Statistik

HA 1.6.2016:
78, 79, 80, 81a, 86, 89

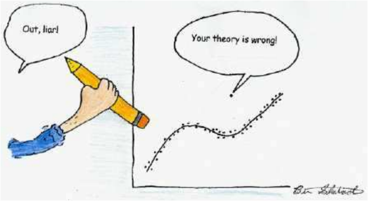
für Betriebswirtschaft, Internationales Management, Wirtschaftsinformatik und Informatik

Sommersemester 2016

Veranstaltungen zur Statistik für BW/IM Sommersemester 2016				
Was?	Wer?	Wann?	Wo?	ab wann?
Vorlesung Statistik	Etschberger	Mi, 14.00-17.00	B2.14	16.03.2016
Vorlesung Statistik PLUS	Etschberger/Jansen	- Blocktermin -	?	?
Übung Statistik	Etschberger	Mi, 17.00-18.30	A1.10	30.03.2016
Übung Statistik	Jansen	Di, 11.30-13.00	W1.06	22.03.2016
Übung Statistik	Jansen	Di, 14.00-15.30	W2.14	22.03.2016
Übung Statistik	Jansen	Mi, 11.30-13.00	W2.11	30.03.2016
Übung Statistik	Jansen	Do, 14.00-15.30	W2.14	31.03.2016
Übung Statistik	Schneller	Do, 14.00-15.30	W3.03	31.03.2016
Übung Statistik	Schneller	Do, 15.30-17.00	W3.03	31.03.2016
Übung Statistik	Wins	Di, 14.00-15.30	J3.19	22.03.2016
Übung Statistik	Wins	Di, 15.30-17.00	J3.19	22.03.2016
Offener Statistikraum	Jansen/Tutoren	Mo, 14.00-17.45	B3.05	04.04.2016
Veranstaltungen für Teilnehmer der WiMa-Klausur im Juli 2016				
Was?	Wer?	Wann?	Wo?	ab wann?
Tutorium Mathematik	Burkart	Do 13.30-15.00	W1.06	07.04.2016
Tutorium Mathematik	Burkart	Do 15.00-16.15	W1.06	07.04.2016
Offener Matheraum	Etschberger/Tutoren	Fr 11.30-15.30	B3.05	29.04.2016

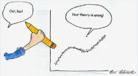
Datum	Statistik für IM/BW	Nr.
Mittwoch, 16. März 2016	Einführung, R Installation, Rstudio Einführung, Skalen	1
Mittwoch, 23. März 2016	univ. deskr. Stat., Quantile, Plots	2
Mittwoch, 30. März 2016	Streuung, Konzentrationsmaße	3
Mittwoch, 6. April 2016	Kontingenztabellen, Mosaikplots, Korrelation	4
Mittwoch, 13. April 2016	Preisindizes, lineare Regression	5
Mittwoch, 20. April 2016	Kombinatorik, Wahrscheinlichkeit	6
Mittwoch, 27. April 2016	Wahrscheinlichkeit	7
Mittwoch, 4. Mai 2016	Pyramid	
Mittwoch, 11. Mai 2016	diskrete Zufallsvariablen	8
Mittwoch, 18. Mai 2016	Stetige ZV, Gleichverteilung	9
Mittwoch, 25. Mai 2016	Normalverteilung, Verteilungsparameter	10
Mittwoch, 1. Juni 2016	Schätzfunktionen und Punktschätzer, Konfidenzintervalle	11
Mittwoch, 8. Juni 2016	Konfidenzintervalle, Tests	12
Mittwoch, 15. Juni 2016	Go-Out-Tag	
Mittwoch, 22. Juni 2016	Tests, WH, (Fragen zur Probekl. In den Übungsgruppen)	13
Mittwoch, 29. Juni 2016	AW Prüfungswoche	

Prof. Dr. Stefan Etschberger
Hochschule Augsburg



Out, liar!
Your theory is wrong!

Grundlagen der induktiven Statistik



- Vollerhebung of unmöglich,
- Deshalb: Beobachte Teilgesamtheit und schließe auf Grundgesamtheit

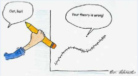
Beispiel

Warensendung von 1000 Stück; darunter M Stück Ausschuss.
M ist unbekannt.
$\rightarrow$ Zufällige Entnahme von $n = 30$ Stück („Stichprobe“).
Darunter 2 Stück Ausschuss.
Denkbare Zielsetzungen:

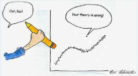
- Schätze M durch eine Zahl (z.B. $\frac{2}{30} \cdot 1000 = 66{,}67$)
- Schätze ein Intervall für M (z.B. $M \in [58; 84]$)
- Teste die Hypothese, dass $M > 50$ ist.

Grundbegriffe

- Grundgesamtheit (G): Menge aller relevanten Merkmalsträger.
- Verteilung von G: $F(x) = P(X \leq x) =$ Wahrscheinlichkeit, dass ein Merkmalsträger ausgewählt wird, der beim untersuchten Merkmal maximal die Ausprägung x aufweist.
- Uneingeschränkte (reine) Zufallsauswahl:
  Jedes Element von G hat die selbe Chance, ausgewählt zu werden.
- Stichprobenumfang (n): Anzahl der Merkmalsträger in der Stichprobe.
- Einfache Stichprobe:
  Uneingeschränkte Zufallsauswahl und unabhängige Ziehung.
  $\rightarrow$ Alle Stichprobenvariablen $X_1, \ldots, X_n$ sind iid.
- Stichprobenergebnis:
  n-Tupel der Realisationen der Stichprobenvariablen, $(x_1, \ldots, x_n)$.



Punkt-Schätzung



theta

- Ein unbekannter Parameter ϑ der Verteilung von G soll auf Basis einer Stichprobe geschätzt werden.
- Zum Beispiel: σ von $N(10; \sigma)$
- Schätzwert: $\hat{\vartheta}$
- Vorgehen: Verwendung einer Schätzfunktion

$$\hat{\Theta} = g(X_1, \ldots, X_n)$$

Beachte: Der Schätzwert $\hat{\vartheta}$ ist die Realisierung der ZV (!) $\hat{\Theta}$.

- Frage: Welche Stichprobenfunktion ist zur Schätzung geeignet?

