Grundlagen der Mathematischen Statistik

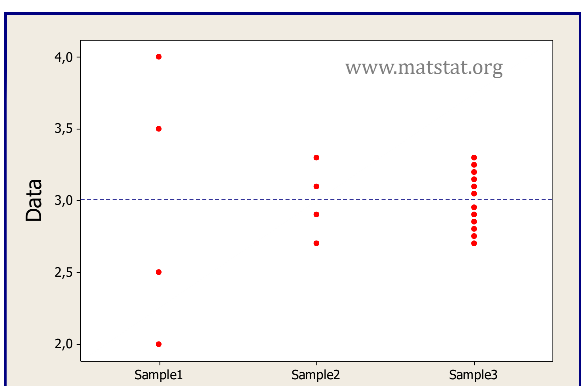
Intervallschätzer

Uwe Menzel, 2018

uwe.menzel@matstat.org

www.matstat.org

Punktschätzung reicht oft nicht



$\mu^* = 3$ $\quad d = \frac{s}{\sqrt{n}} = 0.456$

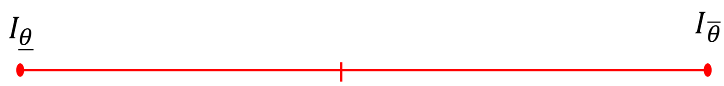
$\mu^* = 3$ $\quad d = \frac{s}{\sqrt{n}} = 0.129$

$\mu^* = 3$ $\quad d = \frac{s}{\sqrt{n}} = 0.0587$

Alle drei Stichproben ergeben dieselbe Punktschätzung für den Mittelwert ($\mu^* = 3$), aber mit verschiedener Unsicherheit (der Standardfehler für die Schätzung liegt zwischen 0.46 und 0.059). Ein Punktschätzer enthält keine Information über seine Genauigkeit und ist daher oft nicht ausreichend.

Konfidenzintervall für eine Schätzung

- Ein einziger Wert beschreibt eine Schätzung für einen Verteilungsparameter θ oft nicht gut genug.
- Es ist besser ein Intervall anzugeben, dessen Breite die Unsicherheit bei der Schätzung beschreibt.
- Ein sog. Konfidenzintervall enthält mit einer gewissen Wahrscheinlichkeit den zu schätzenden Parameter.
- **Genauer**: Ein Intervall I_θ welches mit der Wahrscheinlichkeit $(1-\alpha)$ den gesuchten Verteilungsparameter θ überdeckt, wird Konfidenzintervall (**KI**) für θ mit dem Konfidenzniveau $(1-\alpha)$ genannt.
- typisch: $\alpha = 0.05$ → Konfidenzniveau 0.95 → "95% Konfidenzintervall"



Ausgehend von einer Stichprobe wollen wir ein Intervall finden, das mit 95% Wahrscheinlichkeit den wahren Wert θ umfasst = 95% Konfidenzintervall

www.matstat.org

Konfidenzintervall für die Schätzung von μ in einer Normalverteilung $N(\mu, \sigma)$

Stichprobe: $x = (x_1, x_2, \ldots x_n)$ wobei $X_i \sim N(\mu, \sigma)$

$$\mu^* = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$ Punktschätzer für μ in $N(\mu, \sigma)$

<u>Zur Erinnerung</u>: der Schätzer muss als Zufallsvariable aufgefasst werden

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$ Verteilung für diesen Schätzer (siehe **F5**)

$$\Rightarrow \quad \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$ standardisierte Zufallsvariable

www.matstat.org

KI für die Schätzung von μ in $N(\mu, \sigma)$

Fall 1: σ bekannt

$$X_i \sim N(\mu, \sigma) \quad \rightarrow \quad \bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

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

Wenn σ bekannt ist, ist dies eine sog. Referenzvariable (Pivotvariable).
(eine $N(0,1)$ -Variable wird oft mit Z bezeichnet)

Referenzvariable:

- ist vollständig bekannt bis auf den zu schätzenden Parameter [hier: μ]
- die Verteilung der Variable ist vollständig bekannt, enthält keine unbekannten Parameter [hier: $N(0,1)$]

Zur Erinnerung (siehe **F5**): Wenn Z eine $N(0,1)$-Variable ist, so gilt :

$$\boxed{P\left(-\lambda_{\alpha/2} < Z \leq +\lambda_{\alpha/2}\right) = 1 - \alpha} \quad \text{wenn} \quad Z \sim N(0,1)$$

$\lambda_{\alpha/2}$ = Quantil für die standardisierte Normalverteilung , $N(0,1)$

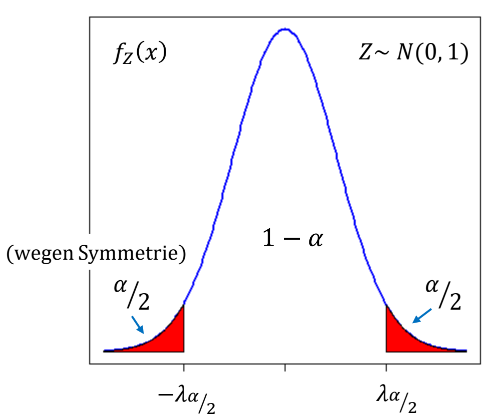
www.matstat.org

KI für die Schätzung von μ in $N(\mu, \sigma)$

Fall 1: σ bekannt

$X_i \sim N(\mu, \sigma)$ [oder ungefähr so wegen Zentralem Grenzwertsatz, ZGS]

$$P\left(-\lambda_{\alpha/2} < Z \leq +\lambda_{\alpha/2}\right) = \Phi\left(\lambda_{\alpha/2}\right) - \Phi\left(-\lambda_{\alpha/2}\right) = \Phi\left(\lambda_{\alpha/2}\right) - \left[1 - \Phi\left(\lambda_{\alpha/2}\right)\right]$$
$$= 2 \cdot \Phi\left(\lambda_{\alpha/2}\right) - 1 = 2 \cdot [1 - \alpha/2] - 1 = \underline{\underline{1 - \alpha}}$$



$$\boxed{P\left(-\lambda_{\alpha/2} < Z \leq +\lambda_{\alpha/2}\right) = 1 - \alpha}$$

$\lambda_{\alpha/2}$ = Quantil für standardisierte Normalverteilung N(0,1)

www.matstat.org

KI für die Schätzung von μ in $N(\mu, \sigma)$

Fall 1: σ bekannt

$X_i \sim N(\mu, \sigma)$ oder ungefähr so (ZGS)

Schritt 1: Um ein Konfidenzintervall zu finden, wird die Referenzvariable zwischen den Quantilen für deren Verteilung eingeschlossen:

$$P\left(-\lambda_{\alpha/2} < Z \leq +\lambda_{\alpha/2}\right) = 1 - \alpha$$

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

$$P\left(-\lambda_{\alpha/2} < \underbrace{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}_{N(0,1)\text{-verteilt}} \leq +\lambda_{\alpha/2}\right) = 1 - \alpha$$

$\lambda_{\alpha/2}$ = Quantil für $N(0,1)$

www.matstat.org

KI für die Schätzung von μ in $N(\mu, \sigma)$

Fall 1: σ bekannt

$X_i \sim N(\mu, \sigma)$ oder ungefähr so (ZGS)

Schritt 2: das Argument in $P(\ldots)$ umgeformen, bis der zu schätzende Parameter zwischen bekannten Größen eingeschlossen ist → $P(\ldots < \mu \leq \cdots)$

$$P\left(-\lambda_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq +\lambda_{\alpha/2}\right) = 1 - \alpha \quad \Big| \cdot \sigma/\sqrt{n}$$

$$P\left(-\lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu \leq +\lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha \quad \Big| -\bar{X}$$

$$P\left(-\bar{X} - \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < -\mu \leq -\bar{X} + \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha \quad \Big| \cdot (-1)$$

$$P\left(\bar{X} + \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \geq \mu > \bar{X} - \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha \quad \text{umdrehen}$$

$$\boxed{P\left(\bar{X} - \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu \leq \bar{X} + \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha}$$

www.matstat.org

KI für die Schätzung von μ in $N(\mu, \sigma)$

Fall 1: σ bekannt

$$\boxed{P\left(\bar{x} - \lambda_{\alpha/2} \cdot \sigma/\sqrt{n} < \mu \leq \bar{x} + \lambda_{\alpha/2} \cdot \sigma/\sqrt{n}\right) = 1 - \alpha}$$

Das bedeutet: Die Wahrscheinlichkeit, dass μ zwischen $\bar{x} - \lambda_{\alpha/2} \cdot \sigma/\sqrt{n}$ und $\bar{x} + \lambda_{\alpha/2} \cdot \sigma/\sqrt{n}$ liegt, beträgt $1 - \alpha$. Dies ist genau das, was wir haben wollten! Die numerischen Werte dieser Grenzen sind bekannt, wenn die Standard-abweichung σ bekannt ist. Das Konfidenzintervall ist also:

$$\boxed{I_\mu = \bar{x} \pm \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}$$

KI für μ i $N(\mu, \sigma)$ mit dem Konfidenzniveau $(1 - \alpha)$ wenn σ bekannt ist

Wenn z.B. $\alpha = 0.05$ erhalten wir also ein KI mit dem Konfidenzniveau 0.95 (95% Konfidenzintervall).

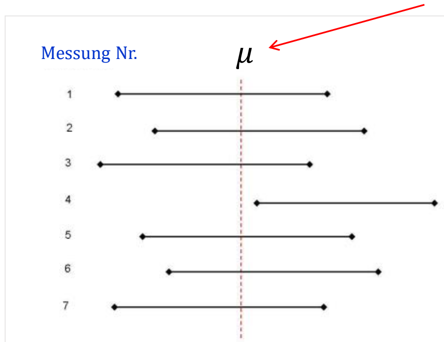
$$I_\mu = \bar{x} \pm \lambda_{0.025} \cdot \frac{\sigma}{\sqrt{n}} = \bar{x} \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

www.matstat.org

Konfidenzintervall: Deutung

Sei $\alpha = 0.05$ $\quad I_\mu = \bar{x} \pm \lambda_{0.025} \cdot \frac{\sigma}{\sqrt{n}} = \bar{x} \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}}$ $\quad$ 95% KI für μ

Die Wahrscheinlichkeit, dass dieses Intervall den wahren Wert μ umfasst, beträgt 95 %.



