



Einführung in die Statistik

Stichproben- umfangsplanung

Contact (1997). Warner Bros. Pictures; South Side Amusement Company.

Überblick

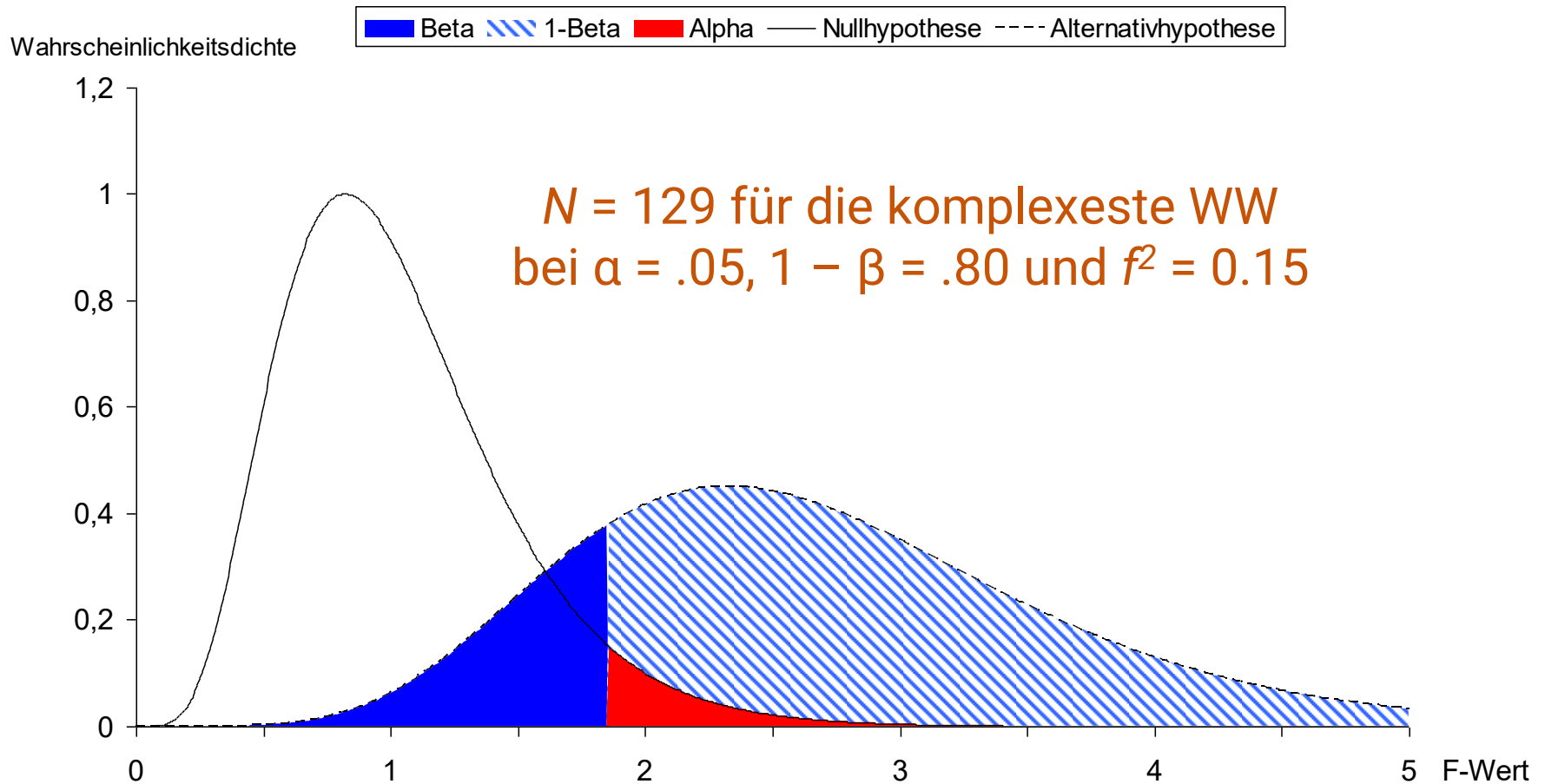
- Einführung
- Signifikanzniveau
- Teststärke
- Effektgröße
- Weitere Einflussgrößen
- Berechnung des Stichprobenumfangs

Einführung (Rey, 2020)

- Stichprobenumfangsplanung sollte im Vorfeld einer Studie erfolgen
- **Forschungspraxis:** Meist keine Stichprobenumfangsplanung
- **Probleme bei fehlender Stichprobenumfangsplanung**
 - **Fehlende Wahrscheinlichkeitsangabe:** Bei nicht signifikantem Ergebnis unklar, ob die Nullhypothese angenommen werden kann
 - **Unzureichende Ökonomie:** Idealerweise sollten genauso viele Versuchspersonen erhoben werden wie benötigt
- **Drei zentrale Einflussgrößen** auf den Stichprobenumfang
 - Signifikanzniveau
 - Teststärke
 - Effektgröße

Beispiel für eine Stichprobenumfangsplanung (Rey, 2020)

