

### Aufgaben zu Kapitel 3

#### Aufgabe 1

- Berechnen Sie einen  $t$ -Test für unabhängige Stichproben für den Vergleich der beiden Verarbeitungsgruppen „strukturell“ und „emotional“ für die abhängige Variable „neutrale Adjektive“.
- Berechnen Sie mit G\*Power den empirischen Effekt.

#### Aufgabe 2

Ein Forscher plant eine Untersuchung, bei der er nur einen kleinen Effekt zwischen den beiden unabhängigen Versuchsgruppen erwartet. Er setzt das  $\alpha$ -Niveau auf 5% und strebt einen  $\beta$ -Fehler an, der nicht mehr als doppelt so groß sein soll. Wie viele Versuchspersonen braucht er, um seine gerichtete Annahme in jedem Fall interpretieren zu können? Wie viele wären es, wäre seine Hypothese ungerichtet gewesen?

#### Aufgabe 3

Ein anderer Forscher glaubt, dass es zwischen zwei unabhängigen Versuchsbedingungen keinen Unterschied gibt. Er möchte also die Nullhypothese interpretieren. Dafür muss er die Wahrscheinlichkeit klein halten (5%), sich zu irren, also zu unrecht die  $H_0$  anzunehmen, obwohl in Wirklichkeit die  $H_1$  gilt. Stattdessen ist er etwas liberaler darin, sich evtl. zu Gunsten der Alternativhypothese zu irren. Er legt die entsprechende Fehlerwahrscheinlichkeit auf 10% fest. Wie viele Versuchspersonen braucht der Forscher, um hinreichend sicher sein zu können, dass nicht zumindest doch ein kleiner Effekt zwischen den Gruppen existiert?

#### Aufgabe 4

Eine Studie führt zu einem nicht signifikanten Ergebnis, obwohl in jeder der beiden unabhängigen Gruppen 100 Personen teilgenommen haben. Wie groß war die Power bei einer ungerichteten Fragestellung, einem  $\alpha$ -Niveau von 5% und einem für inhaltlich relevant erachteten Effekt von  $d = 0,50$ ?

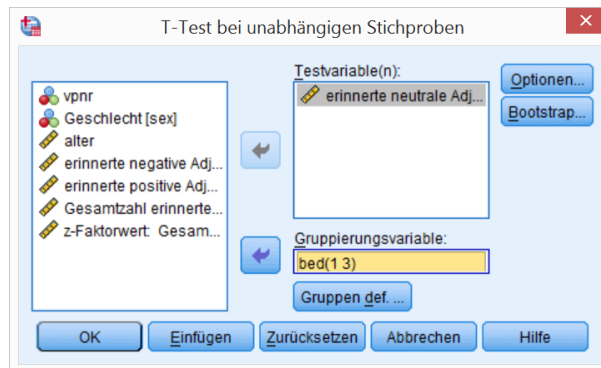
#### Aufgabe 5

- Prüfen Sie mit einem  $t$ -Test für abhängige Stichproben, ob sich im Experiment zu motorischen Fähigkeiten die Anzahl beendeter Sequenzen zwischen dem ersten und dem dritten Messzeitpunkt unterscheiden ( $\alpha = 5\%$ , Datei „Messwiederholung.sav“).
- Wie groß war der empirische Effekt  $d_z$ ?
- Wie groß war die Teststärke, einen Unterschied von  $d_z = 0,5$  zu finden?
- Wie viele Versuchspersonen wären erforderlich gewesen, um einen  $\alpha = \beta$ -Fehler von 5% zu erreichen bei einem angenommenen Effekt von  $d_{unabhängig} = 0,5$ , wenn wir vor der Untersuchung nur von einer Korrelation von  $r = 0,30$  zwischen den Messwertreihen ausgegangen wären?
- Auf welche Zahl hätte sich dieser Bedarf erhöht, wenn die Fragestellung mit unabhängigen Stichproben untersucht worden wäre?

