

Statistik für Digital Humanities

Nichtparametrische Testverfahren

Dr. Jochen Tiepmar

Institut für Informatik
Computational Humanities
Universität Leipzig

04. Januar 2021

[Letzte Aktualisierung: 09/02/2021, 20:05]

- 1 Nichtparametrische Testverfahren
- 2 Wilcoxon's Rangsummen / Mann-Whitney
- 3 Wilcoxon's Vorzeichen-Rang Test
- 4 Kruskal-Wallis Test
- 5 Trendanalyse mit Jonckheere-Terpstra
- 6 Friedmans ANOVA
- 7 Effektstärke bei Friedman und Kruskal-Wallis

- Nichtparametrische / Annahmenfreie Tests
- Weniger Annahmen bezogen auf die Daten Nicht *keine* Annahmen!
- Gerüchtweise weniger Aussagekraft (Power) als parametrische Gegenstücke
 - ... nur bestimmbar bei gegebener Normalverteilung (Typ I / II Fehler)
- Oft über Ranking gelöst Siehe Spearman Korrelation

- Ranking erzeugt künstliche Intervallskalierung
- Rechenschritte:
 - Sortiere Werte von klein nach groß
 - Nummeriere die Werte = Rang
 - Bei gleichen Werten verwende deren Durchschnittsrang Tied Ranks / Gebundene Ränge

- Ranking erzeugt künstliche Intervallskalierung
- Rechenschritte:
 - Sortiere Werte von klein nach groß
 - Nummeriere die Werte = Rang
 - Bei gleichen Werten verwende deren Durchschnittsrang Tied Ranks / Gebundene Ränge

Beispiel $X = \{D, G, F, R, F, R\}$

- Ranking erzeugt künstliche Intervallskalierung
- Rechenschritte:
 - Sortiere Werte von klein nach groß
 - Nummeriere die Werte = Rang
 - Bei gleichen Werten verwende deren Durchschnittsrang Tied Ranks / Gebundene Ränge

Beispiel $X = \{D, G, F, R, F, R\}$

Sortiert	D	F	F	G	R	R
----------	---	---	---	---	---	---

- Ranking erzeugt künstliche Intervallskalierung
- Rechenschritte:
 - Sortiere Werte von klein nach groß
 - Nummeriere die Werte = Rang
 - Bei gleichen Werten verwende deren Durchschnittsrang Tied Ranks / Gebundene Ränge

Beispiel $X = \{D, G, F, R, F, R\}$

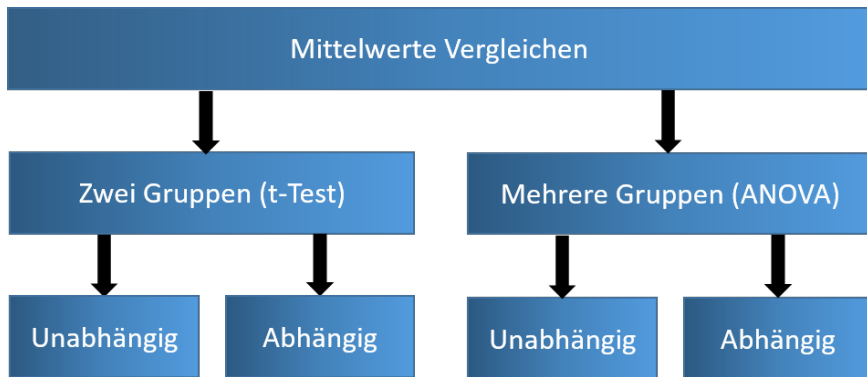
Sortiert	D	F	F	G	R	R
Pot. Rang	1	2	3	4	5	6

- Ranking erzeugt künstliche Intervallskalierung
- Rechenschritte:
 - Sortiere Werte von klein nach groß
 - Nummeriere die Werte = Rang
 - Bei gleichen Werten verwende deren Durchschnittsrang Tied Ranks / Gebundene Ränge

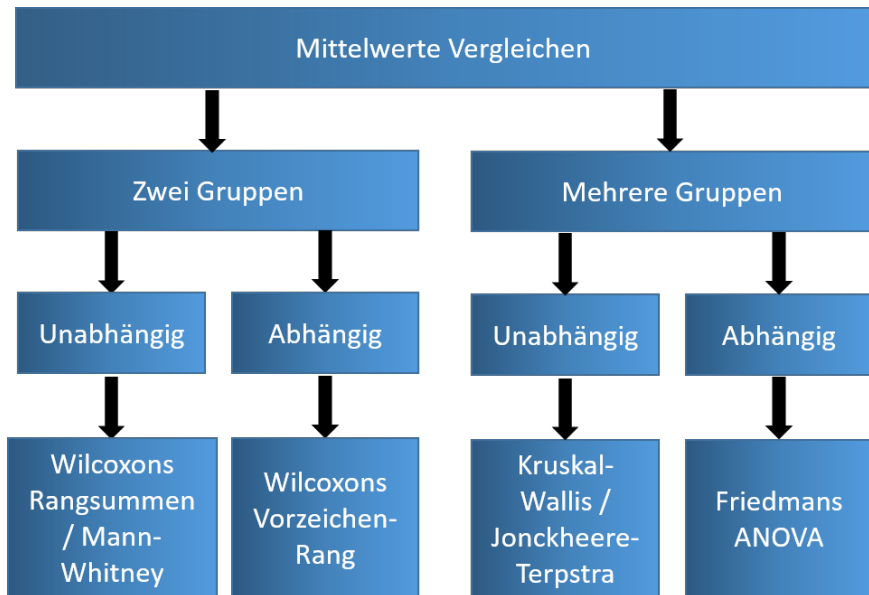
Beispiel $X = \{D, G, F, R, F, R\}$

Sortiert	D	F	F	G	R	R
Pot. Rang	1	2	3	4	5	6
Rang	1	2.5	2.5	4	5.5	5.5

Achtung: Die Abstandsinformation geht damit verloren, deshalb Mediane, Range und am besten Boxplots immer mit angeben



Nichtparametrische Mittelwertvergleiche



Wilcoxon's Rangsummen / Mann-Whitney Test

Mann, H.B. & Whitney, D.R. (1947): *On a test whether one or two random variables are statistically larger than the other*