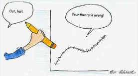
⇒ Kriterien für die Beurteilung/Konstruktion von Schätzfunktionen!

- Im Folgenden: Vorliegen einer einfachen Stichprobe, d.h. $X_1, \ldots, X_n$ iid.

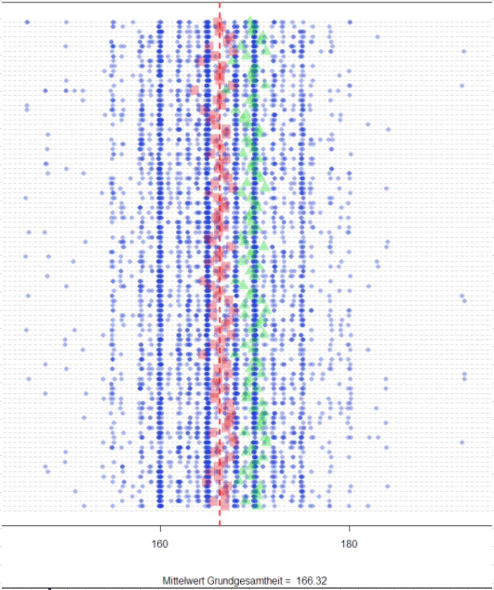
Beispiel

- Schätzen des Mittelwertes einer Grundgesamtheit
- dazu: Einfache Stichprobe vom Umfang 5
- und den beiden Stichprobenfunktionen

$$\hat{\Theta}_1 = \frac{1}{n} \sum_{i=1}^{n} X_i, \qquad \hat{\Theta}_2 = \frac{1}{n-1} \sum_{i=1}^{n} X_i$$



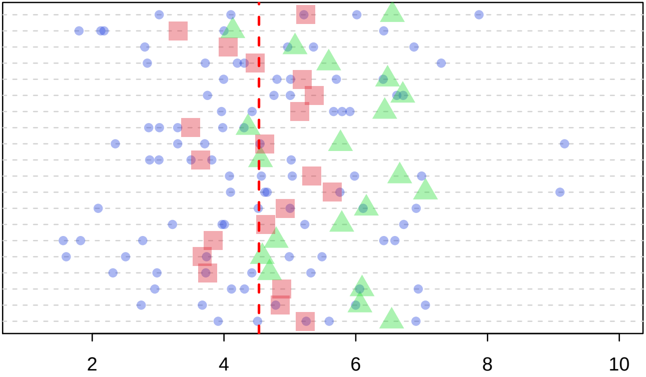
```
# -------------------------------------------------
# Lade Daten
D <- read.csv("http://goo.gl/KCGF16", sep=";", dec=",")

# Groesse der Mütter (Lösche fehlende Werte und korrigiere Tippfehler)
Groesse = na.omit(D$GroesseM)
Groesse[Groesse < 100] = 100 + Groesse[Groesse < 100]

# Arithm. Mittel und (emp.) Standardabweichung als
# Größen der Grundgesamtheit
mu = mean(Groesse)
sigma = sqrt(mean((Groesse-mean(Groesse))^2))
data.frame(sigma, mu)


# -------------------------------------------
zeichneStichproben = function(n=10, N=20,
                              zeichneMu=TRUE, zeichneTheta=TRUE,
                              seed=1,
                              Daten=rnorm(n=100,
                                          sd = runif(1,0,10),
                                          mean = runif(1,0,20))
                              ) {

  # aus der Grundgesamtheit
  mu = mean(Daten)
  sigma = sqrt(mean((Daten - mu)^2))

  A = as.data.frame(matrix(nrow=N,ncol=n))

  set.seed=seed
  for(i in 1:N) {
    A[i,] = sample(Daten, size = n, replace=TRUE)
  }

  min=min(A) - (max(A)-min(A))*0.5
  max=max(A) + (max(A)-min(A))*0.5

  plot(c(min,max), c(1, N), type="n", xlab="", yaxt="n", ylab="")
  abline(h=0:N, col="lightgray", lty=2)
  for(i in 1:N) {
    x = as.numeric(A[i,])
    points(jitter(x), rep(i, n), pch=16, col="#2040dd60")
    if(zeichneTheta) {
      Th1 = mean(x)
      Th2 = mean(x)*n/(n-1)
      points(Th1, i, cex=2, pch=15, col="#dd203060")
      points(Th2, i, cex=2, pch=17, col="#20dd3060")
    }
  }

  if(zeichneMu) {
    abline(v=mu, col="red", lwd=2, lty=2)
    title(sub=paste("Mittelwert Grundgesamtheit = ", round(mu, 2)))
  }
}

n=50
zeichneStichproben(n=n, N=1, zeichneMu=FALSE, zeichneTheta=FALSE, Daten=Groesse)
zeichneStichproben(n=n, N=1, zeichneMu=FALSE, zeichneTheta=TRUE, Daten=Groesse)
zeichneStichproben(n=n, N=5, zeichneMu=FALSE, zeichneTheta=TRUE, Daten=Groesse)
zeichneStichproben(n=n, N=100, zeichneMu=FALSE, zeichneTheta=TRUE, Daten=Groesse)
zeichneStichproben(n=n, N=5, zeichneMu=TRUE, zeichneTheta=TRUE, Daten=Groesse)
zeichneStichproben(n=n, N=100, zeichneMu=TRUE, zeichneTheta=TRUE, Daten=Groesse)
```



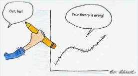
Beispiel

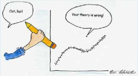
- Schätzen des Mittelwertes einer Grundgesamtheit
- dazu: Einfache Stichprobe vom Umfang 5
- und den beiden Stichprobenfunktionen

$$\hat{\Theta}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \hat{\Theta}_2 = \frac{1}{n-1}\sum_{i=1}^{n} X_i$$



Mittelwert Grundgesamtheit = 4.53





Erwartungstreue und Wirksamkeit

- Eine Schätzfunktion $\hat{\Theta} = g(X_1, \ldots, X_n)$ heißt erwartungstreu oder unverzerrt für ϑ, wenn unabhängig vom numerischen Wert von ϑ gilt:
$$E(\hat{\Theta}) = \vartheta$$

Beispiel

Sind $\hat{\Theta}_1 = \bar{X}$, $\hat{\Theta}_2 = \frac{X_1 + X_n}{2}$, $\hat{\Theta}_3 = \frac{1}{n-1} \sum\limits_{i=1}^{n} X_i$ erwartungstreu für μ?

