2 Vertrauensintervalle (Konfidenzintervalle)

Die bisherigen Punktschätzer sind zwar genau aber unsicher.
Die im Folgenden besprochenen Konfidenzintervalle (Vertrauensintervalle) machen dem gegenüber unscharfe Aussagen aber mit einer gewissen Verlässlichkeit.

Viele Vertrauensintervalle haben die Form:

$$\boldsymbol{Sch\ddot{a}tzung\ des\ Parameters}$$
$$\mp \boldsymbol{Vielfaches\ des\ gesch\ddot{a}tzten\ Standardabweichung\ des\ Parameters}$$

> Definition:
> Gilt für eine Zufallsvariable X für jeden Bereich B
> $P(X \in B) \geq 1 - \alpha$
> dann heisst B Konfidenzintervall (Vertrauensintervall, Prognosebereich) für X zum Niveau α.

Das Vielfache hängt dabei von der Verteilung und der Vertrauenswahrscheinlichkeit ab.

Beispiel:

Ist etwa die Zufallsvariable X etwa $N(5; 2^2)$ normalverteilt (kurz: $X \sim N(5; 2^2)$ dann ist $|X - 5| \leq 1.96 \cdot 2$ oder $1.08 \leq X \leq 8.92$ ein (0.95)-Prognoseintervall.

Dies bedeutet, dass ein Ergebnis (eine Realisation) von X mit einer Wahrscheinlichkeit 95% in diesem Intervall liegen wird. Die Prognose ist dann für die Beobachtung $X = 4$ richtig, für $X = 10$ falsch.

Beispiele für Vertrauensintervalle:

Gesucht ist ein Vertrauensintervall für den
Erwartungswert μ einer normalverteilten Zufallsvariablen

Fall 1:
Die Varianz σ^2 ist bekannt (selten!)

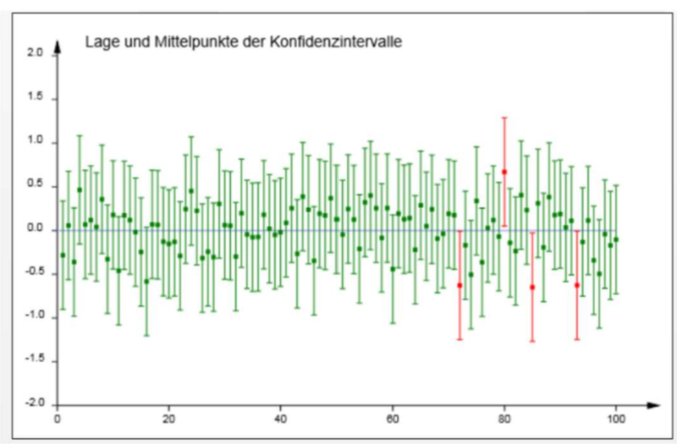
Der zweite Fall führt auf die im nächsten Abschnitt behandelte Student-t-Verteilung.

