

# Statistik II

---

## Einheit 2: Einfaktorielle Varianzanalyse (2)

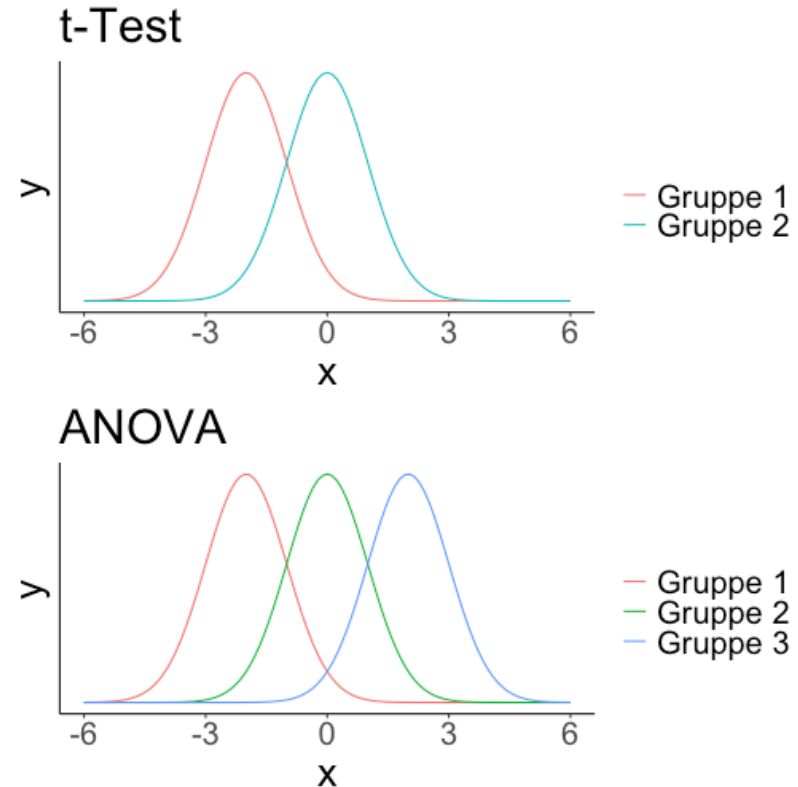
Wintersemester 2025 | Prof. Dr. Stephan Goerigk

# Einfaktorielle Varianzanalyse (ANOVA)

## Determinanten der ANOVA

- Determinanten = Größen, welche die Signifikanz der ANOVA beeinflussen:
  - Signifikanzniveau ( $\alpha$ )
  - Teststärke ( $1 - \beta$ )
  - Effektgröße (Unterschied zwischen Mittelwerten)
  - Stichprobengröße ( $N$ )

→ Wir testen gegen die Nullhypothese ( $H_0$ ) und verwerfen diese bei einem signifikanten Ergebnis



# Einfaktorielle Varianzanalyse (ANOVA)

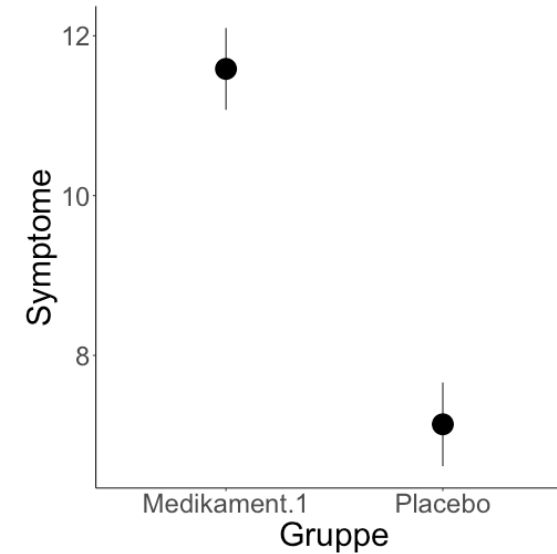
## Determinanten der ANOVA

### Exkurs: Beziehung zwischen F-Wert und t-Wert

- Für den t-Test gelernte Konzepte können vollständig auf die Varianzanalyse übertragen werden
- t-Test = Vergleich 2er Mittelwerte → entspricht ANOVA mit 2-stufigem Faktor (UV)