Wilcoxon, F. (1945): *Individual comparisons by ranking methods*

- Beide machen dasselbe. R verwendet Wilcoxon's Rangsummentest
- Parametrische Entsprechung: unabhängiger t-Test
- Grundprinzip: Vergleiche die Ränge beider Gruppen
 - Wenn ein Unterschied existiert, sollten in einer Gruppe signifikant mehr hohe/niedrige Ränge auftauchen als in der anderen.
- H_0 = Es existiert kein signifikanter Unterschied in den Rängen. Gilt wenn $W > W_{kr}$
- W_{kr} → Wilcoxon/Mann-Whitney Tabelle
- $W_{group} = \sum (ranks_{group}) - Normalrang_{group}$
- $Normalrang = 1 + 2 + 3 + \dots + n = \frac{n*(n+1)}{2}$ mean Rank. Achtung: Nicht mean der Ranks(!)
- $W_{group} = \sum ranks_{group} - \frac{n_{group}*(n_{group}+1)}{2}$
- $W = \min(W_{group})$

Depression (Beck Depression Index) nach Drogenkonsum am Samstag

Proband	Droge	BDI Sonntag	BDI Mittwoch
1	Ecstasy	15	28
2	Ecstasy	35	35
...
9	Ecstasy	13	36
10	Ecstasy	20	35
11	Alkohol	16	5
12	Alkohol	15	6
...
19	Alkohol	18	3
20	Alkohol	18	10

Kompletter Datensatz Siehe folgende Rangtabelle

Wilcoxon's Rangsummen / Mann-Whitney Test Beispiel

Sonntag:

BDI	13	13	14	15	15	15	16	16	16	16	17	18	18	18	19	19	20	20	27	35
Pot. Rang	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Rang	1.5	1.5	3	5	5	5	8.5	8.5	8.5	8.5	11	13	13	13	15.5	15.5	17.5	17.5	19	20
Gruppe	A	E	A	A	A	E	A	A	E	E	E	E	A	A	E	A	E	A	E	E

Mittwoch:

BDI	3	5	6	6	7	8	9	10	17	24	27	28	29	30	32	35	35	35	36	39
Pot. Rang	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Rang	1	2	3.5	3.5	5	6	7	8	9	10	11	12	13	14	15	17	17	17	19	20
Gruppe	A	A	A	A	A	A	A	A	A	E	E	E	E	A	E	E	E	E	E	E

Wilcoxon's Rangsummen / Mann-Whitney Test Beispiel

Mittwoch:

BDI	3	5	6	6	7	8	9	10	17	24	27	28	29	30	32	35	35	35	36	39
Pot. Rang	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Rang	1	2	3.5	3.5	5	6	7	8	9	10	11	12	13	14	15	17	17	17	19	20
Gruppe	A	A	A	A	A	A	A	A	A	E	E	E	E	A	E	E	E	E	E	E

Wilcoxon's Rangsummen / Mann-Whitney Test Beispiel

Mittwoch:

BDI	3	5	6	6	7	8	9	10	17	24	27	28	29	30	32	35	35	35	36	39
Pot. Rang	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Rang	1	2	3.5	3.5	5	6	7	8	9	10	11	12	13	14	15	17	17	17	19	20
Gruppe	A	A	A	A	A	A	A	A	A	E	E	E	E	A	E	E	E	E	E	E

- $W_{group} = \sum ranks_{group} - \frac{n_{group} * (n_{group} + 1)}{2}$
- $W_{Ecstasy} = 151 - 55 = 96$
- $W_{Alcohol} = 59 - 55 = 4$
- $W = \min(W_{group}) = 4 < W_{kr}(10, 10) = 82 \rightarrow H_0$ abgelehnt
→

Mittwoch:

BDI	3	5	6	6	7	8	9	10	17	24	27	28	29	30	32	35	35	35	36	39
Pot. Rang	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Rang	1	2	3.5	3.5	5	6	7	8	9	10	11	12	13	14	15	17	17	17	19	20
Gruppe	A	A	A	A	A	A	A	A	A	E	E	E	E	A	E	E	E	E	E	E

- $W_{group} = \sum ranks_{group} - \frac{n_{group} * (n_{group} + 1)}{2}$
- $W_{Ecstasy} = 151 - 55 = 96$
- $W_{Alcohol} = 59 - 55 = 4$
- $W = \min(W_{group}) = 4 < W_{kr}(10, 10) = 82 \rightarrow H_0$ abgelehnt
→ Unterschied signifikant

Wilcoxon's Rangsummen / Mann-Whitney Test

Wahrscheinlichkeiten

- Berechnung von p folgt der Annahme der Normalverteilung
 - Weder Ränge noch Daten sind hier normalverteilt 😞

Wilcoxon's Rangsummen / Mann-Whitney Test

Wahrscheinlichkeiten

- Berechnung von p folgt der Annahme der Normalverteilung
 - Weder Ränge noch Daten sind hier normalverteilt 😞
- Monte Carlo Methode (exakt)
 - Häufige zufällige Zuordnung der Daten in Gruppen
 - Jeweils Prüfgröße berechnen
 - → Wahrscheinlichkeitsverteilung der Prüfgrößen künstlich erzeugt
 - Nur möglich wenn keine gebundenen Ränge existieren

Wilcoxon's Rangsummen / Mann-Whitney Test

Wahrscheinlichkeiten

- Berechnung von p folgt der Annahme der Normalverteilung
 - Weder Ränge noch Daten sind hier normalverteilt 😞
- Monte Carlo Methode (exakt)
 - Häufige zufällige Zuordnung der Daten in Gruppen
 - Jeweils Prüfgröße berechnen
 - → Wahrscheinlichkeitsverteilung der Prüfgrößen künstlich erzeugt
 - Nur möglich wenn keine gebundenen Ränge existieren
- Normal-Approximation
 - Normalverteilung der Stichprobenverteilung angenommen
 - $SE_{Teststatistik} \rightarrow zScore \rightarrow p$
 - Optional Stetigkeitskorrektur, da Ränge immer um 1 oder 0.5 steigen

Wilcoxon's Rangsummen / Mann-Whitney Test

Wahrscheinlichkeiten

- Berechnung von p folgt der Annahme der Normalverteilung
 - Weder Ränge noch Daten sind hier normalverteilt 😞
- Monte Carlo Methode (exakt)
 - Häufige zufällige Zuordnung der Daten in Gruppen
 - Jeweils Prüfgröße berechnen
 - → Wahrscheinlichkeitsverteilung der Prüfgrößen künstlich erzeugt
 - Nur möglich wenn keine gebundenen Ränge existieren
- Normal-Approximation
 - Normalverteilung der Stichprobenverteilung angenommen
 - $SE_{Teststatistik} \rightarrow zScore \rightarrow p$
 - Optional Stetigkeitskorrektur, da Ränge immer um 1 oder 0.5 steigen
- Monte Carlo besser als Normalenabschätzung aber aufwendiger
- Stetigkeitskorrektur hat wenig Einfluß auf Ergebnis
- R verwendet per default Normal-Approximation mit Stetigkeitskorrektur ab $n = 40$

