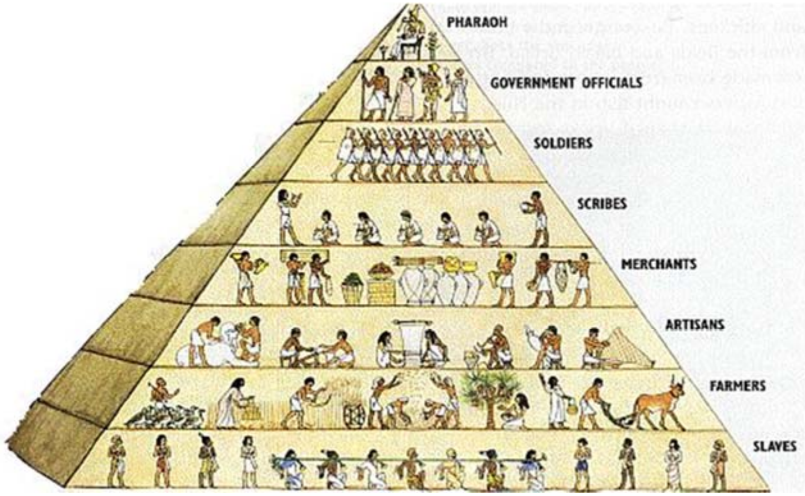
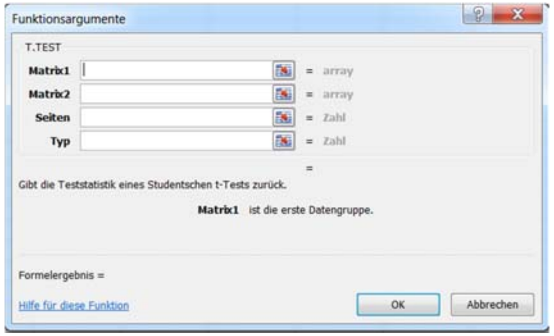
# Nichtparametrische Methoden




KAD 2017.02.07

# =T.TEST( )




**Matrix1** ist die erste Datengruppe.

**Matrix2** ist die zweite Datengruppe.

**Seiten** bestimmt die Anzahl der Endflächen.

**Typ** bestimmt die Form des durchzuführenden t-Tests.

Parameter

| Ist Typ gleich | Wird folgender Test ausgeführt |
|---|---|
| 1 | Gepaart ← **Einstichproben t-Test** |
| 2 | Zwei Stichproben, gleiche Varianz (homoskedastisch) |
| 3 | Zwei Stichproben, ungleiche Varianz (heteroskedastisch) |



# Test auf Varianzgleichheit: *F*-test

Nullhypothese: Die Varianzen sind gleich

Parameter: $F = \dfrac{s_1^2}{s_2^2}$ ; $s_1 > s_2$

Bei der Gültigkeit der Nullhypothese *F* folgt eine *F*-Verteilung mit $n_1$-1 und $n_2$-1 Freiheitsgrade

Bemerkung: Tabelle zum einseitigen Test
wir brauchen einen zweiseitigen Test

$$\boxed{F = \frac{s_1^2}{s_2^2}}$$

**der kritische Wert (aus der Tabelle)**

$F < F_{n_1-1, n_2-1; 5\%}$ wir verwerfen die *Nullhypothese* nicht
**d.h. die Varianzen sind gleich**

$F > F_{n_1-1, n_2-1; 5\%}$ wir verwerfen die *Nullhypothese*
**d.h. die Varianzen sind nicht gleich**

## Übersicht der Testmethode

| Verteilung / Stichproben | *normalverteilte Daten* | *die Verteilung der Daten ist unbekannt* |
|---|---|---|
| *eine Stichprobe* | Einstichproben t-Test | Wilcoxon Test |
| *zwei Stichproben* | Zweistichproben t-Test | Mann-Whittney U-Test |
| *mehrere Stichproben* | ANOVA (Varianzanalyse) | Kruskal-Wallis Test |



## Nichtparametrische Methoden

**Bedingungen der *t*-Tests**

- kontinuierliches Merkmal  (z.B. Körperhöhe, Körpertemperatur...)
- die Daten müssen normalverteilt sein