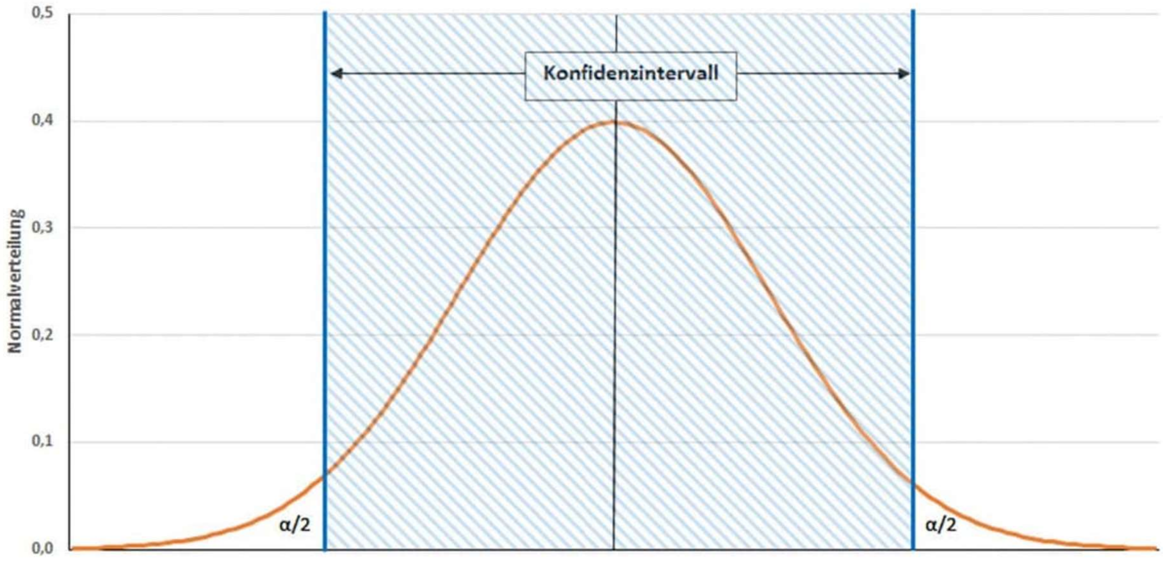
Die standardisierte Variable
$$\frac{\bar{X}-\mu}{\sigma}\cdot\sqrt{n}\sim N(0;1^2)$$
ist dann Standard-normalverteilt.
Gemäss den Eigenschaften von Zufallsvariablen ist nämlich die Varianz des Mittelwerts von n unabhängigen Messungen nicht σ^2 (wie bei Einzelmessungen) sondern nur noch $\frac{\sigma^2}{n}$.
Die Breite des Prognoseintervalles kann also mit einer Erhöhung der Anzahl Messungen verkleinert werden. Bei einer Verkleinerung des Stichprobenumfangs vergrössert sich der Standardfehler und damit die Länge des Vertrauensintervalls.

Das 95%-Vertrauensintervall für den unbekannten Mittelwert μ von Stichproben vom Umfang n zu

$$\left[\bar{X}-1.96\cdot\frac{\sigma}{\sqrt{n}},\bar{X}+1.96\cdot\frac{\sigma}{\sqrt{n}}\right]$$
Das Vertrauensintervall hat also die Länge
$$L=2\cdot z_{1-\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$$

Zur Illustration:
Mit standardnormalverteilten Zufallsvariablen wurden Stichproben vom Umfang $n = 10$ erzeugt und für jede der Mittelwert bestimmt. Dieser Vorgang wurde $k = 100$ mal ausgeführt. So entstanden die 100 abgebildeten Vertrauensintervalle der Länge
$$L=2\cdot\frac{1.96}{\sqrt{10}}\approx 1.24$$
Im Beispiel enthalten die vier rot dargestellten Vertrauensintervalle den (ausnahmsweise) bekannten Mittelwert $\mu = 0$ nicht.





$\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ $\qquad$ $\bar{x}$ $\qquad$ $\bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

Beispiel:

Eine Käserei kauft eine neue Abfüllmaschine, Damit eine Käserei ein eigenes Joghurt herstellen kann, schafft eine neue Abfüllmaschine an. Aus zahlreichen Versuche ist bekannt, dass die Gewichte der Joghurts normalverteilt sind mit dem Mittelwert $\mu = 500.00$ g und der Standardabweichung $\sigma = 2.00$ g. Gesucht ist ein 95%-Vertrauensintervall für den Mittelwert der Joghurtgewichte.
Zur Kontrolle kauft die Käserei 3 zufällig ausgewählte Joghurt und berechnet den Mittelwert der Gewichte. Liegt das mittlere Gewicht 498.0 noch im Vertrauensintervall?

$$\left[\boldsymbol{\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}}\right]$$

Vertrauensintervall				
	95%			
Faktor	1.96			
µ	σ	n		
500.00	2.00	3		
Vertauensintervall		**497.74**	**502.26**	

498.0 liegt zwar noch im Intervall. Angesichts der kleinen Stichprobe ist das Resultat aber wenig aussagekräftig.