Auf "lange Sicht" (mit vielen Stichproben) enthalten 95 von 100 berechneten Konfidenzintervallen den wahren Parameter.

www.matstat.org

Allgemeine Methode zur Herleitung eines KI

1. Finde eine Referenzvariable
2. Schließe die Referenzvariable zwischen deren Quantilgrenzen ein
3. Schließe den zu schätzenden Parameter durch Umformen in der Klammer $P(\dots)$ zwischen bekannten Größen ein

1) $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \Longrightarrow \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ Referenzvariable wenn σ bekannt ist

2) $P\left(-\lambda_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq +\lambda_{\alpha/2}\right) = 1-\alpha$ gilt für eine $N(0,1)$-Variable

3) $P\left(\bar{X} - \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu \leq \bar{X} + \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$ nach Umformen

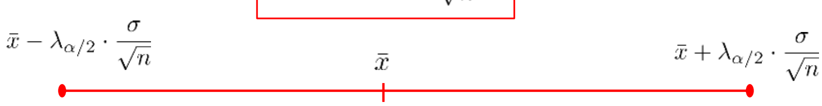
Der wahre Erwartungswert μ wird mit Wahrscheinlichkeit $1-\alpha$ eingeschlossen von:

$$I_\mu = \bar{x} \pm \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

KI für μ mit Konfidenzniveau $1-\alpha$

Breite des Konfidenzintervalls

$$\boxed{I_\mu = \bar{x} \pm \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}$$



$$\boxed{b = 2 \cdot \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}$$

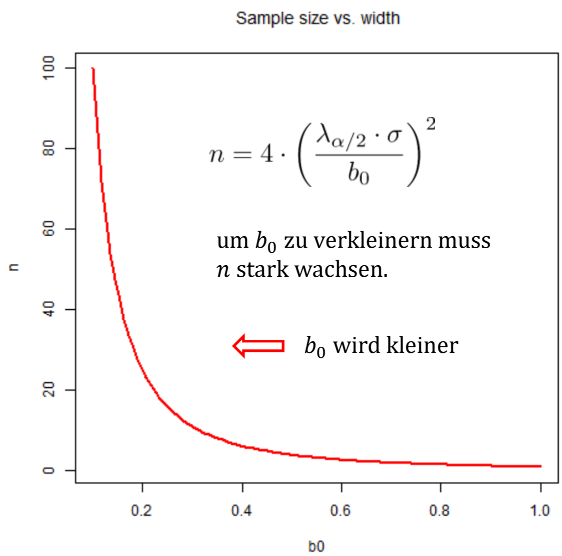
Breite des KI für μ in $N(\mu,\sigma)$ wenn σ bekannt ist; Konfidenzniveau $1-\alpha$

Wird n größer, so wird das Konfidenzintervall schmaler. Dies ist wünschenswert (größere Genauigkeit), wird aber wahrscheinlich auch teurer, denn mehr Messwerte müssen gesammelt werden. Wir können fordern, dass das KI nicht breiter als ein gewisser Wert b_0 wird … und davon ausgehend die erforderliche Stichprobengröße n bestimmen. Umformen oben stehender Formel (wir setzen $b = b_0$) ergibt:

$$\boxed{n = 4 \cdot \left(\frac{\lambda_{\alpha/2} \cdot \sigma}{b_0}\right)^2}$$

$\lambda_{\alpha/2}$ ergibt sich durch Wahl von α
σ bekannt gemäß unserer Annahme
b_0 wird nach Wunsch festgesetzt

Breite des Konfidenzintervalls



n_b0.R

www.matstat.org

Zusammenfassung

Konfidenzintervall für die Schätzung von μ in einer Normalverteilung $N(\mu, \sigma)$
wenn σ bekannt ist

Für die Herleitung eines Konfidenzintervalles für μ haben wir:

- $\bar{X}$ als Punktschätzer für μ verwendet
- die Verteilung für $\bar{X}$ ermittelt
- eine Referenzvariable gefunden
- die Referenzvariable zwischen den entsprechenden Quantilen eingeschlossen
- in der Klammer für $P(\dots)$ umgeformt, um den zu schätzenden Parameter zwischen bekannten Größen einzuschließen

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i \Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \Rightarrow \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

$$\Rightarrow P\left(-\lambda_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \leq +\lambda_{\alpha/2}\right) = 1 - \alpha$$

$$\Rightarrow P\left(\bar{X} - \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu \leq \bar{X} + \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$\Rightarrow I_\mu = \bar{x} \pm \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

KI für μ mit Konfidenzniveau $1 - \alpha$

www.matstat.org

Allgemeine Methode zur Herleitung eines Konfidenzintervalles

1. Finde eine Referenzvariable
2. Schließe die Referenzvariable zwischen deren Quantilgrenzen ein
3. Schließe den zu schätzenden Parameter durch Umformen in der Klammer $P(\ldots)$ zwischen bekannten Größen ein

Dies ist bereits das wichtigste "Handwerkszeug" um Konfidenzintervalle herzuleiten. Es können jedoch verschiedenene Situationen auftreten:

- σ kann unbekannt sein (und ist es auch meistens)
- es liegt eine andere Verteilung als Normalverteilung vor
- ....

Im Folgenden sollen nun verschiedene solcher Situationen behandelt werden →

www.matstat.org

KI für Schätzung von μ in $N(\mu,\sigma)$

Fall 2: σ unbekannt, große Stichprobe

$X_i \sim N(\mu,\sigma)$ oder ungefähr so (ZGS)

$$\mu^* = \bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$ Verteilung für Schätzung für μ

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

Dies ist **keine** Referenzvariable, denn σ ist unbekannt (wir wollen μ schätzen, also darf nur μ die einzige Unbekannte in dem Ausdruck sein). Wenn jedoch n groß ist, kann σ mit s approximiert werden, ohne einen großen Fehler zu machen. (s ist ja eine konsistente Schätzung für σ, d.h. s liegt mit großer Wahrscheinlichkeit nahe σ, wenn n groß ist, Vorlesung **F8**, "Punktschätzer"):

$$\sigma \to s \quad \Longrightarrow \quad \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim N(0,1)$$ Referenzvariable $n \geq 30$

→ Allgemeine Methode zur Herleitung eines Konfidenzintervalles

www.matstat.org

KI für Schätzung von μ in $N(\mu,\sigma)$

Fall 2: σ unbekannt, große Stichprobe

1. Finde eine Referenzvariable $X_i \sim N(\mu,\sigma)$ oder ZGS
2. Schließe die Referenzvariable zwischen deren Quantilgrenzen ein
3. Schließe den zu schätzenden Parameter durch Umformen in der Klammer $P(\dots)$ zwischen bekannten Größen ein

$$\boxed{I_\mu = \bar{x} \pm \lambda_{\alpha/2} \cdot \frac{s}{\sqrt{n}}}$$

KI für μ in $N(\mu,\sigma)$ mit Konfidenzniveau $(1-\alpha)$ wenn σ unbekannt ist, aber die Stichprobe (n) groß ist.

$n \geq 30$

I = Punktschätzer ± Quantil · Standardfehler

<u>Beispiel</u>: Stichprobe: $n = 32$ $\bar{x} = 42250$ $s = 164$
gesucht: 95% KI für μ

$\alpha = 0.05 \rightarrow \lambda_{\alpha/2} = \lambda_{0.025} = 1.96$

$n = 31$ also $\sigma \rightarrow s$

$$I_\mu = 42250 \pm 1.96 \cdot \frac{614}{\sqrt{32}} = 42250 \pm 212 = (42030; 42470)$$

KI für Schätzung von μ in $N(\mu,\sigma)$

Fall 3: σ unbekannt, kleine Stichprobe $X_i \sim N(\mu,\sigma)$

$$\mu^* = \bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$ Verteilung für den Schätzer für μ

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

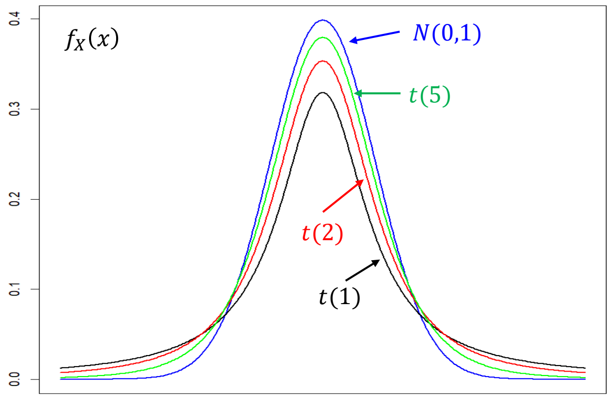
Dies ist **keine** Referenzvariable, denn σ ist unbekannt. Außerdem ist die Stichprobe zu klein um σ einfach mit s zu ersetzen, ohne dass sich die Verteilung des Ausdrucks verändert. Wenn n klein ist und $\sigma \rightarrow s$ ersetzt wird, ändert sich daher die Verteilung zur Studentschen t-Verteilug (Herleitung siehe Anhang):

$$\Rightarrow \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t(n-1)$$ Referenzvariable

Das Argument für die Verteilung (hier: $n-1$) wird "Freiheitsgrade" genannt. Die Studentsche t-Verteilung gleicht der Normalverteilung, ist aber ist etwas breiter, wodurch die größere Unsicherheit widergespiegelt wird (wir haben ja durch die kleinere Stichprobe eine schlechtere Approximation von σ durch s). Je größer n wird, desto schmaler wird die t-Verteilung; eine größere Stichprobe wird also wiederum mit größerer Genauigkeit der Schätzung belohnt.

