

Hypothesenprüfungen 2. Zweistichproben t -Test, F -Test, Varianzanalyse



KAD 2012.10.25

Schätzungen und Hypothesenprüfungen

Schätzungen

Wie gross ist eine Grösse?

Punktschätzungen

ein Wert ist gegeben und
nichts über die Sicherheit

Parameter der Stichprobe \rightarrow \bar{x}
Parameter der Population \rightarrow μ ($n \rightarrow \infty$)
 $s \rightarrow \sigma$ ($n \rightarrow \infty$)

Intervall- schätzungen

ein Intervall ist mit einem
Konfidenzniveau gegeben

95 % Konfidenzintervall für den Erwartungswert:

$$\bar{x} \pm 2 s_{\bar{x}}$$

(95 %) Referenzintervall:

$$\bar{x} \pm 2 s$$

Hypothesenprüfungen

Beantwortung einer Entscheidungsfrage

z. B. t -Tests \leftarrow

ja oder nicht mit einem Signifikanzniveau

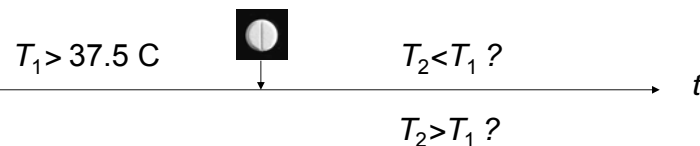
Typische Entscheidungsfragen in der Medizin

Ist die Therapie erfolgreich?

(Gibt es eine Änderung in der erwarteten Richtung?)

Hat eine Behandlung eine Wirkung?

Verändert/Verkleinert ein Fiebermittel die Körpertemperatur?



Die Nullhypothese

Es gibt keine Wirkung der Behandlung.

Die Wirkung der Behandlung ist Null (Nullhypothese, H_0).

Das Fiebermittel verändert die Körpertemperatur nicht.

Wenn die H_0 richtig ist, wir kennen die Verteilung.

Die Alternativhypothese

Es gibt eine Wirkung der Behandlung.

Die Wirkung der Behandlung ist nicht Null
(Alternativhypothese, H_1).

Das Fiebermittel verändert die Körpertemperatur.

Man unterscheidet als **Gegensatzpaar**
Nullhypothese und Alternativhypothese.

Entweder H_0 oder H_1 ist richtig.

Nehmen wir an, dass H_0 richtig ist!

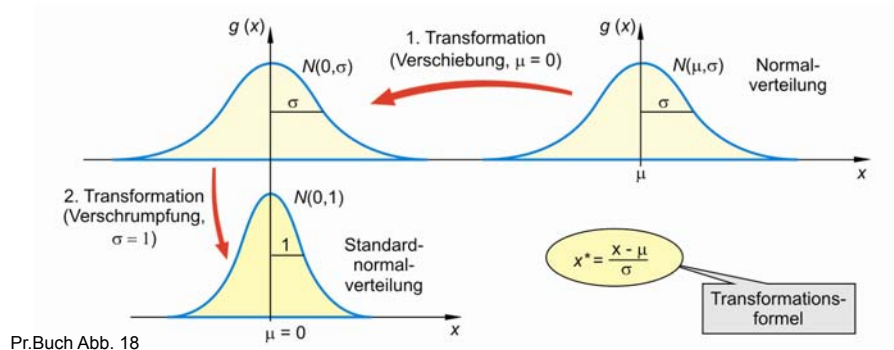
Wenn Ergebnisse mit dieser Voraussetzung nicht passen:
ablehnen wir $H_0 \rightarrow H_1$ ist richtig

Transformation einer Normalverteilung mit allgemeiner Lage und Breite in eine Standardnormalverteilung