### Beispiel aus Einheit 1 (2 Stufen):

Vergleich Placebo vs. Medikament 1



# Einfaktorielle Varianzanalyse (ANOVA)

## Determinanten der ANOVA

### Exkurs: Beziehung zwischen F-Wert und t-Wert

```
t.test(Symptome ~ Gruppe,  
      data = df[df$Gruppe == "Placebo" | df$Gruppe == "Medikament.1",],  
      var.equal = T)
```

```
##  
##      Two Sample t-test  
##  
## data:  Symptome by Gruppe  
## t = 6.0718, df = 98, p-value = 2.403e-08  
## alternative hypothesis: true difference in means between group Medikament.1 and group Placebo is not equal to 0  
## 95 percent confidence interval:  
##  2.994253 5.901785  
## sample estimates:  
## mean in group Medikament.1      mean in group Placebo  
##           11.585633           7.137614
```

# Einfaktorielle Varianzanalyse (ANOVA)

## Determinanten der ANOVA

### Exkurs: Beziehung zwischen F-Wert und t-Wert

```
anova(lm(Symptome ~ Gruppe,  
        data = df[df$Gruppe == "Placebo" | df$Gruppe == "Medikament.1",]))
```

```
## Analysis of Variance Table  
##  
## Response: Symptome  
##           Df  Sum Sq Mean Sq F value    Pr(>F)  
## Gruppe      1   494.62   494.62   36.867 2.403e-08 ***  
## Residuals  98  1314.82    13.42  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Determinanten der ANOVA

### Exkurs: Beziehung zwischen F-Wert und t-Wert

$$F_{(1;98)} = \frac{\hat{\sigma}_{zwischen}^2}{\hat{\sigma}_{innerhalb}^2} = \frac{494.62}{13.42} = 36.87$$

$$t_{(98)} = 6.0718$$

$$6.0718^2 = 36.87$$

- Das Quadrat des t-Werts entspricht dem F-Wert einer einfaktoriellen ANOVA mit zwei Stufen

→ Die Varianzanalyse ist eine Verallgemeinerung des t-Tests

## Determinanten der ANOVA

### Effektstärke

- Das Maß für den Populationseffekt in der Varianzanalyse heißt  $\Omega^2$

$$\Omega^2 = \frac{\sigma_{\text{systematisch}}^2}{\sigma_{\text{Gesamt}}^2}$$

- $\Omega^2$  Gibt den Anteil systematischer Varianz an der Gesamtvarianz an
- Schätzer für den Populationseffekt  $\Omega^2$  ist  $\omega^2$  (klein Omega-Quadrat)
- Die Schätzung erfolgt über  $f^2$

$$f^2 = \frac{(F_{df_{\text{Zähler}}; df_{\text{Nenner}}} - 1) \cdot df_{\text{Zähler}}}{N}$$

$$\omega^2 = \frac{f^2}{1 + f^2}$$

# Einfaktorielle Varianzanalyse (ANOVA)

## Determinanten der ANOVA

### Effektstärke

```
anova(lm(Symptome ~ Gruppe, data = df))
```

```
## Analysis of Variance Table
##
## Response: Symptome
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Gruppe      2   582.4  291.201   20.558 1.341e-08 ***
## Residuals 147 2082.2   14.165
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$f^2 = \frac{(F_{df_{Zähler}; df_{Nenner}} - 1) \cdot df_{Zähler}}{N} = \frac{(20.558 - 1) \cdot 2}{150} = 0.26$$



## Determinanten der ANOVA

### Effektstärke

$$\omega^2 = \frac{f^2}{1 + f^2} = \frac{0.26}{1 + 0.26} = 0.2063$$

- Der Anteil der Effektvarianz des Faktors Gruppe beträgt 20.63%
- Anders ausgedrückt: der Faktor Gruppe klärt circa 21% der Gesamtvarianz auf
- Dies entspricht einem großen Effekt.

### Cut-offs:

Effektstärke	Omega-Quadrat
Kleiner Effekt	0.01
Mittlerer Effekt	0.06
Großer Effekt	0.14

## Determinanten der ANOVA

### Effektstärke

- Ein weiteres, häufig verwendetes Effektmaß ist  $\eta^2$  (Eta-Quadrat)
- Es gibt den Anteil der aufgeklärten Varianz der Messwerte auf Ebene der Stichprobe an
- Es wird aus dem Verhältnis von Quadratsummen, anstelle von Varianzen berechnet

$$\eta^2 = \frac{QS_{zwischen}}{QS_{zwischen} + QS_{innerhalb}}$$

- Berechnung kann ebenfalls über  $f^2$  erfolgen (wir schreiben für die Stichprobe  $f_s^2$ )

$$f_s^2 = \frac{(F_{df_{Zähler}; df_{Nenner}}) \cdot df_{Zähler}}{df_{Nenner}}$$

$$\eta^2 = \frac{f_s^2}{1 + f_s^2}$$

## Determinanten der ANOVA

### Effektstärke

$$f_s^2 = \frac{20.558 \cdot 2}{147} = 0.2797$$

$$\eta^2 = \frac{0.2797}{1 + 0.2797} = 0.2185$$

- Wert fällt im Vergleich zum wahren Effekt auf Populationsebene zu groß aus (Überschätzung)
- $\omega^2$  liefert genauere Schätzung (Daten innerhalb der Stichprobe sind überangepasst)
- Auf der Stichprobenebene klärt der Faktor Gruppe ca. 22% der Varianz der Messwerte auf
- Cut-offs für  $\eta^2$  entsprechen denen für  $\Omega^2$

