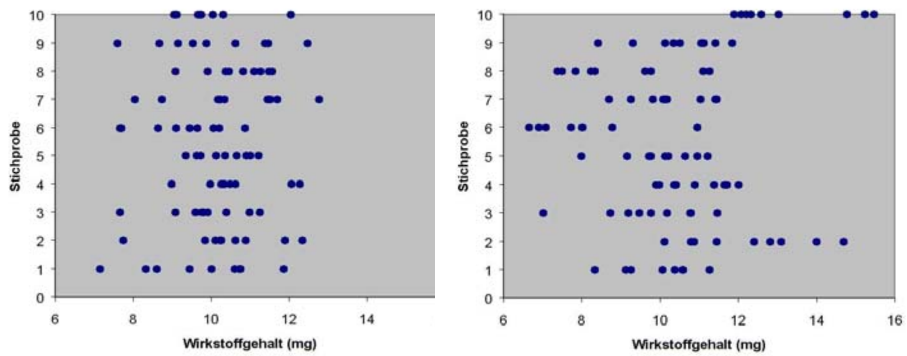
Grundlagen der Biostatistik und Informatik

Hypothesenprüfungen III.
ANOVA,
Nichtparametrische Methoden

dr László Smeller
Semmelweis Universität
2013

Vergleich von mehreren Stichproben



Vergleich von mehreren Stichproben

Bonferroni - Problem

Vergleich von mehreren Stichproben

Paarweise Vergleichungen:

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

Parametrische Methode: ANOVA
(**AN**alysis **O**f **VA**riance)

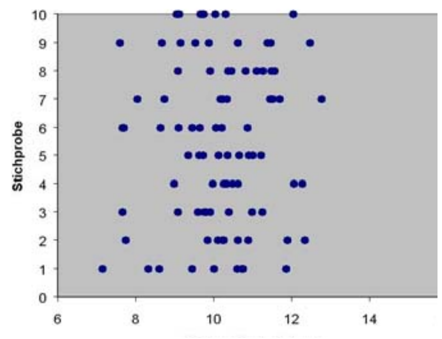
ANOVA

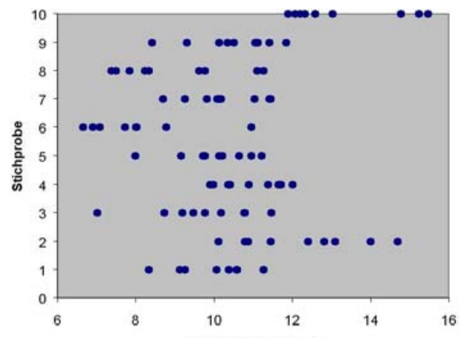
Vorbedingungen:

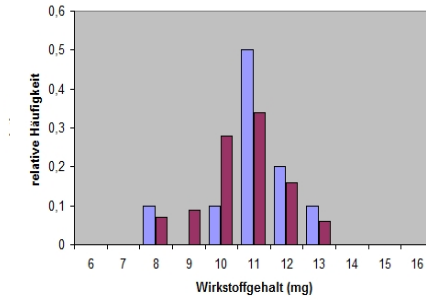
- Unabhängigkeit der Stichproben
- Normalverteilung
- Gleiche Streuungen

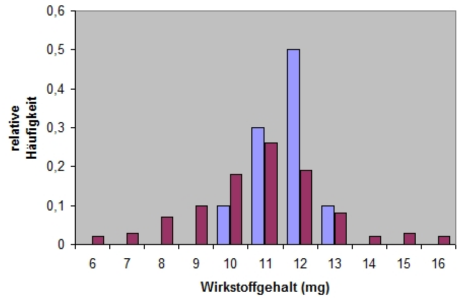
H_0: Alle Stichproben stammen aus der selben Grundgesamtheit