Wilcoxon's Rangsummen / Mann-Whitney Test

Wahrscheinlichkeiten

- Berechnung von p folgt der Annahme der Normalverteilung
 - Weder Ränge noch Daten sind hier normalverteilt 😞
- Monte Carlo Methode (exakt)
 - Häufige zufällige Zuordnung der Daten in Gruppen
 - Jeweils Prüfgröße berechnen
 - → Wahrscheinlichkeitsverteilung der Prüfgrößen künstlich erzeugt
 - Nur möglich wenn keine gebundenen Ränge existieren
- Normal-Approximation
 - Normalverteilung der Stichprobenverteilung angenommen
 - $SE_{Teststatistik} \rightarrow zScore \rightarrow p$
 - Optional Stetigkeitskorrektur, da Ränge immer um 1 oder 0.5 steigen
- Monte Carlo besser als Normalenabschätzung aber aufwendiger
- Stetigkeitskorrektur hat wenig Einfluß auf Ergebnis
- R verwendet per default Normal-Approximation mit Stetigkeitskorrektur ab $n = 40$
- ... Die Experten sind sich uneinig

```
sundayBDI<-c(16, 15, 20, 15, 16, 13, 14, 19, 18, 18, 15, 35, 16, 18, 19, 17,  
            27, 16, 13, 20)  
wedsBDI<-c(5, 6, 30, 8, 9, 7, 6, 17, 3, 10, 28, 35, 35, 24, 39, 32, 27, 29,  
           36, 35)  
  
drug<-gl(2, 10, labels = c("Alcohol", "Ecstasy"))  
drugData<-data.frame(drug, sundayBDI, wedsBDI)  
wilcox.test(wedsBDI ~ drug, data = drugData, exact=FALSE, correct=FALSE)
```

Wilcoxon rank sum test

data: wedsBDI by drug

W = 4, p-value = 0.0004943

alternative hypothesis: true location shift is not equal to 0

$p < 0.05 \rightarrow H_0$ abgewiesen

Wilcoxon's Rangsummen / Mann-Whitney Test in R

```
sundayBDI<-c(16, 15, 20, 15, 16, 13, 14, 19, 18, 18, 15, 35, 16, 18, 19, 17,  
            27, 16, 13, 20)  
wedsBDI<-c(5, 6, 30, 8, 9, 7, 6, 17, 3, 10, 28, 35, 35, 24, 39, 32, 27, 29,  
           36, 35)  
  
drug<-gl(2, 10, labels = c("Alcohol", "Ecstasy"))  
drugData<-data.frame(drug, sundayBDI, wedsBDI)  
wilcox.test(wedsBDI ~ drug, data = drugData, exact=FALSE, correct=FALSE)
```

Wilcoxon rank sum test

data: wedsBDI by drug

W = 4, p-value = 0.0004943

alternative hypothesis: true location shift is not equal to 0

$p < 0.05 \rightarrow H_0$ abgewiesen \rightarrow Unterschied signifikant

Unschön: R wählt nicht das kleinste W sondern abhängig von Inputreihenfolge

```
sundayBDI<-c(15, 35, 16, 18, 19, 17,27, 16, 13, 20, 16, 15, 20, 15, 16, 13,  
            14, 19, 18, 18)  
wedsBDI<-c(28, 35, 35, 24, 39, 32, 27, 29, 36, 35, 5, 6, 30, 8, 9, 7, 6, 17,  
           3, 10)  
drug<-gl(2, 10, labels = c("Ecstasy", "Alcohol"))  
drugData<-data.frame(drug, sundayBDI, wedsBDI)  
wilcox.test(wedsBDI ~ drug, data = drugData, exact=FALSE, correct=FALSE)
```

```
Wilcoxon rank sum test
```

```
data: wedsBDI by drug  
W = 96, p-value = 0.0004943  
alternative hypothesis: true location shift is not equal to 0
```

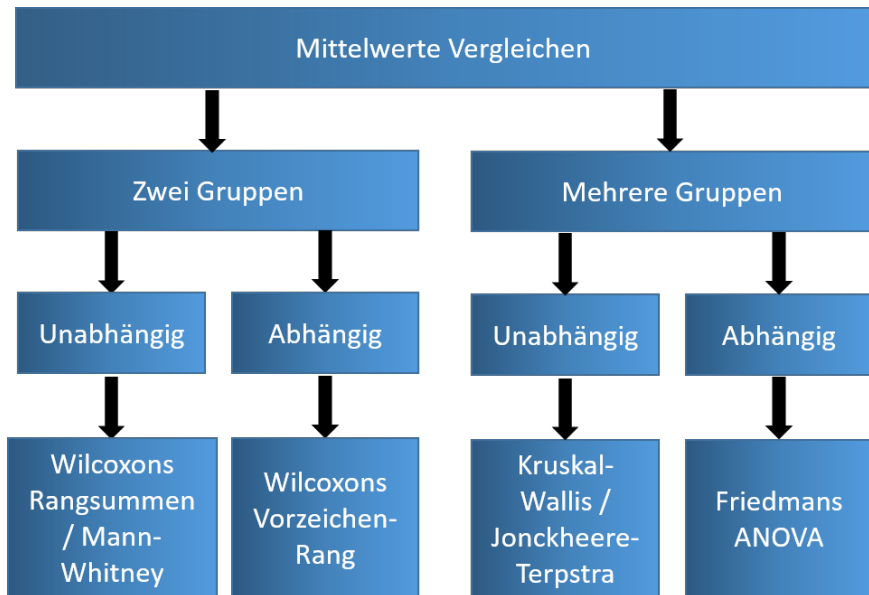
... aber p ändert sich dadurch nicht, also eigentlich egal

Folgendes Skript generiert r für beide Arten von Wilcoxonmodell

```
rFromWilcox<-function(wilcoxModel, n){  
  z<- qnorm(wilcoxModel$p.value/2)  
  r<- z/ sqrt(N)  
  cat(wilcoxModel$data.name, "Effect Size, r = ", r)  
}
```

Interpretation Siehe Vorlesung zu Korrelation

Nichtparametrische Testverfahren



Wilcoxon's Vorzeichen-Rang Test

Wilcoxon, F. (1945): *Individual comparisons by ranking methods*

- Signed Rank Test
- Parametrische Entsprechung: abhängiger t-Test
- Wir interessieren uns also für die Differenz der Ränge von einer Gruppe zur anderen
- H_0 = Die Differenz ist nicht signifikant. Abschätzung mittels z-Score

Wilcoxon's Vorzeichen-Rang Test

Wilcoxon, F. (1945): *Individual comparisons by ranking methods*

- Signed Rank Test
- Parametrische Entsprechung: abhängiger t-Test
- Wir interessieren uns also für die Differenz der Ränge von einer Gruppe zur anderen
- H_0 = Die Differenz ist nicht signifikant. Abschätzung mittels z-Score

