Schwankungsintervalle für $\overline{X}$

- *Eine* Verwendungsmöglichkeit für Verteilung von $\overline{X}$:

  Berechnung von (festen) Intervallen mit der Eigenschaft, dass die Stichprobenziehung mit einer vorgegebenen Wahrscheinlichkeit zu einer Realisation von $\overline{X}$ führt, die in dieses berechnete Intervall fällt.

  Solche Intervalle heißen **Schwankungsintervalle**.
- Gesucht sind also Intervallgrenzen $g_u < g_o$ von Intervallen $[g_u, g_o]$ mit

$$P_{\overline{X}}([g_u, g_o]) = P\{\overline{X} \in [g_u, g_o]\} \overset{!}{=} p_S$$

  für eine vorgegebene Wahrscheinlichkeit $p_S \in (0,1)$.
- Aus bestimmten Gründen (die später verständlich werden) gibt man nicht p_S vor, sondern die Gegenwahrscheinlichkeit $\alpha := 1 - p_S$, d.h. man fordert

$$P_{\overline{X}}([g_u, g_o]) = P\{\overline{X} \in [g_u, g_o]\} \overset{!}{=} 1 - \alpha$$

  für ein vorgegebenes $\alpha \in (0,1)$.
  $1 - \alpha$ wird dann auch **Sicherheitswahrscheinlichkeit** genannt.

- Eindeutigkeit für die Bestimmung von g_u und g_o erreicht man durch die Forderung von **Symmetrie** in dem Sinn, dass die untere bzw. obere Grenze des Intervalls jeweils mit einer Wahrscheinlichkeit von $\alpha/2$ unter- bzw. überschritten werden soll, d.h. man fordert genauer

$$P\{\overline{X} < g_u\} \overset{!}{=} \frac{\alpha}{2} \qquad \text{und} \qquad P\{\overline{X} > g_o\} \overset{!}{=} \frac{\alpha}{2} .$$

- Unter Verwendung der Verteilungseigenschaft

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

  erhält man also exakt bzw. näherungsweise

$$\begin{aligned} P\{\overline{X} < g_u\} &= P\left\{\frac{\overline{X} - \mu}{\sigma}\sqrt{n} < \frac{g_u - \mu}{\sigma}\sqrt{n}\right\} \overset{!}{=} \frac{\alpha}{2} \\ \Leftrightarrow \quad \frac{g_u - \mu}{\sigma}\sqrt{n} &= \Phi^{-1}\left(\frac{\alpha}{2}\right) \\ \Rightarrow \quad g_u &= \mu + \frac{\sigma}{\sqrt{n}} \cdot \Phi^{-1}\left(\frac{\alpha}{2}\right) \end{aligned}$$

  als untere Intervallgrenze.

- Analog erhält man exakt bzw. näherungsweise

$$\begin{aligned} P\{\overline{X} > g_o\} &= P\left\{\frac{\overline{X} - \mu}{\sigma}\sqrt{n} > \frac{g_o - \mu}{\sigma}\sqrt{n}\right\} \overset{!}{=} \frac{\alpha}{2} \\ \Leftrightarrow \quad \frac{g_o - \mu}{\sigma}\sqrt{n} &= \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \\ \Rightarrow \quad g_o &= \mu + \frac{\sigma}{\sqrt{n}} \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) . \end{aligned}$$

  als die obere Intervallgrenze.
- Als Abkürzung für p-Quantile der Standardnormalverteilung (also Funktionswerte von Φ^{-1} an der Stelle $p \in (0,1)$) verwenden wir:

$$N_p := \Phi^{-1}(p)$$

- Man erhält also insgesamt als symmetrisches Schwankungsintervall für $\overline{X}$ exakt bzw. näherungsweise das Intervall

$$\left[\mu + \frac{\sigma}{\sqrt{n}} \cdot N_{\frac{\alpha}{2}}, \mu + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right] .$$