a) $\hat{\Theta}_1$: $\quad E(\bar{X}) = \mu$
$\Rightarrow \hat{\Theta}_1$ ist erwartungstreu.

b) $\hat{\Theta}_2$: $\quad E\left(\frac{X_1 + X_n}{2}\right) = \frac{1}{2}[E(X_1) + E(X_n)] = \frac{1}{2}(\mu + \mu) = \mu$
$\Rightarrow \hat{\Theta}_2$ ist erwartungstreu.

c) $\hat{\Theta}_3$: $E\left(\frac{1}{n-1} \sum\limits_{i=1}^{n} X_i\right) = \frac{1}{n-1} \sum\limits_{i=1}^{n} E(X_i) = \frac{1}{n-1} \sum\limits_{i=1}^{n} \mu = \frac{n}{n-1} \mu \neq \mu$
$\Rightarrow \hat{\Theta}_3$ ist nicht erwartungstreu

Erwartungstreue von $\hat{\Theta}_i$: großes Theta

$$E[\hat{\Theta}_1] = E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} \underbrace{E[X_i]}_{=\mu} = \frac{1}{n}\cdot(\mu+\mu+\dots+\mu)$$

$$= \frac{1}{n}\cdot n\cdot\mu = \mu \quad \Rightarrow \hat{\Theta}_1 \text{ ist erwartungstreu}$$

$$E[\hat{\Theta}_2] = E\left[\tfrac{1}{2}(X_1+X_n)\right] = \tfrac{1}{2}\cdot\Big(\underbrace{E[X_1]}_{=\mu}+\underbrace{E[X_n]}_{=\mu}\Big)$$

$$= \tfrac{1}{2}\cdot(\mu+\mu) = \mu \quad \Rightarrow \hat{\Theta}_2 \text{ ist erw. treu}$$

$$E[\hat{\Theta}_3] = E\left[\frac{1}{n-1}\sum_{i=1}^{n} X_i\right] = \frac{1}{n-1}\cdot\Big(\underbrace{E[X_1]}_{=\mu}+\dots+\underbrace{E[X_n]}_{=\mu}\Big)$$

$$= \frac{1}{n-1}\cdot n\cdot\mu = \frac{n}{n-1}\mu > \mu \quad \Rightarrow \hat{\Theta}_3 \text{ ist nicht erw. treu}$$

Vorüberlegung: $\mathrm{Var}[\bar{X}] = E[(\bar{X} - E[\bar{X}])^2]$

$$= \mathrm{Var}\left[\frac{1}{n}\sum X_i\right] = \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} \underbrace{\mathrm{Var}[X_i]}_{=\sigma^2}$$

da Stichprobe iid

(Varianz der G.G.)

$$= \frac{1}{n^2}\cdot(\sigma^2+\sigma^2+\dots+\sigma^2) = \frac{1}{n^2}\cdot n\cdot\sigma^2 = \frac{\sigma^2}{n}$$

Gesucht: Erwartungstreue Schätzfunktion $\hat{\Theta}$ für $\mathrm{Var}[X_i] = \sigma^2$ aus Grundges.

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

$$? \; E[\hat{\Theta}] \overset{?}{=} \sigma^2 \; ?$$

$$E[\hat{\Theta}] = E\left[\frac{1}{n}\sum (X_i - \bar{X})^2\right]$$

$$= \frac{1}{n} E\left[\sum (X_i - \mu + \mu - \bar{X})^2\right]$$

$$= \frac{1}{n} E\left[\sum \left((X_i-\mu)^2 + 2\cdot(X_i-\mu)(\mu-\bar{X}) + (\mu-\bar{X})^2\right)\right]$$

$$= \frac{1}{n}\cdot E\Big[\sum (X_i-\mu)^2 + 2\cdot(\mu-\bar{X})\cdot\underbrace{\sum (X_i-\mu)}_{=n\cdot(\bar{X}-\mu)} + n(\mu-\bar{X})^2\Big]$$

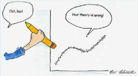
$$= \frac{1}{n}\sum_{i=1}^{n}\underbrace{\mathrm{Var}[X_i]}_{=\sigma^2} + E\Big[\underbrace{-2n(\bar{X}-\mu)^2 + n(\bar{X}-\mu)^2}_{=-n(\bar{X}-\mu)^2}\Big]$$

$$= \frac{1}{n}\, n\cdot\sigma^2 - \mathrm{Var}[\bar{X}] = \sigma^2 - \frac{\sigma^2}{n} = \sigma^2\cdot\left(1-\frac{1}{n}\right)$$

$$\Rightarrow \quad E[\hat{\Theta}] = \sigma^2\cdot\frac{n-1}{n}$$

$\Rightarrow$ Erwartungstreue Schätzfunktion für σ^2 ist

$$\boxed{s^2 = \frac{1}{n-1}\sum (X_i - \bar{X})^2}$$



Erwartungstreue und Wirksamkeit

- Welche der erwartungstreuen Schätzfunktionen $\widehat{\Theta}_1, \widehat{\Theta}_2$ ist „besser“?
- Von zwei erwartungstreuen Schätzfunktionen $\widehat{\Theta}_1, \widehat{\Theta}_2$ für ϑ heißt $\widehat{\Theta}_1$ wirksamer als $\widehat{\Theta}_2$, wenn unabhängig vom numerischen Wert von ϑ gilt:

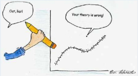
$$\operatorname{Var}(\widehat{\Theta}_1) < \operatorname{Var}(\widehat{\Theta}_2)$$

Beispiel: ($\widehat{\Theta}_1 = \bar{X}$, $\widehat{\Theta}_2 = \frac{X_1 + X_n}{2}$)
Wegen

$$\left.\begin{aligned} \operatorname{Var}(\widehat{\Theta}_1) &= \operatorname{Var}(\bar{X}) &&= \frac{\sigma^2}{n} \\ \operatorname{Var}(\widehat{\Theta}_2) &= \operatorname{Var}\left(\frac{X_1+X_n}{2}\right) = \frac{1}{4}(\sigma^2 + \sigma^2) &&= \frac{\sigma^2}{2} \end{aligned}\right\} \Rightarrow \operatorname{Var}(\widehat{\Theta}_1) < \operatorname{Var}(\widehat{\Theta}_2)$$

(falls $n > 2$) ist $\widehat{\Theta}_1$ wirksamer als $\widehat{\Theta}_2$.