### Cut-offs:

Effektstärke	Omega-Quadrat
Kleiner Effekt	0.01
Mittlerer Effekt	0.06
Großer Effekt	0.14

## Determinanten der ANOVA

### Residualvarianz

- Je kleiner die Residualvarianz, desto größer die Teststärke (Wahrscheinlichkeit, dass Test signifikant wird)
- Schätzung der Residualvarianz durch die Varianz innerhalb steht im Nenner des F-Bruchs → bei kleineren Werten wird F-Wert größer
- Die Größe der Residualvarianz wird häufig auch "Rauschen" genannt

Analogie aus der Akustik:

- Bei lauten Nebengeräuschen sind leise Töne schwerer zu hören
- Das Signal wird vom Rauschen verdeckt

→ ANOVA: Rauschen = Residualvarianz; Signal = gesuchter Effekt (z.B. Gruppenunterschied)

## Determinanten der ANOVA

### Größe des Effekts

- Umso größer der gesuchte Populationseffekt, desto größer die Teststärke
- Je mehr sich die Populationsmittelwerte der Gruppen unterscheiden (systematischer Einfluss auf die AV), desto wahrscheinlicher ist ein signifikantes Ergebnis
- Dies entspricht einem deutlichen Signal, welches auch noch bei starkem Rauschen hörbar ist

## Determinanten der ANOVA

### Stichprobenumfang

- Umso größer der Stichprobenumfang, desto größer die Teststärke

### Gründe:

#### 1. Varianz zwischen

- Varianz zwischen hängt proportional von der Anzahl der Personen in einer Bedingung ab
- Die größer  $N$ , desto größer die Varianz zwischen (steht im Zähler des F-Bruchs)
- größere Varianz zwischen  $\rightarrow$  größerer F-Wert  $\rightarrow$  signifikanteres Ergebnis

#### 2. Freiheitsgrade Varianz innerhalb

- Erhöhen sich mit steigenden Stichprobenumfang  $\rightarrow$  Verkleinerung des kritischen F-Werts (man kommt leichter über die Signifikanzschwelle)

## Determinanten der ANOVA

### $\alpha$ -Fehler

- Je größer (weniger streng) das a priori festgelegte Signifikanzniveau, desto größer die Teststärke
- Durch die Erhöhung des  $\alpha$ -Fehlers steigt die Wahrscheinlichkeit die Alternativhypothese fälschlich zu wählen (Fehler 1. Art).

### **ABER:**

- Gleichzeitig erhöht sich die Wahrscheinlichkeit, einen Effekt zu finden, falls er wirklich existiert.

## Determinanten der ANOVA

### Stichprobenumfangsplanung

- Einer der wichtigsten Schritte vor der Durchführung einer Untersuchung
- Nur so kann gewährleistet werden, dass die Interpretation eines Untersuchungsergebnisses korrekt ablaufen kann

Probleme bei Auslassen der Stichprobenumfangsplanung:

- Stichprobenumfang zu klein. Teststärke ist so klein, dass ein nicht signifikantes Ergebnis nicht interpretierbar ist (**underpowered**)
- Stichprobenumfang zu groß. Auch kleine Effekte werden signifikant, die für eine vernünftige inhaltliche Interpretation zu klein sind (**overpowered**)