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



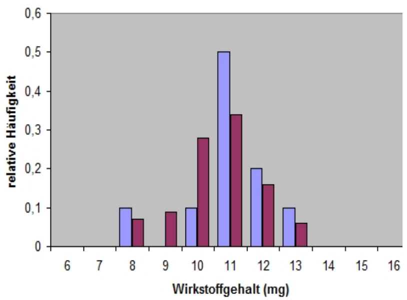


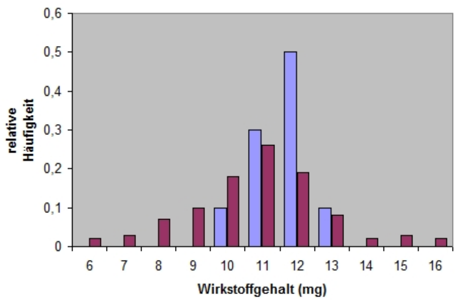




ANOVA

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





ANOVA

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 $S_i^2 << S_g^2 \rightarrow$ 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

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

ANOVA

Varianz zwischen den Stichproben:

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

h: Anzahl der Stichproben

n_j: Anzahl der Elementen in der *j*-ten Stichprobe

$\bar{\bar{x}}$: Durchschnitt von allen Elementen

$\bar{x}_j$: 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

$$\boldsymbol{F = \frac{s_g^2}{s_i^2}}$$

↓

F – Test

FG: h-1; N-h

Nicht Signifikant,
keine Unterschiede
H_0: OK

Signifikant,
Es gibt Unterschiede
$\cancel{H_0}$

↓

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

(Paarweise Vergleichungen,
z.B: gegen Kontrolle)

Nichtparametrische Methoden

Übersicht der Teste

Verteilung / Stichproben	Normalverteilte Daten	Die Verteilung der Daten ist unbekannt
Eine Stichprobe	Einstichproben t-Test	Vorzeichentest
		Wilcoxon Test
Zwei Stichproben	Zweistichproben t-test	Mann-Whittney U-Test
Mehrere Stichproben	ANOVA (Varianzanalyse)	Kruskal-Wallis Test

Nichtparametrische Methoden

Als Erinnerung: Bedingungen der t-Teste:
• kontinuierliches Merkmal (z.B. Körperhöhe, Körpertemperatur...)
• die Daten müssen eine Normalverteilung folgen!

Was passiert, wenn die Vorbedingungen nicht gelten?

•nur ordinale Daten (Ordinalskala)
•keine Normalverteilung

→ **Nichtparametrische Methoden**

z. B. Schmerzmittel – Wie es schmerzt? Kann nur auf einer ordinalen Skala gemessen werden:

1, 2, 3, 4, 5 oder

Nichtparametrische Metoden

Vorteile:
•Verteilungsunabhängigkeit
• Ordinal-, Intervall-, Verhältnisskalen

Nachteile:
• Datenreduktion, Informationsverlust
• größere Wahrscheinlichkeit der Fehler 2. Art
• Nur größere Unterschiede können detektiert werden als bei den parametrischen Teste

Eine Stichprobe: Vorzeichentest

Datenpaaren, d.h. gepaarter Test (Änderung oder Unterschied)

Datenreduktion:

Alternativen (einander ausschließende Ereignisse)

z.B.: Verbesserung oder Verschlechterung des Krankheitszustandes

Erfolg- Misserfolg

⇒ **Binomialverteilung**

Hat das Medikament eine Wirkung? D.h.: Sind signifikant mehr Fallen mit Verbesserung als mit Verschlechterung? Haben die zwei Alternativen unterschiedliche Wahrscheinlichkeit?

H_0: Die zwei Alternativen haben dieselbe Wahrscheinlichkeit.

Analogie: Münzenexperiment: Kopf oder Zahl

Vorzeichentest: Beispiel des Kopfschmerzes

Kopfschmerzen vor und nach der Einnahme des Medikamentes werden an einer relativen Skala gegeben.

Der Kopfschmerz erniedrigt in ***k*** aus ***n*** Fällen (in ***n-k*** Fällen es erhöht sich. Die „keine Änderung“ Fällen werden nicht beachtet.)

Ist die Änderung signifikant?

