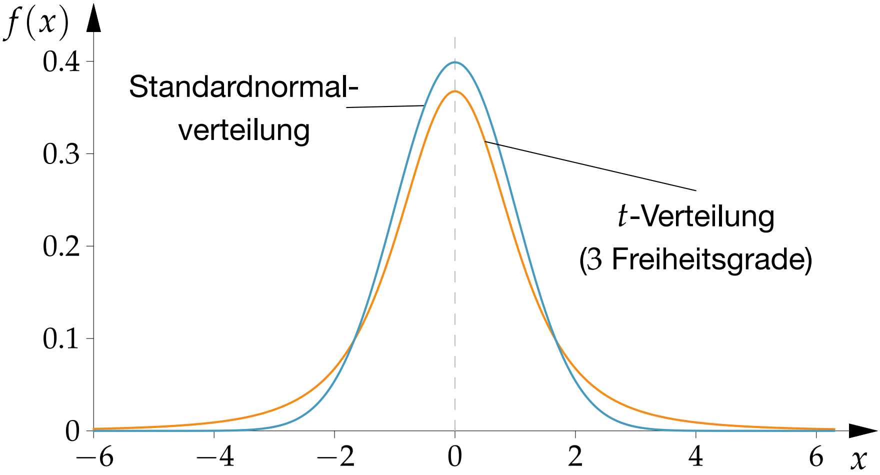


Abb.: Wahrscheinlichkeitsdichte der $N(0,1)$-Verteilung und der t-Verteilung mit $n = 3$ Freiheitsgraden

Eigenschaften der t-Verteilung

- Die Dichtekurve ist symmetrisch um Null.
- Mit abnehmender Anzahl der Freiheitsgrade sinkt das Maximum der t-Verteilung.
- Im Vergleich zur $N(0,1)$-Verteilung ist mehr Wahrscheinlichkeit in den Ausläufen und weniger im zentralen Teil konzentriert.
- Für $n \longrightarrow \infty$ konvergiert die Dichtekurve gegen die Dichte der $N(0,1)$-Verteilung (ab $n > 30$ ist die Approximation bereits sehr gut).

Tabelle der t-Verteilung

D Students t-Verteilung

Tabelliert sind die Quantile für n Freiheitsgrade. Für das Quantil $t_{1-\alpha}(n)$ gilt $F(t_{1-\alpha}(n)) = 1 - \alpha$. Links vom Quantil $t_{1-\alpha}(n)$ liegt die Wahrscheinlichkeitsmasse $1 - \alpha$.

Ablesebeispiel: $t_{0.99}(20) = 2.528$

Die Quantile für $0 < 1 - \alpha < 0.5$ erhält man aus $t_\alpha(n) = -t_{1-\alpha}(n)$

Approximation für $n > 30$: $t_\alpha(n) \approx z_\alpha$ (z_α ist das (α)-Quantil der Standardnormalverteilung)