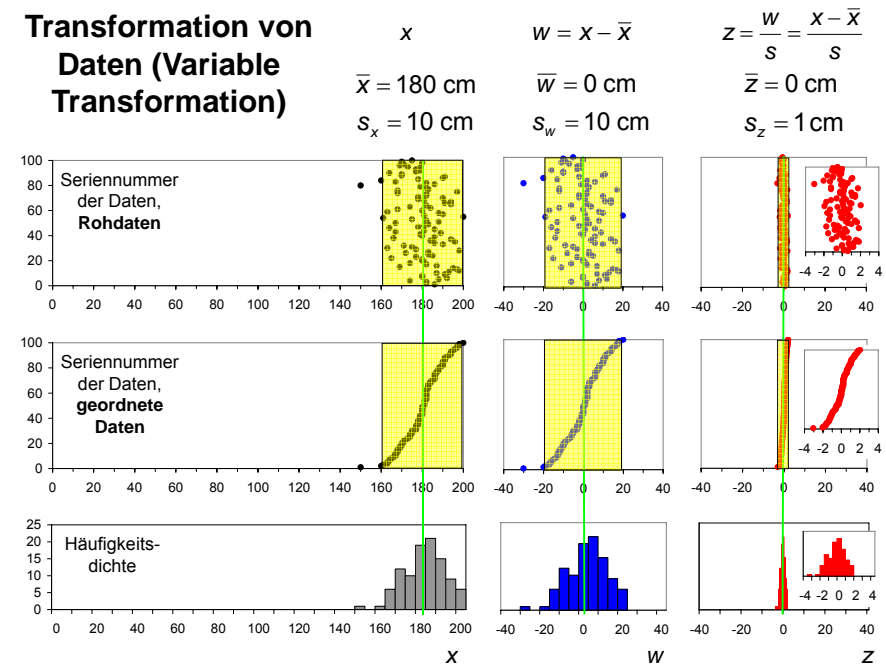
Mit welcher Verteilung sollen wir unsere Stichprobe vergleichen?

Die Standardnormalverteilung hat eine ausgezeichnete Rolle zwischen der Normalverteilungen.

Alle Normalverteilungen können in Standardnormalverteilung transformiert werden.



Transformation von Daten (Variable Transformation)



Einstichproben t-Test

Variable	x	$w = x - \mu$	$z = \frac{x - \mu}{\sigma}$
Verteilung	$N(\mu, \sigma)$	$N(0, \sigma)$	$N(0, 1)$

Wenn die originale Variable x zu einer **Normalverteilung** mit Parameter μ und σ gehört, dann gehört die transformierte Variable z zu der Standardnormalverteilung.

Wenn H_0 richtig ist, kennen wir den Wert von μ , aber σ nicht.

Die durchgeführte Transformation:

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}}$$

Variable

x

$$z = \frac{x - \mu}{\sigma}$$

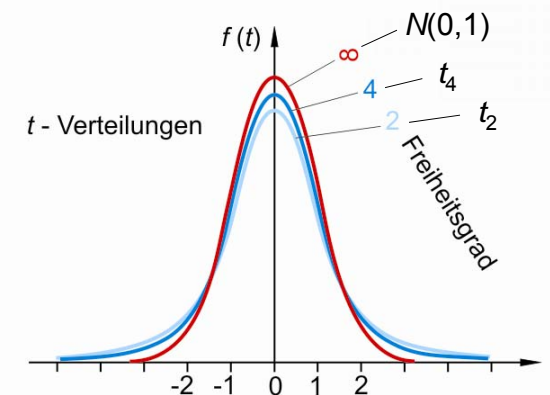
$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}}$$

Verteilung

$N(\mu, \sigma)$

$N(0, 1)$

t_{n-1}

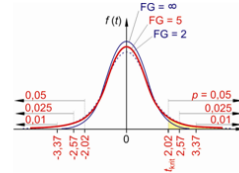


1. STATISTISCHE TABELLEN

t-VERTEILUNG

t-Verteilungsfamilie

Freiheitsgrad (FG)	p (Irrtumswahrscheinlichkeit, einseitiger Test)						
	0,4	0,25	0,1	0,05	0,025	0,01	0,005
	p (Irrtumswahrscheinlichkeit, zweiseitiger Test)						
	0,8	0,5	0,2	0,1	0,05	0,02	0,01
1	0,325	1,000	3,078	6,314	12,70	31,82	63,65
2	0,289	0,816	1,886	2,920	4,303	6,965	9,925
3	0,277	0,765	1,638	2,353	3,182	4,541	5,841
4	0,271	0,741	1,533	2,132	2,776	3,747	4,604
5	0,267	0,727	1,476	2,015	2,571	3,365	4,032
6	0,265	0,718	1,440	1,943	2,447	3,143	3,707
7	0,263	0,711	1,415	1,895	2,365	2,998	3,499
25	0,256	0,684	1,316	1,708	2,060	2,485	2,787
26	0,256	0,684	1,315	1,706	2,056	2,479	2,779
27	0,256	0,684	1,314	1,703	2,052	2,473	2,771
28	0,256	0,683	1,313	1,701	2,048	2,467	2,763
29	0,256	0,683	1,311	1,699	2,045	2,462	2,756
30	0,256	0,683	1,310	1,697	2,042	2,457	2,750
40	0,255	0,681	1,303	1,684	2,021	2,423	2,704
60	0,255	0,679	1,296	1,671	2,000	2,390	2,66
120	0,254	0,677	1,289	1,658	1,980	2,358	2,617
∞	0,250	0,674	1,282	1,645	1,960	2,326	2,576



„Glockenkurven“

Je grösser ist der Freiheitsgrad, desto schmaler ist die Kurve.