Die Studentsche t-Verteilung

Dichtefunktion für die t-Verteilung für verschiedene Freiheitsgrade, Vergleich mit $N(0,1)$



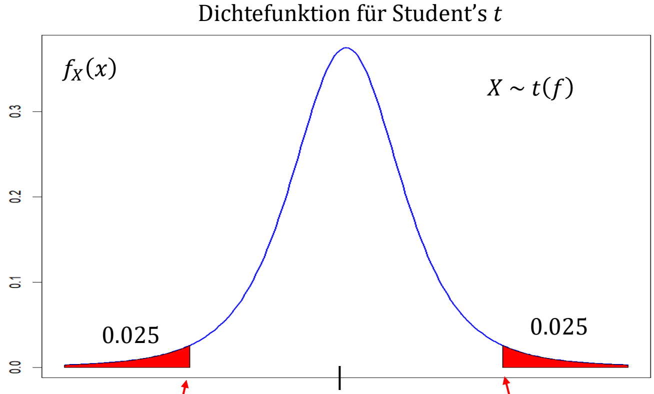
$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-(\frac{\nu+1}{2})}$$

Dichtefunktion
ν = Anzahl der Freiheitsgrade

www.matstat.org

Quantile für die t-Verteilung

Dichtefunktion für Student's t



$t_{0.975}(f) = -\,t_{0.025}(f)$
(symmetrisch um 0)

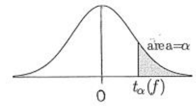
$t_{0.025}(f)$

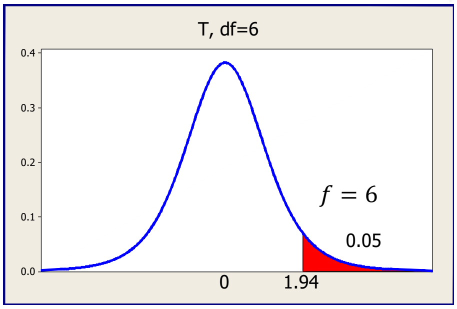
! Die Quantile für t hängen von der Anzahl der Freiheitsgrade (f) ab.

www.matstat.org

Quantile für die t-Verteilung

$$P(X > t_\alpha(f)) = \alpha$$
$$X \in t(f)$$





f \ α	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
1	3.08	6.31	12.71	31.82	63.66	318.31	636.62
2	1.89	2.92	4.30	6.96	9.92	22.33	31.60
3	1.64	2.35	3.18	4.54	5.84	10.21	12.92
4	1.53	2.13	2.78	3.75	4.60	7.17	8.61
5	1.48	2.02	2.57	3.36	4.03	5.89	6.87
6	1.44	1.94	2.45	3.14	3.71	5.21	5.96
7	1.41	1.89	2.36	3.00	3.50	4.79	5.41
8	1.40	1.86	2.31	2.90	3.36	4.50	5.04
9	1.38	1.83	2.26	2.82	3.25	4.30	4.78
10	1.37	1.81	2.23	2.76	3.17	4.14	4.59
11	1.36	1.80	2.20	2.72	3.11	4.02	4.44
12	1.36	1.78	2.18	2.68	3.05	3.93	4.32
13	1.35	1.77	2.16	2.65	3.01	3.85	4.22
14	1.35	1.76	2.14	2.62	2.98	3.79	4.14
15	1.34	1.75	2.13	2.60	2.95	3.73	4.07
16	1.34	1.75	2.12	2.58	2.92	3.69	4.01
17	1.33	1.74	2.11	2.57	2.90	3.65	3.97
18	1.33	1.73	2.10	2.55	2.88	3.61	3.92
19	1.33	1.73	2.09	2.54	2.86	3.58	3.88
20	1.33	1.72	2.09	2.53	2.85	3.55	3.85
21	1.32	1.72	2.08	2.52	2.83	3.53	3.82
22	1.32	1.72	2.07	2.51	2.82	3.50	3.79
23	1.32	1.71	2.07	2.50	2.81	3.48	3.77
24	1.32	1.71	2.06	2.49	2.80	3.47	3.75
25	1.32	1.71	2.06	2.49	2.79	3.45	3.73
26	1.31	1.71	2.06	2.48	2.78	3.43	3.71
27	1.31	1.70	2.05	2.47	2.77	3.42	3.69
28	1.31	1.70	2.05	2.47	2.76	3.41	3.67
29	1.31	1.70	2.05	2.46	2.76	3.40	3.66
30	1.31	1.70	2.04	2.46	2.75	3.39	3.65
40	1.30	1.68	2.02	2.42	2.70	3.31	3.55
60	1.30	1.67	2.00	2.39	2.66	3.23	3.46
120	1.29	1.66	1.98	2.36	2.62	3.16	3.37
∞	1.28	1.64	1.96	2.33	2.58	3.09	3.29 $= \lambda_\alpha$

Die t-Verteilung ist etwas breiter als N, aber je größer n wird, desto ähnlicher werden sie (siehe letzte Reihe der Tabelle).

www.matstat.org

KI für Schätzung von μ in $N(\mu, \sigma)$

Fall 3: σ unbekannt, kleine Stichprobe

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t(n-1)$$ Referenzvariable $X_i \sim N(\mu, \sigma)$

1. Finde eine Referenzvariable
2. Schließe die Referenzvariable zwischen deren Quantilgrenzen ein
3. Schließe den zu schätzenden Parameter durch Umformen in der Klammer $P(\dots)$ zwischen bekannten Größen ein

$$P\left(-t_{\alpha/2}(n-1) < \frac{\bar{X} - \mu}{s/\sqrt{n}} \leq t_{\alpha/2}(n-1)\right) = 1 - \alpha$$

$$P\left(\bar{X} - t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}} < \mu \leq \bar{X} + t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}\right) = 1 - \alpha$$

$\Rightarrow$ $\boxed{I_\mu = \bar{x} \pm t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}}$

KI für μ i $N(\mu, \sigma)$ mit Konfidenzniveau $1 - \alpha$ wenn σ unbekannt ist und die Stichprobe (n) klein ist.

www.matstat.org

KI für Schätzung von μ in $N(\mu, \sigma)$

Fall 3: σ unbekannt, kleine Stichprobe

$X_i \sim N(\mu, \sigma)$

$$I_\mu = \bar{x} \pm t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}$$

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

I = Punktschätzer $\pm$ Quantil $\cdot$ Standardfehler

$\bar{x} - t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}$ $\qquad \bar{x} \qquad$ $\bar{x} + t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}$

Wir erhalten $\bar{x}, s, n$ aus der Stichprobe und $t_{\alpha/2}(n-1)$ mit Hilfe einer Tabelle.

! Achtung: $X_i \sim N(\mu, \sigma)$ wurde vorausgesetzt, um die Formel für das Konfidenzintervall herzuleiten! Eventuell kann mit dem Zentralen Grenzwertsatz begründet werden, dass X_i ungefähr normalverteilt ist.

www.matstat.org

KI für Schätzung von μ in $N(\mu, \sigma)$

Fall 3: σ unbekannt, kleine Stichprobe

Beispiel: $\qquad X_i \sim N(\mu, \sigma)$

Wir wissen (oder haben Grund zur Annahme) dass die X_i normalverteilt sind.

$$\vec{x} = (15.7, 14.3, 12.65, 13.2, 16.85, 16.05, 16.55, 16.05, 16.6, 17.05)$$

Suche: 95% Konfidenzintervall für μ

a) ... wenn bekannt ist, dass $\sigma = 1.56$

$D = \sigma/\sqrt{n} \approx 0.5 \qquad \lambda_{\alpha/2} = \lambda_{0.025} = 1.96$

$I_\mu = \bar{x} \pm \lambda_{\alpha/2} \cdot D = 15.5 \pm 1.96 \cdot 0.5 \approx 15.5 \pm 1 = \underline{\underline{(14.5, 16.5)}}$

b) ... wenn σ unbekannt ist

$s = 1.56 \qquad \mathrm{d} = s/\sqrt{n} \approx 0.5 \qquad t_{\alpha/2}(f) = t_{0.025}(9) = 2.26$

$I_\mu = \bar{x} \pm t_{\alpha/2}(n-1) \cdot d = 15.5 \pm 2.26 \cdot 0.5 \approx 15.5 \pm 1.1 = \underline{\underline{(14.4, 16.6)}}$

www.matstat.org

KI für Schätzung von $\Delta\mu$ für zwei N-Variablen

Fall 1: σ_X und σ_Y bekannt

2 Populationen, 2 Stichproben:

$$\left.\begin{aligned} X &= (x_1, x_2, \cdots, x_{n_x}) \\ Y &= (y_1, y_2, \cdots, y_{n_y}) \end{aligned}\right\} \text{2 Populationen}$$

Punktschätzer für Unterschied $\Delta\mu = \mu_X - \mu_Y$ zwischen den beiden Erwartungswerten:

$\Delta\mu^* = \bar{X} - \bar{Y}$ (erwartungstreu, konsistent)

wir wissen dass: $\bar{X} \sim N\left(\mu_x, \frac{\sigma_x}{\sqrt{n_x}}\right)$ und $\bar{Y} \sim N\left(\mu_y, \frac{\sigma_y}{\sqrt{n_y}}\right)$

$$\Rightarrow \quad \bar{X} - \bar{Y} \sim N\left(\underbrace{\mu_x - \mu_y}_{\Delta\mu = \mu_x - \mu_y}, \underbrace{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}_{D = \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}\right)$$

$$\Rightarrow \quad \bar{X} - \bar{Y} \sim N(\Delta\mu, D)$$

KI für Schätzung von $\Delta\mu$ für zwei N-Variablen

Fall 1: σ_X und σ_Y bekannt

$\bar{X} - \bar{Y} \sim N(\Delta\mu, D)$

$$\Rightarrow \quad \frac{\bar{X} - \bar{Y} - \Delta\mu}{D} \sim N(0,1)$$

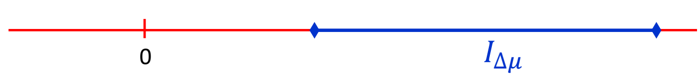
dies ist eine Referenzvariable wenn σ_X und σ_Y bekannt sind.

$$D = \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$$

1. Finde eine Referenzvariable
2. Schließe die Referenzvariable zwischen deren Quantilgrenzen ein
3. Schließe den zu schätzenden Parameter durch Umformen in der Klammer $P(\ldots)$ zwischen bekannten Größen ein

$$\boxed{I_{\Delta\mu} = \bar{x} - \bar{y} \pm \lambda_{\alpha/2} \cdot D}$$

KI für $\Delta\mu$ zwischen zwei Normalverteilungen, Konfidenzniveau $1 - \alpha$, wenn σ_X und σ_Y bekannt sind.