n	0.6	0.8	0.9	0.95	0.975	0.99	0.995	0.999	0.9995
1	0.3249	1.3764	3.0777	6.3138	12.706	31.821	63.657	318.31	636.62
2	0.2887	1.0607	1.8856	2.9200	4.3027	6.9646	9.9248	22.327	31.599
3	0.2767	0.9785	1.6377	2.3534	3.1824	4.5407	5.8409	10.215	12.924
4	0.2707	0.9410	1.5332	2.1318	2.7764	3.7469	4.6041	7.1732	8.6103
5	0.2672	0.9195	1.4759	2.0150	2.5706	3.3649	4.0321	5.8934	6.8688
6	0.2648	0.9057	1.4398	1.9432	2.4469	3.1427	3.7074	5.2076	5.9588
7	0.2632	0.8960	1.4149	1.8946	2.3646	2.9980	3.4995	4.7853	5.4079
8	0.2619	0.8889	1.3968	1.8595	2.3060	2.8965	3.3554	4.5008	5.0413
9	0.2610	0.8834	1.3830	1.8331	2.2622	2.8214	3.2498	4.2968	4.7809
10	0.2602	0.8791	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437	4.5869
11	0.2596	0.8755	1.3634	1.7959	2.2010	2.7181	3.1058	4.0247	4.4370
12	0.2590	0.8726	1.3562	1.7823	2.1788	2.6810	3.0545	3.9296	4.3178
13	0.2586	0.8702	1.3502	1.7709	2.1604	2.6503	3.0123	3.8520	4.2208
14	0.2582	0.8681	1.3450	1.7613	2.1448	2.6245	2.9768	3.7874	4.1405
15	0.2579	0.8662	1.3406	1.7531	2.1314	2.6025	2.9467	3.7328	4.0728
16	0.2576	0.8647	1.3368	1.7459	2.1199	2.5835	2.9208	3.6862	4.0150
17	0.2573	0.8633	1.3334	1.7396	2.1098	2.5669	2.8982	3.6458	3.9651
18	0.2571	0.8620	1.3304	1.7341	2.1009	2.5524	2.8784	3.6105	3.9216
19	0.2569	0.8610	1.3277	1.7291	2.0930	2.5395	2.8609	3.5794	3.8834
20	0.2567	0.8600	1.3253	1.7247	2.0860	2.5280	2.8453	3.5518	3.8495
21	0.2566	0.8591	1.3232	1.7207	2.0796	2.5176	2.8314	3.5272	3.8193
22	0.2564	0.8583	1.3212	1.7171	2.0739	2.5083	2.8188	3.5050	3.7921
23	0.2563	0.8575	1.3195	1.7139	2.0687	2.4999	2.8073	3.4850	3.7676
24	0.2562	0.8569	1.3178	1.7109	2.0639	2.4922	2.7969	3.4668	3.7454
25	0.2561	0.8562	1.3163	1.7081	2.0595	2.4851	2.7874	3.4502	3.7251
26	0.2560	0.8557	1.3150	1.7056	2.0555	2.4786	2.7787	3.4350	3.7066
27	0.2559	0.8551	1.3137	1.7033	2.0518	2.4727	2.7707	3.4210	3.6896
28	0.2558	0.8546	1.3125	1.7011	2.0484	2.4671	2.7633	3.4082	3.6739
29	0.2557	0.8542	1.3114	1.6991	2.0452	2.4620	2.7564	3.3962	3.6594
30	0.2556	0.8538	1.3104	1.6973	2.0423	2.4573	2.7500	3.3852	3.6460
∞	0.2533	0.8416	1.2816	1.6449	1.9600	2.3263	2.5758	3.0903	3.2906

Abb.: t-Verteilung (Aus: Fahrmeir u. a. 2016)

7. Schätzverfahren

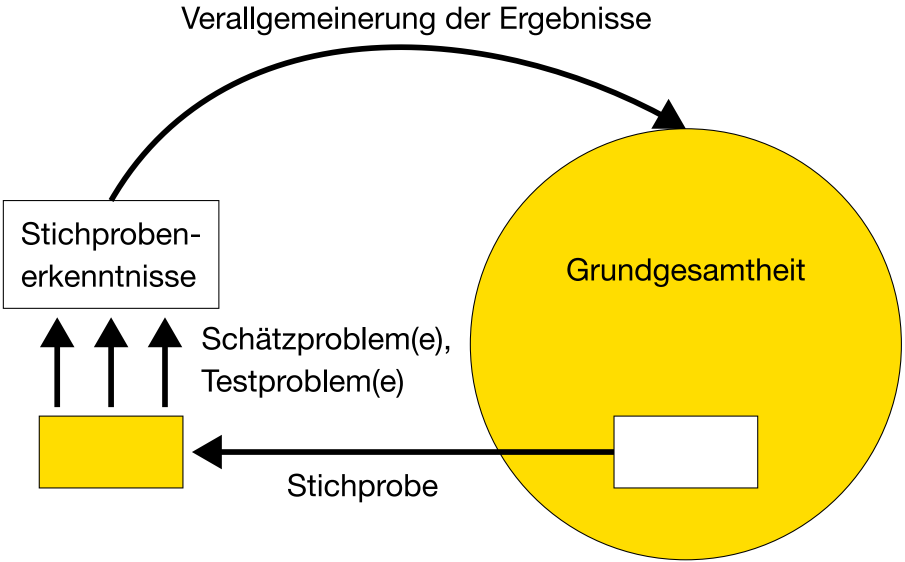
Rückblick

- deskriptive Statistik
  - Beschreibung, grafische Aufbereitung und Komprimierung eines oder mehrerer Merkmale.
  - Basiert rein auf Daten.
  - Trifft keine Verallgemeinerungen!
- Wahrscheinlichkeitstheorie und -verteilungen
  - Gegenstand sind Ereignisse, Wahrscheinlichkeiten, Zufallsvariablen und deren Verteilungen.
  - Formuliert allgemeine Gesetze für das Verhalten von Merkmalen in Grundgesamtheiten.
  - Basiert nur auf mathematischen Modellen (keine Daten!).