Vertrauensintervall für die Binomialverteilung

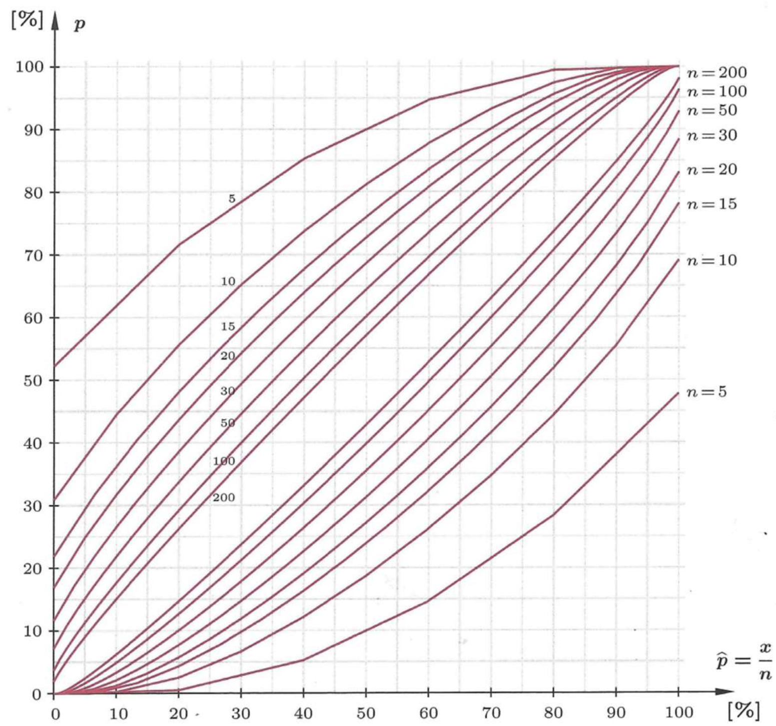
Ist eine Zufallsvariable X binomialverteilt $B(n, p)$, dann gilt bekanntlich für die Schätzer des Erwartungswerts und der Varianz:

Erwartungswert $E(X) = \hat{p} = \frac{x}{n}$ und

Varianz $V(X) = \frac{\hat{p}\cdot(1-\hat{p})}{n}$

Vgl. dazu die Abbildung der Konfidenzellipse in der DMK-Formelsammlung

Tests und Vertrauensintervalle bei der Binomialverteilung



Kurven zur genäherten Bestimmung von Verwerfungsbereichen und Vertrauensintervallen bei der Binomialverteilung

Ablesebeispiele

Eine Münze wird $n = 30$-mal geworfen. Welche Ergebnisse führen zur Verwerfung der Nullhypothese $p = 0.5$ auf dem 5%-Niveau? Die horizontale Gerade $p = 50\%$ schneidet die beiden Kurven für $n = 30$ in den Punkten ($\frac{x}{n} = 0.305$, $p = 50\%$), bzw. ($\frac{x}{n} = 0.695$, $p = 50\%$).
Der Test auf dem 5%-Niveau verwirft daher die Nullhypothese, falls $x \leqslant 30 \cdot 0.305 \Leftrightarrow x \leqslant 9$ oder falls $x \geqslant 30 \cdot 0.695 \Leftrightarrow x \geqslant 21$.

Bei $n = 20$ Wiederholungen ist ein Ereignis $x = 5$-mal eingetreten. Man gebe ein Vertrauensintervall für die unbekannte Wahrscheinlichkeit des Ereignisses an. Die vertikale Gerade $\frac{x}{n} = 0.25$ schneidet die beiden Kurven für $n = 20$ in den Punkten ($\frac{x}{n} = 0.25$, $p = 8\%$) bzw. und ($\frac{x}{n} = 0.25$, $p = 49\%$). Daher ist das gesuchte Vertrauensintervall mit Vertrauenswahrscheinlichkeit 95% gleich (0.08, 0.49).

Für Stichprobengrössen, bei denen keine Kurven eingezeichnet sind, muss man zwischen den beiden nächstliegenden Kurven interpolieren.