! Wenn ein KI für eine Differenz zwischen 2 Erwartungswerten **nicht** den Nullpunkt enthält, dann weiß man mit $(1-\alpha) \cdot 100\%$ Wahrscheinlichkeit, dass die Erwartungswerte der beiden Populationen wirklich unterschiedlich sind. Wenn das KI hingegen die 0 enthält, dann können wir nicht sagen, ob es einen Unterschied gibt – Null ist ja in diesem Fall ein möglicher Unterschied!

KI für Schätzung von $\Delta\mu$ für zwei N-Variablen

Fall 2: σ_X und σ_Y unbekannt, aber große Stichproben

$n_x \geq 30 \;;\; n_y \geq 30$ "Daumenregeln", manchmal sogar noch kleiner

$\sigma_x \to s_x \quad \sigma_y \to s_y$ wegen Konsistenz der Schätzung kein großer Fehler

$D \to d \qquad d = \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}$ Standardfehler

$\Rightarrow \frac{\bar{X} - \bar{Y} - \Delta\mu}{d} \sim N(0,1)$ dies ist eine Referenzvariable denn s_X und s_Y sind bekannt (und eine gute Schätzung für σ_X bzw. σ_Y, so dass sich die Verteilung nicht verändert)

1. Finde eine Referenzvariable
2. Schließe die Referenzvariable zwischen deren Quantilgrenzen ein
3. Schließe den zu schätzenden Parameter durch Umformen in der Klammer $P(\ldots)$ zwischen bekannten Größen ein

$$\boxed{I_{\Delta\mu} = \bar{x} - \bar{y} \pm \lambda_{\alpha/2} \cdot d}$$

KI für $\Delta\mu$ zwischen zwei Normalverteilungen, Konfidenzniveau $1-\alpha$, σ_X und σ_Y unbekannt, aber große Stichproben

KI für Schätzung von $\Delta\mu$ für zwei N-Variablen

Fall 2: σ_X und σ_Y unbekannt, aber große Stichproben

$$\boxed{I_{\Delta\mu} = \bar{x} - \bar{y} \pm \lambda_{\alpha/2} \cdot d}$$

KI für $\Delta\mu$ zwischen zwei Normalverteilungen, Konfidenzniveau $1-\alpha$, σ_X und σ_Y unbekannt, aber große Stichproben

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n_x}(x_i - \bar{x})^2 \qquad s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n_y}(y_i - \bar{y})^2$$

Ex1: Populationen A und B; eine Stichprobe von jeder Population genommen:

$$\left.\begin{array}{lll} \bar{x}_A = 165.09 & s_A = 60.96 & n_A = 22 \\ \bar{x}_B = 153.10 & s_B = 43.08 & n_B = 30 \end{array}\right\} \text{gegeben}$$

Suche: 99% KI für Unterschied zwischen μ_A und μ_B

$$d = \sqrt{\frac{60.96^2}{22} + \frac{43.08^2}{30}} = 15.19 \qquad \alpha = 0.01 \qquad \lambda_{\alpha/2} = \lambda_{0.005} = 2.58$$

$$I_{\Delta\mu} = \bar{x}_A - \bar{x}_B \pm \lambda_{\alpha/2} \cdot d = 11.99 \pm 2.58 \cdot 15.19 = \underline{\underline{(-27; 51)}}$$

Resultat: Kein signifikanter Unterschied zwischen den Erwartungswerten auf 99% Konfidensniveau, denn das Konfidenzintervall enthält die Null.

KI für Schätzung von $\Delta\mu$ für zwei N-Variablen

Fall 2: σ_X und σ_Y unbekannt, aber große Stichproben

Ex2: Populationen A und B; eine Stichprobe von jeder Population liegt vor:

$$A = (195, 240, 154, 95, 65, 82, 132, 155, 125, 119, 155, 345, 145, 200, 130, 223, 145, 207, 183, 190, 137, 210)$$

$$B = (88, 73, 165, 188, 145, 158, 195, 165, 140, 145, 203, 196, 230, 225, 128, 190, 170, 158, 72, 135, 105, 155, 165, 120 138, 125, 188, 145, 208, 75$$

Suche: 99% KI für die Differenz der Erwartungswerte

$$\bar{x}_A = 165.09 \quad s_A = 60.96 \quad \bar{x}_B = 153.1 \quad s_B = 43.08 \quad \bar{x}_A - \bar{x}_B = 11.99$$

$$d = \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}} = \sqrt{\frac{60.96^2}{22} + \frac{43.08^2}{30}} = 15.19 \quad \text{Standardfehler}$$

$$1 - \alpha = 0.99 \quad \Rightarrow \quad \alpha = 0.01 \qquad \lambda_{\alpha/2} = \lambda_{0.005} = 2.58$$

$$I_{\Delta\mu} = \bar{x}_A - \bar{x}_B \pm \lambda_{\alpha/2} \cdot d = 11.99 \pm 2.58 \cdot 15.19 = \underline{\underline{(-27, 51)}}$$

www.matstat.org

KI für Schätzung von $\Delta\mu$ für zwei N-Variablen

Fall 3: σ_X und σ_Y unbekannt, aber gleich(!), kleine Stichproben

$$\left.\begin{array}{ll} x = (x_1, x_2, \ldots x_{n_X}) & \text{mit } X_i \sim N(\mu_x, \sigma) \\ y = (y_1, y_2, \ldots y_{n_Y}) & \text{mit } Y_i \sim N(\mu_y, \sigma) \end{array}\right\}$$ zwei Stichproben von verschiedenen Populationen

$(\Delta\mu)^* = \bar{X} - \bar{Y}$ Punktschätzer für $\Delta\mu$ $\qquad \sigma_x = \sigma_y = \sigma$

σ wird ersetzt durch $s_p = \sqrt{\dfrac{(n_x - 1) \cdot s_x^2 + (n_y - 1) \cdot s_y^2}{(n_x - 1) + (n_y - 1)}}$ "pooled variance"

Standardfehler: $\quad d = s_p \cdot \sqrt{\dfrac{1}{n_x} + \dfrac{1}{n_y}}$

Man kann zeigen: $\dfrac{\bar{X} - \bar{Y} - \Delta\mu}{d} \sim t(n_x + n_y - 2)$ $\quad$ t-Verteilung mit $f = n_x + n_y - 2$

(siehe Anhang)

Stichproben sind klein → t-Verteilung, wie oben, aber mit veränderter Anzahl von Freiheitsgraden

www.matstat.org

KI für Schätzung von $\Delta\mu$ für zwei N-Variablen

Fall 3: σ_X und σ_Y unbekannt, aber gleich, kleine Stichproben

$$s_p = \sqrt{\frac{(n_x - 1)\cdot s_x^2 + (n_y - 1)\cdot s_y^2}{(n_x - 1) + (n_y - 1)}} \qquad d = s_p \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}$$

$$\frac{\bar{X} - \bar{Y} - \Delta\mu}{d} \sim t(f)$$ Referenzvariable

$$f = n_x + n_y - 2$$ Anzahl der Freiheitsgrade

$$\boxed{I_{\Delta\mu} = \bar{x} - \bar{y} \pm t_{\alpha/2}(f) \cdot d}$$

KI für $\Delta\mu$ für zwei Normalverteilungen, Konfidenzniveau $1-\alpha$, σ_X und σ_Y unbekannt aber gleich, kleine Stichproben

Vergleich mit großer Stichprobe ($n_{x,y} \geq 30$): $t_{\alpha/2}(f)$ ist größer als $\lambda_{\alpha/2}$ → das Konfidenzintervall wird also breiter. Dies widerspiegelt die Tatsache, dass wir eine größere Unsicherheit aufgrund der kleinen Stichproben haben.

www.matstat.org

KI für Schätzung der Differenz von Mittelwerten für zwei gepaarte Stichproben von N

Unterschied zwischen unabhängigen und gepaarten Stichproben:

- bei gepaarten Stichproben sind jeweils zwei Werte zusammengehörig, sie sind **nicht unabhängig**

<u>Beispiel</u>: Blutzuckergehalt von Diabetespatienten vor und nach einer Behandlung:

Unabhängige Stichproben:

Gruppe 1, vorher	70.4	80.1	65.8	66.7	74.3	72.1	70.9	85.4	89.3	90.4	78.9	77.0
Gruppe 2, nachher	70.9	65.8	66.7	74.3	87.2	89.3	93.4	82.1	74.1	70.9		

Gepaarte Stichproben:

Person	A	B	C	D	E	F	G	H
vorher	78.1	66.9	74.3	72.5	90.9	78.3	68.4	72.5
nachher	79.2	67.0	77.1	73.3	92.0	78.1	68.4	72.9

Der Blutzuckergehalt einer Person *nach* der Behandlung ist nicht unabhängig vom Blutzuckergehalt dieser Person *vor* der Behandlung.