Wichtige Stichprobenfunktionen

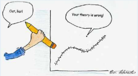


- Gegeben: Einfache Stichprobe $X_1, \ldots, X_n$, Beliebige Verteilung, mit $E(X_i) = \mu$, $\text{Var}(X_i) = \sigma^2$

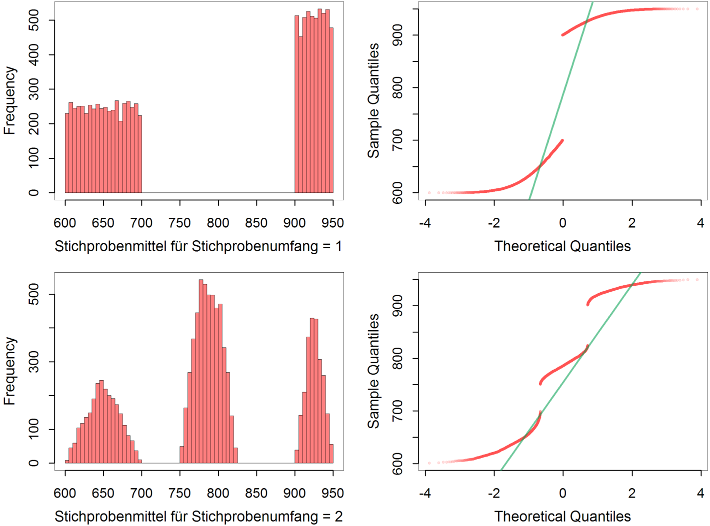
Stichprobenfunktion V	**Bezeichnung**	**E(V)**	**Var(V)**
$\sum_{i=1}^{n} X_i$	Merkmalssumme	$n\mu$	$n\sigma^2$
$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$	Stichprobenmittel	μ	$\frac{\sigma^2}{n}$
$\frac{\bar{X}-\mu}{\sigma}\sqrt{n}$	Gauß-Statistik	0	1
$\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$	mittlere quadratische Abweichung bezüglich μ	σ^2	
$\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$	mittlere quadratische Abweichung	$\frac{n-1}{n}\sigma^2$	
$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$	Stichprobenvarianz	σ^2	
$S = \sqrt{S^2}$	Stichproben-Standardabweichung		
$\frac{\bar{X}-\mu}{S}\sqrt{n}$	t-Statistik		

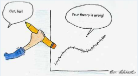
$E[\bar{x}] = \mu$
$\text{Std}[\bar{x}] = \sqrt{\text{Var}[\bar{x}]} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$

Auswirkungen der Stichprobengröße

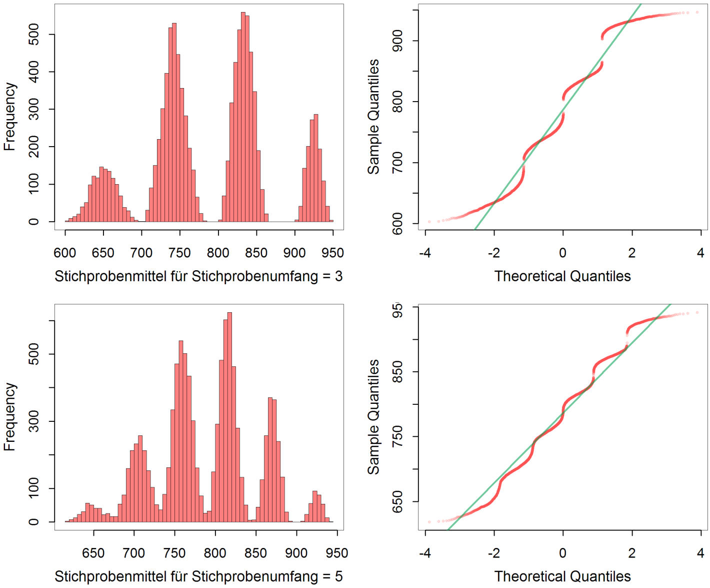


Ziehen von 10.000 Stichproben (jeweils vom Umfang n) und Berechnung der Stichprobenmittel (Verteilung: zwei überlagerte Gleichverteilungen):

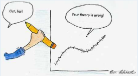


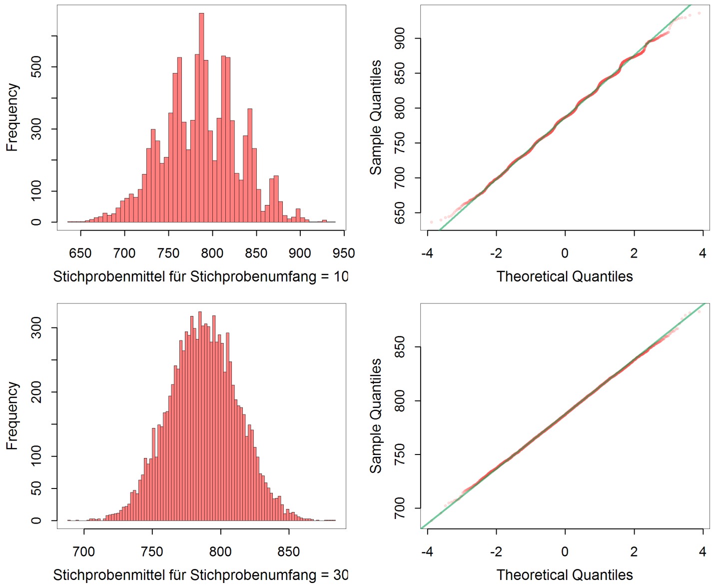


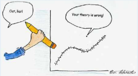
Auswirkungen der Stichprobengröße



Auswirkungen der Stichprobengröße





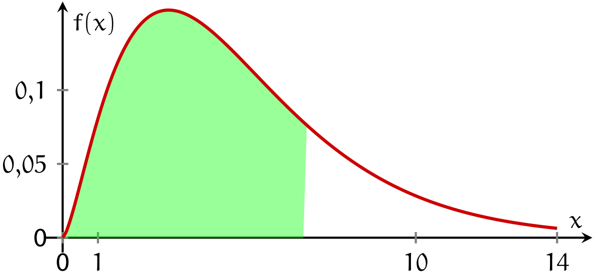


Chi-Quadrat-Verteilung

- Sind $X_1, \ldots, X_n$ iid $N(0;1)$-verteilte Zufallsvariablen, so wird die Verteilung von

$$Z = \sum_{i=1}^{n} X_i^2$$

als Chi-Quadrat-Verteilung mit n Freiheitsgraden bezeichnet.