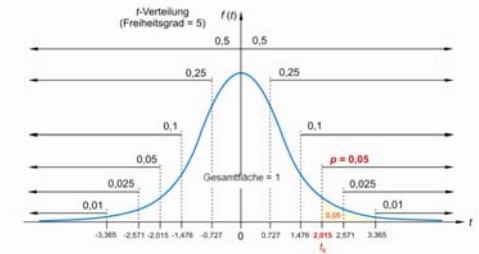
$$t_{\infty} \equiv N(0, 1)$$

Pr.Buch Anhang S.27.2

Kann der (aus der Stichprobe kalkulierte) t -Wert der t -Verteilung (mit entsprechendem Freiheitsgrad) gehören?

Alle Werte können zu der t -Verteilung gehören. Aber: Wenn der t -Wert gross ist, dann ist die Wahrscheinlichkeit klein.

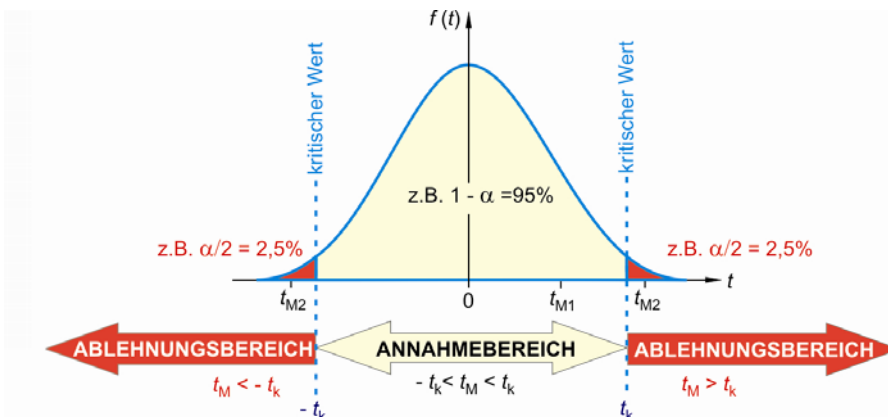
Deswegen benutzen wir nicht die gesamte t -Verteilung, sondern eine abgestutzte t -Verteilung!



akzeptierbare Irrtumswahrscheinlichkeit in der Medizin: kleiner oder gleich 5 %

Zweiseitiger t-Test

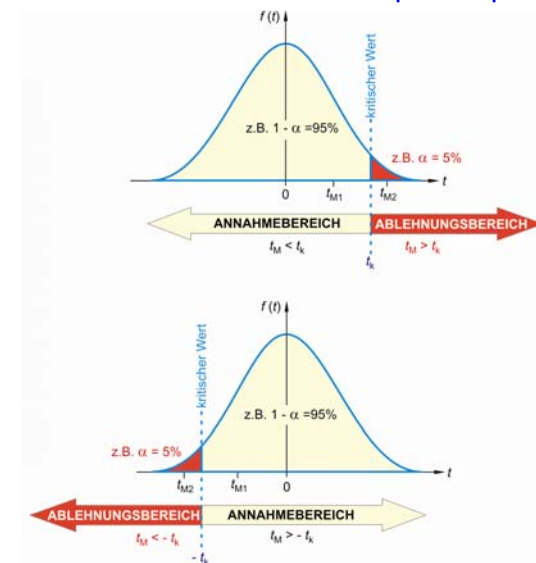
Verändert ein Fiebermittel die Körpertemperatur?



Pr.Buch Abb. 22

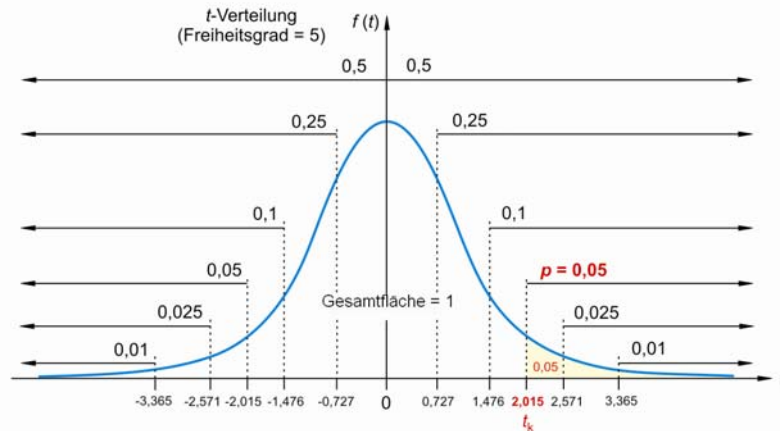
Einseitiger t-Test

Verkleinert ein Fiebermittel die Körpertemperatur?