- 4 x 3 x 3-faktorielles, univariates Versuchsdesign ohne MW



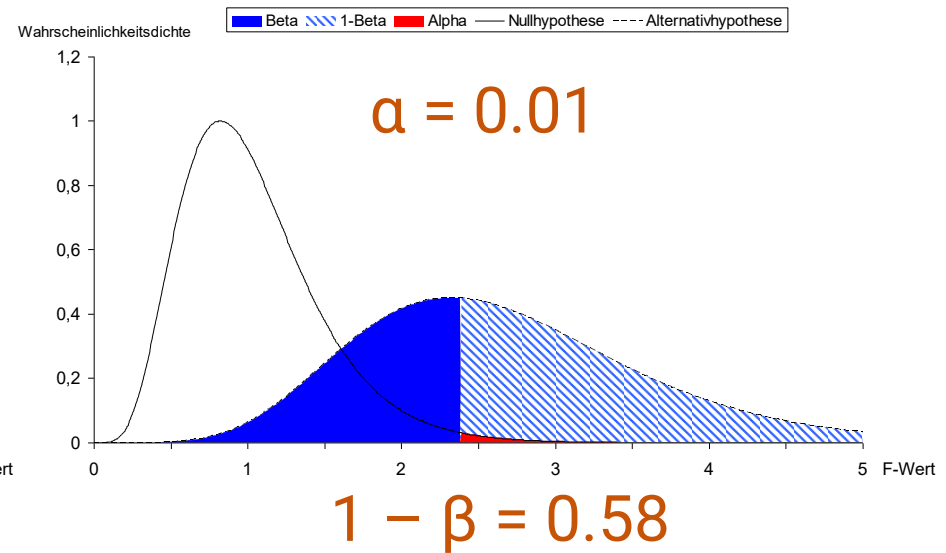
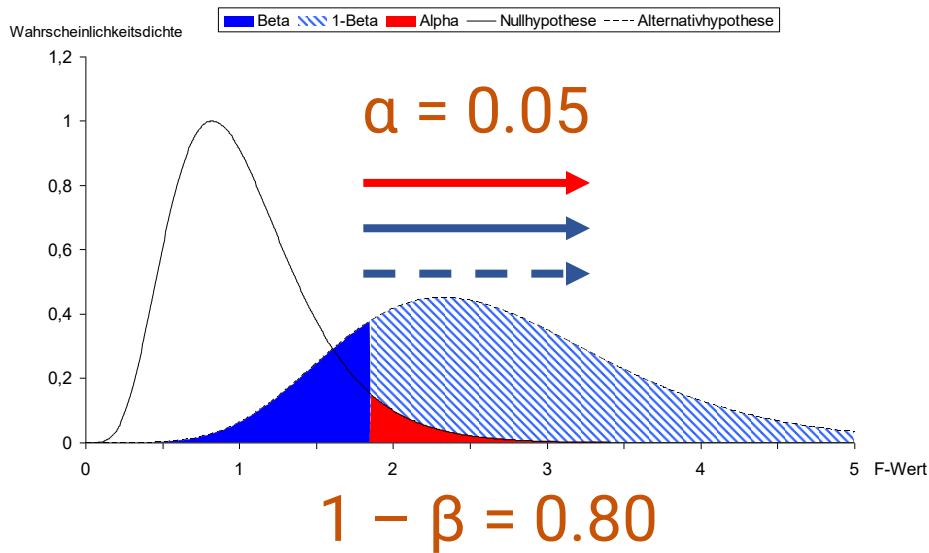
Signifikanzniveau (z. B. Bortz & Schuster, 2010; Döring & Bortz, 2015)

- **Statistische Signifikanz:** Wahrscheinlichkeit, dass das gefundene Ergebnis oder extremere Ergebnisse bei Gültigkeit der Nullhypothese eintreten
- **Konventionen** für die Entscheidungsfindung auf Basis der berechneten Wahrscheinlichkeit (p -Wert)

	Nicht signifikant	Signifikant	Sehr signifikant	Hoch signifikant
p -Wert	> 5%	$\leq 5\%$	$\leq 1\%$	$\leq 0.1\%$
Abkürzung	n.s.	*	**	***

Einfluss des Signifikanzniveaus auf die Teststärke (Rey, 2020)

- Je kleiner das Signifikanzniveau, desto kleiner die Teststärke ($1 - \beta$)
- Veränderung des Signifikanzniveaus wirkt sich nicht auf die Kurvenverläufe der zentralen und nonzentralen Verteilung aus
- Beispiel für den Einfluss des Signifikanzniveaus auf die Teststärke:



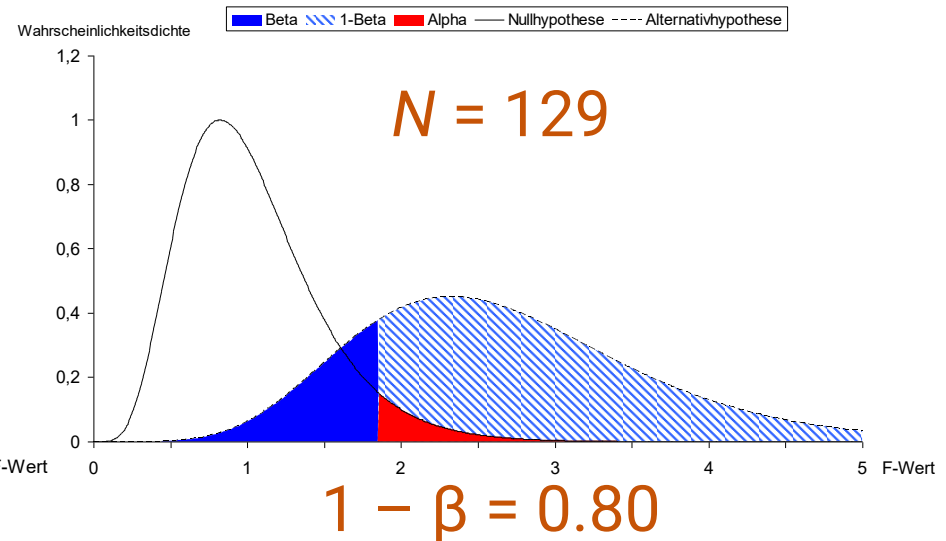
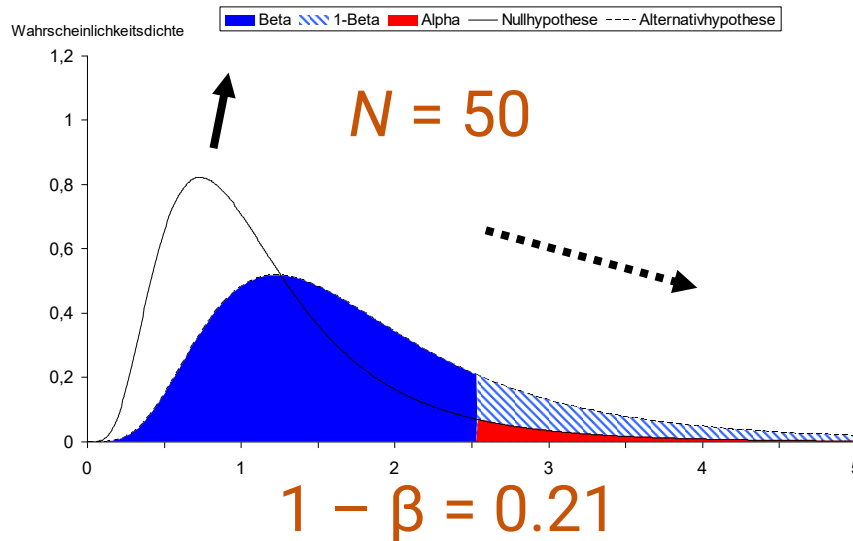
Einfluss des Signifikanzniveaus auf den Stichprobenumfang