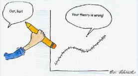


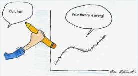
- Kurzschreibweise: $Z \sim \chi^2(n)$
- **Beispiel:** $\chi^2(30)$: $x_{0,975} = 46{,}98$

Quantilstabelle der χ^2-Verteilung mit n Freiheitsgraden

$\alpha \backslash n$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0.005	0.00	0.01	0.07	0.21	0.41	0.68	0.99	1.34	1.73	2.16	2.60	3.07	3.56	4.07	4.60
0.01	0.00	0.02	0.11	0.30	0.55	0.87	1.24	1.65	2.09	2.56	3.05	3.57	4.11	4.66	5.23
0.025	0.00	0.05	0.22	0.48	0.83	1.24	1.69	2.18	2.70	3.25	3.82	4.40	5.01	5.63	6.26
0.05	0.00	0.10	0.35	0.71	1.15	1.64	2.17	2.73	3.33	3.94	4.57	5.23	5.89	6.57	7.26
0.1	0.02	0.21	0.58	1.06	1.61	2.20	2.83	3.49	4.17	4.87	5.58	6.30	7.04	7.79	8.55
0.2	0.06	0.45	1.01	1.65	2.34	3.07	3.82	4.59	5.38	6.18	6.99	7.81	8.63	9.47	10.31
0.25	0.10	0.58	1.21	1.92	2.67	3.45	4.25	5.07	5.90	6.74	7.58	8.44	9.30	10.17	11.04
0.4	0.28	1.02	1.87	2.75	3.66	4.57	5.49	6.42	7.36	8.30	9.24	10.18	11.13	12.08	13.03
0.5	0.45	1.39	2.37	3.36	4.35	5.35	6.35	7.34	8.34	9.34	10.34	11.34	12.34	13.34	14.34
0.6	0.71	1.83	2.95	4.04	5.13	6.21	7.28	8.35	9.41	10.47	11.53	12.58	13.64	14.69	15.73
0.75	1.32	2.77	4.11	5.39	6.63	7.84	9.04	10.22	11.39	12.55	13.70	14.85	15.98	17.12	18.25
0.8	1.64	3.22	4.64	5.99	7.29	8.56	9.80	11.03	12.24	13.44	14.63	15.81	16.98	18.15	19.31
0.9	2.71	4.61	6.25	7.78	9.24	10.64	12.02	13.36	14.68	15.99	17.27	18.55	19.81	21.06	22.31
0.95	3.84	5.99	7.81	9.49	11.07	12.59	14.07	15.51	16.92	18.31	19.68	21.03	22.36	23.68	25.00
0.975	5.02	7.38	9.35	11.14	12.83	14.45	16.01	17.53	19.02	20.48	21.92	23.34	24.74	26.12	27.49
0.99	6.63	9.21	11.34	13.28	15.09	16.81	18.48	20.09	21.67	23.21	24.73	26.22	27.69	29.14	30.58
0.995	7.88	10.60	12.84	14.86	16.75	18.55	20.28	21.95	23.59	25.19	26.76	28.30	29.82	31.32	32.80

$\alpha \backslash n$	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
0.005	5.14	5.70	6.26	6.84	7.43	8.03	8.64	9.26	9.89	10.52	11.16	11.81	12.46	13.12	13.79
0.01	5.81	6.41	7.01	7.63	8.26	8.90	9.54	10.20	10.86	11.52	12.20	12.88	13.56	14.26	14.95
0.025	6.91	7.56	8.23	8.91	9.59	10.28	10.98	11.69	12.40	13.12	13.84	14.57	15.31	16.05	16.79
0.05	7.96	8.67	9.39	10.12	10.85	11.59	12.34	13.09	13.85	14.61	15.38	16.15	16.93	17.71	18.49
0.1	9.31	10.09	10.86	11.65	12.44	13.24	14.04	14.85	15.66	16.47	17.29	18.11	18.94	19.77	20.60
0.2	11.15	12.00	12.86	13.72	14.58	15.44	16.31	17.19	18.06	18.94	19.82	20.70	21.59	22.48	23.36
0.25	11.91	12.79	13.68	14.56	15.45	16.34	17.24	18.14	19.04	19.94	20.84	21.75	22.66	23.57	24.48
0.4	13.98	14.94	15.89	16.85	17.81	18.77	19.73	20.69	21.65	22.62	23.58	24.54	25.51	26.48	27.44
0.5	15.34	16.34	17.34	18.34	19.34	20.34	21.34	22.34	23.34	24.34	25.34	26.34	27.34	28.34	29.34
0.6	16.78	17.82	18.87	19.91	20.95	21.99	23.03	24.07	25.11	26.14	27.18	28.21	29.25	30.28	31.32
0.75	19.37	20.49	21.60	22.72	23.83	24.93	26.04	27.14	28.24	29.34	30.43	31.53	32.62	33.71	34.80
0.8	20.47	21.61	22.76	23.90	25.04	26.17	27.30	28.43	29.55	30.68	31.79	32.91	34.03	35.14	36.25
0.9	23.54	24.77	25.99	27.20	28.41	29.62	30.81	32.01	33.20	34.38	35.56	36.74	37.92	39.09	40.26
0.95	26.30	27.59	28.87	30.14	31.41	32.67	33.92	35.17	36.41	37.65	38.89	40.11	41.34	42.56	43.77
0.975	28.85	30.19	31.53	32.85	34.17	35.48	36.78	38.08	39.36	40.65	41.92	43.19	44.46	45.72	46.98
0.99	32.00	33.41	34.81	36.19	37.57	38.93	40.29	41.64	42.98	44.31	45.64	46.96	48.28	49.59	50.89
0.995	34.27	35.72	37.16	38.58	40.00	41.40	42.80	44.18	45.56	46.93	48.29	49.64	50.99	52.34	53.67





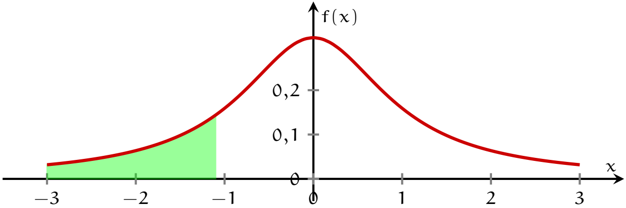
Testverteilungen: t-Verteilung

- Ist $X \sim N(0;1)$, $Z \sim \chi^2(n)$, X, Z unabhängig, so wird die Verteilung von

$$T = \frac{X}{\sqrt{\frac{1}{n} Z}}$$

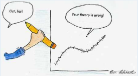
als t-Verteilung mit n Freiheitsgraden bezeichnet.