Bemerkungen

- Die bekannte Symmetrieeigenschaft

$$\Phi(x) = 1 - \Phi(-x) \qquad \text{bzw.} \qquad \Phi(-x) = 1 - \Phi(x)$$

  für alle $x \in \mathbb{R}$ überträgt sich auf die Quantile N_p der Standardnormalverteilung in der Form

$$N_p = -N_{1-p} \qquad \text{bzw.} \qquad N_{1-p} = -N_p$$

  für alle $p \in (0,1)$.
- Üblicherweise sind nur die Quantile für $p \geq \frac{1}{2}$ in Tabellen enthalten. Man schreibt daher das Schwankungsintervall meist in der Form

$$\left[\mu - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}, \mu + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right] .$$

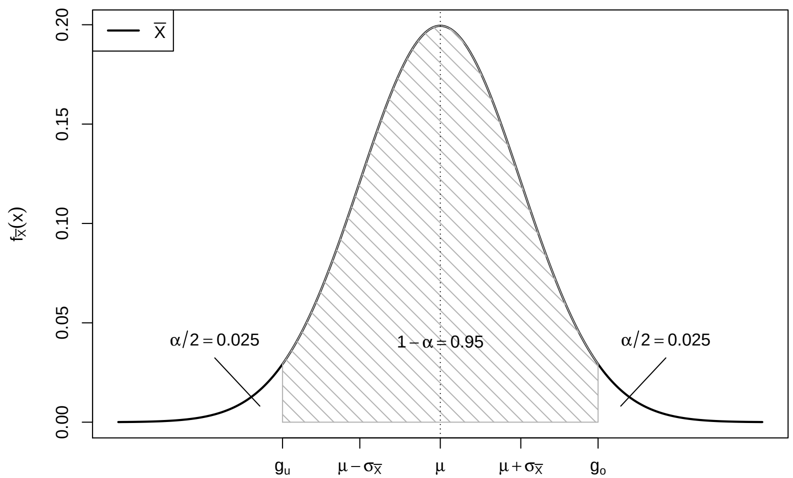
  In dieser Gestalt wird (noch klarer) deutlich, dass symmetrische Schwankungsintervalle für $\overline{X}$ ebenfalls (!) stets symmetrisch um μ sind.
- In der Literatur werden anstelle der Abkürzung N_p für die Quantile der Standardnormalverteilung häufig auch die Abkürzungen z_p oder λ_p verwendet.
- Geläufige Sicherheitswahrscheinlichkeiten sind z.B. $1 - \alpha \in \{0.90, 0.95, 0.99\}$.

Beispiel: Schwankungsintervall

- *Aufgabenstellung:*
  - Es gelte $Y \sim N(50, 10^2)$.
  - Zu Y liege eine einfache Stichprobe $X_1, \ldots, X_{25}$ der Länge $n = 25$ vor.
  - Gesucht ist ein (symmetrisches) Schwankungsintervall für $\overline{X}$ zur Sicherheitswahrscheinlichkeit $1 - \alpha = 0.95$.
- *Lösung:*
  - Es gilt also $\mu := E(Y) = 50$, $\sigma^2 := \text{Var}(Y) = 10^2$, $n = 25$ und $\alpha = 0.05$.
  - Zur Berechnung des Schwankungsintervalls
$$\left[\mu - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}, \mu + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right]$$
benötigt man also nur noch das $1 - \alpha/2 = 0.975$-Quantil $N_{0.975}$ der Standardnormalverteilung. Dies erhält man mit geeigneter Software (oder aus geeigneten Tabellen) als $N_{0.975} = 1.96$.
  - Insgesamt erhält man also das Schwankungsintervall
$$\left[50 - \frac{10}{\sqrt{25}} \cdot 1.96, 50 + \frac{10}{\sqrt{25}} \cdot 1.96\right] = [46.08, 53.92] \ .$$
  - Die Ziehung einer Stichprobenrealisation führt also mit einer Wahrscheinlichkeit von 95% zu einer Realisation $\overline{x}$ von $\overline{X}$ im Intervall $[46.08, 53.92]$.