Für grössere n (Faustregel $n > 30$ oder $\frac{9}{\hat{p}\cdot(1-\hat{p})}$) ist X angenähert normalverteilt und für das Vertrauensintervall ergibt sich damit

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

Beispiele

Eine Urne enthält schwarze und weisse Kugeln, deren Anteile unbekannt sind. Um den Anteil p an schwarzen Kugeln zu schätzen, werden der Urne 100 Kugeln mit Zurücklegen gezogen. Gesucht ist ein 95%-Vertrauensintervall für p, wenn beim Versuch 68 schwarze und 32 weisse gezogen werden.

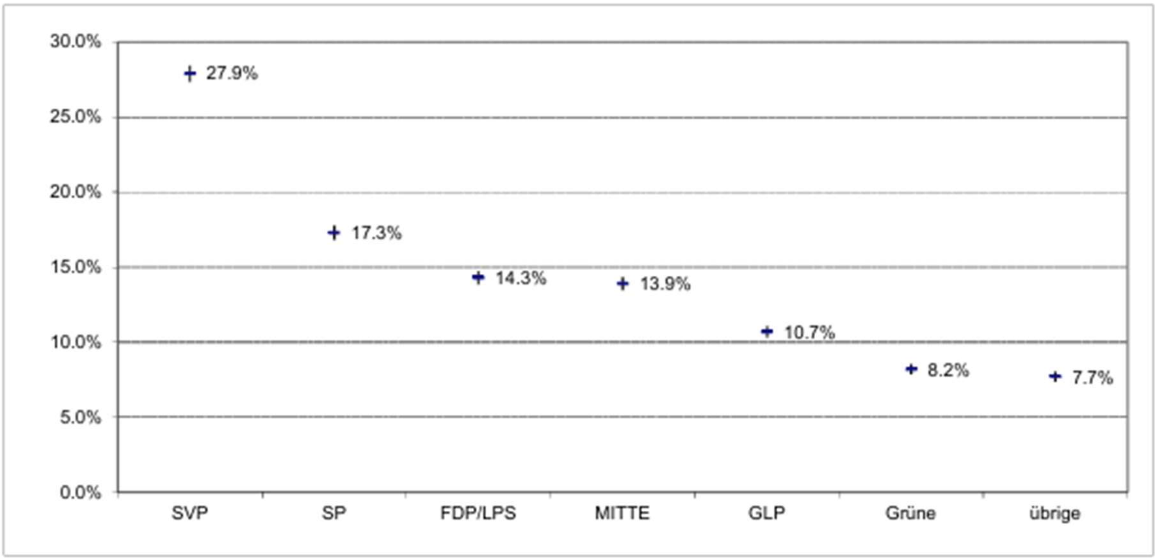
Urne

Farbe	schwarz	weiss	n
Anzahl gezogen	68	32	100
Vertrauensintervall		95%	
u		1.96	
x/n	0.68		
1-x/n	0.32		
d	0.0914		
95%Vertr.intervall		**0.589**	**0.771**

Umfragen vor Wahlen

Auf der folgenden Seite sind die Ergebnisse der letzten Nationalratswahlen 2019 zusammengestellt. Am 10./11. Juli 2023 wurde eine Umfrage über die nächsten Nationalratswahlen im Oktober 2023 durchgeführt. Die relativen Häufigkeiten ergeben Vertrauensintervalle für die Anteile der verschiedenen Parteien.
Zusätzlich sind noch die definitiven Parteistärken bei der Wahl 2023.10. (grün) angegeben.

Vertrauensintervalle		**Wähleranteile**		**Umfrage Nationalratswahlen 2023 11.07.2023**			**NR_Wahl 2023**
n	25688	x/n	1-x/n	d	**95%-Vertrauensintervall**		effektiv
SVP		**0.279**	0.721	0.0055	**27.4%**	**28.4%**	27.90%
SP		**0.173**	0.827	0.0046	**16.8%**	**17.8%**	18.30%
FDP/LPS		**0.143**	0.857	0.0043	**13.9%**	**14.7%**	14.30%
MITTE		**0.139**	0.861	0.0042	**13.5%**	**14.3%**	14.10%
GLP		**0.107**	0.893	0.0038	**10.3%**	**11.1%**	9.80%
Grüne		**0.082**	0.918	0.0034	**7.9%**	**8.5%**	7.60%
übrige		**0.077**	0.923	0.0033	**7.4%**	**8.0%**	8.10%