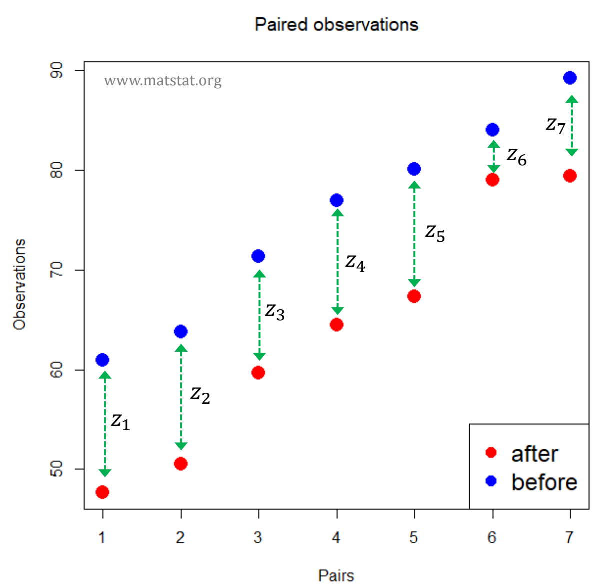
www.matstat.org

KI für Schätzung der Differenz von Mittelwerten für zwei gepaarte Stichproben von N

Person	A	B	C	D	E	F	G	H
vorher	78.1	66.9	74.3	72.5	90.9	78.3	68.4	72.5
nachher	79.2	67.0	77.1	73.3	92.0	78.1	68.4	72.9

- Die Observationen sind nicht unabhängig, die oben gezeigte Methode für die Berechnung des KI für zwei unabhängige Gruppen kann also nicht angewendet werden.
- Durch das Sammeln von gepaarten Observationen kann die in die Berechnung eingehende Varianz im Vergleich zu zwei unabhängigen Stichproben verringert werden, da die Varianz innerhalb der Paare (vertikal in der Tabelle) meist kleiner ist als die Varianz zwischen den Paaren (horizontal in der Tabelle), siehe weiter unten …
- Es wird nach wie vor gefordert, dass die Paare unabhängig voneinander sind (es muss also z. B. Person A von Person B unabhängig sein, usw.).

Idee: nutze die Differenzen z_i innerhalb der Paare, um $\Delta\mu$ zwischen den beiden Gruppen zu schätzen



Annahmen zum Modell für KI für gepaarte Stichproben

- alle Observationen kommen von einer Normalverteilung
- es gibt eine systematische Verschiebung $\Delta\mu$ zwischen beiden Gruppen, die nur durch rein zufälliges (normalverteiltes) Rauschen überlagert wird.
- innerhalb der Gruppen ist die Varianz gleich, d.h. wir haben σ_x^2 in Gruppe X und σ_y^2 in Gruppe Y.
- **zwischen** den Paaren besteht Unabhängigkeit der Beobachtungen

$$\left.\begin{aligned} X_i &\sim N(\mu_i, \sigma_x) \\ Y_i &\sim N(\mu_i + \Delta\mu, \sigma_y) \end{aligned}\right\}$$ systematische Verschiebung $\Delta\mu$ zwischen beiden Gruppen

$Z_i = Y_i - X_i$ wird berechnet

$Z_i \sim N(\Delta\mu, \sigma_z)$ Verteilung der Z_i

$\bar{Z} \sim N\left(\Delta\mu, \frac{\sigma_z}{\sqrt{n}}\right)$ Verteilung für Mittelwert der Z_i

etwas mehr dazu im Anhang!

$$\sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2 - 2 \cdot \underbrace{\rho \cdot \sigma_x \sigma_y}_{C(X,Y)}}$$ Standardabweichung für σ_z (unbekannt)

$C(X,Y) = \rho \cdot \sigma_x \sigma_y$ Kovarianz zwischen X und Y

KI für $\Delta\mu$, gepaarte Stichproben

$$Z_i \sim N(\Delta\mu, \sigma_z) \qquad \bar{Z} \sim N\left(\Delta\mu, \frac{\sigma_z}{\sqrt{n}}\right)$$

Dies stellt ein uns schon bekanntes Problem dar: die Schätzung des Erwartungswertes (hier: $\Delta\mu$), einer normalverteilten Zufallsvariable (hier: Z_i), mit unbekannter Standardabweichung (hier: σ_Z) und kleiner Stichprobe!
Da σ_Z unbekannt ist, ist die entsprechende Referenzvariable t-verteilt:

$$\frac{\bar{Z} - \Delta\mu}{s_z/\sqrt{n}} \sim t(n-1)$$ Referenzvariable

1. Finde eine Referenzvariable
2. Schließe die Referenzvariable zwischen deren Quantilgrenzen ein
3. Schließe den zu schätzenden Parameter durch Umformen in der Klammer $P(\ldots)$ zwischen bekannten Größen ein

$$\boxed{I_{\Delta\mu} = \bar{z} \pm t_{\alpha/2}(n-1) \cdot \frac{s_z}{\sqrt{n}}}$$ KI für Differenz der Erwartungswerte zweier gepaarter Stichproben von N, Konfidenzniveau $1 - \alpha$

$$z_i = y_i - x_i \qquad \bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i \qquad s_z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i - \bar{z})^2$$

KI für Schätzung der Differenz von Mittelwerten für zwei gepaarte Stichproben von N

Beispiel: zwei Personen A und B führten Messungen an 11 Objekten aus, wir wollen wissen ob diese "systematisch verschieden" messen:

	1	2	3	4	5	6	7	8	9	10	11
B	21.6	22.9	20	23.6	15.7	17.5	27.3	19.8	16.4	14.7	20.1
A	20.2	22	19.7	21.4	16.3	17	24.5	15.6	16	13.2	19
z	1.4	0.9	0.3	2.2	-0.6	0.5	2.8	4.2	0.4	1.5	1.1

Es wird angenommen, dass die Messungen Observationen von $N(\mu_i\,,\sigma_x)$ bzw. $N(\mu_i+\Delta\mu,\sigma_y)$ sind, wobei $\Delta\mu$ der systematische Unterschied beider Gruppen ist.

$$\bar{z}=1.34 \qquad f=n-1=10 \qquad t_{\alpha/2}(f)=t_{0.025}(10)=2.23$$

$$s=\sqrt{\tfrac{1}{n-1}\sum_{i=1}^{n}(z_i-\bar{z})^2}=1.33 \qquad d=s/\sqrt{11}=0.401$$

$$I_\Delta=\left(\bar{z}\pm t_{\alpha/2}(f)d\right)=(1.34\pm 2.23\cdot 0.401)=(1.34\pm 0.89)=\underline{\underline{(0.45,2.23)}}$$

Es ergibt sich tatsächlich ein signifikanter Unterschied auf 95% Konfidenzniveau (KI beinhaltet nicht die Null).

www.matstat.org

KI für Schätzung von p in $Bin(n,p)$

(KI für Schätzung einer Proportion)

$$\left.\begin{array}{l}\bullet\ x = \text{Anzahl der "geglückten" Bernoulli-Versuche}\\ \bullet\ n = \text{Gesamtanzahl der Bernoulli-Versuche}\end{array}\right\}\ \text{aus der Stichprobe}$$

$$p^*=\frac{X}{n} \qquad \text{Punktschätzer für } p \qquad X\sim Bin(n,p)$$

$$X \sim Bin(n,p)$$

$$X \sim N\left(n\cdot p,\sqrt{n\cdot p\cdot(1-p)}\right) \quad \text{wenn} \quad V(X)=n\cdot p\cdot(1-p)>5$$

Normalapproximation! Die Stichprobe muss groß sein, um diese Näherung zu rechtfertigen (wie z.B. in einer Meinungsumfrage). Zur Erinnerung: Beispiel Münzwurf, um p zu schätzen (Vorlesung **F8**), wo schlappe 10 Würfe kein überzeugendes Resultat erbrachten!

Wir wenden nun die Regeln für die Lineartransformation von normal-verteilten Zufallsvariablen (Vorlesung **F5**) an →

$$p^*=\frac{X}{n}\sim N\left(p,\sqrt{\frac{p\cdot(1-p)}{n}}\right) \quad \text{Verteilung für Punktschätzer } p^* \text{ für } p$$

$$p^*=\frac{X}{n}\sim N(p,D) \quad \text{mit} \quad D=\sqrt{\frac{p\cdot(1-p)}{n}} \quad \text{Abkürzung}$$

KI für Schätzung von p in $Bin(n,p)$
(KI für Schätzung einer Proportion)

$$p^* = \frac{X}{n} \sim N(p, D) \quad \text{mit} \quad D = \sqrt{\frac{p\cdot(1-p)}{n}}$$ Abkürzung

Standardiserung ergibt: $$\frac{\frac{X}{n}-p}{D} \sim N(0,1)$$

Die Standardabweichung D wird nun mit dem Standardfehler d approximiert indem p mit dessen Schätzung, also mit $p^* = x/n$ ersetzt wird. Dies kann gemacht werden, weil die Schätzung erwartungstreu und konsistent ist; und weil die Stichprobe (n) groß ist:

$$p \to p^* = \frac{x}{n} \quad \Longrightarrow \quad D \to d \qquad d = \sqrt{\frac{\frac{x}{n}\cdot(1-\frac{x}{n})}{n}}$$

$$\frac{\frac{X}{n}-p}{d} \sim N(0,1)$$ Referenzvariable

1. Finde eine Referenzvariable
2. Schließe die Referenzvariable zwischen deren Quantilgrenzen ein
3. Schließe den zu schätzenden Parameter durch Umformen in der Klammer $P(\ldots)$ zwischen bekannten Größen ein

www.matstat.org

KI für Schätzung von p in $Bin(n,p)$
(KI für Schätzung einer Proportion)

$$\frac{\frac{X}{n}-p}{d} \sim N(0,1)$$ Referenzvariable $$d = \sqrt{\frac{\frac{x}{n}\cdot(1-\frac{x}{n})}{n}}$$

$$\boxed{I_p = \frac{x}{n} \pm \lambda_{\alpha/2}\cdot d}$$

KI für Schätzung von p in $Bin(n,p)$ (Schätzung einer Proportion). Große Stichprobe erforderlich.
Konfidenzniveau $1-\alpha$
Punktschätzer $\mp$ Quantil · Standardfehler