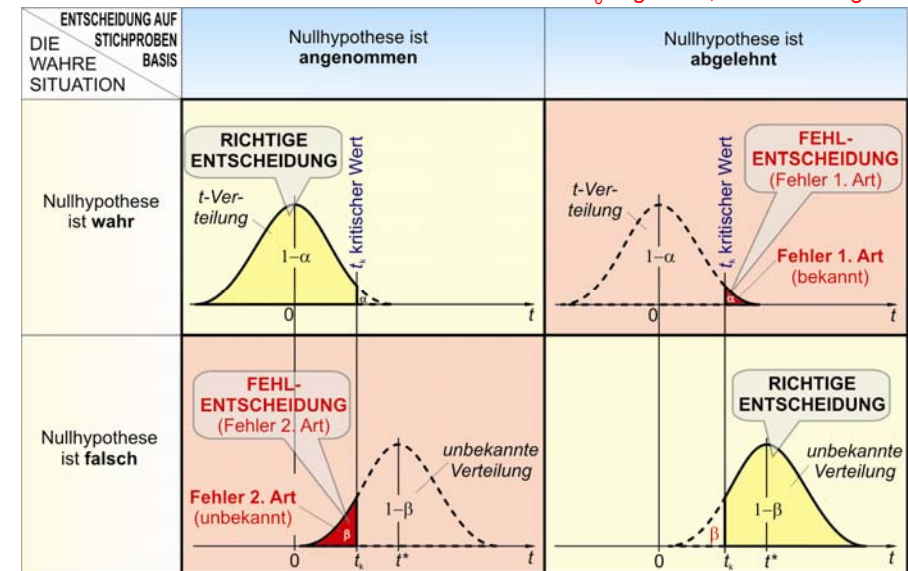
Pr.Buch Abb. 23

t-Verteilungskurve mit Freiheitsgrad 5. Die kritischen Werte und Wahrscheinlichkeiten des einseitigen t-tests



Pr.Buch Abb. 25

H_0 abgelehnt, obwohl richtig



H_0 angenommen, obwohl falsch

Pr.Buch Abb. 24

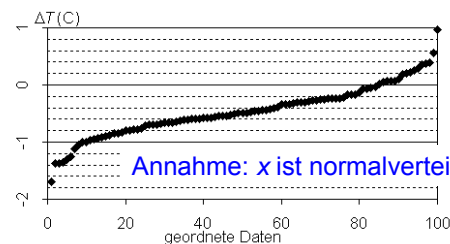
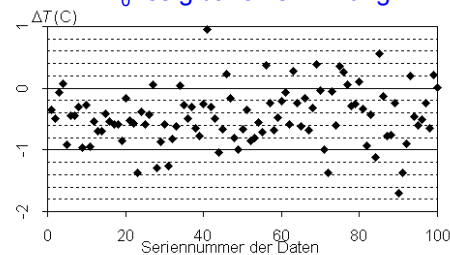
x = Temperaturdifferenzen

-0.35	-0.52	0.96	-0.07	-0.34
-0.50	-0.57	-0.31	-0.58	-0.94
-0.07	-1.37	-0.49	0.28	-0.43
0.07	-0.39	-1.05	-0.24	-1.12
-0.92	-0.58	-0.66	-0.62	0.56
-0.44	-0.43	0.23	-0.16	-0.13
-0.45	0.06	-0.16	-0.69	-0.78
-0.30	-1.30	-0.80	-0.31	-0.77
-0.97	-0.88	-1.00	0.38	-0.24
-0.28	-0.58	-0.67	-0.04	-1.69
-0.95	-1.26	-0.34	-1.00	-1.37
-0.55	-0.82	-0.86	-1.36	-0.90
-0.70	-0.61	-0.80	-0.05	0.19
-0.70	0.05	-0.55	-0.60	-0.46
-0.41	-0.27	-0.71	0.36	-0.61
-0.54	-0.50	0.38	0.26	-0.50
-0.59	-0.31	-0.24	0.06	-0.24
-0.59	-0.65	-0.68	-0.29	-0.65
-0.85	-0.77	-0.47	-0.25	0.21
-0.17	-0.26	-0.22	0.10	0.01

Beispiel: Einstichproben t-Test

Verändert ein Fiebermittel die Körpertemperatur?