Liegt bei der nächsten Umfrage das Ergebnis einer Partei ausserhalb des 95%-Intervalls, dann sind zwei Interpretationen möglich:
a) die Umfrage zeigt bei dieser Partei ein ungewöhnliches (signifikantes) Ergebnis.
b) der Anteil der Partei hat sich verändert.

Wahrscheinlichkeit einer Mädchengeburt in der Schweiz
Quelle: Ineichen

Von den n = 2 644 757 Geburten in der Schweiz (1871 – 1900) waren k = 1 285 086 Mädchengeburten. Gesucht ist ein 95%-Vertrauensintervall für die (unbekannte) Wahrscheinlichkeit einer Mädchengeburt.

Die Anzahl der Mädchengeburten k ist eine binomialverteilte Zufallsvariable, die durch die Normalverteilung angenähert werden kann.

Im Geburtenbeispiel ergibt sich als 95%-Vertrauensintervall für die Wahrscheinlichkeit einer Mädchengeburt

$[0.4853, 0.4865]$
Zum Vergleich:

Im Jahr 2023 wurden bei insgesamt 80'024 Geburten 38976 Mädchen.

Damit ergibt sich als 95%-Vertrauensintervall: [0.4836, 0.4904]

Bemerkung:
Die Vertrauensintervalle weltweit unterscheiden sich auffällig

Ergänzung:

Um die unbekannte Wahrscheinlichkeit einer Konfidenzintervalls mit einer bestimmten Genauigkeit d zu schätzen ist ein Stichprobenumfang n nötig.

Für die Länge des Intervalls ε gilt:

$$d = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}$$

quadriert

$$d^2 = z_{1-\frac{\alpha}{2}}^2 \cdot \frac{\hat{p} \cdot (1-\hat{p})}{n}$$

und damit

$$n \geq z_{1-\frac{\alpha}{2}}^2 \cdot \frac{\hat{p} \cdot (1-\hat{p})}{d^2}$$

Wegen $\hat{p} \cdot (1-\hat{p}) \leq \frac{1}{4}$ gilt die ungenauere Abschätzung

$$n \geq \cdot \frac{z_{1-\frac{\alpha}{2}}^2}{4d^2}$$

Beispiele:

zulässiger Fehler d	0.05	0.02	0.01
notwendiger Stichprobenumfang	400	2500	10000

Übungsaufgabe

Die SVP hat bei den Nationalratswahlen 2019 einen Anteil von 25.6 % der Stimmenden erhalten. Wie gross ist der Stichprobenumfang zu wählen, damit der Wähleranteil mit 95% Sicherheit auf 1.5% genau geschätzt werden kann?

Fall 2:

Die Varianz σ^2 muss mit einer Stichprobe geschätzt werden.

Die t-Verteilung

Die t-Verteilung ersetzt die Normalverteilung, wenn die Varianz der Grundgesamtheit nicht bekannt ist und durch eine Schätzung aus einer Stichprobe ersetzt wird. Es ist plausibel, dass die geschätzte Varianz dieser Verteilung insbesondere bei kleinen Stichproben grösser als die der Normalverteilung ist.

Entdeckt hat diese Verteilung 1908 der Statistiker William S. Gosset (1876 – 1937), der als Biometriker in der Bierbrauerei Guiness arbeitete. Da er kein Geschäftsgeheimnis verraten durfte, publizierte er seine Arbeit unter dem Pseudonym «Student».
Die Dichte der nur vom Stichprobenumfang n abhängigen t-Verteilung ist

$$\varphi(t) = c \cdot \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}.$$

In der Abbildung ist zu erkennen, dass die t-Verteilung für kleine Stichproben von der Normalverteilung abweicht und sich bei grossem Stichprobenumfang asymptotisch der Normalverteilung nähert.