Induktive Statistik

- auch schließende, analytische oder beurteilende Statistik genannt
- Ziel: Ziehen von Schlüssen aus **Stichproben** über das Verhalten von **Merkmalen in der Grundgesamtheit** mit Hilfe der **Wahrscheinlichkeitstheorie**
- Teilgebiete der induktiven Statistik
  - Schätzverfahren: einen Parameter mit vorgegebener Ungenauigkeit schätzen
  - Testverfahren: eine Nullhypothese mit vorgegebener Unsicherheit ablehnen

Strategie der induktiven Statistik



Deskriptive vs. induktive Statistik

- Die deskriptive Statistik befasst sich mit der Untersuchung und Beschreibung möglichst der ganzen Grundgesamtheit.
- Die induktive Statistik untersucht demgegenüber nur einen Teil (Stichprobe), der für die Grundgesamtheit, deren Eigenschaften uns interessieren, charakteristisch oder repräsentativ sein soll.

Merkmale als Zufallsvariablen

- Untersuchte Merkmale werden als Zufallsvariablen aufgefasst.
- Die Beobachtung eines Merkmals wird als Ergebnis eines Zufallsexperiments interpretiert.

Stichprobenvariablen und -werte

- Ausgangspunkt ist eine Zufallsvariable X.
- Die wiederholte Beobachtung von X wird durch Zufallsvariablen oder **Stichprobenvariablen** $X_1, \ldots, X_n$ gekennzeichnet.
- Für die Stichprobenvariablen gilt:
  - $X_1, \ldots, X_n$ haben die selbe Verteilung wie X.
  - $X_1, \ldots, X_n$ sind unabhängig.
- Die **Stichprobenwerte** $x_1, \ldots, x_n$ sind Realisierungen der Stichprobenvariablen $X_1, \ldots, X_n$.
- $X_1, \ldots, X_n$ ist eine **mathematische Stichprobe**, $x_1, \ldots, x_n$ eine **empirische Stichprobe**.

Zufallsstichproben

- Die induktive Statistik setzt stets Zufallsstichproben voraus!
- Eine **Zufallsstichprobe** ist Teil einer Grundgesamtheit, die durch einen Auswahlprozess mit Zufallsprinzip aus dieser entnommen und stellvertretend, repräsentativ für die Grundgesamtheit ist.
- Statistische Methoden der Stichprobenziehung sind ein eigenes Teilgebiet der Statistik.

Einfache Zufallsstichproben

- Bei einer **einfachen Zufallsstichprobe** besitzt jedes Element in der Grundgesamtheit dieselbe bekannte Wahrscheinlichkeit, für die Stichprobe ausgewählt zu werden.
- Die Elemente der Grundgesamtheit müssen nummerierbar sein.

- Aus der Menge $\{1, 2, \ldots, 80\}$ sollen 10 Zahlen zufällig gezogen werden:

```
sample(x = c(1:80), size = 10,
    replace = FALSE)
 [1] 68 39  1 34 43 14 59 51 21 54
```

Mit `replace = FALSE` wird verhindert, dass einzelne Zahlen mehrfach auftreten.

Weitere Stichprobenverfahren

- **Geschichtete Stichprobe:** Grundgesamtheit wird in homogene Teilgesamtheiten (Schichten) zerlegt. Aus jeder Schicht zieht man dann eine einfache Zufallsstichprobe.
  - Beispiele: Schichtung nach Alter, Schichtung nach Schulabschuss
- **Klumpenstichprobe:** Grundgesamtheit ist in „naturgegebene“ Teilgesamtheiten (Klumpen) zerlegt. Aus der Menge von Klumpen zieht man eine einfache Zufallsstichprobe.
  - Beispiel: Klasse einer Schule bei der Befragung von Schülern