William Sealy Gosset
1876 – 1937



- Kurzschreibweise: $T \sim t(n)$
- **Beispiel:** $t(10)$ $x_{0,6} = 0{,}260$, $x_{0,5} = 0$, $x_{0,1} = -x_{0,9} = -1{,}372$

Quantilstabelle der t-Verteilung mit n Freiheitsgraden

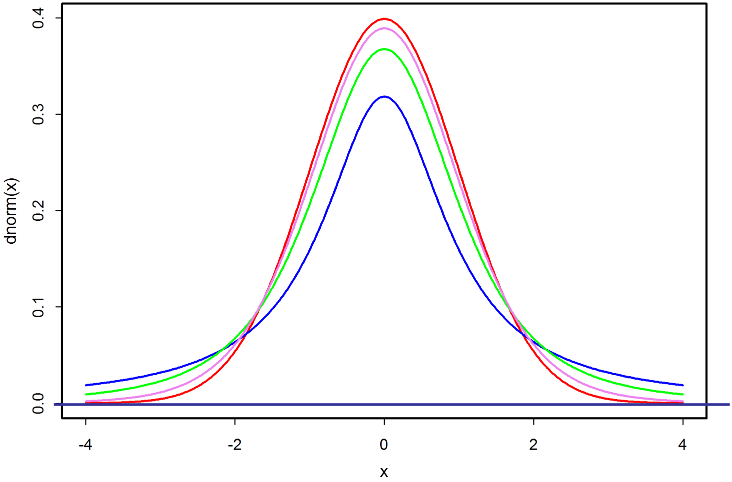


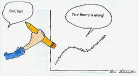
$\alpha \setminus n$	0.6	0.75	0.8	0.9	0.95	0.975	0.99	0.995
1	0.325	1.000	1.376	3.078	6.314	12.706	31.820	63.657
2	0.289	0.816	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.765	0.979	1.638	2.353	3.183	4.541	5.841
4	0.271	0.741	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.727	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.718	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.711	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.706	0.889	1.397	1.860	2.306	2.897	3.355
9	0.261	0.703	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.700	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.698	0.875	1.363	1.796	2.201	2.718	3.106
12	0.259	0.696	0.873	1.356	1.782	2.179	2.681	3.054
13	0.259	0.694	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.692	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.691	0.866	1.341	1.753	2.131	2.603	2.947
16	0.258	0.690	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.689	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.688	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.688	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.687	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.686	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.686	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.685	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.685	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.684	0.856	1.316	1.708	2.059	2.485	2.787
26	0.256	0.684	0.856	1.315	1.706	2.055	2.479	2.779
27	0.256	0.684	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.683	0.855	1.312	1.701	2.048	2.467	2.763
29	0.256	0.683	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.683	0.854	1.310	1.697	2.042	2.457	2.750

t-Verteilung vs. Normalverteilung

Dichtefunktion

- t-Verteilung mit 1 (blau), 3 (grün) und 10 (lila) Freiheitsgraden
- Standardnormalverteilung (rot)



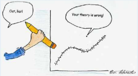


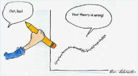
Intervall-Schätzung

- Für einen unbekannten Verteilungsparameter ϑ soll auf Basis einer Stichprobe ein Intervall geschätzt werden.
- Verwendung der Stichprobenfunktionen V_u, V_o, so dass $V_u \leqq V_o$ und

$$P(V_u \leqq \vartheta \leqq V_o) = 1 - \alpha$$

  stets gelten.
  $[V_u; V_o]$ heißt Konfidenzintervall (KI) für ϑ zum Konfidenzniveau $1 - \alpha$.
- Beachte: Das Schätzintervall $[v_u; v_o]$ ist Realisierung der Zufallsvariablen (!) V_u, V_o.
  ➟ Irrtumswahrscheinlichkeit α (klein, i.d.R. $\alpha \leqq 0{,}1$)
- Frage: Welche Konfidenzintervalle sind zur Schätzung geeignet?
  ➟ Hängt von Verteilung von G sowie vom unbekannten Parameter (μ, σ^2) ab!
- Im Folgenden: Einfache Stichprobe $X_1, \ldots, X_n$ mit $E(X_i) = \mu$, $Var(X_i) = \sigma^2$

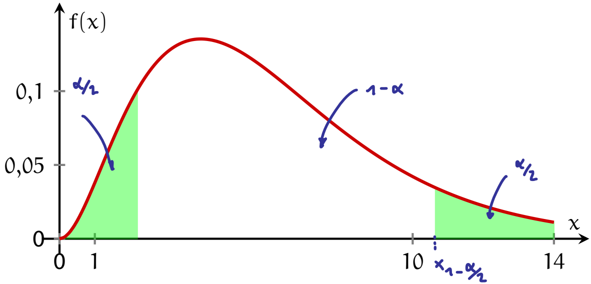




Wichtiger Spezialfall: Symmetrische Konfidenzintervalle

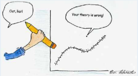
- Symmetrisch heißt **nicht**, dass die Dichte symmetrisch ist, sondern
- übereinstimmende Wahrscheinlichkeiten für Über-/Unterschreiten des Konfidenzintervalls, d.h.

$$P(V_u > \vartheta) = P(V_o < \vartheta) = \frac{\alpha}{2}$$



- **Wichtig:** Eine Verkleinerung von α bewirkt eine Vergrößerung des Konfidenzintervalls.