Beispiel: Anteil minderwertiger Komponenten in einem Produktionsprozess: eine Stichprobe von 250 Komponenten wurde genommen. Ergebnis: 42 davon hatten eine zu kurze Lebensdauer. Wir suchen ein 95% KI für den Anteil der fehlerhaften Komponenten (in der Gesamtpopulation).

$$p^* = \frac{x}{n} = \frac{42}{250} = 0.168$$ Punktschätzer für den Anteil fehlerhafter Komponenten

$$d = \sqrt{\frac{0.168\cdot(1-0.168)}{250}} = 0.023 \qquad \alpha = 0.05 \qquad \lambda_{\alpha/2} = \lambda_{0.025} = 1.96$$

$$I_p = \frac{x}{n} \pm \lambda_{\alpha/2}\cdot d = 0.168 \pm 1.96\cdot 0.023 = 0.168 \pm 0.046 = \underline{\underline{(0.12; 0.21)}}$$

www.matstat.org

KI für Schätzung von p in $Bin(n, p)$
(KI für Schätzung einer Proportion)

$$I_p = \frac{x}{n} \pm \lambda_{\alpha/2} \cdot d$$ KI für die Schätzung einer Proportion

$$b = 2 \cdot \lambda_{\alpha/2} \cdot d = 2 \cdot \lambda_{\alpha/2} \cdot \sqrt{\frac{p^* \cdot (1-p^*)}{n}}$$ Breite des Konfidenzintervalles:

Wenn die Breite b_0 des Konfidenzintervalles vor Entnahme der Stichprobe justiert werden soll, muss die Stichprobengröße entsprechend gewählt werden. Umformung ergibt:

$$n_0 = 4 \cdot \left(\frac{\lambda_{\alpha/2}}{b_0} \right)^2 \cdot p^*(1-p^*)$$

Wir kennen jedoch p^* nicht im Voraus. Wenn wir keinen Anhaltspunkt haben, wie groß p^* sein könnte, müssen wir "vorsichtshalber" den größtmöglichen Wert für $p^* \cdot (1-p^*)$ einsetzen, welcher $1/_4$ beträgt (1. Ableitung Null setzen!)

$$n_0 = \left(\frac{\lambda_{\alpha/2}}{b_0} \right)^2 \quad \text{oder} \quad b_0 = \frac{\lambda_{\alpha/2}}{\sqrt{n_0}}$$

Wenn b_0 verkleinert werden soll, muss die Stichprobe wesentlich vergrößert werden : $n_0 \sim b_0^{-2}$ (b_0 wird halbiert wenn n_0 vervierfacht wird).

www.matstat.org

KI für Schätzung des Unterschiedes zwischen 2 Proportionen
(Schätzung von Δp für zwei Bin)

- **Daten**: 2 Stichproben aus 2 Populationen: $x_1, x_2 =$ jeweilige Anzahl der "Erfolge"
- z.B. Meinungsumfrage (Ja/Nein) in der Stadt (x_1) und auf dem Land (x_2)
- z.B. Anteil fehlerhafter Komponenter in Fabrik A (x_1) und in Fabrik B (x_2)

$$p_1^* = \frac{X_1}{n_1} \qquad p_2^* = \frac{X_2}{n_2}$$ Punktschätzer für p_1 und p_2

$$(\Delta p)^* = \frac{X_1}{n_1} - \frac{X_2}{n_2}$$ Punktschätzer für Δp (erwartungstreu, konsistent)

$$X_1 \sim Bin(n_1, p_1) \qquad X_2 \sim Bin(n_2, p_2)$$

Sowohl n_1 als auch n_2 müssen groß sein → Normalapproximation, Lineartransformation (wie oben):

$$\frac{X_1}{n_1} \sim N\left(p_1, \sqrt{\frac{p_1 \cdot (1-p_1)}{n_1}}\right) \qquad \frac{X_2}{n_2} \sim N\left(p_2, \sqrt{\frac{p_2 \cdot (1-p_2)}{n_2}}\right)$$

$$\Rightarrow \quad \frac{X_1}{n_1} - \frac{X_2}{n_2} \sim N\left(p_1 - p_2, \sqrt{\frac{p_1 \cdot (1-p_1)}{n_1} + \frac{p_2 \cdot (1-p_2)}{n_2}}\right)$$

www.matstat.org

KI für Schätzung des Unterschiedes zwischen 2 Proportionen
(Schätzung von Δp für zwei Bin)

$$\Rightarrow \quad \frac{X_1}{n_1} - \frac{X_2}{n_2} \sim N\left(\underbrace{p_1 - p_2}_{\Delta p}, \underbrace{\sqrt{\frac{p_1 \cdot (1-p_1)}{n_1} + \frac{p_2 \cdot (1-p_2)}{n_2}}}_{D}\right)$$

Die Standardabweichung D wird mit dem Standardfehler d approximiert, indem beide p_i mit dem entsprechenden Schätzer ersetzt werden, also $p_i^* = x_i/n_i$ $(i = 1,2)$.

$$\frac{X_1}{n_1} - \frac{X_2}{n_2} \sim N(\Delta p, d) \qquad d = \sqrt{\frac{\frac{x_1}{n_1} \cdot (1-\frac{x_1}{n_1})}{n_1} + \frac{\frac{x_2}{n_2} \cdot (1-\frac{x_2}{n_2})}{n_2}}$$

$$\frac{\frac{X_1}{n_1} - \frac{X_2}{n_2} - \Delta p}{d} \sim N(0,1)$$

Referenzvariable → ”allgemeine Methode”

$$\boxed{I_{\Delta p} = \frac{x_1}{n_1} - \frac{x_2}{n_2} \pm \lambda_{\alpha/2} \cdot d}$$

KI für Schätzung des Unterschiedes zwischen zwei Proportionen (Δp für zwei Bin).
Beide Stichproben müssen groß sein.
Konfidenzniveau $(1-\alpha)$.
Punktschätzer $\pm$ Quantil $\cdot$ Standardfehler

KI für Schätzung von μ in $Po(\mu)$

Stichprobe: $(x_1, x_2, \cdots, x_n)$ = Anzahl der eingetroffenen Ereignisse (”events”) in n beobachteten (gleich großen) Intervallen

$\mu^* = \bar{X}$ Punktschätzer (siehe Tabelle Vorlesung **F8**) $\quad X_i \sim Po(\mu)$

$X_i \sim N(\mu, \sqrt{\mu})$ Normalapproximation (Vorlesung **F7**)
$\mu > 15$ (viele Ereignisse in jedem Intervall gefordert)

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

Verteilung für den Mittelwert aus n Observationen

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

Referensvariable (wenn μ im Nenner duch Schätzer ersetzt wird → Standardfehler)

$$\boxed{I_\mu = \bar{x} \pm \lambda_{\alpha/2} \cdot \sqrt{\frac{\bar{x}}{n}}}$$

KI für Erwartungswert μ in Poisson-Verteilung.
$\mu > 15$ erforderlich. Konfidenzniveau $1-\alpha$.
Punktschätzer $\pm$ Quantil $\cdot$ Standardfehler.

KI für Schätzung der Standardabweichung σ in $N(\mu, \sigma)$

Stichprobe $x = (x_1, x_2, \ldots x_n)$

Wenn alle X_i die gleiche Verteilung $X_i \sim N(\mu, \sigma)$ haben und unabhängig sind, dann gilt **allgemein**:

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

$\chi^2(n-1)$: Chi-Quadrat-Verteilung, für $n-1$ Freiheitsgrade

$$(n-1)\frac{S^2}{\sigma^2} \sim \chi^2(n-1) \quad \text{weil} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

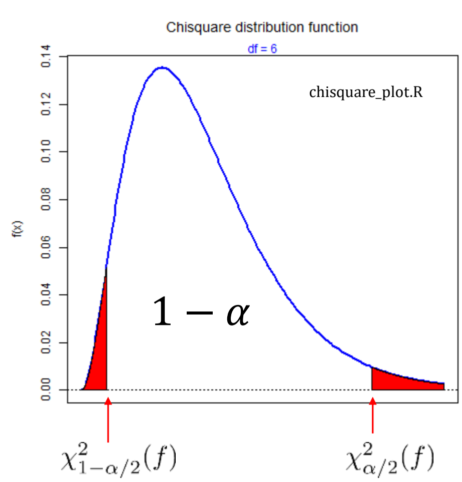
Dies ist bereits eine Referenzvariable (nur σ unbekannt, welches wir schätzen wollen).

1. Finde eine Referenzvariable
2. Schließe die Referenzvariable zwischen deren Quantilgrenzen ein
3. Schließe den zu schätzenden Parameter durch Umformen in der Klammer $P(\ldots)$ zwischen bekannten Größen ein

Achtung!: Chi-Quadrat-Verteilung ist **nicht** symmetrisch!