H_0: Das Medikament ist unwirksam, d.h. Erniedrigung und Erhöhung des Kopfschmerzes sind gleich wahrscheinlich.

Beispiel1.: Kopfschmerzen sinkt bei 9 aus 10 Patienten.

Beispiel2.: Kopfschmerzen sinkt bei 7 aus 9 Patienten.

Bei Gültigkeit der Nullhypothese: Analogie mit dem Münzenexperiment:

Experiment: *k*-mal Kopf aus *n* Versuche.

Analogie für Beispiel 1.: 9-mal Kopf aus 10 Versuche

Analogie für Beispiel2.: 7-mal Kopf aus 9 Versuche

Vorzeichentest: Analogie mit dem Münzenexperiment

Bei Münzenexperiment kann man die Wahrscheinlichkeit der unterschiedlichen Fällen ausrechnen (Binomialverteilung!):

Anzahl von Experimenten

Anzahl von Kopf	1	2	3	4	5	6	7	8	9	10
0	50.0%	25.0%	12.5%	6.3%	3.1%	1.6%	0.8%	0.4%	0.2%	0.1%
1	50.0%	50.0%	37.5%	25.0%	15.6%	9.4%	5.5%	3.1%	1.8%	1.0%
2		25.0%	37.5%	37.5%	31.3%	23.4%	16.4%	10.9%	7.0%	4.4%
3			12.5%	25.0%	31.3%	31.3%	27.3%	21.9%	16.4%	11.7%
4				6.3%	15.6%	23.4%	27.3%	27.3%	24.6%	20.5%
5					3.1%	9.4%	16.4%	21.9%	24.6%	24.6%
6						1.6%	5.5%	10.9%	16.4%	20.5%
7							0.8%	3.1%	7.0%	11.7%
8								0.4%	1.8%	4.4%
9									0.2%	1.0%
10										0.1%

Binomialverteilung

$$p_k = \binom{n}{k} p^k q^{n-k}$$

Vorzeichentest: Anwendung der Binomialverteilung

Bei Gültigkeit der H_0 gibt dieselbe Tabelle die Wahrscheinlichkeit der entsprechenden Fällen:

Anzahl von Patienten

Anzahl der Verbesserungen	1	2	3	4	5	6	7	8	9	10
0	50.0%	25.0%	12.5%	6.3%	3.1%	1.6%	0.8%	0.4%	0.2%	0.1%
1	50.0%	50.0%	37.5%	25.0%	15.6%	9.4%	5.5%	3.1%	1.8%	1.0%
2		25.0%	37.5%	37.5%	31.3%	23.4%	16.4%	10.9%	7.0%	4.4%
3			12.5%	25.0%	31.3%	31.3%	27.3%	21.9%	16.4%	11.7%
4				6.3%	15.6%	23.4%	27.3%	27.3%	24.6%	20.5%
5					3.1%	9.4%	16.4%	21.9%	24.6%	24.6%
6						1.6%	5.5%	10.9%	16.4%	20.5%
7							0.8%	3.1%	7.0%	11.7%
8								0.4%	1.8%	4.4%
9									0.2%	1.0%
10										0.1%

Binomialverteilung

$$p_k = \binom{n}{k} p^k q^{n-k}$$

Irrtumswahrscheinlichkeit=5% (2,5%+2,5%)

Vorzeichentest: Beispiel 3, Überlebenszeit

Überlebenszeit bei behandelten Ratten mit einem Tumor. (Tage)

168, 190, 280, 221, 110, 165, 179, 250, 195, 276

Es ist bekannt dass die Überlebenszeit der nicht behandelten Ratten mit dieser Tumorart 170 Tage beträgt (Median!)

Kann der Median der Überlebenszeiten der behandelten Ratten 170 Tage sein?

H_0: Median der Überlebenszeiten der behandelten Ratten beträgt 170 Tage.

168, 150, 280, 221, 230, 165, 179, 250, 195, 276

\- - + + + - + + + +

Bei Gültigkeit der H_0 die Daten sind >170 Tage zu 50% Wahrsch.