Beispiel: Schwankungsintervall (Grafische Darstellung)

Im Beispiel: $\overline{X} \sim N\left(50, \frac{10^2}{25}\right)$



Konfidenzintervalle

- Schwankungsintervalle für $\overline{X}$ zu gegebenem Erwartungswert μ und gegebener Varianz σ^2 von Y eher theoretisch interessant.
- In praktischen Anwendungen der schließenden Statistik: μ (und eventuell auch σ^2) unbekannt!
- Ziel ist es, über die (bereits diskutierte) Parameterpunktschätzung durch $\overline{X}$ hinaus *mit Hilfe der Verteilung von* $\overline{X}$ eine Intervallschätzung von μ zu konstruieren, die bereits Information über die Güte der Schätzung enthält.
- Ansatz zur Konstruktion dieser Intervallschätzer ähnlich zum Ansatz bei der Konstruktion von (symmetrischen) Schwankungsintervallen.
- Idee: Verwende die Kenntnis der Verteilung von $\overline{X}$ (abhängig vom unbekannten μ), um zufällige (von der Stichprobenrealisation abhängige) Intervalle zu konstruieren, die den wahren Erwartungswert μ mit einer vorgegebenen Wahrscheinlichkeit überdecken.
- Konfidenzintervalle nicht nur für den Erwartungswert μ einer Verteilung möglich; hier allerdings Beschränkung auf Konfidenzintervalle für μ.

Konfidenzintervalle für μ bei bekannter Varianz σ^2

- Für die (festen!) Schwankungsintervalle $\left[\mu - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}, \mu + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right]$ für $\overline{X}$ zur Sicherheitswahrscheinlichkeit $1 - \alpha$ auf Grundlage der exakten oder näherungsweise verwendeten Standardnormalverteilung der Größe $\frac{\overline{X}-\mu}{\sigma}\sqrt{n}$ gilt nach Konstruktion
$$P\left\{\overline{X} \in \left[\mu - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}, \mu + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right]\right\} = 1 - \alpha \ .$$
- Idee: Auflösen dieser Wahrscheinlichkeitsaussage nach μ, das heißt, Suche von **zufälligen** Intervallgrenzen $\mu_u < \mu_o$ mit der Eigenschaft
$$P\left\{\mu \in [\mu_u, \mu_o]\right\} = P\left\{\mu_u \le \mu \le \mu_o\right\} \stackrel{!}{=} 1 - \alpha \ .$$
(bzw. genauer $P\{\mu < \mu_u\} \stackrel{!}{=} \frac{\alpha}{2}$ und $P\{\mu > \mu_o\} \stackrel{!}{=} \frac{\alpha}{2}$).
- Solche Intervalle $[\mu_u, \mu_o]$ nennt man dann **(zweiseitige) Konfidenzintervalle für μ zum Konfidenzniveau (zur Vertrauenswahrscheinlichkeit)** $1 - \alpha$.

Man erhält

$$
\begin{aligned}
& P\left\{\overline{X} \in \left[\mu - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}, \mu + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right]\right\} = 1-\alpha \\
\Leftrightarrow\quad & P\left\{\mu - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}} \leq \overline{X} \leq \mu + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right\} = 1-\alpha \\
\Leftrightarrow\quad & P\left\{-\overline{X} - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}} \leq -\mu \leq -\overline{X} + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right\} = 1-\alpha \\
\Leftrightarrow\quad & P\left\{\overline{X} + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}} \geq \mu \geq \overline{X} - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right\} = 1-\alpha \\
\Leftrightarrow\quad & P\left\{\overline{X} - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}} \leq \mu \leq \overline{X} + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right\} = 1-\alpha \\
\Leftrightarrow\quad & P\left\{\mu \in \left[\overline{X} - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}, \overline{X} + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right]\right\} = 1-\alpha
\end{aligned}
$$

und damit das Konfidenzintervall

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

zum Konfidenzniveau $1-\alpha$ für μ.