Schätzen

- **Schätzen:** Festlegen von Werten für unbekannte Parameter des Merkmals in der Grundgesamtheit mittels einer Stichprobe
  - **Punktschätzung:** Verfahren, bei dem ein unbekannter Parameter durch einen einzigen Wert geschätzt wird
  - **Intervallschätzung:** Verfahren, dessen Ergebnis ein (Vertrauens-) Bereich ist

Punktschätzungen

- Ziel: Gebe einen möglichst genauen Näherungswert für unbekannte Parameter eines Merkmals in der Grundgesamtheit an.
- Parameter treten auf als:
  - Kenngrößen einer beliebigen, unbekannten Verteilung, z. B. Erwartungswert, Varianz, Standardabweichung
  - spezifische Parameter eines für das Merkmal angenommenen Verteilungsmodells, z. B. Parameter λ einer Exponentialverteilung

Schätzfunktionen und -werte

- **Schätzfunktion** oder **Schätzer**: Vorschrift (Formel), nach der aus Stichprobenvariablen $X_1, \ldots, X_n$ ein angenäherter Wert für den unbekannten Parameter berechnet wird.
- Schätzfunktionen sind als Funktionen von Zufallsvariablen selbst wieder Zufallsvariablen.
- Bsp.: Schätzfunktion für den Erwartungswert $E(X)$:

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

- Eine **Schätzung** oder ein **Schätzwert** ist der Wert, den die Schätzfunktion in Abhängigkeit von der jeweiligen Stichprobe annimmt.

Bsp. Schätzfunktionen und -werte

- Frage: Wie groß ist die zu erwartende Körpergröße μ (in Zentimeter) männlicher Studierender an der HSWT?
  - Schätzfunktion für den Erwartungswert $\mu = E(X)$:

    $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$
  - angenommene Stichprobe ($n = 21$):

    $x = 190, 170, \ldots, 181$
  - berechneter Schätzwert:

    $\bar{x} = 183.0 = \frac{1}{21} \left(190 + 170 + \ldots + 181\right)$
- Antwort: Die zu erwartende Körpergröße μ beträgt 183.0 cm.

Kriterien zur Güte einer Schätzung

- **Erwartungstreue:** Der Erwartungswert der Schätzfunktion ist gleich dem wahren Wert des zu schätzenden Parameters.
- **Konsistenz:** Je größer der Stichprobenumfang n, desto genauer ist Schätzung.
- **Effizienz:** Die Schätzfunktion liefert auch schon bei einer kleinen Stichprobe einen brauchbaren Schätzwert.
- **Suffizienz:** Alle Informationen in der Stichprobe über den Parameter werden genutzt.

„Gute“ Schätzfunktionen

- Schätzfunktion für den Erwartungswert $\mu = E(X)$:

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

- Schätzfunktion für die Varianz $\sigma^2 = Var(X)$:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

- Schätzfunktion für die Standardabweichung $\sigma = \sigma(X)$:

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2}$$

Bem. „Gute“ Schätzfunktionen

- Die Größen $\overline{X}$, S^2 und S sind universelle Schätzfunktionen für die Parameter μ, σ^2 und σ. Sie gelten unabhängig von der zugrunde liegenden Verteilung.

  Voraussetzung: Die Stichprobenvariablen sind unabhängig und identisch verteilt und die entsprechenden Parameter existieren (d. h. sind endlich).

Bem. Notation

- Zufallsvariablen werden mit Großbuchstaben bezeichnet, ihre Realisierungen mit Kleinbuchstaben.
- Beispiele:

Zielgröße (Parameter)	Schätzfunktion/ Schätzer (Zufallsvariable)	Schätzwert/ Schätzung (Realisierung)
μ	$\overline{X}$	$\bar{x}$
σ	S	s
β_0, β_1	B_0, B_1	b_0, b_1

Bem. Punktschätzungen

- Punktschätzungen schwanken von Zufallsstichprobe zu Zufallsstichprobe.
- Wie vertrauenswürdig sind die Informationen (Parameter), die aus den Daten gewonnen werden?
- Ziel: Bestimme ein Intervall, das den wahren aber unbekannten Parameter mit einer vorgegebenen Wahrscheinlichkeit überdeckt!