< 170 Tage zu 50% Wahrsch

Vorzeichentest: Anwendung der Binomialverteilung

Bei Gültigkeit der H_0 gibt diese Tabelle die Wahrscheinlichkeit der entsprechenden Fällen:

Anzahl von Ratten

Anzahl der Ratten mit Überelebenszeit > 170 Tage	1	2	3	4	5	6	7	8	9	10
0	50.0%	25.0%	12.5%	6.3%	3.1%	1.6%	0.8%	0.4%	0.2%	0.1%
1	50.0%	50.0%	37.5%	25.0%	15.6%	9.4%	5.5%	3.1%	1.8%	1.0%
2		25.0%	37.5%	37.5%	31.3%	23.4%	16.4%	10.9%	7.0%	4.4%
3			12.5%	25.0%	31.3%	31.3%	27.3%	21.9%	16.4%	11.7%
4				6.3%	15.6%	23.4%	27.3%	27.3%	24.6%	20.5%
5					3.1%	9.4%	16.4%	21.9%	24.6%	24.6%
6						1.6%	5.5%	10.9%	16.4%	20.5%
7							0.8%	3.1%	7.0%	11.7%
8								0.4%	1.8%	4.4%
9									0.2%	1.0%
10										0.1%

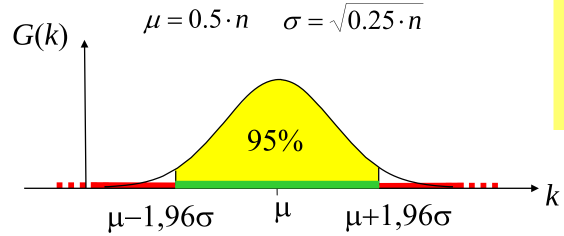
Binomialverteilung

$$p_k = \binom{n}{k} p^k q^{n-k}$$

Irrtumswahrscheinlichkeit=5%   (2,5%+2,5%)

Vorzeichentest: Annäherung bei n > 20

Annäherung bei $n > 20$ mit Normalverteilung:



Siehe: Binomial-verteilung:

$\boldsymbol{\mu = p \cdot n}$

$\boldsymbol{\sigma = \sqrt{p \cdot q \cdot n}}$

z. B. 100 Patienten, 56 Verbesserungen, 34 Verschlechterungen

H_0: ?; μ_0=? μ = ?; σ = ?; Entscheidung?

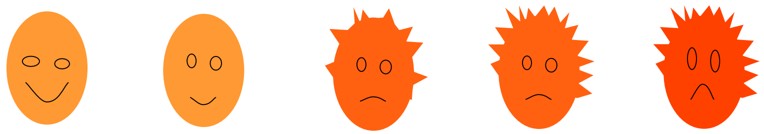
Analogie zu Einstichproben t-Test

(Lösung: 56+34=90   μ=45   σ =Wurzel(90/4)=4,74 μ+1,96*σ=45+9,3=54,3 <56 => signifikant (5% Irrt.w.)!

Prinzip der Rang Teste

Rang: Position eines Wertes innerhalb einer nach der Größe sortierten Wertereihe

z.B. Kopfschmerzen:



1 2 3 4 5

Mit Hilfe der Ränge führt man eine Gleichverteilung ein!

Rang Test Methode – Verbundene Ränge

Wenn zwei oder mehrere ursprüngliche Daten gleich sind:

originale Daten	3, 7, 1, 13, 13, 16
geordnete Daten	1, 3, 7, 13, 13, 16
Ränge	1, 2, 3, 4.5, 4.5, 6

Verbundene Ränge:

die bekommen den Durschnittsrang

Durchschnitt der Ränge

In steigende Reihe geordnete Daten:	x_1 , x_2 … $x_{(n-1)/2}$, $\boxed{x_{(n+1)/2}}$ …. x_{n-1} ,x_n
Ränge:	1, 2 … (n-1)/2 (n+1)/2 … n-1, n

(n ist ungerade)

Durchschnitt der Ränge: $\bar{R} = \frac{1}{n}\sum_{i=1}^{n} i = \frac{1}{n}\frac{n(n+1)}{2} = \frac{n+1}{2}$

Durchschnittlicher Rang = Rang des Medians

Wenn n ist gerade:

Median= $(x_{n/2} + x_{n/2+1})/2$

Durchschnittlicher Rang= (n+1)/2

Rangteste testen den Median!