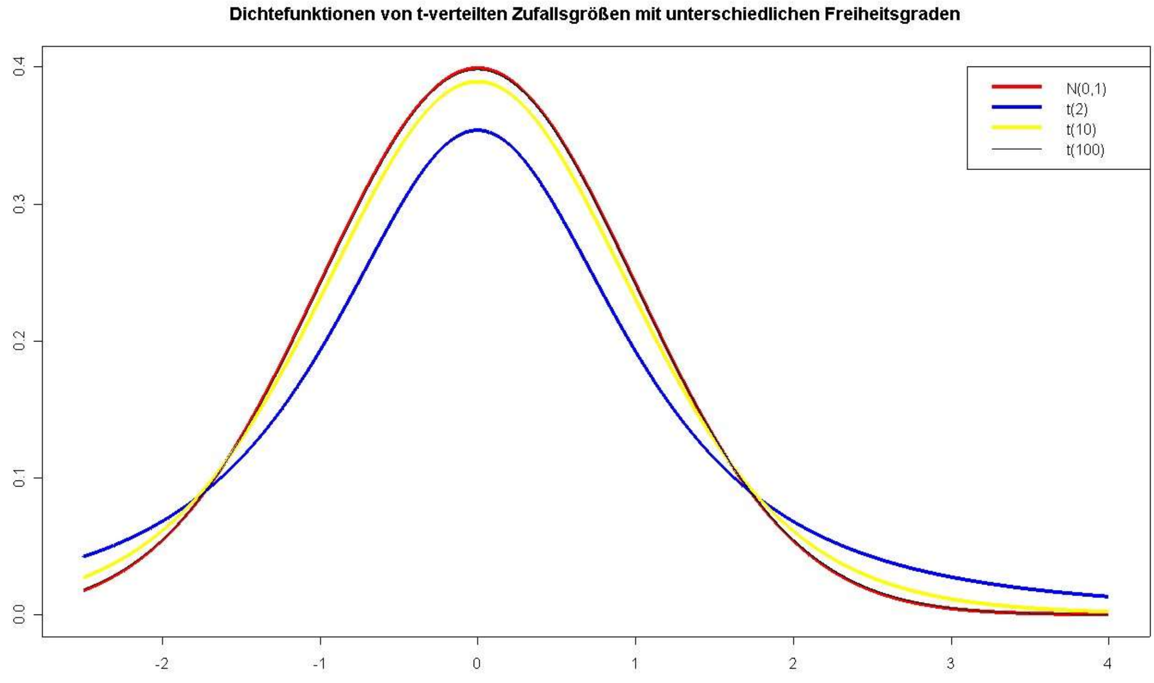


Tabelle der t-Verteilung

Anzahl k der Freiheitsgrade	Werte der Verteilungsfunktion 0,90	0,95	0,975	0,995	0,9995
1	3,078	6,314	12,71	63,66	636,6
2	1,886	2,920	4,303	9,925	31,59
3	1,638	2,353	3,182	5,841	12,82
4	1,533	2,132	2,776	4,604	8,610
5	1,476	2,015	2,571	4,032	6,869
6	1,440	1,943	2,447	3,707	5,959
7	1,415	1,895	2,365	3,499	5,408
8	1,397	1,860	2,306	3,355	5,041
9	1,383	1,833	2,262	3,250	4,781
10	1,372	1,812	2,228	3,169	4,587
11	1,363	1,796	2,201	3,106	4,437
12	1,356	1,782	2,179	3,055	4,318
13	1,350	1,771	2,160	3,012	4,221
14	1,345	1,761	2,145	2,977	4,140
15	1,341	1,753	2,131	2,947	4,073
16	1,337	1,746	2,120	2,921	4,015
17	1,333	1,740	2,110	2,898	3,965
18	1,330	1,734	2,101	2,878	3,922
19	1,328	1,729	2,093	2,861	3,883
20	1,325	1,725	2,086	2,845	3,850
21	1,323	1,721	2,080	2,831	3,819
22	1,321	1,717	2,074	2,819	3,792
23	1,319	1,714	2,069	2,807	3,768
24	1,318	1,711	2,064	2,797	3,745
25	1,316	1,708	2,060	2,787	3,725
26	1,315	1,706	2,056	2,779	3,707
27	1,314	1,703	2,052	2,771	3,690
28	1,313	1,701	2,048	2,763	3,674
29	1,311	1,699	2,045	2,756	3,659
30	1,310	1,697	2,042	2,750	3,546
40	1,303	1,684	2,021	2,704	3,551
60	1,296	1,671	2,000	2,660	3,460
120	1,289	1,658	1,980	2,617	3,373
∞	1,282	1,645	1,960	2,576	3,290