## Lösungen

### Aufgabe 1

- a) „Analysieren“ → „Mittelwerte vergleichen“ → „t-Test bei unabhängigen Stichproben“. Testvariable ist „erinnerte neutrale Adjektive“, Gruppenvariable „Verarbeitungsbedingung“ und davon die beiden Stufen „1“ (strukturell) und „3“ (emotional).



Die Analyse liefert den folgenden Output:

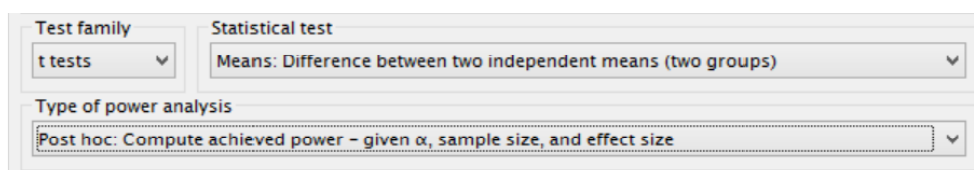
Gruppenstatistik				
Verarbeitungsbedingung	H	Mittelwert	Standardabweichung	Standardfehler Mittelwert
erinnerte neutrale Adjektive	50	1,96	1,428	,202
emotional	50	4,06	2,064	,292

Test bei unabhängigen Stichproben										
		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit					95% Konfidenzintervall der Differenz	
		F	Sig.	t	df	Sig. (2-seitig)	Mittelwertdifferenz	Standardfehler der Differenz	Unterer	Oberer
erinnerte neutrale Adjektive	Varianzgleichheit angenommen	3,609	,060	-5,916	98	,000	-2,100	,355	-2,804	-1,396
	Varianzgleichheit nicht angenommen			-5,916	87,156	,000	-2,100	,355	-2,806	-1,394

Der Mittelwertsvergleich mit bloßem Auge im Feld Gruppenstatistiken zeigt einen Unterschied zwischen den beiden Gruppen. Der Levene-Test auf Varianzhomogenität liefert ein marginal signifikantes Ergebnis. Dies drückt sich u.a. in den reduzierten Freiheitsgraden in der Zeile „Varianzen sind nicht gleich“ aus. Doch unabhängig von einer möglichen Korrektur der Freiheitsgrade wird der Unterschied zwischen den Gruppen hoch signifikant. Entsprechend der theoretischen Vorhersage erinnern Personen in der Verarbeitungsbedingung „emotional“ mehr Adjektive als Personen in der Bedingung „strukturell“.

- b) Wählen Sie das folgende Menü und dann „Determine“, um die Effektstärke zu berechnen. G\*Power ermittelt einen empirischen Effekt von  $d = 1,18$ . Dies ist nach den Konventionen von Cohen (1988) ein großer Effekt.



## Aufgaben mit SPSS & G\*Power

Rasch, Friese, Hofmann & Naumann (2014). *Quantitative Methoden. Band 1* (4. Auflage). Heidelberg: Springer.

The screenshot shows the G\*Power dialog box for a two-group mean difference test. The 'n1 = n2' radio button is selected. The input fields are: Mean group 1 (1.96), Mean group 2 (4.06), and SD  $\sigma$  within each group (0.5). The calculated 'Effect size d' is 1.183284. Buttons for 'Calculate', 'Calculate and transfer to main window', and 'Close' are visible.

Achtung: Bitte beachten Sie, dass G\*Power für die korrekte Berechnung Punkte an Stelle von Kommata erwartet.

### Aufgabe 2

Er braucht 858 Versuchspersonen, fast 430 Personen in jeder Zelle. Würde er eine ungerichtete Hypothese verfolgen, wären es sogar 1054 Personen insgesamt. Dieses Ergebnis macht deutlich, wie dramatisch die benötigte Anzahl Versuchspersonen bei kleinen angenommenen Effekten ansteigt, wenn eine hohe Teststärke erreicht werden soll.