Berechnung:

- Teststatistik berechnen
 - Ränge der Differenzen berechnen, Differenz=0 → Exclude
 - T_+ , T_- : Ränge der positiven und negativen Differenzen summieren
 - $T = \min(T_+, T_-)$

Wilcoxon's Vorzeichen-Rang Test

Wilcoxon, F. (1945): *Individual comparisons by ranking methods*

- Signed Rank Test
- Parametrische Entsprechung: abhängiger t-Test
- Wir interessieren uns also für die Differenz der Ränge von einer Gruppe zur anderen
- H_0 = Die Differenz ist nicht signifikant. Abschätzung mittels z-Score

Berechnung:

- Teststatistik berechnen
 - Ränge der Differenzen berechnen, Differenz=0 → Exclude
 - T_+ , T_- : Ränge der positiven und negativen Differenzen summieren
 - $T = \min(T_+, T_-)$
- Signifikanz bestimmen:
 - $\bar{T} = \frac{n*(n+1)}{4}$ Mittelwert der Zufallsverteilung
 - $SE_{\bar{T}} = \sqrt{\frac{n*(n+1)*(n+2)}{24}}$
 - $z = \frac{T - \bar{T}}{SE_{\bar{T}}}$
 - $z > 1.96 \rightarrow H_0$ abgewiesen → Differenz signifikant mit $p < 0.05$

Wilcoxon's Vorzeichen-Rang Test Beispiel

Sonntag:

BDI	13	13	14	15	15	15	16	16	16	16	17	18	18	18	19	19	20	20	27	35
Pot. Rang	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Rang	1.5	1.5	3	5	5	5	8.5	8.5	8.5	8.5	11	13	13	13	15.5	15.5	17.5	17.5	19	20
Gruppe	A	E	A	A	A	E	A	A	E	E	E	E	A	A	E	A	E	A	E	E

Mittwoch:

BDI	3	5	6	6	7	8	9	10	17	24	27	28	29	30	32	35	35	35	36	39
Pot. Rang	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Rang	1	2	3.5	3.5	5	6	7	8	9	10	11	12	13	14	15	17	17	17	19	20
Gruppe	A	A	A	A	A	A	A	A	A	E	E	E	E	A	E	E	E	E	E	E

Wilcoxon's Vorzeichen-Rang Test Beispiel

Ecstasy											
BDI Sonntag	15	35	16	18	19	17	27	16	13	20	
BDI Mittwoch	28	35	35	24	39	32	27	29	36	35	

Wilcoxon's Vorzeichen-Rang Test Beispiel

Ecstasy											
BDI Sonntag	15	35	16	18	19	17	27	16	13	20	
BDI Mittwoch	28	35	35	24	39	32	27	29	36	35	
Differenz	13	0	19	6	20	15	0	13	23	15	

Wilcoxon's Vorzeichen-Rang Test Beispiel

Ecstasy											
BDI Sonntag	15	35	16	18	19	17	27	16	13	20	
BDI Mittwoch	28	35	35	24	39	32	27	29	36	35	
Differenz	13	0	19	6	20	15	0	13	23	15	
Vorzeichen	+	XX	+	+	+	+	XX	+	+	+	

Wilcoxon's Vorzeichen-Rang Test Beispiel

Ecstasy											
BDI Sonntag	15	35	16	18	19	17	27	16	13	20	
BDI Mittwoch	28	35	35	24	39	32	27	29	36	35	
Differenz	13	0	19	6	20	15	0	13	23	15	
Vorzeichen	+	XX	+	+	+	+	XX	+	+	+	
Rang	2.5		6	1	7	4.5		2.5	8	4.5	sum

Wilcoxon's Vorzeichen-Rang Test Beispiel

Ecstasy											
BDI Sonntag	15	35	16	18	19	17	27	16	13	20	
BDI Mittwoch	28	35	35	24	39	32	27	29	36	35	
Differenz	13	0	19	6	20	15	0	13	23	15	
Vorzeichen	+	XX	+	+	+	+	XX	+	+	+	
Rang	2.5		6	1	7	4.5		2.5	8	4.5	sum
Positive Ränge	2.5		6	1	7	4.5		2.5	8	4.5	36
Negative Ränge											0
Alkohol

Teststatistik berechnen

Wilcoxon's Vorzeichen-Rang Test Beispiel

Teststatistik berechnen

$$T_E = \min(36, 0) = 0$$

Signifikanz bestimmen:

$$\overline{T_E} = \frac{8*(8+1)}{4} = 18$$

$$SE_{\overline{T_E}} = \sqrt{\frac{8*(8+1)*(8+2)}{24}} = 7.14 \quad z_E = \frac{T_E - \overline{T_E}}{SE_{\overline{T_E}}} = \frac{0-18}{7.14} = -2.52$$

$$z_A = \dots \ddot{U}bungsgelegenheit \dots = -1.99$$

Wilcoxon's Vorzeichen-Rang Test Beispiel

Teststatistik berechnen

$$T_E = \min(36, 0) = 0$$

Signifikanz bestimmen:

$$\overline{T_E} = \frac{8 \cdot (8+1)}{4} = 18$$

$$SE_{\overline{T_E}} = \sqrt{\frac{8 \cdot (8+1) \cdot (8+2)}{24}} = 7.14 \quad z_E = \frac{T_E - \overline{T_E}}{SE_{\overline{T_E}}} = \frac{0 - 18}{7.14} = -2.52$$

$$z_A = \dots \text{Übungsgelegenheit} \dots = -1.99$$

Interpretation

Sowohl z_E als auch z_A liegen über 1.96 $\rightarrow H_0$ für beide abgewiesen \rightarrow Differenz zwischen Sonntag und Mittwoch ist für beide signifikant

Wilcoxon's Vorzeichen-Rang Test in R

```
sundayBDI<-c(16, 15, 20, 15, 16, 13, 14, 19, 18, 18, 15, 35, 16, 18, 19, 17,  
27, 16, 13, 20)  
wedsBDI<-c(5, 6, 30, 8, 9, 7, 6, 17, 3, 10, 28, 35, 35, 24, 39, 32, 27, 29,  
36, 35)  
  
drug<-gl(2, 10, labels = c("Alcohol", "Ecstasy"))  
drugData<-data.frame(drug, sundayBDI, wedsBDI)  
alcoholData<-subset(drugData, drug == "Alcohol")  
ecstasyData<-subset(drugData, drug == "Ecstasy")  
wilcox.test(alcoholData$wedsBDI, alcoholData$sundayBDI, paired = TRUE,  
correct= FALSE, exact=FALSE)  
wilcox.test(ecstasyData$wedsBDI, ecstasyData$sundayBDI, paired = TRUE,  
correct= FALSE, exact=FALSE)
```