Aus der Prüfgrösse $\frac{\bar{X}-\mu}{\sigma}$ der Normalverteilung wird die Prüfgrösse $\frac{\bar{X}-\mu}{\hat{\sigma}}$ der t-Verteilung, indem man σ durch die Schätzung $\hat{\sigma}$ ersetzt.
Die standardisierte Variable

$$\frac{\bar{X}-\mu}{\hat{\sigma}} \cdot \sqrt{n}$$

hat dann eine t-Verteilung mit dem Parameter $n-1$ Diese Zahl heisst Freiheitsgrad ν.
Für den Erwartungswert μ erhält man das folgende beidseitige 95%-Vertrauensintervall:

$$\left[\bar{X} - \frac{\hat{\sigma}}{\sqrt{n}} \cdot t(n-1), \bar{X} + \frac{\hat{\sigma}}{\sqrt{n}} \cdot t(n-1)\right]$$

Beispiel:
Es wird angenommen, die Temperatur in einem Schmelzofen habe bei $n = 6$ Temperaturmessungen die Werte $\bar{X} = 1073$ und $\hat{\sigma} = 5$ ergeben. Gesucht ist ein 95%-Vertrauensintervall für den Erwartungswert μ.

Den $n = 6$ Messungen entsprechen $6 - 1 = 5$ Freiheitsgrade.
Da sich die 5% verteilen auf die beiden Grenzen des Vertrauensintervalls verteilen ist der t-Wert
$t(5)_{0.975} = 2.571$)
Für μ erhält man damit das folgende Vertrauensintervall

$$1073 - \frac{5}{\sqrt{6}} \cdot 2.571 \leq \mu \leq 1073 + \frac{5}{\sqrt{6}} \cdot 2.571$$

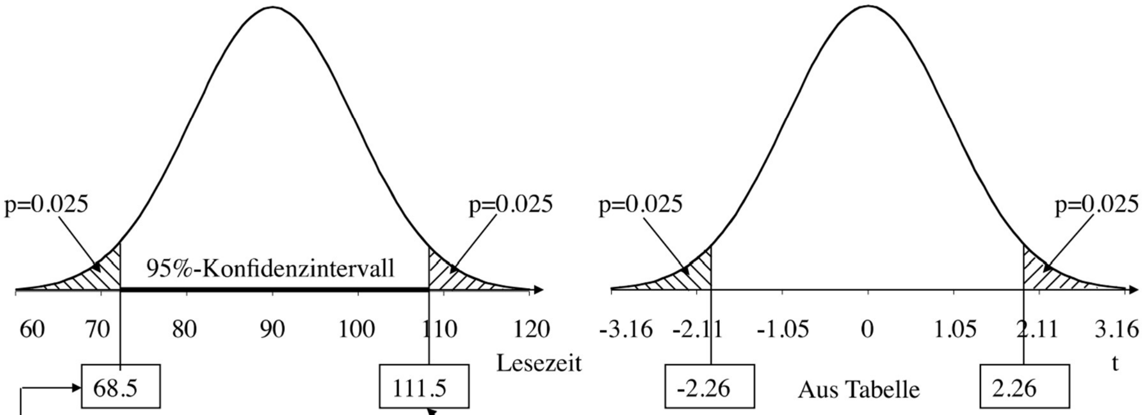
oder

$$1067.75 \leq \mu \leq 1078.25$$

Im Fall 1 (Die Varianz der Population bekannt) hätte sich der deutlich kleinere Wert 1.96 und damit ein kleineres Vertrauensintervall ergeben. Grund dafür ist die kleine Anzahl der Messungen für die Schätzung.