- Je kleiner das Signifikanzniveau, desto größer der benötigte Stichprobenumfang
- Beispiel: Stichprobenumfänge für ein 2 x 2 faktorielles Design mit einem angenommenen mittleren Effekt von $f = 0.25$ und einer gewünschten Teststärke ($1 - \beta$) von 0.80:

Signifikanzniveau	Stichprobenumfang
5.0%	128
1.0%	191
0.1%	279

Einfluss des Stichprobenumfangs auf die Teststärke (Rey, 2020)

- Je mehr Versuchspersonen, desto größer die Teststärke ($1 - \beta$)
- Gründe hierfür:
 - Veränderung der zentralen Verteilung
 - Veränderung der nonzentralen Verteilung
- Beispiel für den Einfluss des Stichprobenumfangs auf die Teststärke:



Einfluss der Teststärke auf den Stichprobenumfang

- Je größer die (gewünschte) Teststärke, desto größer der benötigte Stichprobenumfang
- Beispiel: Stichprobenumfänge für ein 2 x 2 faktorielles Design mit einem angenommenen mittleren Effekt von $f = 0.25$ und einem Signifikanzniveau von 5%:

Teststärke	Stichprobenumfang
80%	128
95%	210
99%	296

Effektgröße (Rey, 2020)

- **Effektgröße:** Statistischer Kennwert, der angibt, wie gut das gesuchte Muster in den Zahlen erkennbar ist bzw. wie stark sich das gesuchte Signal vom Umgebungsrauschen unterscheidet
- **Vorteile durch Angabe von (standardisierten) Effektgrößen**
 - Leichtere Interpretierbarkeit der Ergebnisse
 - Mögliche Vergleichbarkeit von Studien
 - Einfache Aggregation für Metaanalysen
- **Verschiedene Effektgrößen** wie d , η_p^2 , f , f^2 oder ω_p^2
- **Umrechnung:** Jede Effektgröße kann in jede andere Effektgröße umgerechnet werden
- **Webseite zur Umrechnung:** www.psychometrica.de/effektstaerke

Effektgröße (Cohen, 1988)

- Konventionen zur Einschätzung verschiedener Effektgrößen

Effektgröße	Kleiner Effekt	Mittlerer Effekt	Großer Effekt
d	0.20	0.50	0.80
η_p^2	0.01	0.06	0.14
f	0.10	0.25	0.40
f^2	0.02	0.15	0.35
ω_p^2	0.01	0.06	0.14

- Wichtige Hinweise:
 - Konventionen dienen nur zur groben Orientierung
 - Die Konventionen sind numerisch nicht äquivalent

Effektgröße (z. B. Rasch, Frieze, Hofmann & Naumann, 2021)

- **Effektgröße d :** Standardisierte Distanz zwischen zwei Stichprobenmittelwerten

- **Formel:**
$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\hat{\sigma}_{x_1}^2 + \hat{\sigma}_{x_2}^2}{2}}}$$

$\bar{x}_1$	= Mittelwert der Stichprobe 1
$\bar{x}_2$	= Mittelwert der Stichprobe 2
$\hat{\sigma}_{x_1}^2$	= Geschätzte Varianz der Stichprobe 1
$\hat{\sigma}_{x_2}^2$	= Geschätzte Varianz der Stichprobe 2

- **Beispiel** zur Berechnung (vgl. Folien zur letzten Sitzung):

$$d = \frac{8.0 - 2.0}{\sqrt{\frac{0.625 + 0.625}{2}}} = \frac{6.0}{\sqrt{0.625}} \approx 7.59$$

- **Wertebereich** von d reicht von $-$ Unendlich bis $+$ Unendlich

Effektgröße (z. B. Rasch, Frieze, Hofmann & Naumann, 2021)

- **Partielles Eta-Quadrat (η_p^2)**: Effektgröße, die den Anteil der aufgeklärten Variabilität der Messwerte relativ zur Summe der Quadratsummen des systematischen Effekts und des Residuums auf der Ebene der Stichprobe angibt

- **Formel:**
$$\eta_p^2 = \frac{QS_{Effekt}}{QS_{Effekt} + QS_{Innerhalb}}$$

- **Beispiel** zur Berechnung (vgl. Folien zur letzten Sitzung):

$$\eta_p^2 = \frac{QS_A}{QS_A + QS_{Innerhalb}} = \frac{11.25}{11.25 + 10} \approx 0.53$$

- **SPSS** verwendet dieses Maß als Effektgröße
- **Problem:** Maß überschätzt den Anteil an aufgeklärter Varianz in der Population

Effektgröße (z. B. Rasch, Frieze, Hofmann & Naumann, 2021)