Konfidenzintervall für μ bei Normalverteilung mit bekanntem σ^2

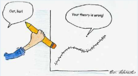


Vorgehensweise:

1. Festlegen des Konfidenzniveaus $1-\alpha$
2. Bestimmung des $\left(1-\frac{\alpha}{2}\right)$-Fraktils c der $N(\overset{\mu}{0}, \overset{\sigma}{1})$-Verteilung
3. Berechnen des Stichprobenmittels $\bar{x}$
4. Berechnen des Wertes $\frac{\sigma c}{\sqrt{n}}$
5. Ergebnis der Intervall-Schätzung:

$$\left[\bar{x} - \frac{\sigma c}{\sqrt{n}}\,;\quad \bar{x} + \frac{\sigma c}{\sqrt{n}}\right]$$

Intervallschätzung: Beispiel



Beispiel

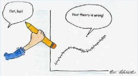
Normalverteilung mit $\sigma = 2{,}4$
$(x_1, \ldots, x_9)$ = (184.2, 182.6, 185.3, 184.5, 186.2, 183.9, 185.0, 187.1, 184.4)

Gesucht: Konfidenzintervall für μ zum Konfidenzniveau $1 - \alpha = 0{,}99$

1. $1 - \alpha = 0{,}99$
2. $N(0;1)$: $c = x_{1-\frac{\alpha}{2}} = x_{1-\frac{0{,}01}{2}} = x_{0{,}995} = 2{,}576$ (Tab. 3; Interpolation)
3. $\bar{x} = \frac{1}{9}(184{,}2 + \cdots + 184{,}4) = 184{,}8$
4. $\frac{\sigma c}{\sqrt{n}} = \frac{2{,}4 \cdot 2{,}576}{\sqrt{9}} = 2{,}06$
5. $\mathrm{KI} = [184{,}8 - 2{,}06; 184{,}8 + 2{,}06] = [182{,}74; 186{,}86]$

Interpretation: Mit 99 % Wahrscheinlichkeit ist $\mu \in [182{,}74; 186{,}86]$.

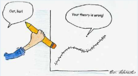
Wichtige Fraktilswerte



Wichtige $N(0; 1)$-Fraktilswerte:

α	x_α
0,9	1,281552
0,95	1,644854
0,975	1,959964
0,99	2,326348
0,995	2,575829

(I.d.R. genügen drei Nachkommastellen.)



Intervalllänge

- Bei bekannter Standardabweichung gilt offenkundig

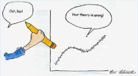
$$L = V_o - V_u = \frac{2\sigma c}{\sqrt{n}}$$

- Welcher Stichprobenumfang n sichert eine vorgegebene (Maximal-)Länge L? $\Rightarrow$ Nach n auflösen! $\Rightarrow$

$$n \geqq \left(\frac{2\sigma c}{L}\right)^2$$

- Eine Halbierung von L erfordert eine Vervierfachung von n!
- Angewendet auf letztes **Beispiel**:

$$L = 4 \Rightarrow n \geqq \left(\tfrac{2 \cdot 2{,}4 \cdot 2{,}576}{4}\right)^2 = 9{,}556 \Rightarrow n \geqq 10$$
$$L = 2 \Rightarrow n \geqq \left(\tfrac{2 \cdot 2{,}4 \cdot 2{,}576}{2}\right)^2 = 38{,}222 \Rightarrow n \geqq 39$$



Konfidenzintervall für μ bei Normalverteilung mit unbekanntem σ^2

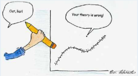
z.B. Stichprobenumfang 10 bedeutet 10 − 1 = 9 Freiheitsgrade

- Vorgehensweise:
  1. Festlegen des Konfidenzniveaus $1-\alpha$
  2. Bestimmung des $\left(1-\frac{\alpha}{2}\right)$-Fraktils c der $t(n-1)$-Verteilung
  3. Berechnen des Stichprobenmittels $\bar{x}$ und der Stichproben-Standardabweichung s
  4. Berechnen des Wertes $\frac{sc}{\sqrt{n}}$
  5. Ergebnis der Intervall-Schätzung:

$$\left[\bar{x}-\frac{sc}{\sqrt{n}}\,;\quad \bar{x}+\frac{sc}{\sqrt{n}}\right]$$

- Zu Schritt 2: Falls $n-1>30$ wird die $N(0;1)$-Verteilung verwendet.

Konfidenzintervall



Beispiel:

Wie das letzte Beispiel, jedoch σ unbekannt.

n=9 bedeutet
9-1=8 F.G.

1. $1-\alpha = 0{,}99$
2. $t(8)$: $c = x_{1-\frac{\alpha}{2}} = x_{1-\frac{0{,}01}{2}} = x_{0{,}995} = 3{,}355$ (Tab. 4)
3. $\bar{x} = \frac{1}{9}\,(184{,}2 + \cdots + 184{,}4) = 184{,}8$
   $s = \sqrt{\frac{1}{8}\,[(184{,}2^2 + \cdots + 184{,}4^2) - 9 \cdot 184{,}8^2]} = 1{,}31$
4. $\frac{sc}{\sqrt{n}} = \frac{1{,}31 \cdot 3{,}355}{\sqrt{9}} = 1{,}47$
5. $\text{KI} = [184{,}8 - 1{,}47; 184{,}8 + 1{,}47] = [183{,}33; 186{,}27]$