Wilcoxon signed rank test

data: alcoholData\$wedsBDI and alcoholData\$sundayBDI

V = 8, p-value = 0.04657

alternative hypothesis: true location shift is not equal to 0

Wilcoxon signed rank test

data: ecstasyData\$wedsBDI and ecstasyData\$sundayBDI

V = 36, p-value = 0.01151

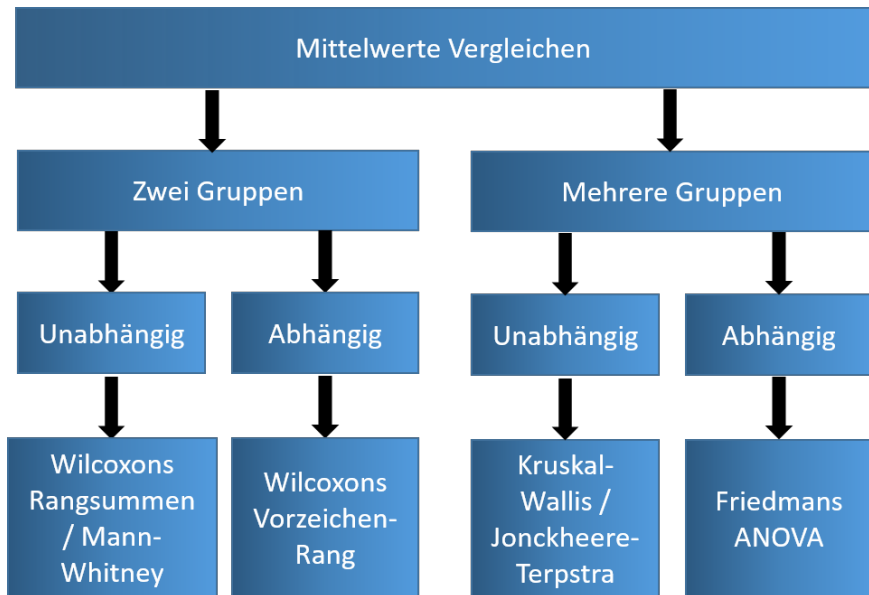
alternative hypothesis: true location shift is not equal to 0

Folgendes Skript generiert r für beide Arten von Wilcoxonmodell

```
rFromWilcox<-function(wilcoxModel, n){  
  z<- qnorm(wilcoxModel$p.value/2)  
  r<- z/ sqrt(N)  
  cat(wilcoxModel$data.name, "Effect Size, r = ", r)  
}
```

Interpretation Siehe Vorlesung zu Korrelation

Nichtparametrische Testverfahren



Kruskal, W.H. & Wallis, W.A. (1952): *Use of ranks in one-criterion variance analysis*

- ANOVA relativ robust und Welchs F hilft bei heterogenen Varianzen
Welch, B.L. (1951): *On the comparison of several mean values: An alternative approach*
- Kruskal-Wallis H als nichtparametrische Alternative für unabhängige ANOVA mittels Ranking
- Omnibustest
- *Nullhypothese* : Die Gruppen sind nicht signifikant verschieden. Gilt bei $H < H_{kr}$
- $H = \frac{12}{n*(n+1)} * \sum \left(\frac{\text{Rangsumme}_{group}^2}{n_{group}} \right) - 3 * (n + 1)$ Rangsumme manchmal als R abgekürzt
- H_{kr} folgt χ^2 -Verteilung mit $df = \text{Gruppenanzahl} - 1$

Kruskal-Wallis Beispiel

Zeichenlänge des Dokumententitels pro Autor (wie bei ANOVA)

Autor 1	Autor 2	Autor 3
30	40	10
35	35	25
45	50	5

Zeichenlänge des Dokumententitels pro Autor (wie bei ANOVA)

Autor 1	Rang	Autor 2	Rang	Autor 3	Rang
30	4	40	7	10	2
35	5.5	35	5.5	25	3
45	8	50	9	5	1

Kruskal-Wallis Beispiel

Zeichenlänge des Dokumententitels pro Autor (wie bei ANOVA)

Autor 1	Rang	Autor 2	Rang	Autor 3	Rang
30	4	40	7	10	2
35	5.5	35	5.5	25	3
45	8	50	9	5	1
sum	17.5	sum	21.5	sum	6

$$H = \frac{12}{n*(n+1)} * \sum \left(\frac{Rangsumme_{group}^2}{n_{group}} \right) - 3 * (n + 1)$$

Kruskal-Wallis Beispiel

Zeichenlänge des Dokumententitels pro Autor (wie bei ANOVA)

Autor 1	Rang	Autor 2	Rang	Autor 3	Rang
30	4	40	7	10	2
35	5.5	35	5.5	25	3
45	8	50	9	5	1
sum	17.5	sum	21.5	sum	6

$$H = \frac{12}{n*(n+1)} * \sum \left(\frac{Rangsumme_{group}^2}{n_{group}} \right) - 3 * (n + 1)$$

$$H = \frac{12}{9*(10)} * \left(\frac{306.25}{3} + \frac{462.25}{3} + \frac{36}{3} \right) - 3 * (10) \approx$$

$$0.133 * (102.083 + 154.083 + 12) - 30 \approx 0.133 * 268.166 - 30 \approx 5.666$$

Zeichenlänge des Dokumententitels pro Autor (wie bei ANOVA)

Autor 1	Rang	Autor 2	Rang	Autor 3	Rang
30	4	40	7	10	2
35	5.5	35	5.5	25	3
45	8	50	9	5	1
sum	17.5	sum	21.5	sum	6

$$H = \frac{12}{n*(n+1)} * \sum \left(\frac{Rangsumme_{group}^2}{n_{group}} \right) - 3 * (n + 1)$$

$$H = \frac{12}{9*(10)} * \left(\frac{306.25}{3} + \frac{462.25}{3} + \frac{36}{3} \right) - 3 * (10) \approx$$

$$0.133 * (102.083 + 154.083 + 12) - 30 \approx 0.133 * 268.166 - 30 \approx 5.666$$

$H = 5.666 < 5.99 = H_{kr}(2) \rightarrow$ Nullhypothese bestätigt \rightarrow Es existieren keine signifikante Unterschiede zwischen den Gruppen.