Intervallschätzungen

- $(\mathbf{1}-\boldsymbol{\alpha})$**-Konfidenzintervall (KI):** Intervall, das den interessierenden (unbekannten) Parameter mit der Wahrscheinlichkeit $1-\alpha$ überdeckt
- **Irrtumswahrscheinlichkeit** $\boldsymbol{\alpha}$**:** Wahrscheinlichkeit, dass das bestimmte Intervall den wahren Wert nicht überdeckt (übliche Werte: $\alpha = 0.10$, $\alpha = 0.05$, $\alpha = 0.01$)
- **Sicherheitswahrscheinlichkeit** oder **Vertrauenswahrscheinlichkeit:** $1-\alpha$
- Beispiel: $95\,\%$-KI: Intervall, das den wahren aber unbekannten Parameter mit einer Wahrscheinlichkeit von $95\,\%$ überdeckt

Konfidenzintervalle für μ

- gesucht: Konfidenzintervall für den unbekannten Erwartungswert μ einer normalverteilten Zufallsvariable X
- Fallunterscheidung:
  1. Fall: σ bekannt
  2. Fall: σ unbekannt

Konfidenzintervalle für μ (σ bekannt) (1)

- Seien $X_1, \ldots, X_n$ unabhängige Wiederholungen einer $N(\mu, \sigma)$-verteilten Zufallsvariable X.
- Dann gilt:
  - Die Schätzfunktion für μ ist das arithmetische Mittel $\overline{X}$.
  - Die Zufallsvariable $\overline{X}$ ist $N(\mu, \frac{\sigma}{\sqrt{n}})$-verteilt.
  - Durch Standardisierung von $\overline{X}$ ergibt sich

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

Konfidenzintervalle für μ (σ bekannt) (2)

- Der zweiseitig beschränkte Bereich, in dem $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ mit der Wahrscheinlichkeit $1-\alpha$ liegt, ist

$$P\left(-z_{1-\alpha/2} \leq \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{1-\alpha/2}\right) = 1-\alpha\,,$$

wobei $z_{1-\alpha/2}$ das $(1-\alpha/2)$-Quantil der Standardnormalverteilung bezeichnet.

- Wenn σ bekannt ist, erhält man $(1-\alpha)$-Konfidenzintervalle für μ bei normalverteiltem Merkmal X durch

$$\left[\overline{X} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right],$$

  wobei $z_{1-\alpha/2}$ das $(1-\alpha/2)$-Quantil der Standardnormalverteilung bezeichnet.

Bsp. Konfidenzintervalle für μ (σ bekannt)

- Ausgangspunkt: Der Schätzwert für den unbekannten Erwartungswert μ der Körpergröße X (in Zentimeter) weiblicher Studierender beträgt

$$\bar{x} = 169.5 = \tfrac{1}{20}\left(170 + 173 + \ldots + 171\right).$$

- Frage: Wie lautet das $95\,\%$-Konfidenzintervall für den unbekannten Erwartungswert μ?
- Annahme: Die Körpergröße X ist normalverteilt mit Erwartungswert μ und bekannter Standardabweichung $\sigma = 10\,$cm.

Tabelle der $N(0,1)$-Verteilung

A Standardnormalverteilung

Tabelliert sind die Werte der Verteilungsfunktion $\Phi(z) = P(Z \leq z)$ für $z \geq 0$.

Ablesebeispiel: $\Phi(1.75) = 0.9599$

Funktionswerte für negative Argumente: $\Phi(-z) = 1 - \Phi(z)$

Die z-Quantile ergeben sich genau umgekehrt. Beispielsweise ist $z(0.9599) = 1.75$ und $z(0.9750) = 1.96$.

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

Abb.: $N(0,1)$-Verteilung (Aus: Fahrmeir u. a. 2016)

- Seien $X_1, \ldots, X_n$ unabhängige Wiederholungen einer $N(\mu, \sigma)$-verteilten Zufallsvariable X.
- Wenn σ unbekannt ist, lässt $\overline{X}$ sich standardisieren:

$$\frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \quad \text{mit} \quad S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2}$$

Diese standardisierte Zufallsvariable folgt einer t-Verteilung mit $n-1$ Freiheitsgraden.

- Wenn σ unbekannt ist, ergeben sich $(1-\alpha)$-Konfidenzintervalle für μ bei normalverteiltem Merkmal X durch

$$\left[\overline{X} - t_{1-\alpha/2}(n-1) \cdot \frac{S}{\sqrt{n}}, \overline{X} + t_{1-\alpha/2}(n-1) \cdot \frac{S}{\sqrt{n}}\right],$$

wobei $t_{1-\alpha/2}(n-1)$ das $(1-\alpha/2)$-Quantil der t-Verteilung mit $n-1$ Freiheitsgraden bezeichnet.