**Nichtparametrische Methoden**

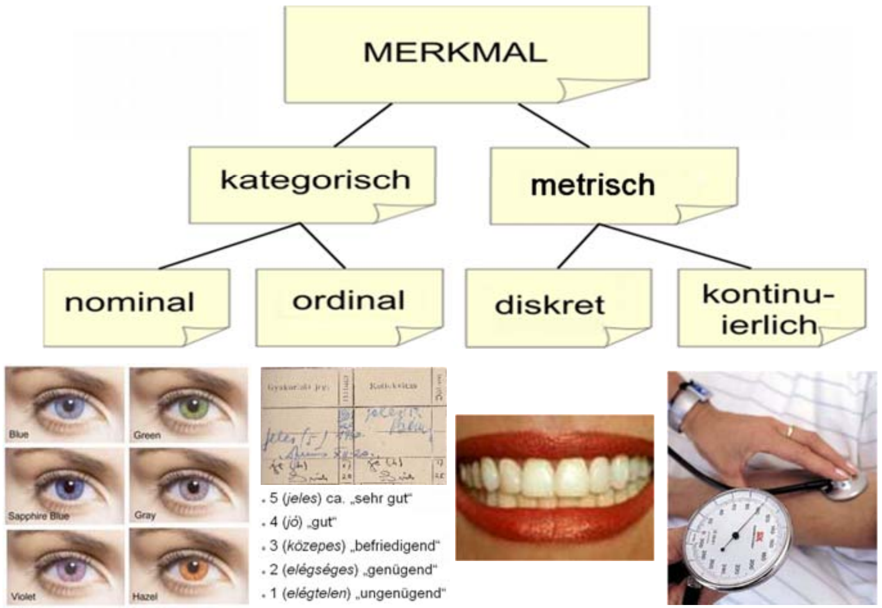
- nur ordinale Daten (Ordinalskala)
- keine Normalverteilung  (auch bei unbekannter Verteilung möglich)

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

1, 2, 3, 4, 5 oder



**Klassifizierung der Merkmale**




**Vorteile:**

- Verteilungsunabhängigkeit
- Ordinal-, Intervall-, Verhältnisskalen

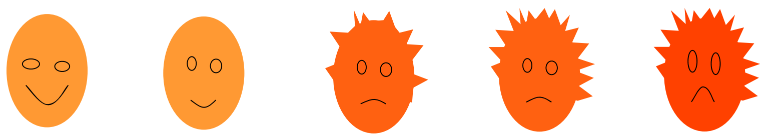
**Nachteile:**

- Datenreduktion, Informationsverlust
- größere Wahrscheinlichkeit der Fehler 2. Art:
  nur größere Unterschiede können statistisch bewiesen werden

## Prinzip der Rang-Tests

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!



**Ergänzungsmaterial**

## 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 Durchschnittsrang



**Ergänzungsmaterial**

## 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!



**Ergänzungsmaterial**

## Eine Stichprobe: Wilcoxon-Vorzeichen Rangtest

eine Stichprobe (gepaarte Daten)

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!

**Ergänzungsmaterial**

## 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



**Ergänzungsmaterial**

## 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!):

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

$\bar{R}$: Durchschnitt der Ränge

$s$: Standardabweichung der Ränge

$n$: Anzahl der Daten

$n-1$: Freiheitsgrad

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

**Ränge mit Vorzeichen:**

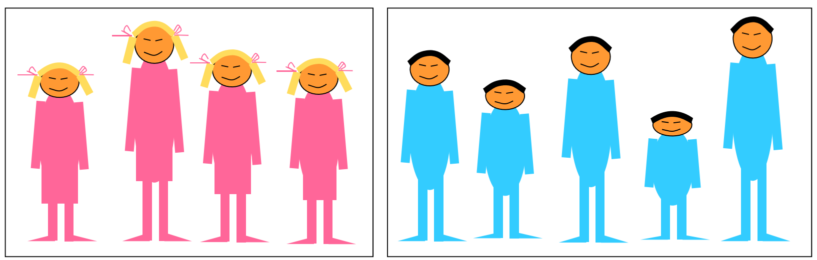
**-1, -2, +3, -4, +5, +6, +7, +8, +9, +10** → Durchschnitt: 4.10
Standardabw.: 4.91

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

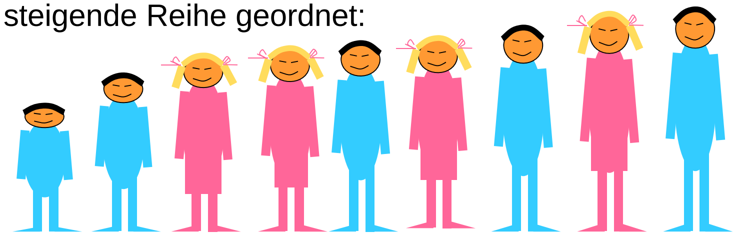
$t_{9;5\%}=2{,}26$ (aus der Tabelle)  $p < 5\%$ (mit Excel)



## Vergleich von zwei Stichproben: Mann-Whitney Test



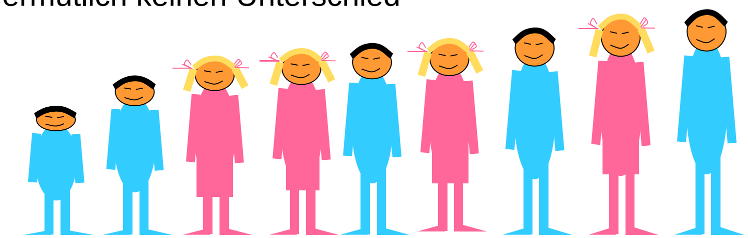
In steigende Reihe geordnet:



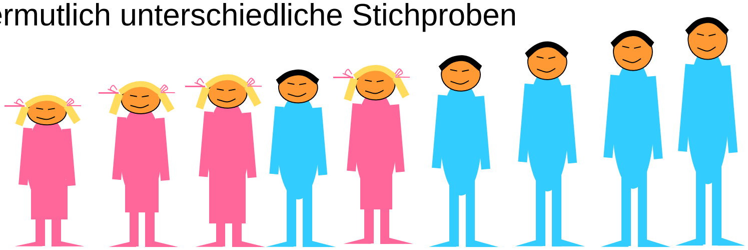
Ränge: 1 2 3 4 5 6 7 8 9



vermutlich keinen Unterschied



vermutlich unterschiedliche Stichproben



1 2 3 4 5 6 7 8 9

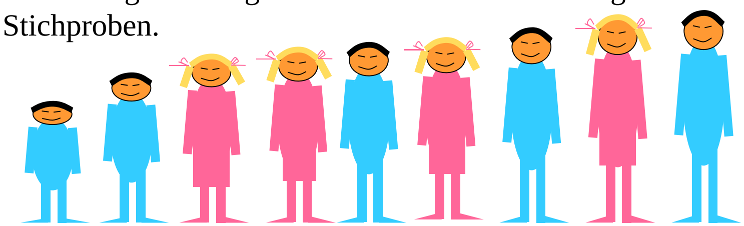
## Mann – Whitney *U* Test (Annäherung)

(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: $T_1$.

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





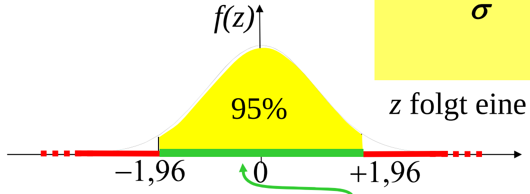
## Mann – Whitney *U* Test: Annäherung

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:

$$\mu = \frac{n_1(n_1+n_2+1)}{2} \qquad \sigma = \sqrt{n_1 n_2(n_1+n_2+1)/12}$$

$$z = \frac{T_1-\mu}{\sigma} = \frac{T_1 - \frac{n_1(n_1+n_2+1)}{2}}{\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}}$$

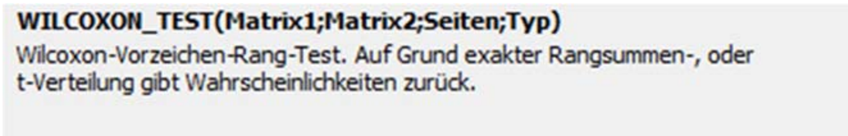



$z$ folgt eine Standard-Normalverteilung (wenn $H_0$ gültig ist)

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



## Mit der erweiterten Excel Funktionen

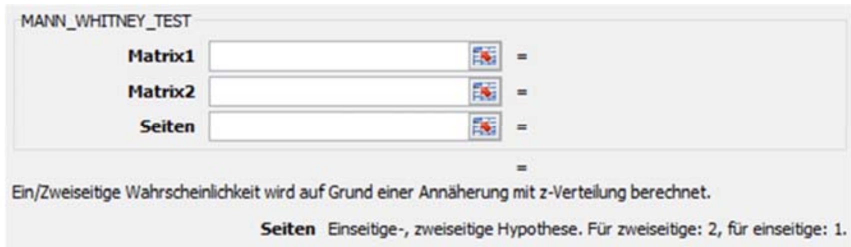


**WILCOXON_TEST(Matrix1;Matrix2;Seiten;Typ)**
Wilcoxon-Vorzeichen-Rang-Test. Auf Grund exakter Rangsummen-, oder t-Verteilung gibt Wahrscheinlichkeiten zurück.

WILCOXON_TEST
Matrix1 =
Matrix2 =
Seiten =
Typ =
=
Wilcoxon-Vorzeichen-Rang-Test. Auf Grund exakter Rangsummen-, oder t-Verteilung gibt Wahrscheinlichkeiten zurück.
**Typ** Exakt oder t-Verteilung; exakt:TRUE, t-Verteilung:FALSE.



**MANN_WHITNEY_TEST(Matrix1;Matrix2;Seiten)**
Ein/Zweiseitige Wahrscheinlichkeit wird auf Grund einer Annäherung mit z-Verteilung berechnet.



MANN_WHITNEY_TEST
Matrix1 =
Matrix2 =
Seiten =
=
Ein/Zweiseitige Wahrscheinlichkeit wird auf Grund einer Annäherung mit z-Verteilung berechnet.
**Seiten** Einseitige-, zweiseitige Hypothese. Für zweiseitige: 2, für einseitige: 1.

Die Funktion liefert eine Wahrscheinlickeit, den Fehler von 1.Art.

$p < 0.05$ (oder die max. annehmbare Irrtumswahrscheinlichkeit = Signifikanzniveau): Ablehnen der Nullhypothese

$p > 0.05$: Annehmen der Nullhypothese