## Determinanten der ANOVA

### Stichprobenumfangsplanung (in R)

```
pwr::pwr.anova.test(k=3, f= 0.2, sig.level = 0.05, power = 0.9)
```

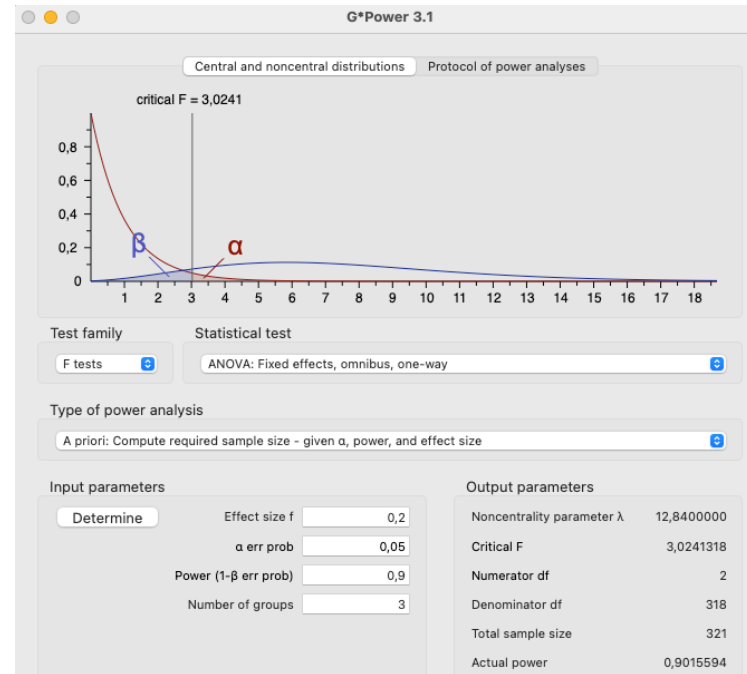
```
##  
##      Balanced one-way analysis of variance power calculation  
##  
##              k = 3  
##              n = 106.455  
##              f = 0.2  
##      sig.level = 0.05  
##              power = 0.9  
##  
## NOTE: n is number in each group
```

- Bei Annahme eines Signifikanzniveaus von  $\alpha = 0.05$  wären für eine ANOVA mit 3 Gruppen und einer großen Effektstärke (0.20) wären für eine Teststärke  $(1 - \beta)$  von 90% lediglich N=106 Personen zum Nachweisen eines signifikanten Effekts notwendig gewesen.

# Einfaktorielle Varianzanalyse (ANOVA)

## Determinanten der ANOVA

### Stichprobenumfangplanung (in GPower)



## Post-Hoc-Analysen

- $H_0$  der ANOVA: Alle Mittelwerte sind gleich. (Soll verworfen werden)
  - Signifikantes Ergebnis führt zur Ablehnung der  $H_0$  und zur vorübergehende Annahme der  $H_1$ 
    - ABER: Alternativhypothese ist vollkommen unspezifisch
    - $H_1$  macht keine Aussage darüber, welche der Gruppen sich voneinander unterscheiden
    - $H_1$  umfasst alle Möglichkeiten, welche nicht der  $H_0$  entsprechen
  - Signifikante ANOVA: Es ist lediglich gesichert, dass Gruppe mit kleinstem Mittelwert sich signifikant von Gruppe mit größtem Mittelwert unterscheidet.
  - ABER: Global formulierte  $H_1$  erlaubt viele Kombinationen (bei 3 Gruppen bereits 18 Möglichkeiten)
- In den meisten Untersuchungen ist jedoch die **genaue Struktur** der  $H_1$  von Interesse
- **Post-Hoc Verfahren** analysieren die genaue Struktur der  $H_1$  (welche Gruppen sich genau unterscheiden)

## Post-Hoc-Analysen

- Die meisten Post-Hoc Verfahren ermöglichen einen paarweisen Vergleich der Gruppenmittelwerte.
- Post-Hoc Tests sind mehr oder weniger streng.
- Es gibt zahlreiche Verfahren, die sich vor allem durch unterschiedliche Risiken für den Fehler 1. und 2. Art auszeichnen (Auswahl):
  - Bonferroni
  - Tukey HSD
  - Holm
  - Sidak
  - Games-Howell
  - Benjamini & Hochberg (aka False Discovery Rate Correction, FDR)
  - Benjamini & Yekutieli

## Post-Hoc-Analysen

### Tukey HSD

- Tukey HSD Test ermöglicht paarweisen Mittelwertsvergleich ohne  $\alpha$ -Fehlerkumulierung und Verringerung der Teststärke
- Wie groß muss die Differenz zwischen zwei Gruppenmittelwerten sein, damit diese auf einem kumulierten  $\alpha$ -Niveau signifikant ist, das nicht die zuvor festgelegte Grenze überschreitet (i.d.R. 0.05)?
- Die Teststärke des Tukey HSD Tests ist mindestens so hoch wie die des getesteten Haupteffekts der ANOVA.
- HSD = "Honest significant difference"

→ Ist die tatsächliche Differenz zwischen den Gruppen größer als der kritische Wert des Tukey HSD Tests besteht ein signifikanter Unterschied