- Weitere Effektgrößen: f^2 und ω_p^2

- Formeln:
$$f^2 = \frac{(F - 1) \cdot df_{\text{Zähler}}}{N}$$

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

- Beispiel zur Berechnung (vgl. Folien zur letzten Sitzung):

$$f_A^2 = \frac{(F_A - 1) \cdot df_A}{N} = \frac{(18 - 1) \cdot 1}{20} = 0.85 \quad \omega_p^2 = \frac{f_A^2}{1 + f_A^2} = \frac{0.85}{1 + 0.85} \approx 0.46$$

- GPower nutzt u. a. f und f^2 für die Stichprobenumfangsplanung
- Keine Überschätzung: Im Gegensatz zu η_p^2 überschätzen f^2 und ω_p^2 nicht den Anteil an aufgeklärter Varianz in der Population
- Zusammenhang zwischen f und d : $d = 2 \cdot f$ bzw. $d = 2 \cdot \sqrt{\frac{\omega^2}{1 - \omega^2}}$

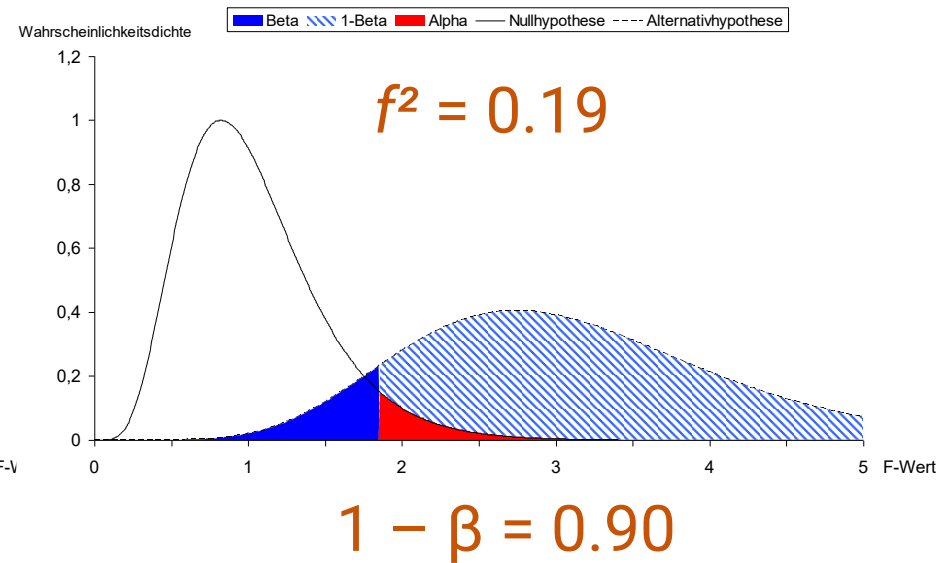
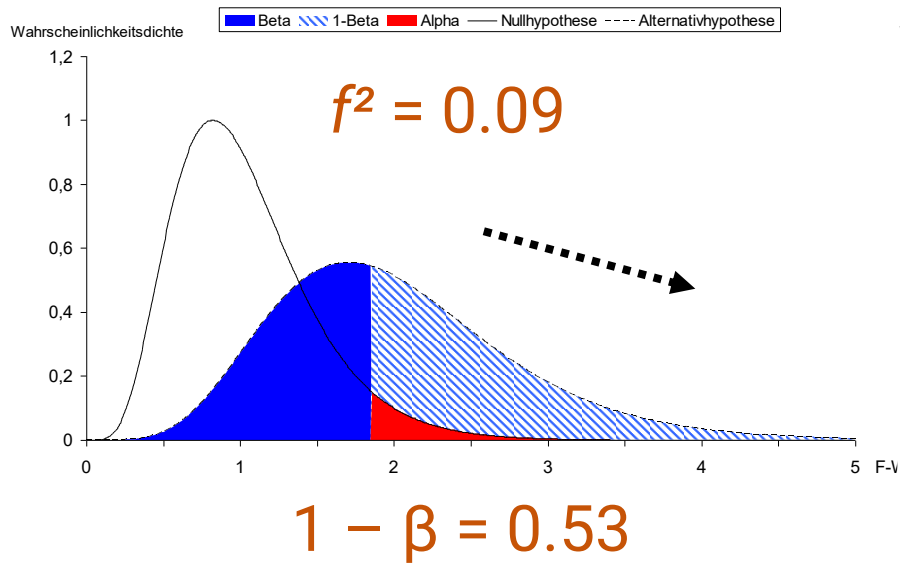
Umrechnung von Effektgrößen

Bitte rechnen Sie die Effektgröße $d = 0.5$ in f^2 und ω_p^2 um! Welche Werte sind richtig?

- ☐ A: $f^2 = 4.00$ und $\omega_p^2 = 0.80$
- ☐ B: $f^2 = 0.25$ und $\omega_p^2 = 0.20$
- ☐ C: $f^2 = 1.00$ und $\omega_p^2 = 0.50$
- ☐ D: $f^2 = 0.06$ und $\omega_p^2 = 0.06$

Einfluss der Effektgröße auf die Teststärke (Rey, 2020)

- Je größer die Effektgröße, desto größer die Teststärke ($1 - \beta$)
- Effektgröße wirkt sich (nur) auf die nonzentrale Verteilung aus
 - Nonzentrale Verteilung „wandert“ nach rechts
 - Nonzentrale Verteilung „verflacht“
- Beispiel für den Einfluss der Effektgröße auf die Teststärke:



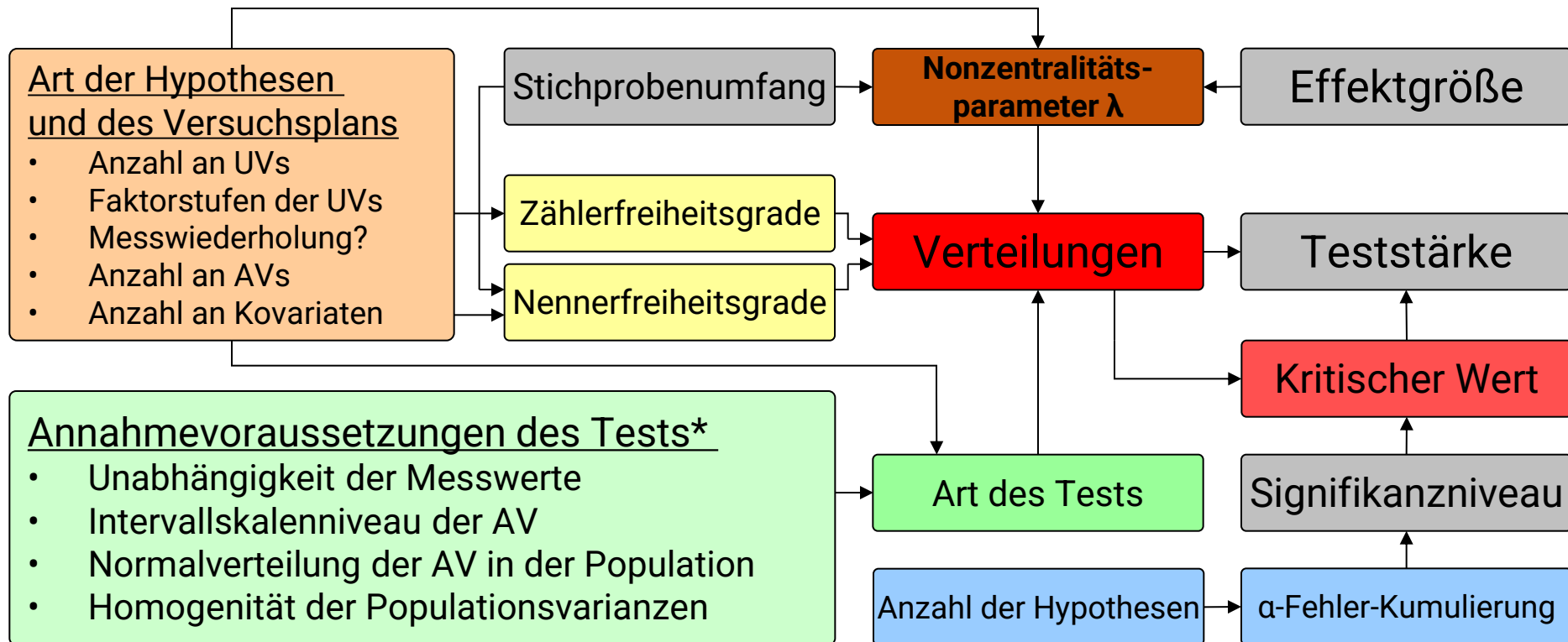
Einfluss der Effektgröße auf den Stichprobenumfang

- Je größer die Effektgröße, desto kleiner der benötigte Stichprobenumfang
- Beispiel: Stichprobenumfänge für ein 2 x 2 faktorielles Design mit einer gewünschten Teststärke ($1 - \beta$) von 0.80 und einem Signifikanzniveau von 5%:

Effektgröße (f)	Stichprobenumfang
0.10	787
0.25	128
0.40	52

Weitere Einflussgrößen (Rey, 2020)

- Einflussgrößen neben den drei zentralen Kenngrößen (Stichprobenumfang, Signifikanzniveau & Effektgröße) auf die Teststärke:

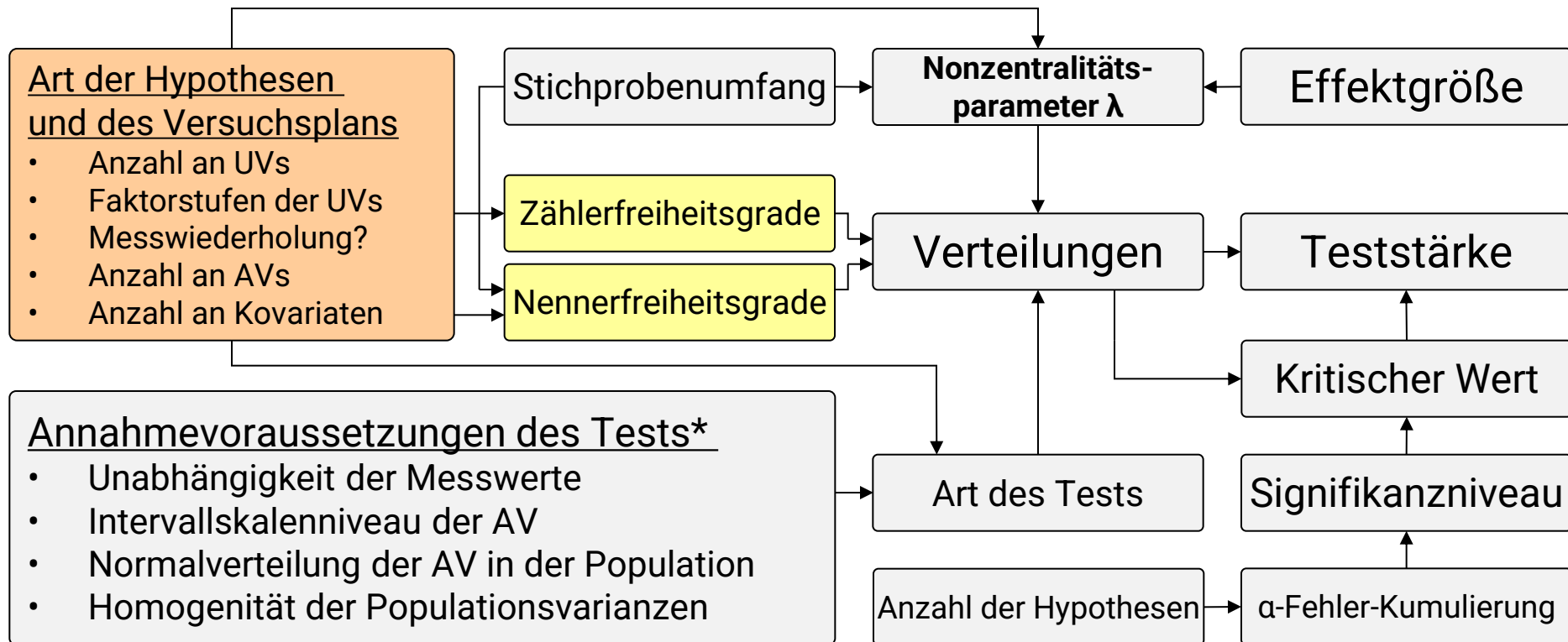


* Hier beispielhaft: Annahmeveraussetzungen der ANOVA ohne MW

Quelle: Angelehnt an Rey (2020)