Kruskal-Wallis Beispiel

Zeichenlänge des Dokumententitels pro Autor (wie bei ANOVA)

Autor 1	Rang	Autor 2	Rang	Autor 3	Rang
30	4	40	7	10	2
35	5.5	35	5.5	25	3
45	8	50	9	5	1
sum	17.5	sum	21.5	sum	6

$$H = \frac{12}{n*(n+1)} * \sum \left(\frac{Rangsumme_{group}^2}{n_{group}} \right) - 3 * (n + 1)$$

$$H = \frac{12}{9*(10)} * \left(\frac{306.25}{3} + \frac{462.25}{3} + \frac{36}{3} \right) - 3 * (10) \approx$$

$$0.133 * (102.083 + 154.083 + 12) - 30 \approx 0.133 * 268.166 - 30 \approx 5.666$$

$H = 5.666 < 5.99 = H_{kr}(2) \rightarrow$ Nullhypothese bestätigt \rightarrow Es existieren keine signifikante Unterschiede zwischen den Gruppen. ANOVA kam hier auf signifikante Unterschiede.

Kruskal-Wallis Test in R

```
titlelength<-c(30,35,45,40,35,50,10,25,5)
group<-gl(3,3,labels=c("autor1","autor2","autor3"))
df<-data.frame(group,titlelength)
kruskal.test(titlelength~group, data=df)
```

Kruskal-Wallis rank sum test

```
data: titlelength by group
Kruskal-Wallis chi-squared = 5.8039, df = 2, p-value = 0.05492
```

```
# Wenn  $p < 0.05$  existieren signifikante Unterschiede.
```

```
# Bzgl der Differenz zum Beispiel: Rundungsfehler erklären teilweise  
( $12 / (9 * 10) * (17.5^2 / 3 + 21.5^2 / 3 + 12) - (3 * (9 + 1)) = 5.7556$ )
```

```
Wo die restliche Differenz herkommt, weiß ich im Moment nicht.
```

Problem: Omnibus & Familienbezogener Fehler

Lösung:

Problem: Omnibus & Familienbezogener Fehler
Lösung: Kontrastierung & Post Hoc Tests

Siegel, S. & Castellan, N.J. (1988): *Nonparametric statistics for the behavioural science*

- Post-Hoc Test für nichtparametrische Daten
- Grundidee: Vergleiche für alle Paare der Gruppen die Differenz der gemittelten Ränge mit dem Zufallsgrenzwert.
Differenz > KritischerWert → *Signifikant*
- $diff_{kr} = z_{\frac{\alpha}{k*(k-1)}} * \sqrt{\frac{n*(n+1)}{12} * (\frac{1}{n_u} + \frac{1}{n_v})}$ mit $k =$ Anzahl der Gruppen
- $diff_{kr}$ ist bei gleichen Gruppengrößen wiederverwendbar, sonst für jeweils Gruppen u und v nochmal berechnen

Beispiel Post Hoc Tests

Autor 1	Rang	Autor 2	Rang	Autor 3	Rang
30	4	40	7	10	2
35	5.5	35	5.5	25	3
45	8	50	9	5	1
mean	5.833	mean	7.167	mean	2

Beispiel Post Hoc Tests

Autor 1	Rang	Autor 2	Rang	Autor 3	Rang
30	4	40	7	10	2
35	5.5	35	5.5	25	3
45	8	50	9	5	1
mean	5.833	mean	7.167	mean	2

Paar	R1	R2	R1-R2
A1-A2	5.833	7.167	1.334
A2-A3	7.167	2	5.167
A1-A3	5.833	2	3.833

$$diff_{kr} = z_{\frac{\alpha}{k*(k-1)}} * \sqrt{\frac{n*(n+1)}{12} * (\frac{1}{n_u} + \frac{1}{n_v})} = z_{\frac{0.05}{3*(2)}} * \sqrt{\frac{9*(9+1)}{12} * (\frac{1}{3} + \frac{1}{3})} = z_{0.0083} * \sqrt{5} = 2.4 * 2.236 \approx \underline{5.3664}$$

$z_{0.0083} \rightarrow$ z-Tabelle smaller portion

Beispiel Post Hoc Tests

Autor 1	Rang	Autor 2	Rang	Autor 3	Rang
30	4	40	7	10	2
35	5.5	35	5.5	25	3
45	8	50	9	5	1
mean	5.833	mean	7.167	mean	2

Paar	R1	R2	R1-R2
A1-A2	5.833	7.167	1.334
A2-A3	7.167	2	5.167
A1-A3	5.833	2	3.833

$$diff_{kr} = z_{\frac{\alpha}{k*(k-1)}} * \sqrt{\frac{n*(n+1)}{12} * (\frac{1}{n_u} + \frac{1}{n_v})} = z_{\frac{0.05}{3*(2)}} * \sqrt{\frac{9*(9+1)}{12} * (\frac{1}{3} + \frac{1}{3})} =$$

$$z_{0.0083} * \sqrt{5} = 2.4 * 2.236 \approx \underline{5.3664}$$

$$z_{0.0083} \rightarrow z\text{-Tabelle smaller portion}$$

→ Keine der Differenzen überschreitet den Kritischen Wert → Alle
 unsignifikant

Post Hoc Tests in R

```
library(pgirmess)
titlelength<-c(30,35,45,40,35,50,10,25,5)
group<-gl(3,3,labels=c("autor1","autor2","autor3"))
df<-data.frame(titlelength,group)
kruskalmc(titlelength~group, data=df)
```

Multiple comparison test after Kruskal-Wallis
p.value: 0.05

Comparisons

	obs.dif	critical.dif	difference
autor1-autor2	1.33333333333333	5.3531015691555	FALSE
autor1-autor3	3.83333333333333	5.3531015691555	FALSE
autor2-autor3	5.16666666666667	5.3531015691555	FALSE

Zielgerichtete Post Hoc Tests

Statt Kreuzvergleich lieber Vergleich gegen eine Kontrollgruppe (Bspw A2)
Reduziert Einfluss auf α

Paar	R1	R2	R1-R2
A1-A2	5.833	7.167	1.334
A2-A3	7.167	2	5.167

$$diff_{kr} = 5.011929$$

Zielgerichtete Post Hoc Tests

Statt Kreuzvergleich lieber Vergleich gegen eine Kontrollgruppe (Bspw A2)
Reduziert Einfluss auf α