## Post-Hoc-Analysen

### Tukey HSD

- Die HSD ergibt sich über den Wert  $q$
- $q$  hat beim Vergleich mehrerer Mittelwerte die Funktion des t-Werts (ist auch ähnlich definiert)

Für jeden paarweisen Vergleich gilt:

$$q_{r;df_{innerhalb}} = \frac{\bar{x}_2 - \bar{x}_1}{\sqrt{\frac{\hat{\sigma}_{innerhalb}^2}{n}}}$$

- $r$  = Zahl der Mittelwerte

## Post-Hoc-Analysen

### Tukey HSD

#### Bestimmung der kleinsten noch signifikanten Differenz (Krit. Wert)

- $q$ -Wert bezieht sich auf multiple Mittelwertsvergleiche  $\rightarrow$  hat eigene Verteilung (studentized range)
- studentized range: ermöglicht Bestimmung kritischen  $q$ -Werts in Abhängigkeit von der Anzahl der betrachteten Mittelwerte (t-Test kann dies nicht)
- Dadurch wird  $\alpha$ -Fehlerkumulierung verhindert

$$HSD = q_{r;df_{innerhalb}} \cdot \sqrt{\frac{\hat{\sigma}_{innerhalb}^2}{n}}$$

## Post-Hoc-Analysen

### Tukey HSD

#### Bestimmung der kleinsten noch signifikanten Differenz (Krit. Wert)

- kritische q-Werte stehen in eigener Tabelle
- Die Werte in der Tabelle sind abhängig von:
  - Anzahl der zu vergleichenden Gruppen (also der paarweisen Vergleiche)
  - festgelegtem Signifikanzniveau
  - den Fehlerfreiheitsgraden ( $df_{innerhalb}$ )



# Einfaktorielle Varianzanalyse (ANOVA)

## Post-Hoc-Analysen (1. Wert für $\alpha = 0.05$ , 2. Wert für 0.01)

df innerhalb	k=Gruppen								
df for Error Term	2	3	4	5	6	7	3	9	10
5	3.64 5.70	4.60 6.98	5.22 7.80	5.67 8.42	6.03 8.91	6.33 9.32	6.58 9.67	6.80 9.97	6.99 10.24
6	3.46 5.24	4.34 6.33	4.90 7.03	5.30 7.96	5.63 7.97	5.90 8.32	6.12 8.61	6.32 8.87	6.49 9.10
7	3.34 4.95	4.16 5.92	4.68 6.54	5.06 7.01	5.36 7.37	5.61 7.68	5.82 7.94	6.00 8.17	6.16 8.37
8	3.26 4.75	4.04 5.64	4.53 6.20	4.89 6.62	5.17 6.96	5.40 7.24	5.60 7.47	5.77 7.68	5.92 7.86
9	3.20 4.60	3.95 5.43	4.41 5.96	4.76 6.35	5.02 6.66	5.24 6.91	5.43 7.13	5.59 7.33	5.74 7.49
10	3.15 4.48	3.88 5.27	4.33 5.77	4.65 6.14	4.91 6.43	5.12 6.67	5.30 6.87	5.46 7.05	5.60 7.21
11	3.11 4.39	3.82 5.15	4.26 5.62	4.57 5.97	4.82 6.25	5.03 6.48	5.20 6.67	5.35 6.84	5.49 6.99
12	3.08 4.32	3.77 5.05	4.20 5.50	4.51 5.84	4.75 6.10	4.95 6.32	5.12 6.51	5.27 6.67	5.39 6.81
13	3.06 4.26	3.73 4.96	4.15 5.40	4.45 5.73	4.69 5.98	4.88 6.19	5.05 6.37	5.19 6.53	5.32 6.67
14	3.03 4.21	3.70 4.89	4.11 5.32	4.41 5.63	4.64 5.88	4.83 6.08	4.99 6.26	5.13 6.41	5.25 6.54
15	3.01 4.17	3.67 4.84	4.08 5.25	4.37 5.56	4.59 5.80	4.78 5.99	4.94 6.16	5.08 6.31	5.20 6.44
16	3.00 4.13	3.65 4.79	4.05 5.19	4.33 5.49	4.56 5.72	4.74 5.92	4.90 6.08	5.03 6.22	5.15 6.35
17	2.98 4.10	3.63 4.74	4.02 5.14	4.30 5.43	4.52 5.66	4.70 5.85	4.86 6.01	4.99 6.15	5.11 6.27
18	2.97 4.07	3.61 4.70	4.00 5.09	4.28 5.38	4.49 5.60	4.67 5.79	4.82 5.94	4.96 6.08	5.07 6.20
19	2.96 4.05	3.59 4.67	3.98 5.05	4.25 5.33	4.47 5.55	4.65 5.73	4.79 5.89	4.92 6.02	5.04 6.14
20	2.95 4.02	3.58 4.64	3.96 5.02	4.23 5.29	4.45 5.51	4.62 5.69	4.77 5.84	4.90 5.97	5.01 6.09
24	2.92 3.96	3.53 4.55	3.90 4.91	4.17 5.17	4.37 5.37	4.54 5.54	4.68 5.69	4.81 5.81	4.92 5.92
30	2.89 3.89	3.49 4.45	3.85 4.80	4.10 5.05	4.30 5.24	4.46 5.40	4.60 5.54	4.72 5.65	4.82 5.76
40	2.86 3.82	3.44 4.37	3.79 4.70	4.04 4.93	4.23 5.11	4.39 5.26	4.52 5.39	4.63 5.50	4.73 5.60
60	2.83 3.76	3.40 4.28	3.74 4.59	3.98 4.82	4.16 4.99	4.31 5.13	4.44 5.25	4.55 5.36	4.65 5.45
120	2.80 3.70	3.36 4.20	3.68 4.50	3.92 4.71	4.10 4.87	4.24 5.01	4.36 5.12	4.47 5.21	4.56 5.30
infinity	2.77 3.64	3.31 4.12	3.63 4.40	3.86 4.60	4.03 4.76	4.17 4.88	4.29 4.99	4.39 5.08	4.47 5.16