Interpretation: Mit 99 % Wahrscheinlichkeit ist $\mu \in [183{,}33; 186{,}27]$.

```
# ------------------------------------------------------
# Lade Daten
D <- read.csv("http://goo.gl/KCGF16", sep=";", dec=",")

# Groesse der Mütter (Lösche fehlende Werte und korrigiere Tippfehler)
Groesse = na.omit(D$GroesseM)
Groesse[Groesse < 100] = 100 + Groesse[Groesse < 100]

# Arithm. Mittel und (emp.) Standardabweichung als
# Größen der Grundgesamtheit
mu = mean(Groesse)
```

```
# ---------------------------------------------------------------------------------
# Konfidenzintervall für µ bei unbekannter Varianz

N = 20            # Anzahl Stichproben
n = 5             # Umfang einzelne Stichprobe
conf.level=0.80  # Konfidenzniveau
mu = mean(Groesse)

A.Stichproben = matrix(nrow=N, ncol=n)
set.seed(3)
for(i in 1:N) {
  A.Stichproben[i,] = sample(Groesse, size = n, replace=TRUE)
}

A.Stichproben[1,]

A.KI = matrix(nrow=N,ncol=2)
alpha = (1-conf.level)
x.c = qt(p=1-alpha/2, df = n-1)
for (i in 1:N) {
  x = A.Stichproben[i,]
  s = sd(x)
  x.m = mean(x)
  A.KI[i,] = x.m + s*x.c*c(-1,1)/sqrt(n)
}

A

A.KI[20,]


# ------------------------------------------------
min=min(A.Stichproben) - (max(A.Stichproben)-min(A.Stichproben))*0.5
max=max(A.Stichproben) + (max(A.Stichproben)-min(A.Stichproben))*0.5

plot(c(min,max), c(1, N), type="n", xlab="", yaxt="n", ylab="")
abline(h=1:N, col="lightgray", lty=2)
KI.falsch = 0
for(i in 1:N) {
  x.data = as.numeric(A.Stichproben[i,])
  x.KI = as.numeric(A.KI[i,])
  points(jitter(x.data), rep(i, n), pch=16, cex=1,col="#2040dd60")
  Farbe = "#dd203060"
  if (mu>=A.KI[i,1] & mu<=A.KI[i,2]) {Farbe = "#20dd3060"}
  else {KI.falsch = KI.falsch + 1}
  points(x.KI, rep(i, 2), pch=15, cex=2,col=Farbe)
  lines(x.KI, rep(i, 2), lwd=5, pch=16, cex=2,col=Farbe)
  text(min,i,labels = round(x.KI[1],2), cex=0.8, col="darkgreen")
  text(max,i,labels = round(x.KI[2],2), cex=0.8, col="darkgreen")
}

# Zeichne mu
abline(v=mu, col="#20dd3099", lwd=2, lty=2)
title(sub=paste("Mittelwert Grundgesamtheit = ", round(mu, 2),
                "; Anzahl falsch: ", KI.falsch,
                "; Fehlerquote = ", KI.falsch/N))
```

```
      [,1] [,2] [,3] [,4] [,5]
 [1,]  165  168  165  169  165
 [2,]  168  162  172  160  175
 [3,]  168  163  171  165  161
 [4,]  162  160  165  170  162
 [5,]  162  166  160  173  159
 [6,]  164  155  165  165  166
 [7,]  170  167  174  158  175
 [8,]  179  170  158  165  170
 [9,]  170  176  168  155  163
[10,]  165  155  165  170  163
[11,]  162  169  158  160  162
[12,]  165  168  165  165  162
[13,]  163  163  160  175  160
[14,]  167  174  170  156  173
[15,]  172  160  170  173  170
[16,]  172  172  172  176  155
[17,]  150  165  165  160  158
[18,]  170  169  165  173  175
[19,]  169  165  162  162  170
[20,]  177  165  160  160  164
```

① $1-\alpha = 0.80$

② $t(4): \; x_{0.9} = c = 1.533$

↳ 4 Freiheitsgr.

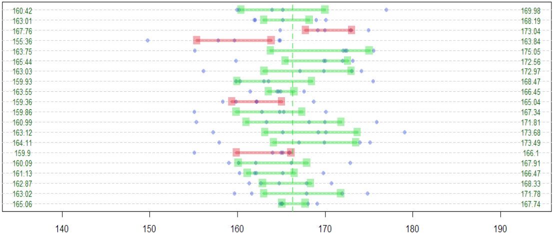
③ $\bar{x} = \dots$

$s = \dots$

④ $\frac{s \cdot 1.533}{\sqrt{5}} \approx \dots$

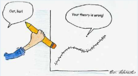
⑤ $KI = [\bar{x} \pm \dots]$

↓n \ α →	0.6	0.75	0.8	0.9
1	0.325	1.000	1.376	3.078
2	0.289	0.816	1.061	1.886
3	0.277	0.765	0.979	1.638
4	0.271	0.741	0.941	1.533
5	0.267	0.727	0.920	1.476



Mittelwert Grundgesamtheit = 166.32 ; Anzahl falsch: 4 ; Fehlerquote = 0.2

R Beispiel




```r
x <- c(184.2, 182.6, 185.3, 184.5, 186.2,
       183.9, 185.0, 187.1, 184.4)
t.test(x,conf.level=.99)
```

```
##
##  One Sample t-test
##
## data:  x
## t = 422.1129, df = 8, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 99 percent confidence interval:
##  183.331 186.269
## sample estimates:
## mean of x
##     184.8
```

Verteilungsfunktion Φ der Standardnormalverteilung

Dabei bedeutet $\Phi(x)$ zum Beispiel: $\Phi(2{,}13) = \Phi(2{,}1 + 0{,}03) = 0{,}9834$. Diesen Wert findet man in der Zeile mit $x_1 = 2{,}1$ und der Spalte mit $x_2 = 0{,}03$.

$x_1 \backslash x_2$	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1	0.84134	0.84375	0.84614	0.84850	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99897	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976

α-Fraktile der t-Verteilung mit n Freiheitsgraden

$\downarrow n \setminus \alpha \rightarrow$	0.6	0.75	0.8	0.9	0.95	0.975	0.99	0.995
1	0.325	1.000	1.376	3.078	6.314	12.706	31.820	63.657
2	0.289	0.816	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.765	0.979	1.638	2.353	3.183	4.541	5.841
4	0.271	0.741	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.727	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.718	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.711	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.706	0.889	1.397	1.860	2.306	2.897	3.355
9	0.261	0.703	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.700	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.698	0.875	1.363	1.796	2.201	2.718	3.106
12	0.259	0.696	0.873	1.356	1.782	2.179	2.681	3.054
13	0.259	0.694	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.692	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.691	0.866	1.341	1.753	2.131	2.603	2.947
16	0.258	0.690	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.689	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.688	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.688	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.687	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.686	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.686	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.685	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.685	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.684	0.856	1.316	1.708	2.059	2.485	2.787
26	0.256	0.684	0.856	1.315	1.706	2.055	2.479	2.779
27	0.256	0.684	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.683	0.855	1.312	1.701	2.048	2.467	2.763
29	0.256	0.683	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.683	0.854	1.310	1.697	2.042	2.457	2.750