Bsp. Konfidenzinterv. für μ (σ unbek.)

- Ausgangspunkt: Der Schätzwert für den unbekannten Erwartungswert μ der Körpergröße X (in Zentimeter) männlicher Studierender beträgt

$$\bar{x} = 183.0 = \frac{1}{21}(190 + 170 + \ldots + 181) .$$

- Frage: Wie lautet das 95 %-Konfidenzintervall für den unbekannten Erwartungswert μ?
- Annahme: Die Körpergröße X ist normalverteilt mit Erwartungswert μ und unbekannter Standardabweichung σ.
- Stichprobenstandardabweichung (als Schätzwert für σ):

$$s = 7.7 = \sqrt{\frac{1}{20}\left((190 - 183.0)^2 + \ldots + (181 - 183.0)^2\right)}$$

t-Verteilung (1)

Tabelliert sind die Quantile für n Freiheitsgrade. Für das Quantil $t_{1-\alpha}(n)$ gilt $F(t_{1-\alpha}(n)) = 1-\alpha$. Links vom Quantil $t_{1-\alpha}(n)$ liegt die Wahrscheinlichkeitsmasse $1-\alpha$.

Ablesebeispiel: $t_{0.99}(20) = 2.528$

Die Quantile für $0 < 1-\alpha < 0.5$ erhält man aus $t_\alpha(n) = -t_{1-\alpha}(n)$

Approximation für $n > 30$: $t_\alpha(n) \approx z_\alpha$ (z_α ist das (α)-Quantil der Standardnormalverteilung)

n	0.6	0.8	0.9	0.95	0.975	0.99	0.995	0.999	0.9995
1	0.3249	1.3764	3.0777	6.3138	12.706	31.821	63.657	318.31	636.62
2	0.2887	1.0607	1.8856	2.9200	4.3027	6.9646	9.9248	22.327	31.599
3	0.2767	0.9785	1.6377	2.3534	3.1824	4.5407	5.8409	10.215	12.924
4	0.2707	0.9410	1.5332	2.1318	2.7764	3.7469	4.6041	7.1732	8.6103
5	0.2672	0.9195	1.4759	2.0150	2.5706	3.3649	4.0321	5.8934	6.8688
6	0.2648	0.9057	1.4398	1.9432	2.4469	3.1427	3.7074	5.2076	5.9588
7	0.2632	0.8960	1.4149	1.8946	2.3646	2.9980	3.4995	4.7853	5.4079
8	0.2619	0.8889	1.3968	1.8595	2.3060	2.8965	3.3554	4.5008	5.0413
9	0.2610	0.8834	1.3830	1.8331	2.2622	2.8214	3.2498	4.2968	4.7809
10	0.2602	0.8791	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437	4.5869
11	0.2596	0.8755	1.3634	1.7959	2.2010	2.7181	3.1058	4.0247	4.4370
12	0.2590	0.8726	1.3562	1.7823	2.1788	2.6810	3.0545	3.9296	4.3178
13	0.2586	0.8702	1.3502	1.7709	2.1604	2.6503	3.0123	3.8520	4.2208
14	0.2582	0.8681	1.3450	1.7613	2.1448	2.6245	2.9768	3.7874	4.1405
15	0.2579	0.8662	1.3406	1.7531	2.1314	2.6025	2.9467	3.7328	4.0728