KI für Schätzung der Standardabweichung σ in $N(\mu, \sigma)$

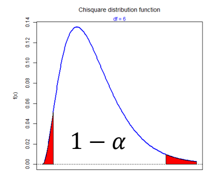
Achtung!: Chi-Quadrat-Verteilung ist **nicht** symmetrisch:



www.matstat.org

KI für Schätzung der Standardabweichung σ in $N(\mu, \sigma)$

$$P\left(\chi^2_{1-\alpha/2}(f) < (n-1)\frac{S^2}{\sigma^2} \le \chi^2_{\alpha/2}(f)\right) = 1 - \alpha$$



$\Downarrow$ umformen ...

$$P\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2}(f)}} < \sigma \le \sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(f)}}\right) = 1 - \alpha \qquad f = n - 1$$

$$\boxed{I_{\sigma^2} = \left(\frac{f \cdot s^2}{\chi^2_{\alpha/2}(f)}; \frac{f \cdot s^2}{\chi^2_{1-\alpha/2}(f)}\right)}$$

KI für Schätzung der Varianz σ^2 in der Normalverteilung $N(\mu, \sigma)$

$$\boxed{I_{\sigma} = \left(\sqrt{\frac{f}{\chi^2_{\alpha/2}(f)}} \cdot s; \sqrt{\frac{f}{\chi^2_{1-\alpha/2}(f)}} \cdot s\right)}$$

KI für Schätzung der Standardabweichung σ in einer Normalverteilung $N(\mu, \sigma)$

www.matstat.org

KI für Schätzung der Standardabweichung σ in $N(\mu, \sigma)$

Beispiel: Stichprobe $x = (x_1, x_2, \ldots x_n)$ $\quad X_i \sim N(\mu, \sigma)$

$s = 1.56$; $n = 10$ aus der Stichprobe

Suche: 95% KI für σ

$f = n - 1 = 9$; $\alpha = 0.05$

Tabelle

Tabelle

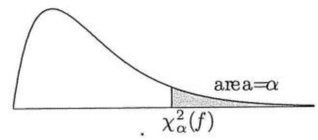
$$\chi^2_{1-\alpha/2}(f) = \chi^2_{0.975}(9) = 2.7 \qquad \chi^2_{\alpha/2}(f) = \chi^2_{0.025}(9) = 19$$

$$I_{\sigma} = \left(\sqrt{\frac{9}{19}} \cdot 1.56\ ;\ \sqrt{\frac{9}{2.7}} \cdot 1.56\right) = \underline{\underline{(1.07\ ;\ 2.85)}}$$

Der wahre Wert der Standardabweichung wird mit der Wahrscheinlichkeit 95% vom Intervall (1.07 ; 2.85) umfasst.

www.matstat.org

Tabell 4. χ^2**-fördelningen**
$P(X > \chi^2_\alpha(f))$, där $X \in \chi^2(f)$.



wir finden:

$\chi^2_{0.975}(9) = 2.7$

$\chi^2_{0.025}(9) = 19$

f \ α	0.9995	0.999	0.995	0.99	0.975	0.95	0.05	0.025	0.01	0.005	0.001	0.0005
1	0.00	0.00	0.00	0.00	0.00	0.00	3.84	5.02	6.63	7.88	10.8	12.1
2	0.00	0.00	0.01	0.02	0.05	0.10	5.99	7.38	9.21	10.6	13.8	15.2
3	0.02	0.02	0.07	0.11	0.22	0.35	7.81	9.35	11.3	12.8	16.3	17.7
4	0.06	0.09	0.21	0.30	0.48	0.71	9.49	11.1	13.3	14.9	18.5	20.0
5	0.16	0.21	0.41	0.55	0.83	1.15	11.1	12.8	15.1	16.7	20.5	22.1
6	0.30	0.38	0.68	0.87	1.24	1.64	12.6	14.4	16.8	18.5	22.5	24.1
7	0.48	0.60	0.99	1.24	1.69	2.17	14.1	16.0	18.5	20.3	24.3	26.0
8	0.71	0.86	1.34	1.65	2.18	2.73	15.5	17.5	20.1	22.0	26.1	27.9
9	0.97	1.15	1.73	2.09	2.70	3.33	16.9	19.0	21.7	23.6	27.9	29.7
10	1.26	1.48	2.16	2.56	3.25	3.94	18.3	20.5	23.2	25.2	29.6	31.4
11	1.59	1.83	2.60	3.05	3.82	4.57	19.7	21.9	24.7	26.8	31.3	33.1
12	1.93	2.21	3.07	3.57	4.40	5.23	21.0	23.3	26.2	28.3	32.9	34.8
13	2.31	2.62	3.57	4.11	5.01	5.89	22.4	24.7	27.7	29.8	34.5	36.5
14	2.70	3.04	4.07	4.66	5.63	6.57	23.7	26.1	29.1	31.3	36.1	38.1
15	3.11	3.48	4.60	5.23	6.26	7.26	25.0	27.5	30.6	32.8	37.7	39.7
16	3.54	3.94	5.14	5.81	6.91	7.96	26.3	28.8	32.0	34.3	39.3	41.3
17	3.98	4.42	5.70	6.41	7.56	8.67	27.6	30.2	33.4	35.7	40.8	42.9
18	4.44	4.90	6.26	7.01	8.23	9.39	28.9	31.5	34.8	37.2	42.3	44.4
19	4.91	5.41	6.84	7.63	8.91	10.1	30.1	32.9	36.2	38.6	43.8	46.0
20	5.40	5.92	7.43	8.26	9.59	10.9	31.4	34.2	37.6	40.0	45.3	47.5

www.matstat.org

Einseitiges KI für Schätzung von μ in $N(\mu, \sigma)$

σ unbekannt, kleine Stichprobe, nach links begrenzt

$$\boxed{\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t(f)}$$

Referenzvariable $\quad X_i \sim N(\mu, \sigma)$

$f = n - 1$

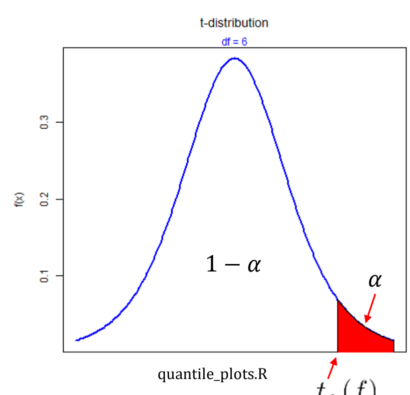
Es gilt (siehe Bild):

$$P\left(\frac{\bar{X} - \mu}{s/\sqrt{n}} \leq t_\alpha(f)\right) = 1 - \alpha$$

Nach Umformen wird das:

$$P\left(\mu > \bar{X} - t_\alpha(f) \cdot \frac{s}{\sqrt{n}}\right) = 1 - \alpha$$

Die Wahrscheinlichkeit, dass μ größer ist als dieser Ausdruck beträgt also $1 - \alpha$



www.matstat.org

Einseitiges KI für Schätzung von μ in $N(\mu, \sigma)$

σ unbekannt, kleine Stichprobe, nach links begrenzt

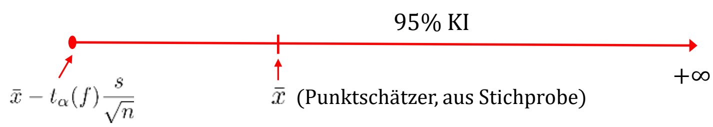
$$P\left(\mu > \bar{X} - t_\alpha(f) \cdot \frac{s}{\sqrt{n}}\right) = 1 - \alpha$$

Die Wahrscheinlichkeit, dass μ größer ist als dieser Ausdruck beträgt also $1 - \alpha$

$$\boxed{I_\mu = \left(\bar{x} - t_\alpha(f) \frac{s}{\sqrt{n}}\ ;\ +\infty\right)}$$

Einseitiges KI für Erwartungswert μ in $N(\mu, \sigma)$. Konfidenzniveau $1 - \alpha$

Das bedeutet: Wir wissen mit $(1 - \alpha) \cdot 100\%$ Wahrscheinlichkeit, dass μ größer als die untere Konfidenzgrenze ist.



Einseitiges KI für Schätzung von μ in $N(\mu, \sigma)$

σ unbekannt, kleine Stichprobe, nach rechts begrenzt

$$\boxed{\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t(f)}$$

Referenzvariable $X_i \sim N(\mu, \sigma)$

$f = n - 1$

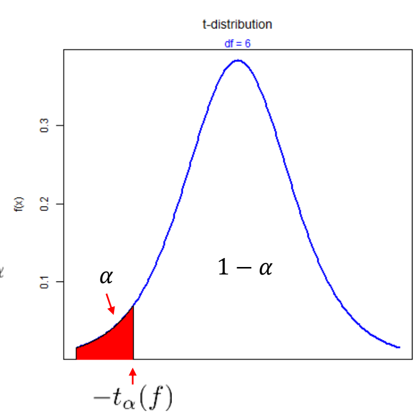
Es gilt (siehe Bild):

$$P\left(\frac{\bar{X} - \mu}{s/\sqrt{n}} > -t_\alpha(f)\right) = 1 - \alpha$$

Nach Umformen wird das:

$$P\left(\mu < \bar{X} + t_\alpha(f) \cdot \frac{s}{\sqrt{n}}\right) = 1 - \alpha$$

Die Wahrscheinlichkeit, dass μ kleiner ist als dieser Ausdruck ist, beträgt also $1 - \alpha$



Einseitiges KI für Schätzung von μ in $N(\mu, \sigma)$

σ unbekannt, kleine Stichprobe, nach rechts begrenzt

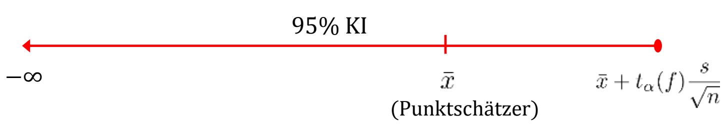
$$P\left(\mu < \bar{X} + t_\alpha(f) \cdot \frac{s}{\sqrt{n}}\right) = 1 - \alpha$$

Die Wahrscheinlichkeit dass μ kleiner als dieser Ausdruck ist, beträgt also $1 - \alpha$

$$\boxed{I_\mu = \left(-\infty \; ; \; \bar{x} + t_\alpha(f) \frac{s}{\sqrt{n}}\right)}$$

Einseitiges KI für Erwartungswert μ in $N(\mu, \sigma)$. Konfidenzniveau $1 - \alpha$

Dies bedeutet: Wir wissen mit $(1 - \alpha) \cdot 100\%$ Wahrscheinlichkeit, dass μ kleiner als die obere Konfidenzgrenze ist.



www.matstat.org

Anhang

Intervallschätzung

Uwe Menzel, 2018

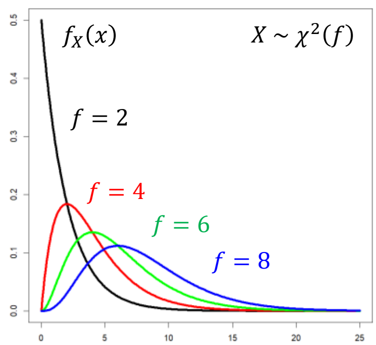
uwe.menzel@matstat.org

www.matstat.org

Die Chi-Quadrat-Verteilung

$$W = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sigma^2} = \frac{(n-1) \cdot S^2}{\sigma^2} \sim \chi^2(n-1)$$

Dies gilt allgemein wenn die X_i unabhängige, normalverteilte Zufallsvariablen mit der Standardabweichung σ sind.