Eine Stichprobe: Wilcoxon-Vorzeichen Rangtest

Eine Stichprobe (Gepaarte Test)

Ordinale Daten

Ist der Median der Datenreihe gleich Null?

(oder ein bestimmter Wert)?

H_0: Der Median der Daten ist Null (oder ein bestimmter Wert).

Die Ränge bekommen Vorzeihen.

Der Durchschnitt der Ränge wird geprüft.

Wenn die Nullhypothese gültig ist, es sind gleich viele und gleich große positive und negative Ränge, Durchschnitt der Ränge ist Null!

Wilcoxon-Vorzeichen Rangtest: Einführung mit einem Beispiel

Überlebenszeit der Ratten:

168, 150, 280, 221, 230, 165, 179, 250, 195, 276

Ist der Median der Überlebenszeiten unterschiedlich von 170 Tage?

H_0: Der Median der Überlebenszeiten beträgt 170 Tage.

Überlebenszeitenunterschiede der Ratten im Vergleich zur 170 Tage:

-2, -20, +110, +51, +60, -5, +9, +80, +25, +106

Geordnet nach Betrag der Änderung:

-2, -5, +9, -20, +25, +51, +60, +80, +106, +110,

Ränge (nach betrag der Änderung):

1, 2, 3, 4, 5, 6, 7, 8, 9, 10

Ränge mit Vorzeichen:

-1, -2, +3, -4, +5, +6, +7, +8, +9, +10

Durchschnitt: 4.10

Standardabw.: 4.91

Wilcoxon Vorzeichen Rangtest: Beispiel der Überlebenszeiten der Ratten

Der Durchschnitt folgt einer Normalverteilung, wenn genug viele Daten sind (Zentraler Grenzwertsatz)

Anwendung der t-Verteilung (Annäherung!):

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

$\bar{R}$: Durchschnitt der Ränge

s: Standardabweichung der Ränge

n: Anzahl der Daten

$n-1$: Freiheitsgrad

Entscheidung: wie beim Einstichproben ***t***-Test

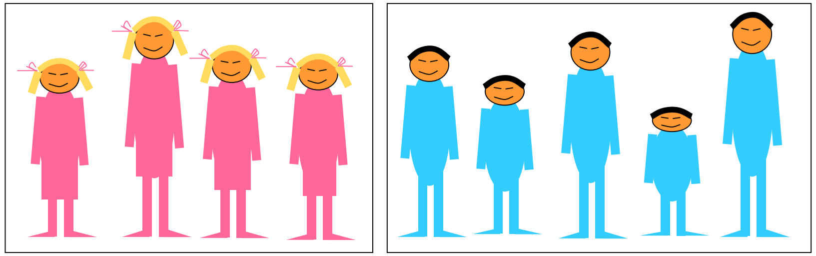
Ränge mit Vorzeichen:

-1, -2, +3, -4, +5, +6, +7, +8, +9, +10 → Durchschnitt: 4.10, Standardabw.: 4.91

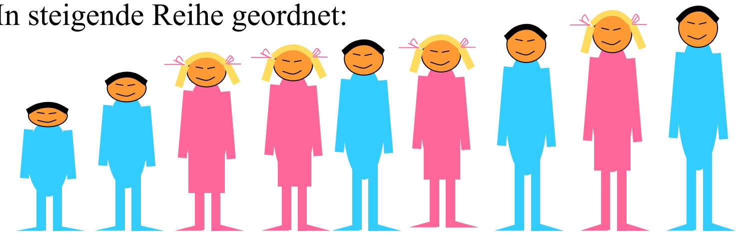
$t_9 = \frac{4,10}{4,91/\sqrt{10}} = 2,64 \Rightarrow t_9 > t_{9;5\%} \Rightarrow H_0$ is abgelehnt