- In der resultierenden Wahrscheinlichkeitsaussage

$$P\left\{\overline{X} - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}} \leq \mu \leq \overline{X} + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right\} = 1-\alpha$$

  sind die **Intervallgrenzen**

$$\mu_u = \overline{X} - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}} \qquad \text{und} \qquad \mu_o = \overline{X} + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}$$

  des Konfidenzintervalls **zufällig** (nicht etwa μ!).
- Ziehung einer Stichprobenrealisation liefert also Realisationen der Intervallgrenzen und damit ein konkretes Konfidenzintervall, welches den wahren (unbekannten) Erwartungswert μ entweder überdeckt oder nicht.
- Die Wahrscheinlichkeitsaussage für Konfidenzintervalle zum Konfidenzniveau $1-\alpha$ ist also so zu verstehen, dass man bei der Ziehung der Stichprobe mit einer Wahrscheinlichkeit von $1-\alpha$ ein Stichprobenergebnis erhält, welches zu einem realisierten Konfidenzintervall führt, das den wahren Erwartungswert überdeckt.

Beispiel: Konfidenzintervall bei bekanntem σ^2

- Die Zufallsvariable Y sei normalverteilt mit unbekanntem Erwartungswert und bekannter Varianz $\sigma^2 = 2^2$.
- Gesucht: Konfidenzintervall für μ zum Konfidenzniveau $1-\alpha = 0.99$.
- Als Realisation $x_1, \ldots, x_{16}$ einer einfachen Stichprobe $X_1, \ldots, X_{16}$ vom Umfang $n = 16$ zu Y liefere die Stichprobenziehung

  18.75, 20.37, 18.33, 23.19, 20.66, 18.36, 20.97, 21.48, 21.15, 19.39, 23.02, 20.78, 18.76, 15.57, 22.25, 19.91 ,

  was zur Realisationen $\overline{x} = 20.184$ von $\overline{X}$ führt.
- Als Realisation des Konfidenzintervalls für μ zum Konfidenzniveau $1-\alpha = 0.99$ erhält man damit insgesamt

$$
\begin{aligned}
& \left[\overline{x} - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}, \overline{x} + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right] \\
= & \left[20.184 - \frac{2}{\sqrt{16}} \cdot 2.576, 20.184 + \frac{2}{\sqrt{16}} \cdot 2.576\right] \\
= & [18.896, 21.472] \ .
\end{aligned}
$$

Verteilung von $\overline{X}$ bei unbekanntem σ^2

- Wie kann man vorgehen, falls die Varianz σ^2 von Y unbekannt ist?
- Naheliegender Ansatz: Ersetzen von σ^2 durch eine geeignete Schätzfunktion.
- Erwartungstreue Schätzfunktion für σ^2 bereits bekannt:

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

- Ersetzen von σ durch $S = \sqrt{S^2}$ möglich, Verteilung ändert sich aber:

Satz 5.1

Seien $Y \sim N(\mu, \sigma^2)$, $X_1, \ldots, X_n$ eine einfache Stichprobe zu Y. Dann gilt mit $S := \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2} = \sqrt{\frac{n}{n-1}(\overline{X^2} - \overline{X}^2)}$