n	0.6	0.8	0.9	0.95	0.975	0.99	0.995	0.999	0.9995
1	0.3249	1.3764	3.0777	6.3138	12.706	31.821	63.657	318.31	636.62
2	0.2887	1.0607	1.8856	2.9200	4.3027	6.9646	9.9248	22.327	31.599
3	0.2767	0.9785	1.6377	2.3534	3.1824	4.5407	5.8409	10.215	12.924
4	0.2707	0.9410	1.5332	2.1318	2.7764	3.7469	4.6041	7.1732	8.6103
5	0.2672	0.9195	1.4759	2.0150	2.5706	3.3649	4.0321	5.8934	6.8688
6	0.2648	0.9057	1.4398	1.9432	2.4469	3.1427	3.7074	5.2076	5.9588
7	0.2632	0.8960	1.4149	1.8946	2.3646	2.9980	3.4995	4.7853	5.4079
8	0.2619	0.8889	1.3968	1.8595	2.3060	2.8965	3.3554	4.5008	5.0413
9	0.2610	0.8834	1.3830	1.8331	2.2622	2.8214	3.2498	4.2968	4.7809
10	0.2602	0.8791	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437	4.5869
11	0.2596	0.8755	1.3634	1.7959	2.2010	2.7181	3.1058	4.0247	4.4370
12	0.2590	0.8726	1.3562	1.7823	2.1788	2.6810	3.0545	3.9296	4.3178
13	0.2586	0.8702	1.3502	1.7709	2.1604	2.6503	3.0123	3.8520	4.2208
14	0.2582	0.8681	1.3450	1.7613	2.1448	2.6245	2.9768	3.7874	4.1405
15	0.2579	0.8662	1.3406	1.7531	2.1314	2.6025	2.9467	3.7328	4.0728
16	0.2576	0.8647	1.3368	1.7459	2.1199	2.5835	2.9208	3.6862	4.0150
17	0.2573	0.8633	1.3334	1.7396	2.1098	2.5669	2.8982	3.6458	3.9651
18	0.2571	0.8620	1.3304	1.7341	2.1009	2.5524	2.8784	3.6105	3.9216
19	0.2569	0.8610	1.3277	1.7291	2.0930	2.5395	2.8609	3.5794	3.8834
20	0.2567	0.8600	1.3253	1.7247	2.0860	2.5280	2.8453	3.5518	3.8495
21	0.2566	0.8591	1.3232	1.7207	2.0796	2.5176	2.8314	3.5272	3.8193
22	0.2564	0.8583	1.3212	1.7171	2.0739	2.5083	2.8188	3.5050	3.7921
23	0.2563	0.8575	1.3195	1.7139	2.0687	2.4999	2.8073	3.4850	3.7676
24	0.2562	0.8569	1.3178	1.7109	2.0639	2.4922	2.7969	3.4668	3.7454
25	0.2561	0.8562	1.3163	1.7081	2.0595	2.4851	2.7874	3.4502	3.7251
26	0.2560	0.8557	1.3150	1.7056	2.0555	2.4786	2.7787	3.4350	3.7066
27	0.2559	0.8551	1.3137	1.7033	2.0518	2.4727	2.7707	3.4210	3.6896
28	0.2558	0.8546	1.3125	1.7011	2.0484	2.4671	2.7633	3.4082	3.6739
29	0.2557	0.8542	1.3114	1.6991	2.0452	2.4620	2.7564	3.3962	3.6594
30	0.2556	0.8538	1.3104	1.6973	2.0423	2.4573	2.7500	3.3852	3.6460
∞	0.2533	0.8416	1.2816	1.6449	1.9600	2.3263	2.5758	3.0903	3.2906

Konfidenzintervalle für μ $(n > 30)$

- approximative $(1-\alpha)$-Konfidenzintervalle für den Erwartungswert μ für ein beliebig verteiltes Merkmal X, aber großen Stichprobenumfang ($n > 30$)
  - Wenn σ bekannt ist, gilt

$$\left[\bar{X} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right] .$$

  - Wenn σ unbekannt ist, gilt

$$\left[\bar{X} - z_{1-\alpha/2} \cdot \frac{S}{\sqrt{n}}, \bar{X} + z_{1-\alpha/2} \cdot \frac{S}{\sqrt{n}}\right] .$$

Breite des Konfidenzintervalls

- Sicherheitswahrscheinlichkeit $1 - \alpha$: je höher die Sicherheitswahrscheinlichkeit, desto größer die Breite des Konfidenzintervalls
- Standardabweichung σ: je homogener die Grundgesamtheit, desto kleiner die Breite des Konfidenzintervalls
- Stichprobenumfang n: je höher der Stichprobenumfang, desto kleiner die Breite des Konfidenzintervalls

Weitere Konfidenzintervalle

- Konfidenzintervalle für
  - σ bzw. σ^2
  - den Korrelationskoeffizienten
  - den Erwartungswert einer Exponentialverteilung
  - Parameter einer linearen Regression
  - $\ldots$