H_0 : es gibt keine Wirkung



Annahme: x ist normalverteilt

Kalkulation:

$$t_{n-1} = \frac{\bar{X} - \mu}{s_{\bar{X}}}, \quad \mu = 0$$

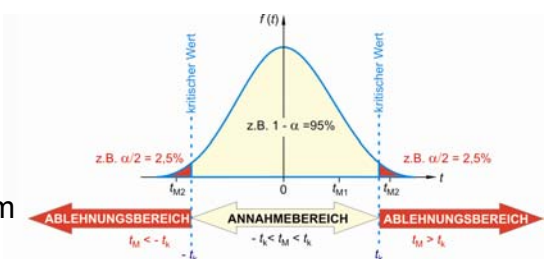
$$t_{n-1} = \frac{\bar{X}}{s_{\bar{X}}} = \frac{\bar{X}}{s} \sqrt{n}$$

zweiseitiger Test

$$|t| > t_{\text{krit}} \rightarrow$$

wir ablehnen die Nullhypothese mit einem Signifikanzniveau von 5%

Anzahl der Daten	n	100
Durchschnitt	avg	-0.457
Standardabweichung	stdev	0.454
Standardfehler	sem	0.045
t-Wert	t	-10.056
Freiheitsgrad	df	99
max. zulässige Irrtumswahrscheinlichkeit	α	0.05
kritischer t-Wert	t_{krit}	1.984



$$|t| > t_{\text{krit}} \rightarrow$$

Das Fiebermittel signifikant verändert (verkleinert) die Körpertemperatur ($p \leq 0.05$).

Im Klammer steht die Irrtumswahrscheinlichkeit. Es gibt die Wahrscheinlichkeit, dass die Nullhypothese richtig ist. In diesem Fall unsere Klassifikation ist falsch (Fehler 1. Art).

Die Nullhypothese abgelehnt wird, obwohl sie richtig ist.

Freiheitsgrad (FG)	p (Irrtumswahrscheinlichkeit, einseitiger Test)						
	0,4	0,25	0,1	0,05	0,025	0,01	0,005
	p (Irrtumswahrscheinlichkeit, zweiseitiger Test)						
	0,8	0,5	0,2	0,1	0,05	0,02	0,01
1	0,325	1,000	3,078	6,314	12,70	31,82	63,65
...
60	0,255	0,679	1,296	1,671	2,000	2,390	2,66
120	0,254	0,677	1,289	1,658	1,980	2,358	2,617
∞	0,250	0,674	1,282	1,645	1,960	2,326	2,576

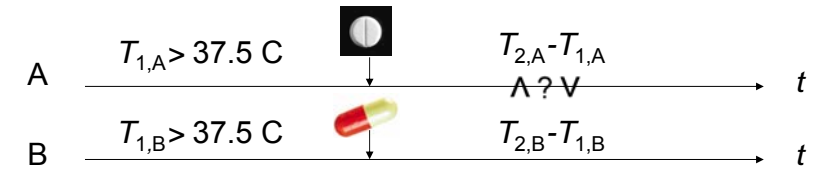
weitere Bemerkungen:

$$|t| = 10.056 > 2.66$$

$$p \leq 0.01 \text{ (zweiseitiger Test)}$$

Typische Entscheidungsfragen in der Medizin

Gibt es einen Unterschied zwischen zwei Therapiemethoden?



Ist der zu vergleichende Parameter unterschiedlich in der zwei Populationen?

Sind die Erwartungswerte in der zwei Populationen unterschiedlich?

Stammen die zwei Stichproben aus einem Population?

Nullhypothese: Es gibt keinen Unterschied, die Erwartungswerte sind gleich: $\mu_1 = \mu_2$

Einstichproben t-Test

$$t_{n-1} = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} = \frac{\bar{x} - \mu_0}{s} \sqrt{n}, \text{ wo } s = \sqrt{\frac{Q}{n-1}} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

Zweistichproben t-Test

$$t_{n_1+n_2-2} = \frac{\bar{x}_1 - \bar{x}_2}{s^*} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}}, \text{ wo } s^* = \sqrt{\frac{Q_1 + Q_2}{n_1 + n_2 - 2}}$$

Vergleichen wir die Formeln!

$$t_{n-1} = \frac{\bar{x}}{s \sqrt{\frac{1}{n}}}$$

Einstichproben

$$t_{n_1+n_2-1} = \frac{\bar{x}_1 - \bar{x}_2}{s^* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Zweistichproben

Einstichproben t-Test

Wirkung?:

Effekt der 5 Kniebeugen auf die Pulsfrequenz

n: Pulsfrequenz (1/30s), 1: vor, 2 nach, d: Differenz

H_0 : keine Wirkung

	n1	n2	dn	Zweistichproben t-Test bei abhängigen Stichproben (Paarvergleichstest)		
m	36	37	1			
m	33	35	2		Variable 1	Variable 2
m	34	37	3	Mittelwert	37.142857	42.285714
m	33	37	4	Varianz	43.516484	80.681319
m	34	38	4	Beobachtungen	14	14
m	37	41	4	Pearson Korrelation	0.9196855	
m	37	46	9	Hypothetische Differenz der Mittelwerte	0	
w	30	29	-1	Freiheitsgrade (df)	13	
w	45	49	4	t-Statistik	-4.9342498	
w	50	55	5	P(T<=t) einseitig	0.0001365	
w	35	40	5	Kritischer t-Wert bei einseitigem t-Test	1.7709334	
w	28	37	9	P(T<=t) zweiseitig	0.000273	
w	39	48	9	Kritischer t-Wert bei zweiseitigem t-Test	2.1603687	
w	49	63	14			

$$|t| = 4.934 > t_{13, \text{krit}(0,05)} = 2.160 \rightarrow H_0 \text{ ist falsch (} p \leq 0.05 \text{)}$$

$$(p = 0.000273)$$

Zweistichproben t-Test

Gibt es einen Unterschied zwischen zwei (Therapie)methoden?

m: männlich, w: weiblich

F-Test !

	n1	n2	dn	Zweistichproben t-Test unter der Annahme gleicher Varianzen		
m	36	37	1			
m	33	35	2		Variable 1	Variable 2
m	34	37	3	Mittelwert	3.8571429	6.4285714
m	33	37	4	Varianz	6.4761905	22.619048
m	34	38	4	Beobachtungen	7	7
m	37	41	4	Gepoolte Varianz	14.547619	
m	37	46	9	Hypothetische Differenz der Mittelwerte	0	
w	30	29	-1	Freiheitsgrade (df)	12	
w	45	49	4	t-Statistik	-1.261283	
w	50	55	5	P(T<t) einseitig	0.1155878	
w	35	40	5	Kritischer t-Wert bei einseitigem t-Test	1.7822875	
w	28	37	9	P(T<t) zweiseitig	0.2311755	
w	39	48	9	Kritischer t-Wert bei zweiseitigem t-Test	2.1788128	
w	49	63	14			

$$|t| = 1.261 < t_{12, \text{krit}(0,05)} = 2.179$$

⇒ H_0 ist richtig

21

	n1	n2	dn	Zweistichproben t-Test bei abhängigen Stichproben (Paarvergleichstest)		
m	36	37	1			
m	33	35	2		Variable 1	Variable 2
m	34	37	3	Mittelwert	37.142857	42.285714
m	33	37	4	Varianz	43.516484	80.681319
m	34	38	4	Beobachtungen	14	14
m	37	41	4	Pearson Korrelation	0.9196855	
m	37	46	9	Hypothetische Differenz der Mittelwerte	0	H_0 : keine Wirkung
w	30	29	-1	Freiheitsgrade (df)	13	
w	45	49	4	t-Statistik	-4.9342498	
w	50	55	5	P(T<t) einseitig	0.0001365	
w	35	40	5	Kritischer t-Wert bei einseitigem t-Test	1.7709334	
w	28	37	9	P(T<t) zweiseitig	0.000273	H_0 ist falsch
w	39	48	9	Kritischer t-Wert bei zweiseitigem t-Test	2.1603687	
w	49	63	14			
	n1	n2	dn	Zweistichproben t-Test unter der Annahme gleicher Varianzen		
					Variable 1	Variable 2
				Mittelwert	3.8571429	6.4285714
				Varianz	6.4761905	22.619048
				Beobachtungen	7	7
				Gepoolte Varianz	14.547619	
				Hypothetische Differenz der Mittelwerte	0	H_0 : keine Differenz zwischen der Wirkungen
				Freiheitsgrade (df)	12	
				t-Statistik	-1.261283	
				P(T<t) einseitig	0.1155878	
				Kritischer t-Wert bei einseitigem t-Test	1.7822875	
				P(T<t) zweiseitig	0.2311755	H_0 ist richtig
				Kritischer t-Wert bei zweiseitigem t-Test	2.1788128	