$t_{9;5\%}$=2,26 (aus der Tabelle) $p < 5\%$ (mit Excel)

Vergleich von zwei Stichproben

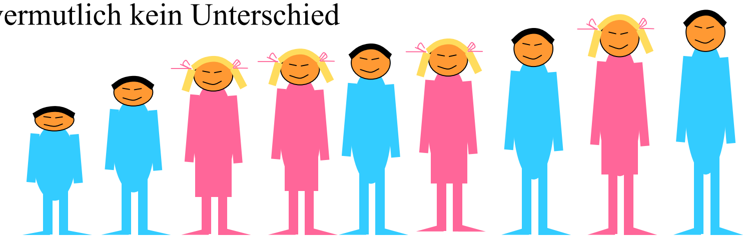


In steigende Reihe geordnet:

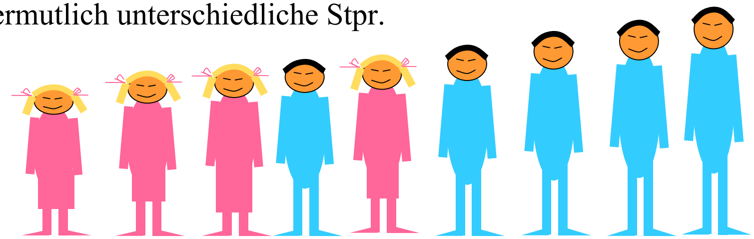


Ränge: 1 2 3 4 5 6 7 8 9

vermutlich kein Unterschied



vermutlich unterschiedliche Stpr.



1 2 3 4 5 6 7 8 9

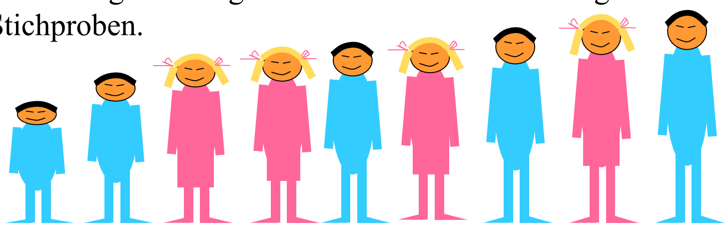
Mann – Whitney *U* Test (Annäherung)

(Auch als Wilcoxon Rank Summe Test genannt)

Vergleich von zwei Stichproben (n_1, n_2)

H_0: Die zwei Stichproben stammen aus der selben Grundgesamtheit

1. Zuordnung der Ränge der in den zwei zusammengeordneten Stichproben.



Ränge: 1 2 3 4 5 6 7 8 9

2. Bestimmung die Summen der Ränge in eine Gruppe: T_1.

   $T_1 = 1+2+5+7+9=24$

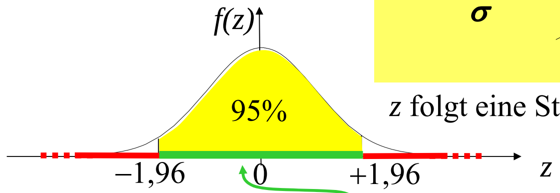
Mann – Whitney *U* Test: Annäherung

Bei Gültigkeit der Nullhypothese folgen die Daten der Gruppe 1 eine Gleichverteilung, mit möglichen werten von $1 \ldots n_1+n_2$)

Erwartungswert und die theoretische Streuung von T_1 können berechnet werden:

$$\boldsymbol{\mu} = \frac{\boldsymbol{n}_1(\boldsymbol{n}_1 + \boldsymbol{n}_2 + 1)}{2} \qquad \boldsymbol{\sigma} = \sqrt{\boldsymbol{n}_1\boldsymbol{n}_2(\boldsymbol{n}_1 + \boldsymbol{n}_2 + 1)/12}$$