Paar	R1	R2	R1-R2
A1-A2	5.833	7.167	1.334
A2-A3	7.167	2	5.167

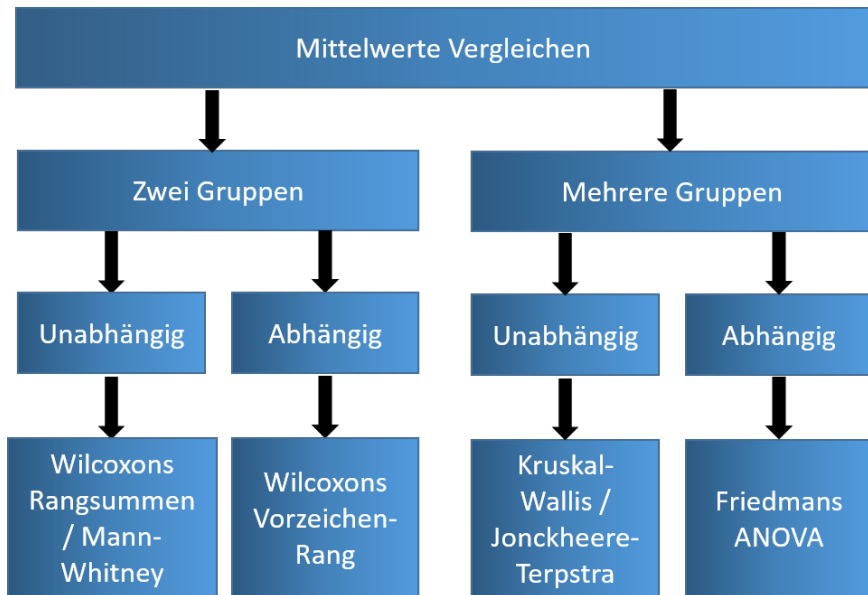
$diff_{kr} = 5.011929$

```
library(pgirmess)
titlelength<-c(30,35,45,40,35,50,10,25,5)
group<-gl(3,3,labels=c("autor1","autor2","autor3"))
df<-data.frame(titlelength,group)
df$group<-factor(df$group, levels=levels(df$group)[c(2,1,3)])
kruskalmc(titlelength~group, data=df, cont='two-tailed')
#two-tailed macht 1. Gruppe zur Kontr.gruppe, factor(.. hat A2 nach vorn sortiert
```

Multiple comparison test after Kruskal-Wallis, treatments vs control (two-tailed)
p.value: 0.05
Comparisons

	obs.dif	critical.dif	difference	
autor2-autor1	1.333333	5.011929	FALSE	
autor2-autor3	5.166667	5.011929	TRUE	<----

Nichtparametrische Testverfahren

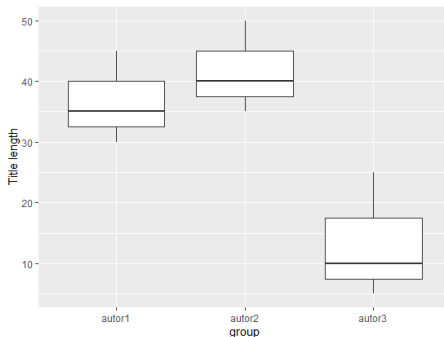


- kurz: Jonckheere-Test
- Omnibus-Test wie Kruskal-Wallis aber nicht auf Unterschied sondern auf Trend
- *Nullhypothese* : Es existiert kein signifikanter Trend
- $R \rightarrow$ Trend signifikant wenn $p < 0.05$
- Berechnung der Trendrichtung mit Gruppenmittelwerten

Jonckheere-Terpstra Test

```
library(clinfun)
titlelength<-c(30,35,45,40,35,50,10,25,5)
group<-gl(3,3,labels=c("autor1","autor2","autor3"))
df1<-data.frame(titlelength,group)
jonckheere.test(df1$titlelength, as.numeric(df1$group))
boxplot<-ggplot(df1, aes(group, titlelength))
boxplot + geom_boxplot()+labs(y="Title length")
```

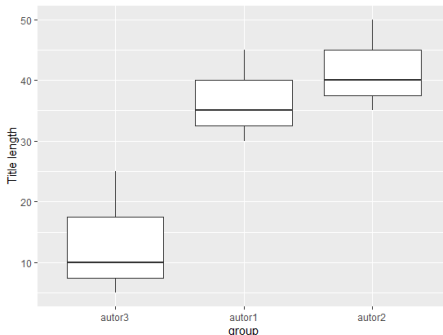
JT = 6.5, p-value = 0.1198



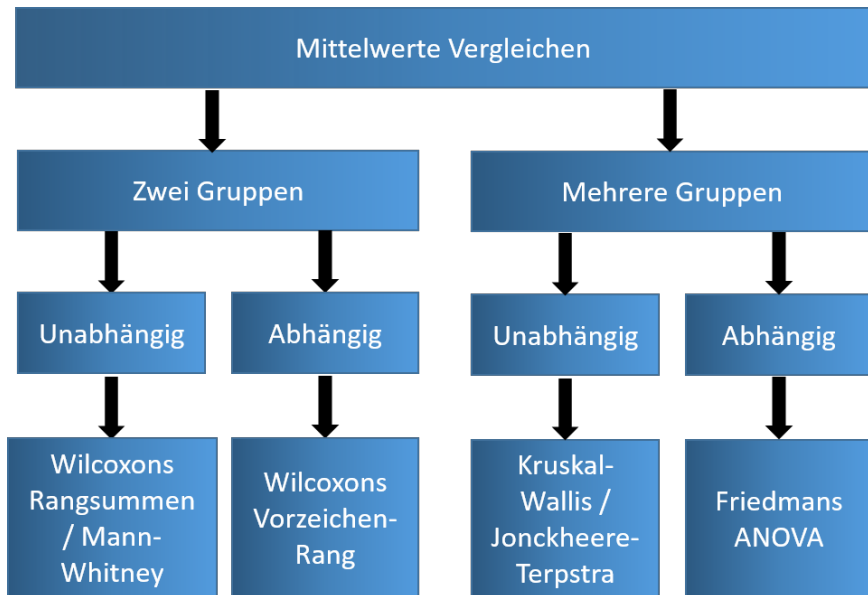
Jonckheere-Terpstra Test

```
library(clinfun)
titlelength<-c(30,35,45,40,35,50,10,25,5)
group<-gl(3,3,labels=c("autor1","autor2","autor3"))
df1<-data.frame(titlelength,group)
df2$group<-factor(df1$group, levels=levels(df1$group)[c(3,1,2)]) #Umsortieren
jonckheere.test(df2$titlelength, as.numeric(df2$group))
boxplot<-ggplot(df2, aes(group, titlelength))
boxplot + geom_boxplot()+labs(y="Title length")
```

JT = 24.5, p-value = 0.01451



Nichtparametrische Testverfahren



Friedman, M. (1937): *The use of ranks to avoid the assumption of normality implicit in the analysis of variance*

- Friedmans ANOVA als nichtparametrische Alternative für abhängige ANOVA mittels Ranking
- Omnibustest
- Ranking pro Person, jede Person beginnt bei 1
- *Nullhypothese* : Die Gruppenunterschiede pro Person sind nicht signifikant. Gilt bei $F_r < F_{rkr}$

Friedman, M. (1937): *The use of ranks to avoid the assumption of normality implicit in the analysis of variance*