Weitere Einflussgrößen (Rey, 2020)

- Einflussgrößen neben den drei zentralen Kenngrößen (Stichprobenumfang, Signifikanzniveau & Effektgröße) auf die Teststärke:



* Hier beispielhaft: Annahmeveraussetzungen der ANOVA ohne MW

Quelle: Angelehnt an Rey (2020)

Weitere Einflussgrößen (Rey, 2020)

- Anzahl und Faktorstufen unabhängiger, hypothesenrelevanter Variablen beeinflusst die Zähler- und Nennerfreiheitsgrade
- Zählerfreiheitsgrade (df_Z):
 - Je größer df_Z , desto mehr Versuchspersonen werden benötigt
 - Formel für die Haupteffekte:
$$df_Z = p - 1 \text{ (bzw. } q - 1 \text{ oder } z - 1)$$
 - Formel für die Wechselwirkungen:
$$df_Z = (p - 1) \cdot (q - 1) \cdot (...) \cdot (z - 1)$$
- Nennerfreiheitsgrade (df_N):
 - Je größer df_N , desto weniger Versuchspersonen werden benötigt
 - Formel für sämtliche Effekte: $df_N = N - p \cdot q \cdot ... \cdot z$

p = Faktorstufen des ersten Faktors q = Faktorstufen des zweiten Faktors z = Faktorstufen des letzten Faktors

Weitere Einflussgrößen (Rey, 2020)

- Veränderung von df_z wirkt sich i.d.R. deutlich stärker auf das benötigte N aus als Veränderung von df_N
- Hieraus und aus den vorherigen Formeln folgt:
 - Hinzunahme von zweifachgestuften UVs beeinflusst das benötigte N kaum
 - Hinzunahme von UVs mit mehreren Faktorstufen führt vor allem in komplexen Versuchsplänen zu erheblich höherem benötigten N
- **Wichtig:** Versuchsplan einschließlich UVs primär nach inhaltlichen Gesichtspunkten und nicht nur nach Stichprobenumfang auswählen

Einfluss des Versuchsdesigns auf den benötigten Stichprobenumfang

Welches Versuchsdesign benötigt Ihrer Vermutung nach den höchsten Stichprobenumfang für die komplexeste Wechselwirkung (falls vorhanden)?

- ☐ A: Ein einfaktorielles, fünffachgestuftes Design
- ☐ B: Ein 2 x 4 faktorielles Design
- ☐ C: Ein 2 x 2 x 2 x 2 faktorielles Design

Berechnung des Stichprobenumfangs (z. B. Rasch, Frieze, Hofmann & Naumann, 2021)

- Formel zur Berechnung von N :

$$N = \frac{\lambda_{df;\alpha;1-\beta}^2}{\phi^2} = \frac{\lambda_{df;\alpha;1-\beta}^2}{\frac{\omega_p^2}{1 - \omega_p^2}}$$

N = Benötigter Stichprobenumfang
 λ = Nonzentralitätsparameter Lambda
 Φ^2 = Phi-Quadrat
 ω_p^2 = Effektgröße partielles Omega-Quadrat

- Beispiel: 2 x 2 faktorielles Design mit $\alpha = .01$, $f = .40$ (großer Effekt) und $1 - \beta = .80$
- Formel (siehe vorherige Folien): $\omega_p^2 = f^2 : (1 + f^2)$
- $\lambda = 11.60$ (laut Tabelle); $f^2 = .16$; $\omega_p^2 \approx .14$; $\Phi^2 = .16$
- Benötigter Stichprobenumfang: $N = 73$ (72.5)

Berechnung des Stichprobenumfangs

Bitte berechnen Sie das benötigte N für die Wechselwirkung zu einem 2×3 faktoriellen Design mit $f = .25$, $\alpha = .01$ und $1 - \beta = .80$.

☐ A: $N = 227$

☐ B: $N = 223$

☐ C: $N = 80$

☐ D: $N = 236$

TPF-3

λ als Funktion der Teststärke und der Zählerfreiheitsgrade $FG(Z)$ für F -, χ^2 - und zweiseitige t -Tests ($FG(Z) = 1$)