$$\boldsymbol{z} = \frac{\boldsymbol{T}_1 - \boldsymbol{\mu}}{\boldsymbol{\sigma}} = \frac{\boldsymbol{T}_1 - \dfrac{\boldsymbol{n}_1(\boldsymbol{n}_1 + \boldsymbol{n}_2 + 1)}{2}}{\sqrt{\dfrac{\boldsymbol{n}_1\boldsymbol{n}_2(\boldsymbol{n}_1 + \boldsymbol{n}_2 + 1)}{12}}}$$



z folgt eine Standard-Normalverteilung (wenn H_0 gültig ist)

z.B. $T_1=24$, $n_1=5$, $n_2=4$ => $\boxed{z = -0{,}245}$ => H_0 wird angenommen

Kruskal – Wallis Test

- Vergleich von mehreren Stichproben
- Mit unbekannter Verteilung der Daten

Ende der Hypothesenprüfungen!

Bemerkung: Vergleich von Hypothesenprüfungen und Schätzungen

zB.: Blutdrucksenker: Blutdruckänderungen (mmHg):

-13, 5, -29, -22, 13, -8, -19, -12

Durchschnitt: -10,625 mmHg

Standardfehler: 4,917 mmHg

Schätzung: Konfidenintervall:

$x \pm 2 s_{\bar{x}}$ -10,6±9,8 mmHg -20,4 … – 0,8 mmHg

enthält Null nicht! => Blutdrucksänkender Effekt!

t-Test:

t = -10,625/4,917=-2,161 $|t| < t_{FG=7\,;5\%}$=2,365

kein signifikanter Effekt!

Lösung des Problems: Genaues Konfidenzintervall

$x \pm 2 s_{\bar{x}}$ ist nur eine **grobe** Annäherung des Konfidenzintervalles.

Das **genaue** Konfidenzintervall für 95% Konfidenzniveau ist:

$$x \pm t_{n-1;5\%} s_{x}$$

Es zählt nur bei kleinen Stichproben (n<20)

FG \ p	0.05	0.02	0.01
2	4.30266	6.96455	9.92499
3	3.18245	4.54071	5.84085
4	2.77645	3.74694	4.60408
5	2.57058	3.36493	4.03212
6	2.44691	3.14267	3.70743
7	2.36462	2.99795	3.49948
8	2.30601	2.89647	3.35538
9	2.26216	2.82143	3.24984
10	2.22814	2.76377	3.16926
11	2.20099	2.71808	3.10582
12	2.17881	2.68099	3.05454
13	2.16037	2.65030	3.01228
14	2.14479	2.62449	2.97685
15	2.13145	2.60248	2.94673
20	2.08596	2.52798	2.84534
50	2.00856	2.40327	2.67779
70	1.99444	2.38080	2.64790
100	1.98397	2.36421	2.62589
unendlich	1.95996	2.32635	2.57583

Bei dem Beispiel des Blutdrucksenkers:

In dem Beispiel des Blutdrucksenkers:

$x \pm t_{n-1;5\%} s_{\bar{x}}$ =

(-10,6±2,365·4,917) mmHg =

(-10,6±11,6) mmHg

d.h. μ ist in: -22,2 … 0,8 mmHg

=> μ kann 0 sein.

Die Schätzung und der *t*-Test geben derselbe Ergebnisse!

FG \ p	0.05	0.02	0.01
2	4.30266	6.96455	9.92499
3	3.18245	4.54071	5.84085
4	2.77645	3.74694	4.60408
5	2.57058	3.36493	4.03212
6	2.44691	3.14267	3.70743
7	2.36462	2.99795	3.49948
8	2.30601	2.89647	3.35538
9	2.26216	2.82143	3.24984
10	2.22814	2.76377	3.16926
11	2.20099	2.71808	3.10582
12	2.17881	2.68099	3.05454
13	2.16037	2.65030	3.01228
14	2.14479	2.62449	2.97685
15	2.13145	2.60248	2.94673
20	2.08596	2.52798	2.84534
50	2.00856	2.40327	2.67779
70	1.99444	2.38080	2.64790
100	1.98397	2.36421	2.62589
unendlich	1.95996	2.32635	2.57583