22

Freiheitsgrad (FG)	p (Irrtumswahrscheinlichkeit, einseitiger Test)						
	0,4	0,25	0,1	0,05	0,025	0,01	0,005
	p (Irrtumswahrscheinlichkeit, zweiseitiger Test)						
	0,8	0,5	0,2	0,1	0,05	0,02	0,01
1	0,325	1,000	3,078	6,314	12,70	31,82	63,65
2	0,289	0,816	1,886	2,920	4,303	6,965	9,925
3	0,277	0,765	1,638	2,353	3,182	4,541	5,841
4	0,271	0,741	1,533	2,132	2,776	3,747	4,604
5	0,267	0,727	1,476	2,015	2,571	3,365	4,032
6	0,265	0,718	1,440	1,943	2,447	3,143	3,707
7	0,263	0,711	1,415	1,895	2,365	2,998	3,499
8	0,262	0,706	1,397	1,860	2,306	2,896	3,355
9	0,261	0,703	1,383	1,833	2,262	2,821	3,250
10	0,260	0,700	1,372	1,812	2,228	2,764	3,169
11	0,260	0,697	1,363	1,796	2,201	2,718	3,106
12	0,259	0,695	1,356	1,782	2,179	2,681	3,055
13	0,259	0,694	1,350	1,771	2,160	2,650	3,012
14	0,258	0,692	1,345	1,761	2,145	2,624	2,977
15	0,258	0,691	1,341	1,753	2,131	2,602	2,947

Einstichproben t-Test

$$|t| = 4.934 > t_{13, \text{krit}(0,05)} = 2.160 \Rightarrow H_0 \text{ ist falsch (} p < 0.05 \text{)}$$

$$|t| = 4.934 > t_{13, \text{krit}(0,01)} = 3.012 \Rightarrow H_0 \text{ ist falsch (} p < 0.01 \text{)}$$

$$|t| = 4.934 \geq t_{13, \text{krit}(0.00027)} = 4.934 \Rightarrow H_0 \text{ ist falsch (} p \leq 0.00027 \text{)}$$

5 Kniebeugen verursachen Veränderung der Pulsfrequenzen mit einem Signifikanzniveau von 5 % (sogar: 1 %, ..., 0.027%)

Zweistichproben t-Test

$$|t| = 1.261 < t_{12, \text{krit}(0,05)} = 2.179 \quad H_0 \text{ ist richtig}$$

Es gibt keine signifikante Differenz zwischen der Pulsfrequenzen in der Männer- und Frauengruppen nach 5 Kniebeugen.

Vergleich der gepaarten-ungepaarten Tests bei Stichproben mit kleinem Umfang

z.B. Körperhöhe

ungepaarter Test
Zweistichproben t -Test

kein signifikanter Unterschied



gepaarter Test:
Einstichproben t -Test

signifikanter Unterschied



25

Frage: sind die Varianzen in zwei Stichproben gleich?

H_0 : die Varianzen sind gleich

Teststatistik (Parameter): $F = \frac{s_1^2}{s_2^2}$, $s_1 > s_2$

F-test

Wenn die Varianzen gleich sind ($H_0 \uparrow$), ist die Teststatistik F verteilt mit n_1-1 und n_2-1 Freiheitsgraden.

$F < F_{n_1-1, n_2-1; 5\%}$

Annahme der Nullhypothese:
die Varianzen sind gleich

man darf den herkömmlichen Zweistichproben t -Test anwenden

$F > F_{n_1-1, n_2-1; 5\%}$

Ablehnung der Nullhypothese:
die Varianzen sind nicht gleich

Man darf den hk Zweistichproben t -Test nicht anwenden.

Umweg: Excel kann eine Datentransformation durchführen:
Zwei Stichproben, ungleiche Varianz (heteroskedastisch) 26

TTEST Funktion im Excel

TTEST

Irrtumswahrscheinlichkeit

Gibt die Teststatistik eines Student'schen t -Tests zurück. Mit Hilfe von TTEST können Sie testen, ob zwei Stichproben aus zwei Grundgesamtheiten mit demselben Mittelwert stammen.

Syntax

TTEST(Matrix1;Matrix2;Seiten;Typ)

Matrix1 ist die erste Datengruppe.

Matrix2 ist die zweite Datengruppe.

Seiten bestimmt die Anzahl der Endflächen (Schwänze). Ist **Seiten** = 1, verwendet TTEST eine Endfläche (einseitiger Test). Ist **Seiten** = 2, verwendet TTEST eine Endfläche (zweiseitiger Test).

Typ bestimmt den Typ des durchzuführenden t -Tests.