## Post-Hoc-Analysen

### Tukey HSD

#### Bestimmung der kleinsten noch signifikanten Differenz (Krit. Wert)

In unserem Beispiel:

- Signifikanzniveau:  $\alpha = .05$
- Anzahl der zu vergleichenden Mittelwerte:  $r = 3$
- Fehlerfreiheitsgrade:  $df_{innerhalb} = 147$

→ In der Tabelle der  $q_{krit}$ -Werte verwenden wir die nächst kleinere enthaltene Anzahl an Fehlerfreiheitsgraden  $df_{innerhalb} = 120$

$$q_{krit}(\alpha=.05;r=3;df_{innerhalb}=120) = 3.36$$

## Post-Hoc-Analysen

### Tukey HSD

#### Bestimmung der kleinsten noch signifikanten Differenz (Krit. Wert)

In unserem Beispiel:

- In jeder Gruppe wurden  $n = 50$  Patient:innen behandelt.
- Die Varianz innerhalb betrug  $\hat{\sigma}_{innerhalb}^2 = 14.165$  (siehe ANOVA Output, z.B. Folie 8)

$$HSD = q_{r;df_{innerhalb}} \cdot \sqrt{\frac{\hat{\sigma}_{innerhalb}^2}{n}} = 3.36 \cdot \sqrt{\frac{14.165}{50}} = 1.79$$

→ Mittelwertspaare, deren Differenz  $HSD = 1.79$  überschreiten sind signifikant.

# Einfaktorielle Varianzanalyse (ANOVA)

## Post-Hoc-Analysen

### Tukey HSD

### Bestimmung der kleinsten noch signifikanten Differenz (Krit. Wert)

In unserem Beispiel:

Mittelwerte:

Gruppe	Mittelwert
Medikament.1	11.59
Medikament.2	10.98
Placebo	7.14

Mittelwertsvergleiche (Differenzen):

Kontrast	Differenz
Medikament.1 - Medikament.2	0.60
Medikament.1 - Placebo	4.45
Medikament.2 - Placebo	3.85

→ Beide Medikamente sind signifikant besser als Placebo (Differenz > HSD=1.79), aber unterscheiden sich nicht signifikant voneinander.

# Einfaktorielle Varianzanalyse (ANOVA)

## Post-Hoc-Analysen

### Tukey HSD

```
emmeans::emmeans(lm(Symptome ~ Gruppe, data = df), pairwise ~ Gruppe, adjust = "tukey")
```

```
## $emmeans
##   Gruppe      emmean    SE  df lower.CL upper.CL
## Medikament.1  11.59 0.532 147    10.53    12.64
## Medikament.2  10.98 0.532 147     9.93    12.04
## Placebo        7.14 0.532 147     6.09     8.19
##
## Confidence level used: 0.95
##
## $contrasts
##   contrast      estimate    SE  df t.ratio p.value
## Medikament.1 - Medikament.2    0.601 0.753 147    0.799  0.7044
## Medikament.1 - Placebo        4.448 0.753 147    5.909 <.0001
## Medikament.2 - Placebo        3.847 0.753 147    5.111 <.0001
##
## P value adjustment: tukey method for comparing a family of 3 estimates
```