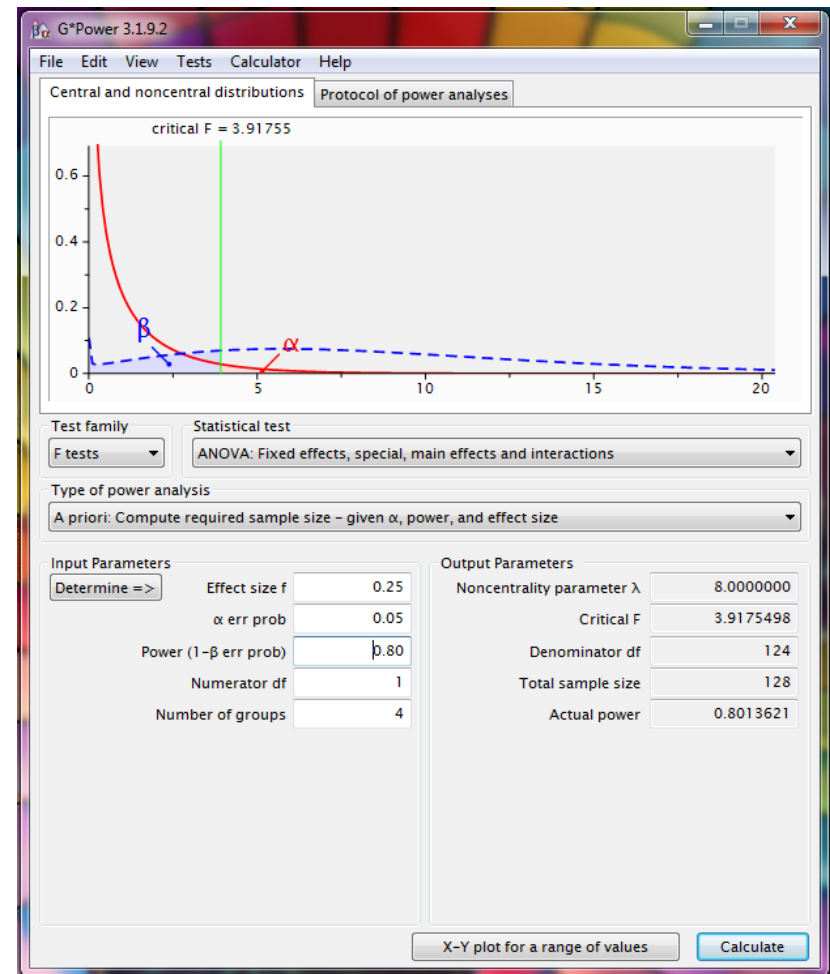
bei $\alpha = 0,01$ (entspricht bei einseitigen t -Tests ($FG(Z) = 1$) $\alpha = 0,005$)

Test- stärke	0,1000	0,5000	0,6667	0,7500	0,8000	0,8500	0,9000	0,9250	0,9500	0,9750	0,9900	0,9925	0,9950	0,9975	0,9990
FG(Z)															
1	1,67	6,63	9,04	10,57	11,60	13,05	14,88	16,12	17,81	20,57	24,03	25,08	26,54	28,97	32,10
2	2,30	8,19	10,92	12,64	13,88	15,40	17,43	18,80	20,65	23,66	27,41	28,55	30,13	32,75	36,11
3	2,76	9,31	10,28	14,10	15,46	17,09	19,25	20,78	22,67	25,86	29,83	31,02	32,68	35,43	38,96
4	3,15	10,23	13,38	15,34	16,75	18,47	20,74	22,27	24,33	27,66	31,79	33,04	34,76	37,63	41,29
5	3,49	11,03	14,35	16,39	17,87	19,66	22,03	23,60	25,76	29,22	33,50	34,79	36,57	39,83	43,30

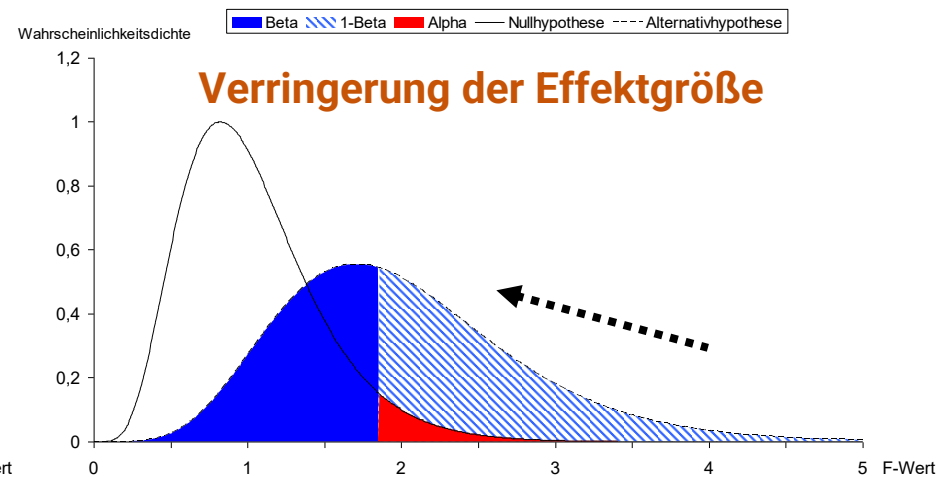
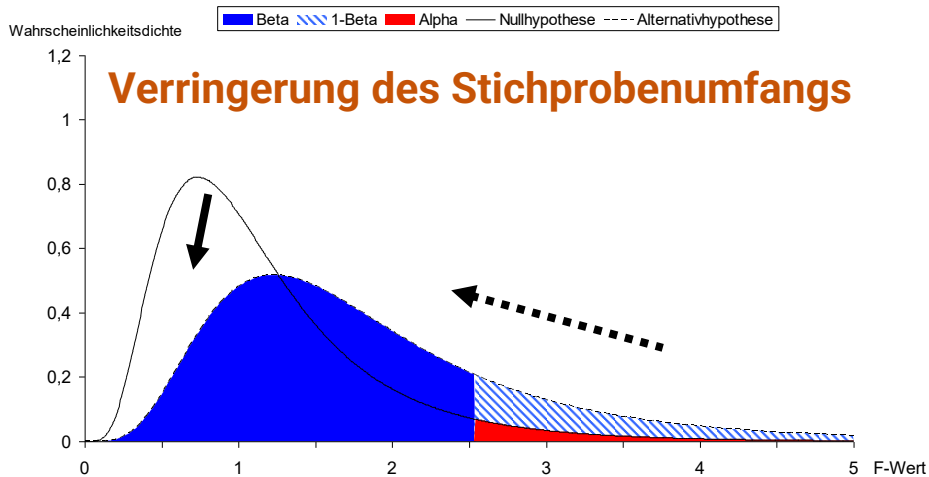
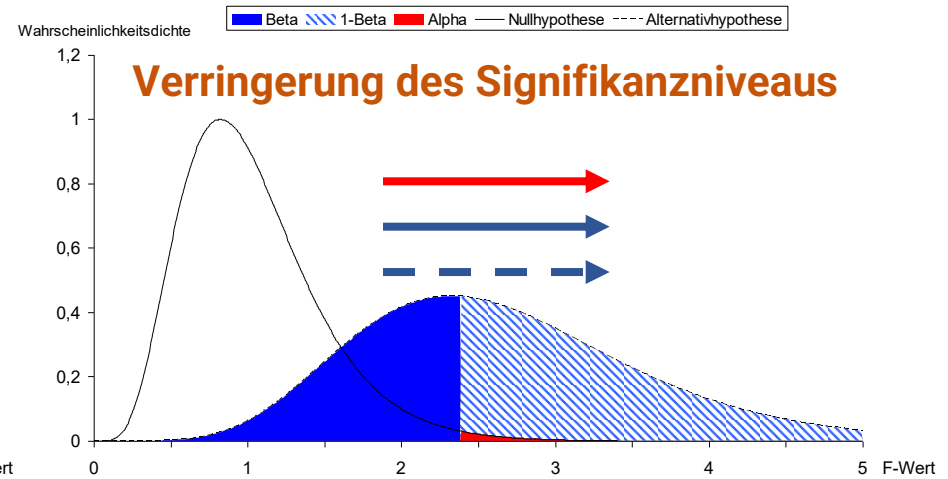
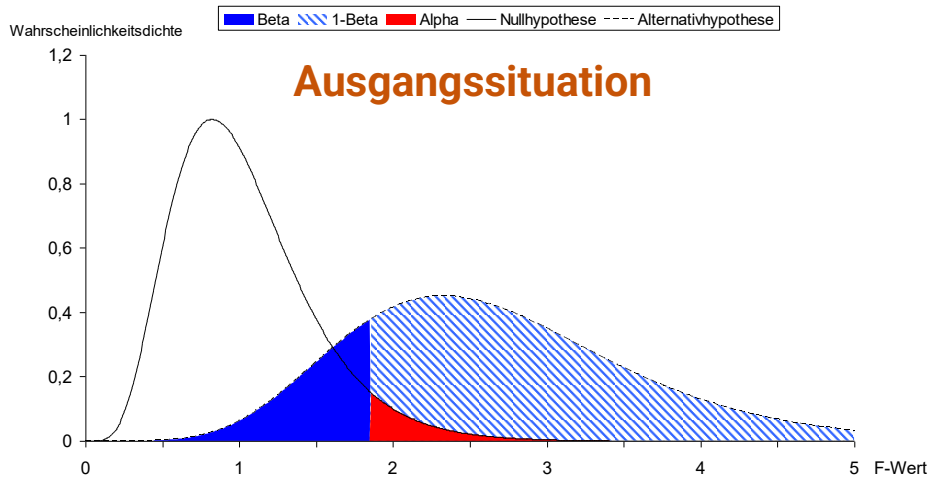
Quelle: Rasch, Frieze, Hofmann und Naumann (2021)