$$\frac{\overline{X} - \mu}{S}\sqrt{n} \sim t(n-1) \ ,$$

*wobei $t(n-1)$ die t-**Verteilung mit** $n-1$ **Freiheitsgraden** bezeichnet.*

Die Familie der $t(n)$-Verteilungen

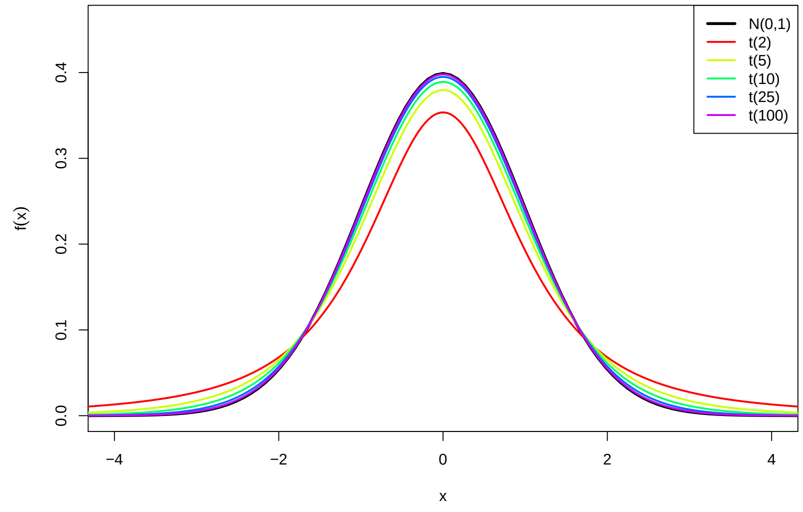
- Die Familie der $t(n)$-Verteilungen mit $n > 0$ ist eine spezielle Familie stetiger Verteilungen. Der Parameter n wird meist „Anzahl der Freiheitsgrade" („degrees of freedom") genannt.
- t-Verteilungen werden (vor allem in englischsprachiger Literatur) oft auch als „Student's t distribution" bezeichnet; „Student" war das Pseudonym, unter dem William Gosset die erste Arbeit zur t-Verteilung in englischer Sprache veröffentlichte.
- $t(n)$-Verteilungen sind für alle $n > 0$ symmetrisch um 0. Entsprechend gilt für p-Quantile der $t(n)$-Verteilung, die wir im Folgendem mit $t_{n;p}$ abkürzen, analog zu Standardnormalverteilungsquantilen

$$t_{n;p} = -t_{n;1-p} \qquad \text{bzw.} \qquad t_{n;1-p} = -t_{n;p}$$

  für alle $p \in (0, 1)$
- Für wachsendes n nähert sich die $t(n)$-Verteilung der Standardnormalverteilung an.

Grafische Darstellung einiger $t(n)$-Verteilungen

für $n \in \{2, 5, 10, 25, 100\}$



- Konstruktion von Konfidenzintervallen für μ bei unbekannter Varianz $\sigma^2 = \text{Var}(Y)$ ganz analog zur Situation mit bekannter Varianz, lediglich
  1. Ersetzen von σ durch $S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$
  2. Ersetzen von $N_{1-\frac{\alpha}{2}}$ durch $t_{n-1;1-\frac{\alpha}{2}}$

  erforderlich.
- Resultierendes Konfidenzintervall:

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

- Benötigte Quantile $t_{n-1;1-\frac{\alpha}{2}}$ können ähnlich wie bei der Standardnormalverteilung z.B. mit der Statistik-Software **R** ausgerechnet werden oder aus geeigneten Tabellen abgelesen werden.
- Mit **R** erhält man z.B. $t_{15;0.975}$ durch

```
> qt(0.975,15)
[1] 2.13145
```

- Mit zunehmendem n werden die Quantile der $t(n)$-Verteilungen betragsmäßig kleiner und nähern sich den Quantilen der Standardnormalverteilung an.

Quantile der t-Verteilungen: $t_{n;p}$

$n\backslash p$	0.85	0.90	0.95	0.975	0.99	0.995	0.9995
1	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	1.386	1.886	2.920	4.303	6.965	9.925	31.599
3	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	1.074	1.341	1.753	2.131	2.602	2.947	4.073
20	1.064	1.325	1.725	2.086	2.528	2.845	3.850
25	1.058	1.316	1.708	2.060	2.485	2.787	3.725
30	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	1.050	1.303	1.684	2.021	2.423	2.704	3.551
50	1.047	1.299	1.676	2.009	2.403	2.678	3.496
100	1.042	1.290	1.660	1.984	2.364	2.626	3.390
200	1.039	1.286	1.653	1.972	2.345	2.601	3.340
500	1.038	1.283	1.648	1.965	2.334	2.586	3.310
1000	1.037	1.282	1.646	1.962	2.330	2.581	3.300
5000	1.037	1.282	1.645	1.960	2.327	2.577	3.292