# Einfaktorielle Varianzanalyse (ANOVA)

## Post-Hoc-Analysen

### Bonferroni

```
emmeans::emmeans(lm(Symptome ~ Gruppe, data = df), pairwise ~ Gruppe, adjust = "bonferroni")
```

```
## $emmeans
##   Gruppe      emmean    SE  df lower.CL upper.CL
## Medikament.1  11.59 0.532 147    10.53    12.64
## Medikament.2  10.98 0.532 147     9.93    12.04
## Placebo        7.14 0.532 147     6.09     8.19
##
## Confidence level used: 0.95
##
## $contrasts
##   contrast      estimate    SE  df t.ratio p.value
## Medikament.1 - Medikament.2    0.601 0.753 147    0.799  1.0000
## Medikament.1 - Placebo        4.448 0.753 147    5.909 <.0001
## Medikament.2 - Placebo        3.847 0.753 147    5.111 <.0001
##
## P value adjustment: bonferroni method for 3 tests
```

# Einfaktorielle Varianzanalyse (ANOVA)

## Post-Hoc-Analysen

### Benjamini & Hochberg (aka False Discovery Rate Correction, FDR)

```
emmeans::emmeans(lm(Symptome ~ Gruppe, data = df), pairwise ~ Gruppe, adjust = "fdr")
```

```
## $emmeans
##   Gruppe      emmean    SE  df lower.CL upper.CL
## Medikament.1  11.59 0.532 147    10.53    12.64
## Medikament.2  10.98 0.532 147     9.93    12.04
## Placebo        7.14 0.532 147     6.09     8.19
##
## Confidence level used: 0.95
##
## $contrasts
##   contrast      estimate    SE  df t.ratio p.value
## Medikament.1 - Medikament.2    0.601 0.753 147    0.799  0.4257
## Medikament.1 - Placebo        4.448 0.753 147    5.909 <.0001
## Medikament.2 - Placebo        3.847 0.753 147    5.111 <.0001
##
## P value adjustment: fdr method for 3 tests
```

## Voraussetzungen der ANOVA

- ANOVA gehört zur den sog. parametrischen Verfahren (wie auch der t-Test)

Es gelten folgende Voraussetzungen:

1. Die abhängige Variable ist intervallskaliert
  - messtheoretisch abgesichert (muss man durch Kenntnis des Messinstruments wissen)
2. Das untersuchte Merkmal ist in der Population normalverteilt
3. Varianzhomogenität (Varianzen sind innerhalb der verglichenen Gruppen ungefähr gleich)
4. Messwerte in allen Bedingungen sind unabhängig voneinander
  - Durch randomisierte Zuweisung der Personen zu den Faktorstufen
  - Falls nicht möglich gegebenenfalls Kontrolle von Störvariablen



## Voraussetzungen der ANOVA

### Varianzhomogenität

```
car::leveneTest(Symptome ~ Gruppe, data = df, center = "mean")
```

```
## Levene's Test for Homogeneity of Variance (center = "mean")
##      Df F value Pr(>F)
## group  2  0.2909  0.748
##      147
```

- Ein signifikanter Levene's Test weist auf Verletzung der Varianzhomogenität hin
- In unserem Beispiel:  $p = .748 > .05$

→ Varianzhomogenität kann als gegeben angesehen werden

## Voraussetzungen der ANOVA

### Normalverteilung

- NV muss in allen Gruppen gegeben sein (hier 3x geprüft)
- Ein signifikanter Shapiro-Wilk Test weist auf Verletzung der NV-Annahme hin
- In unserem Beispiel: alle  $p > .05$  (Tests nicht signifikant)

→ NV-Annahme kann als gegeben angesehen werden

```
by(df$Symptome, df$Gruppe, shapiro.test)
```

```
## df$Gruppe: Medikament.1
##
##      Shapiro-Wilk normality test
##
## data:  dd[x, ]
## W = 0.99073, p-value = 0.9618
##
## -----
## df$Gruppe: Medikament.2
##
##      Shapiro-Wilk normality test
##
## data:  dd[x, ]
## W = 0.97211, p-value = 0.2814
##
## -----
## df$Gruppe: Placebo
##
##      Shapiro-Wilk normality test
##
## data:  dd[x, ]
## W = 0.98928, p-value = 0.9279
```

## Berichten der Ergebnisse nach APA

### Statistischer Bericht: (ausführlich)

Eine einfaktorielle ANOVA wurde durchgeführt, um die Symptomverbesserung zwischen verschiedenen Medikamentengruppen (Medikament 1, Medikament 2 und Placebo) zu vergleichen.