Zu Studienbeginn hat man eine Zufallsstichprobe von 10 Studenten eine Seite Standardprosatext lesen lassen. Die mittlere Lesezeit betrug 90 Sekunden. Die Standardabweichung wird anhand der Stichprobendaten geschätzt $\hat{\sigma} = 30$ Sekunden. Bestimmen Sie das 95%-Konfidenzintervall.

μ bekannt, σ muss geschätzt werden → t-Verteilung



$$\bar{x}_{(97.5\%)} = 90 + 2.26 \cdot \frac{30}{\sqrt{10}} = 111.4$$

$$\bar{x}_{(2.5\%)} = 90 - 2.26 \cdot \frac{30}{\sqrt{10}} = 68.6$$

$$\bar{x}_{(2.5\%;\,97.5\%)} = \bar{x} \pm z \cdot \sigma_{\bar{x}}$$

Lösung mit R:

```
90+qt(p=c(.025,.975), df=9)*30/sqrt(10)
```

91

Übungsaufgaben:
Schafmilch wird hin und wieder wegen ihres hohen Gehalts an Orotsäure als Allerweltsheilmittel beschrieben, obwohl wissenschaftliche Untersuchungen dazu fehlen. Der langjährige mittlere Gehalt in der Schweiz beträgt im Juni 1.87 mg/100 g Milch (Mittelwert der Population). Im Juni 2007 wurde bei zufällig ausgewählten Schafen der Orotsäuregehalt bestimmt:

t-Verteilung		Orotsäure			
			Stichprobenumfang		
1	1.72		Populationsmittelwert		1.870
2	1.97		Stichprobenmittelwert		2.123
3	2.03		Standardabweichung		0.280
4	2.14		Standardfehler		0.089
5	2.21		t-Wert		2.853
6	1.77		Freiheitsgrad n-1		9
7	2.62				
8	2.45				
9	2.27				
10	2.05				

Berechnung des t-Werts:

$$t = \frac{\bar{X} - \mu}{\hat{\sigma}} \cdot \sqrt{n} = \frac{0.2123 - 1.87}{0.28} \cdot \sqrt{10} \approx 2.85$$

Für μ erhält man damit das folgende Vertrauensintervall
$1.87 - 0.089 \cdot 2.85 \leq \mu \leq 1.87 + \cdot\, 0.089 \cdot 2.85$
oder
$1.616 \leq \mu \leq 2.124$
Da das Stichprobenergebnis am Rande des Vertrauensintervalls liegt und der Stichprobenumfang relativ klein ist, ist eine Interpretation schwierig.

In einem Mathematiktest mit Studierenden (MATH99) wurden die folgenden Masszahlen für die erreichte Punktezahl berechnet:

$$\bar{x} = 22.71, \quad \tilde{x} = 23, \quad s^2 = 10.74, \quad s = 3.28 \quad \text{und} \quad n = 87$$

Bei einer Irrtumswahrscheinlichkeit $\alpha = 0.05$ ist das für die Intervallschätzung des Erwartungswertes zu verwendende Quantil der t-Verteilung gegeben als $t_{1-\alpha/2}(n-1) = t_{0.975}(86) = 1.988$ und die Berechnung des $(1-\alpha)$-Konfidenzintervalls kann erfolgen als

$$\left[22.71 - 1.988\frac{3.28}{\sqrt{87}} = 22.0, 22.71 + 1.988\frac{3.28}{\sqrt{87}} = 23.4\right].$$

Der Erwartungswert der Punktezahl liegt also mit einer Wahrscheinlichkeit von 0.95 zwischen 22.0 und 23.4 Punkten, also $P(22.0 \leq \mu \leq 23.4) = 0.95$. Es bleibt ein Restrisiko von $\alpha = 0.05$ bestehen, dass der Parameter μ ausserhalb dieses Intervalls liegt.[a]