Berechnung des Stichprobenumfangs mit GPower

- **Beispiel:** Berechnung des Stichprobenumfangs mit GPower für ein 2 x 2 faktorielles Design
 - ANOVA / *F*-Test
 - Signifikanzniveau: 0.05
 - Angenommene Effektgröße: $f = 0.25$ (mittlerer Effekt)
 - Gewünschte Teststärke: 0.80
 - Zählerfreiheitsgrade (Numerator df): 1
 - Anzahl an Gruppen: 4
- **Benötigter Stichprobenumfang:** 128



Zusammenfassung



Prüfungsliteratur

- Rey, G. D. (2020). *Methoden der Entwicklungspsychologie. Datenerhebung und Datenauswertung* (3., überarbeitete Auflage). Norderstedt BoD.

(Unter-)Kapitel	Taschenbuch	E-Book (ePUB)	Webseite
Stichprobenumfangsplanung	S. 129–151	S. 102–116	S. 102–126

Weiterführende Literatur I

- Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Berlin: Springer.
 - Hypothesentesten (S. 97–115)
- Eid, M., Gollwitzer, M., & Schmitt, M. (2017). *Statistik und Forschungsmethoden* (5. Aufl.). Weinheim: Beltz.
 - Grundlagen der Inferenzstatistik (S. 217–278)
- Leonhart, R. (2022). *Lehrbuch Statistik. Einstieg und Vertiefung* (5. Auflage). Bern: Huber.
 - Einführung in die inferenzstatistische Hypothesenprüfung (S. 195–230)
- Sedlmeier, P., & Renkewitz, F. (2018). *Forschungsmethoden und Statistik: Ein Lehrbuch für Psychologen und Sozialwissenschaftler* (3. Aufl.). München: Pearson.
 - Signifikanztests (S. 369–406)

Weiterführende Literatur II

- Bühner, M., & Ziegler, M. (2017). *Statistik für Psychologen und Sozialwissenschaftler* (2. Aufl.). Hallbergmoos: Pearson Studium.
 - Versuchsplanung mit G*Power und R (S. 237–247)
- Rasch, B., Frieze, M., Hofmann, W., & Naumann, E. (2021). *Quantitative Methoden 1: Einführung in die Statistik für Psychologie, Sozial- & Erziehungswissenschaften* (5. Aufl.). Heidelberg: Springer.
 - Der *t*-Test (S. 35–82)
- Rasch, B., Frieze, M., Hofmann, W., & Naumann, E. (2021). *Quantitative Methoden 2: Einführung in die Statistik für Psychologie, Sozial- & Erziehungswissenschaften* (5. Aufl.). Heidelberg: Springer.
 - Einfaktorielle Varianzanalyse (S. 1–36)
 - Zweifaktorielle Varianzanalyse (S. 41–69)

Weiterführende Literatur III

- Döring, N., & Bortz, J. (2015). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Aufl.). Berlin: Springer.
- O'Keefe, D. J. (2007). Post hoc power, observed power, a priori power, retrospective power, prospective power, achieved power: Sorting out appropriate uses of statistical power analyses. *Communication Methods and Measures*, 1, 291–299.