Vor der Analyse wurden die Voraussetzungen überprüft. Die Annahme der Varianzhomogenität konnte nach Prüfung mit dem Levene Test bestätigt werden ( $p = .748$ ). Auch die Normalverteilungsannahme, welche mit Shapiro Wilk Tests überprüft wurde, war gegeben (Medikament 1:  $p = .962$ ; Medikament 2:  $p = .281$ ; Placebo:  $p = .928$ ).

Die ANOVA ergab einen signifikanten Unterschied in der Symptomverbesserung zwischen den Gruppen,  $F(2, 147) = 20.56, p < .001, \omega^2 = .21$ .

Ein Post-Hoc Test mit Tukey Korrektur zeigte keinen signifikanten Unterschied ( $p = .704$ ) zwischen Medikament 1 und Medikament 2, aber signifikante Unterschiede zwischen Medikament 1 und Placebo ( $4.45, p < .001$ ) und Medikament 2 und Placebo ( $3.85, p < .001$ ).

## Berichten der Ergebnisse nach APA

### Hinweis zum Berichten statistischer Ergebnisse

Die Ergebnisse einer **Varianzanalyse (ANOVA)** werden mit zwei Freiheitsgraden berichtet. Es folgen der  $F$ -Wert, das Signifikanzniveau sowie die Effektstärke (häufig das partielle Eta-Quadrat:  $\eta_p^2$ , oder Omega-Quadrat:  $\omega^2$ ).

$F(df_{\text{Zähler}}, df_{\text{Nenner}})$  = F-Wert,  $p$  = Signifikanz,  $\omega^2$  = Effektstärke

#### Berichten des $p$ -Werts

Ist der  $p$ -Wert kleiner als .001, schreiben Sie  $p < .001$ . Ist er genau .001 oder größer, schreiben Sie den exakten  $p$ -Wert auf drei Nachkommastellen gerundet.

Berichten Sie nur, was die Daten tatsächlich hergeben. Wenn eine Effektstärke berechnet wurde, gehört sie in den Bericht. Wenn Post-Hoc Tests durchgeführt wurden, sollten auch deren Ergebnisse mit angegeben werden.

Dasselbe gilt für die Prüfung der Voraussetzungen: Wenn Sie in der Aufgabe Informationen zu Tests, wie dem Levene-Test oder dem Shapiro-Wilk-Test sehen, dann berichten Sie diese. Wenn keine Testergebnisse angegeben sind, können (und sollen) Sie auch keine angeben.

→ Bleiben Sie also flexibel und passen Sie Ihre Berichterstattung daran an, was Ihnen vorliegt.

# Einfaktorielle Varianzanalyse (ANOVA)

## Berichten der Ergebnisse nach APA

### Statistischer Bericht: (In Ihrer Klausur)

Wenn Sie in Ihrer Klausur eine ANOVA händisch berechnen, könnte Ihr Antwortsatz so aussehen:

Die ANOVA zeigte, dass der Effekt des Medikaments für die Symptomverbesserung der Patient:innen signifikant war,  $F(2, 147) = 20.56, p < .001, \omega^2 = .21$ .

oder

Die ANOVA zeigte, dass es signifikante Mittelwertsunterschiede hinsichtlich der Symptomverbesserung zwischen den Medikamentengruppen gab,  $F(2, 147) = 20.56, p < .001, \omega^2 = .21$ .

- Der t-Test ist ein **Spezialfall der Varianzanalyse**, daher gelten bekannte Zusammenhänge zwischen Stichprobengröße, Effektstärke, Teststärke und Signifikanzniveau auch hier.
- Für eine gut geplante ANOVA ist eine **Stichprobenumfangsplanung** (Poweranalyse) notwendig.
- Ein **signifikantes Ergebnis** der ANOVA bedeutet lediglich, dass sich zumindest eine Stufe des Faktors von einer anderen unterscheidet.
- **Post-Hoc-Analysen** werden zur Untersuchung genutzt, wo genau die Mittelwertsunterschiede liegen.
- Der **Tukey-HSD Test** determiniert die kleinste noch signifikante Differenz, somit ist der paarweise Vergleich der Gruppen im Anschluss an die Durchführung der ANOVA möglich.
- Außer dem Tukey-HSD Test gibt es noch zahlreiche **weitere Optionen** für Post-Hoc-Tests, welche mehr/weniger streng für Fehler 1. und 2. Art kontrollieren.