Dichtefunktion von χ^2 für verschiedene Freiheitsgrade f.
Die Observationen X_i müssen von einer Normalverteilung kommen.

$$X_i \sim N(\mu, \sigma)$$

www.matstat.org

Die t-Verteilung

(Student'sche Verteilung)

Voraussetzungen: $\left.\begin{array}{l} Z \sim N(0,1) \\ W \sim \chi^2(\nu) \end{array}\right\}$ müssen unabhängig sein

Wenn die Zufallsvariablen Z und W obenstehende Verteilung haben, dann hat der folgende Quotient die sogenannte t-Verteilung:

$$\boxed{\frac{Z}{\sqrt{\frac{W}{\nu}}} \sim t(\nu)}$$

t-Verteilung mit ν Freiheitsgraden

www.matstat.org

Student'sche t-Verteilung

- $Z \sim N(0{,}1)$ standardisierte Normalverteilung
- $W \sim \chi^2(\nu)$ Chi-Quadrat-Verteilung, sei $\nu = n-1$ (Freiheitsgrade)
- Z und W unabhängig

$$Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1) \qquad W = (n-1)\frac{S^2}{\sigma^2} \sim \chi^2(n-1)$$

$$\frac{Z}{\sqrt{\frac{W}{\nu}}} = \frac{\frac{(\bar{X}-\mu_0)\cdot\sqrt{n}}{\sigma}}{\sqrt{\frac{(n-1)\cdot S^2}{\sigma^2\cdot(n-1)}}} = \frac{(\bar{X}-\mu_0)\cdot\sqrt{n}}{S} = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$$

Die linke Seite ist $t(n-1)$ – verteilt, also muss das auch für die rechte Seite gelten:

$$\frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t(n-1)$$

Die Zufallsvariablen X_i müssen normalverteilt und unabhängig sein.

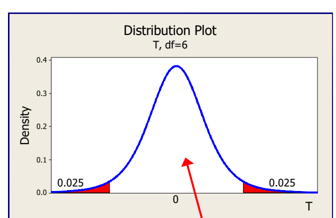
www.matstat.org

Schätzung von μ für $N(\mu, \sigma)$ wenn σ unbekannt ist und n klein ist

$X_i \sim N(\mu, \sigma)$; unabhängig

Referenzvariable:

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



$$P\left(-t_{\alpha/2}(n-1) \leq \frac{\bar{x}-\mu}{s/\sqrt{n}} \leq +t_{\alpha/2}(n-1)\right) = 1-\alpha$$

$$P\left(\bar{x} - t_{\alpha/2}(f)\cdot\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2}(f)\cdot\frac{s}{\sqrt{n}}\right) = 1-\alpha$$

$$I_\mu = \bar{x} \pm t_{\alpha/2}(f)\cdot\frac{s}{\sqrt{n}} \qquad f = n-1 \qquad t_{\alpha/2}(f) > \lambda_{\alpha/2}$$

(breiteres Konfidenzintervall wenn n klein ist und σ unbekannt ist)

www.matstat.org

Schätzung der Differenz zwischen zwei Mittelwerten von N wenn σ_x und σ_y unbekannt, aber gleich sind $(= \sigma)$

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma^2}{n_x} + \frac{\sigma^2}{n_y}}} \sim N(0,1)$$

keine Referenzvariable, denn σ ist unbekannt, $(\mu_x - \mu_y)$ soll geschätzt werden

$$W = \frac{(n_x - 1) \cdot S_x^2 + (n_y - 1) \cdot S_y^2}{\sigma^2} \sim \chi^2 (n_x + n_y - 2)$$

Forderungen: $\sigma_x = \sigma_y = \sigma$; die X_i müssen unabhängig und normalverteilt sein

$$T = \frac{Z}{\sqrt{\frac{W}{n_x + n_y - 2}}} = \cdots = \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{S_p \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim t(n_x + n_y - 2)$$

Referenzvariable

$$S_p^2 = \frac{(n_x - 1) \cdot S_x^2 + (n_y - 1) \cdot S_y^2}{(n_x - 1) + (n_y - 1)}$$

"pooled variance"

Schätzung der Differenz zwischen zwei Mittelwerten von N wenn σ_x und σ_y unbekannt, aber gleich sind $(= \sigma)$

$$\frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{S_p \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim t(f) \qquad f = n_x + n_y - 2$$

$$P\left[-t_{\alpha/2}(f) \leq \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{S_p \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \leq +t_{\alpha/2}(f)\right] = 1 - \alpha$$

$$P\left[\bar{X} - \bar{Y} - t_{\alpha/2}(f) \cdot S_p \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \leq \mu_x - \mu_y \leq \bar{X} - \bar{Y} + t_{\alpha/2}(f) \cdot S_p \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}\right] = 1 - \alpha$$

$$I_{\Delta\mu} = \bar{x} - \bar{y} \pm t_{\alpha/2}(f) \cdot s_p \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}$$

$(1 - \alpha) \cdot 100\%$ KI für Unterschied zwischen den beiden Mittelwerten

$I_{\Delta\mu}$ = Punktschätzer + Quantil · Standardfehler

KI für die Schätzung der Differenz zwischen den den Mittelwerten zweier gepaarter Stichproben von N

$$\left.\begin{array}{l} X_i \sim N(\mu_i, \sigma_x) \\ Y_i \sim N(\mu_i + \Delta\mu, \sigma_y) \end{array}\right\}$$ **Modell**: systematischer Unterschied $\Delta\mu$ zwischen den Gruppen

$$\left.\begin{array}{l} Z_i = Y_i - X_i \\ Z_i \sim N(\Delta\mu, \sigma_z) \end{array}\right\}$$ Differenzen Z_i sowie deren Verteilung

$$\sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2 - 2 \cdot \rho \cdot \sigma_x \sigma_y} \qquad C(X, Y) = \rho \cdot \sigma_x \sigma_y$$ Kovarianz

$\rho =$ Korrelationskoeffizient. Dieser ist positiv, wenn die X_i und Y_i den gleichen Trend haben (also gemeinsam steigen oder fallen). Damit wird die Varianz für die Variable Z kleiner als für unabhängige Gruppen (wo $\rho = 0$ gelten würde).

$$\bar{Z} \sim N\left(\Delta\mu, \frac{\sigma_z}{\sqrt{n}}\right)$$ Punktschätzer, konsistent, erwartungstreu (der Mittelwert der Z_i ist also eine gute Schätzung für $\Delta\mu$)

$$V(\bar{Z}) = \frac{1}{n} \cdot (\sigma_x^2 + \sigma_y^2 - 2 \cdot \rho \cdot \sigma_x \sigma_y) = \frac{\sigma_z^2}{n}$$

www.matstat.org

KI für die Schätzung der Differenz zwischen den Mittelwerten zweier gepaarter Stichproben von N

$$\bar{Z} \sim N\left(\Delta\mu, \frac{\sigma_z}{\sqrt{n}}\right) \Longrightarrow \frac{\bar{Z} - \Delta\mu}{\sigma_z / \sqrt{n}} \sim N(0, 1)$$ **keine** Referenzvariable, σ_z unbekannt (wir werden also zu einer t-Verteilung kommen ...)

$Z_i \sim N(\Delta\mu, \sigma_z)$ und Z_i unabhängig $\Longrightarrow \frac{1}{\sigma_z^2} \cdot \sum_i (Z_i - \bar{Z})^2 \sim \chi^2(n-1)$

$$W = (n-1) \cdot \frac{S_z^2}{\sigma_z^2} \sim \chi^2(n-1) \quad \text{mit} \quad S_z^2 = \frac{1}{n-1} \cdot \sum_i (Z_i - \bar{Z})^2$$

$$\frac{\bar{Z}}{\sqrt{\frac{W}{n-1}}} = \cdots = \frac{\bar{Z} - \Delta\mu}{S_z / \sqrt{n}} \sim t(n-1)$$ Referensvariable (Rechnung analog t-Test für kleine Stichprobe)

www.matstat.org

KI für die Schätzung der Differenz zwischen den Mittelwerten zweier gepaarter Stichproben von *N*

$$\frac{\bar{Z} - \Delta\mu}{S_z/\sqrt{n}} \sim t(n-1)$$ Referensvariable

1. Finde eine Referenzvariable
2. Schließe die Referenzvariable zwischen deren Quantilgrenzen ein
3. Schließe den zu schätzenden Parameter durch Umformen in der Klammer $P(\dots)$ zwischen bekannten Größen ein

$$P\left[-t_{\alpha/2}(n-1) \le \frac{\bar{Z} - \Delta\mu}{S_z/\sqrt{n}} \le +t_{\alpha/2}(n-1)\right] = 1-\alpha$$

$$P\left[\bar{Z} - t_{\alpha/2}(f) \cdot \frac{S_z}{\sqrt{n}} \le \Delta\mu \le \bar{Z} + t_{\alpha/2}(f) \cdot \frac{S_z}{\sqrt{n}}\right] = 1-\alpha$$

$$I_{\Delta\mu} = \bar{z} \pm t_{\alpha/2}(f) \cdot \frac{s_z}{\sqrt{n}} \qquad f = n-1$$ Freiheitsgrade

www.matstat.org