Beispiel: Konfidenzintervall bei unbekanntem σ^2

- Die Zufallsvariable Y sei normalverteilt mit unbekanntem Erwartungswert und unbekannter Varianz.
- Gesucht: Konfidenzintervall für μ zum Konfidenzniveau $1-\alpha = 0.95$.
- Als Realisation $x_1, \ldots, x_9$ einer einfachen Stichprobe $X_1, \ldots, X_9$ vom Umfang $n = 9$ zu Y liefere die Stichprobenziehung

  28.12, 30.55, 27.49, 34.79, 30.99, 27.54, 31.46, 32.21, 31.73 ,

  was zur Realisationen $\overline{x} = 30.542$ von $\overline{X}$ und zur Realisation $s = 2.436$ von $S = \sqrt{S^2}$ führt.
- Als Realisation des Konfidenzintervalls für μ zum Konfidenzniveau $1-\alpha = 0.95$ erhält man damit insgesamt

$$\begin{aligned}
&\left[\overline{x} - \frac{s}{\sqrt{n}} \cdot t_{n-1;1-\frac{\alpha}{2}}, \overline{x} + \frac{s}{\sqrt{n}} \cdot t_{n-1;1-\frac{\alpha}{2}}\right] \\
=&\left[30.542 - \frac{2.436}{\sqrt{9}} \cdot 2.306, 30.542 + \frac{2.436}{\sqrt{9}} \cdot 2.306\right] \\
=&\,[28.67, 32.414] \ .
\end{aligned}$$

Konfidenzintervalle, falls Y nicht normalverteilt

1. Ist Y nicht normalverteilt, aber die **Varianz** σ^2 von Y **bekannt**, so verwendet man wie bei der Berechnung der Schwankungsintervalle näherungsweise (durch den zentralen Grenzwertsatz gerechtfertigt!) die Standardnormalverteilung als Näherung der Verteilung von $\frac{\overline{X}-\mu}{\sigma}\sqrt{n}$ und erhält so **approximative (näherungsweise)** Konfidenzintervalle

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

   zum (Konfidenz-)Niveau $1-\alpha$.
2. Ist Y nicht normalverteilt und die **Varianz** von Y unbekannt, so verwendet man nun analog als Näherung der Verteilung von $\frac{\overline{X}-\mu}{S}\sqrt{n}$ die $t(n-1)$-Verteilung und erhält so **approximative (näherungsweise)** Konfidenzintervalle

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

   zum (Konfidenz-)Niveau $1-\alpha$.

Spezialfall: Konfidenzintervalle für p, falls $Y \sim B(1,p)$

- Gilt $Y \sim B(1,p)$ für einen unbekannten Parameter $p \in [0,1]$, so können Konfidenzintervalle wegen $p = E(Y) = \mu$ näherungsweise ebenfalls mit Hilfe der Näherung 2 aus Folie 88 bestimmt werden.
- In der „Formel“ für die Berechnung der Konfidenzintervalle ersetzt man üblicherweise $\overline{X}$ wieder durch die in dieser Situation geläufigere (gleichbedeutende!) Notation $\widehat{p}$.
- Die (notwendige) Berechnung von $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$ gestaltet sich hier besonders einfach. Man kann zeigen, dass $S^2 = \frac{n}{n-1}\widehat{p}(1-\widehat{p})$ gilt.
- Man erhält so die *von der Stichprobe nur noch über* $\widehat{p}$ *abhängige* Darstellung

$$\left[\widehat{p} - \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n-1}} \cdot t_{n-1;1-\frac{\alpha}{2}}, \widehat{p} + \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n-1}} \cdot t_{n-1;1-\frac{\alpha}{2}}\right]$$

  für *approximative* Konfidenzintervalle für p zum Niveau $1-\alpha$.
- Die Güte der Näherung hängt von n und p ab. Je größer n, desto besser; je näher p an $\frac{1}{2}$, desto besser.