- Friedmans ANOVA als nichtparametrische Alternative für abhängige ANOVA mittels Ranking
- Omnibustest
- Ranking pro Person, jede Person beginnt bei 1
- *Nullhypothese* : Die Gruppenunterschiede pro Person sind nicht signifikant. Gilt bei $F_r < F_{rkr}$
- $F_r = \frac{12}{n_{group} * k(k+1)} * \sum Rangsumme_{group}^2 - 3n_{group} * (k + 1)$ mit $k = Gruppenanzahl$

Friedman, M. (1937): *The use of ranks to avoid the assumption of normality implicit in the analysis of variance*

- Friedmans ANOVA als nichtparametrische Alternative für abhängige ANOVA mittels Ranking
- Omnibustest
- Ranking pro Person, jede Person beginnt bei 1
- *Nullhypothese* : Die Gruppenunterschiede pro Person sind nicht signifikant. Gilt bei $F_r < F_{rkr}$
- $F_r = \frac{12}{n_{group} * k(k+1)} * \sum Rangsumme_{group}^2 - 3n_{group} * (k + 1)$ mit $k = \text{Gruppenanzahl}$ Es wird oft die Stichprobengröße n notiert (auch Field verwendet N), aber es wird mit n_{group} gerechnet
- F_r folgt χ^2 -Verteilung mit $df = \text{Gruppenanzahl} - 1$ bei $n_{group} \geq 10$

Zeichenlänge

des Dokumententitels pro Autor beim ersten, zweiten und dritten Dokument

Person	D1	D2	D3
Autor 1	30	35	45
Autor 2	40	35	50
Autor 3	10	25	5

Zeichenlänge

des Dokumententitels pro Autor beim ersten, zweiten und dritten Dokument

Person	D1	D2	D3
Autor 1	30	35	45
Autor 2	40	35	50
Autor 3	10	25	5

→ *Rangify* →

Person	D1	D2	D3
Autor 1	1	2	3
Autor 2	2	1	3
Autor 3	2	3	1
sum	5	6	7

Friedmans ANOVA Beispiel

Zeichenlänge

des Dokumententitels pro Autor beim ersten, zweiten und dritten Dokument

Person	D1	D2	D3
Autor 1	30	35	45
Autor 2	40	35	50
Autor 3	10	25	5

→ Rangify →

Person	D1	D2	D3
Autor 1	1	2	3
Autor 2	2	1	3
Autor 3	2	3	1
sum	5	6	7

$$F_r = \left(\frac{12}{n_{group} * k(k+1)} * \sum Rangsumme_{group}^2 \right) - 3n_{group} * (k + 1) =$$
$$\left(\frac{12}{36} * 110 \right) - 36 = \underline{0.667} \quad F_r = 0.667 < F_{rkr} = 5.99 \rightarrow H_0 \text{ kann nicht abgewiesen werden} \rightarrow$$

Friedmans ANOVA Beispiel

Zeichenlänge

des Dokumententitels pro Autor beim ersten, zweiten und dritten Dokument

Person	D1	D2	D3
Autor 1	30	35	45
Autor 2	40	35	50
Autor 3	10	25	5

→ Rangify →

Person	D1	D2	D3
Autor 1	1	2	3
Autor 2	2	1	3
Autor 3	2	3	1
sum	5	6	7

$$F_r = \left(\frac{12}{n_{group} * k(k+1)} * \sum Rangsumme_{group}^2 \right) - 3n_{group} * (k + 1) =$$
$$\left(\frac{12}{36} * 110 \right) - 36 = \underline{0.667} \quad F_r = 0.667 < F_{rkr} = 5.99 \rightarrow H_0 \text{ kann nicht}$$

abgewiesen werden → Unterschiede nicht signifikant

```
D1<-c(30,40,10)
D2<-c(35,35,25)
D3<-c(45,50,5)
df<-data.frame(D1, D2, D3)
```

```
keinefehlwerte<-na.omit(df) # Fehlwerte verwirren den Algorithmus
matrix<-as.matrix(keinefehlwerte) # Algorithmus nimmt nur Matrix als Input
friedman.test(matrix)
```

Friedman rank sum test

```
data: matrix
Friedman chi-squared = 0.66667, df = 2, p-value = 0.7165
# p<0.05 -> Unterschied signifikant
```

Siegel, S. & Castellan, N.J. (1988): *Nonparametric statistics for the behavioural science*

- Im Prinzip wie bei Kruskal Wallis

Post Hoc Tests in R

```
library(pgirmess)
D1<-c(30,40,10)
D2<-c(35,35,25)
D3<-c(45,50,5)
```

```
df<-data.frame(D1, D2, D3)
```

```
keinefehlwerte<-na.omit(df) # Fehlwerte verwirren den Algorithmus
matrix<-as.matrix(keinefehlwerte) # Algorithmus nimmt nur Matrix als Input
friedmanmc(matrix)
```

Multiple comparisons between groups after Friedman test

p.value: 0.05

Comparisons

	obs.dif	critical.dif	difference
1-2	1	5.864029	FALSE
1-3	2	5.864029	FALSE
2-3	1	5.864029	FALSE

- 1 Nichtparametrische Testverfahren
- 2 Wilcoxon's Rangsummen / Mann-Whitney
- 3 Wilcoxon's Vorzeichen-Rang Test
- 4 Kruskal-Wallis Test
- 5 Trendanalyse mit Jonckheere-Terpstra
- 6 Friedmans ANOVA
- 7 Effektstärke bei Friedman und Kruskal-Wallis

- Umrechnung von χ^2 zu r problematisch mit $df > 1$
- Effektstärke bei Omnibus eher wenig hilfreich
- Paarweise Wilcoxon Tests (abhängig / unabhängig beachten) und Berechnung von r damit

- Nichtparametrische Verfahren erlauben die Arbeit mit diesen Daten
 - Mediane, Effektstärke und Ranges mit angeben
 - Boxplot
- Wilcoxon's Rangsummen / Mann-Whitney
 - beide wohl austauschbar
 - Entspricht unabhängigem t-Test
- Wilcoxon's Vorzeichen-Rang
 - Entspricht abhängigem t-Test
- Kruskal-Wallis
 - Entspricht unabhängiger ANOVA
 - Post Hoc Tests zur Lokalisierung der Unterschiede
- Jonckheere-Terpstra
 - Entspricht unabhängiger ANOVA aber berücksichtigt Trends
 - Trendrichtung mittels Gruppenmittelwerten
 - Effektstärke → paarweise Wilcoxon-Tests
- Friedmans ANOVA
 - Entspricht unabhängiger ANOVA
 - Post Hoc Tests zur Lokalisierung der Unterschiede
 - Effektstärke → paarweise Wilcoxon-Tests