IST TYP GLEICH	WIRD FOLGENDER TEST AUSGEFÜHRT
1	Gepaart
2	Zwei Stichproben, gleiche Varianz (homoskedastisch)
3	Zwei Stichproben, ungleiche Varianz (heteroskedastisch)

Übersicht der Testmethoden

Verteilungs- typ Anzahl der Stichproben	normalverteilte Daten	die Verteilung der Daten ist unbekannt
eine Stichprobe	Einstichproben t -Test	Vorzeichnentest Wilcoxon Test
zwei Stichproben	Zweistichproben t -test	Mann-Whitney U-Test
mehrere Stichproben	ANOVA (Varianzanalyse)	Kruskal-Wallis Test

Bonferroni – Problem. Alphafehler-Kumulierung

Vergleich von mehreren Stichproben

Paarweise Vergleichen:

- Hohe Wahrscheinlichkeit des Fehlers von 1. Art
- z.B.: 10 Stichproben, 45 Vergleichen
(10 über 2)
alle mit 5% Irrtumswahrscheinlichkeit
Gesamtirrtumsw.: $\rightarrow 1-(1-0,05)^{45}=90,0\%$

Parametrische Methode: ANOVA
(**AN**alysis **Of** **VA**riance)

ANOVA

Vorbedingungen:

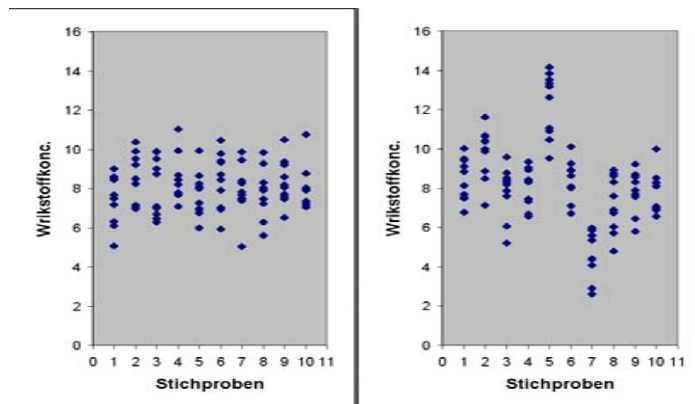
Unabhängigkeit der Stichproben

- Normalverteilung
- gleiche Streuungen

H_0 : Alle Stichproben stammen aus der selben Grundgesamtheit

H_1 : Mindestens *eine* Stichprobe stammt aus einer anderen Grundgesamtheit

Wenn H_0 gültig ist, sollen die Streuungen *zwischen* den Stichproben und *innerhalb* der Stichproben dieselbe sein.



h Stichproben

$$H_0: \mu_1 = \mu_2 = \dots \mu_h$$

Zwei unabhängige Varianzschätzungen:

Varianz innerhalb der Stichproben: S_i^2

Varianz zwischen den Stichproben: S_g^2

Wenn $S_i^2 \ll S_g^2 \rightarrow$ Varianzen sind unterschiedlich $\rightarrow H_0$ ablehnen

Wenn $S_i^2 \approx S_g^2 \rightarrow$ Varianzen sind die

Schätzungen derselben Varianz $\rightarrow H_0$ annehmen

$$F = \frac{S_g^2}{S_i^2} \quad F\text{-Test; Einseitig, Freiheitsgrad: } h-1; N-h$$

ANOVA

Varianz zwischen den Stichproben:

$$s_g^2 = \frac{\sum_{j=1}^h n_j (\bar{x}_j - \bar{x})^2}{h-1} = \frac{Q_g}{h-1}$$

h : Anzahl der Stichproben

n_j : Anzahl der Elementen in der j -ten Stichprobe

: Durchschnitt von allen Elementen

: Durchschnitt in der j -ten Stichprobe

ANOVA

Varianz innerhalb der Stichproben:

$$s_i^2 = \frac{\sum_{j=1}^h Q_j}{N-h} = \frac{\sum_{j=1}^h \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}{N-h} = \frac{Q_i}{N-h}$$

h : Anzahl der Stichproben

n_j : Anzahl der Elementen in der j -ten Stichprobe

x_{ij} : i -ten Element der j -ten Stichprobe

\bar{x}_j : Durchschnitt in der j -ten Stichprobe

N : Gesamte Anzahl der Stichprobenelementen

ANOVA

$$F = \frac{S_g^2}{S_i^2}$$

F -Test

FG: $h-1$; $N-h$

nicht Signifikant,
keine Unterschiede

H_0 : OK

Signifikant,

Es gibt Unterschiede

Post-hoc Test
(Paarweise Vergleichen
der Stichproben, z.B. Min-Max)

(Paarweise Vergleichen,
z.B: gegen Kontrolle)

Ich habe jetzt statistisch
bewiesen, daß alle diese
Kristallkugeln für
Wahrsagen gleich gut
geeignet sind!



Es gibt keinen Unterschied zwischen den Kristallkugeln.