The screenshot shows the G\*Power dialog box for a two-group mean difference test. The 'Type of power analysis' is set to 'A priori: Compute required sample size - given alpha, power, and effect size'. The 'Input Parameters' section includes: Tail(s) One, Effect size d 0.2, alpha err prob 0.05, Power (1 - beta err prob) 0.90, and Allocation ratio N2/N1 1. The 'Output Parameters' section includes: Noncentrality parameter  $\delta$  2.9291637, Critical t 1.6466357, Df 856, Sample size group 1 429, Sample size group 2 429, Total sample size 858, and Actual power 0.9000777.

### Aufgabe 3

Der Forscher möchte den  $\beta$ -Fehler minimieren, während er beim  $\alpha$ -Fehler etwas liberaler ist. Entsprechend legt er den  $\beta$ -Fehler auf 5%, den  $\alpha$ -Fehler auf 10% fest. Da ein Unterschied zwischen den Gruppen unabhängig vom Vorzeichen gegen seine Annahme sprechen würde, muss er einen zweiseitigen Test durchführen. Bei einem existierenden kleinen Effekt von  $d = 0,20$  müsste das Ergebnis also bei 1084 Versuchspersonen signifikant werden. Wird es dies nicht, kann der Forscher mit 95%iger Sicherheit davon ausgehen, dass es keinen Effekt von  $d = 0,20$  oder größer gibt.

Quelle: <http://www.lehrbuch-psychologie.de/qm>

© Rasch, Friese, Hofmann & Naumann

## Aufgaben mit SPSS & G\*Power

Rasch, Friese, Hofmann & Naumann (2014). *Quantitative Methoden. Band 1* (4. Auflage). Heidelberg: Springer.

The screenshot shows the G\*Power interface for a t-test. The 'Test family' is 't tests' and the 'Statistical test' is 'Means: Difference between two independent means (two groups)'. The 'Type of power analysis' is 'A priori: Compute required sample size - given alpha, power, and effect size'. The 'Input Parameters' section includes: 'Determine =>' (checked), 'Tail(s)' set to 'Two', 'Effect size d' set to 0.2, 'alpha err prob' set to 0.10, 'Power (1-beta err prob)' set to 0.95, and 'Allocation ratio N2/N1' set to 1. The 'Output Parameters' section shows: 'Noncentrality parameter delta' as 3.2924155, 'Critical t' as 1.6462631, 'Df' as 1082, 'Sample size group 1' as 542, 'Sample size group 2' as 542, 'Total sample size' as 1084, and 'Actual power' as 0.9500673.

### Aufgabe 4

Das Experiment war gut geplant. Die Power lag bei 94%. Da die Daten trotzdem kein signifikantes Ergebnis zeigten, liegt die Vermutung nahe, dass der angenommene Effekt zu hoch war. Wäre er in Wirklichkeit nur von der Größe  $d = 0,20$ , hätte das Experiment nur eine Power von 29% gehabt! Prüfen Sie es nach!

The screenshot shows the G\*Power interface for a t-test. The 'Test family' is 't tests' and the 'Statistical test' is 'Means: Difference between two independent means (two groups)'. The 'Type of power analysis' is 'Post hoc: Compute achieved power - given alpha, sample size, and effect size'. The 'Input Parameters' section includes: 'Determine =>' (checked), 'Tail(s)' set to 'Two', 'Effect size d' set to 0.5, 'alpha err prob' set to 0.05, 'Sample size group 1' set to 100, and 'Sample size group 2' set to 100. The 'Output Parameters' section shows: 'Noncentrality parameter delta' as 3.5355339, 'Critical t' as 1.9720175, 'Df' as 198, and 'Power (1-beta err prob)' as 0.9404272.

### Aufgabe 5

a) „Analysieren“ → „Mittelwerte vergleichen“ → „t-Test bei verbundenen Stichproben“. Gepaarte Variablen sind „Messung 1“ und „Messung 3“. Die Analyse liefert folgenden Output:

# Aufgaben mit SPSS & G\*Power

Rasch, Friese, Hofmann & Naumann (2014). *Quantitative Methoden. Band 1* (4. Auflage). Heidelberg: Springer.

Statistik für Stichproben mit paarigen Werten

	Mittelwert	H	Standardabweichung	Standardfehler Mittelwert
Paar 1 Anzahl getippter Sequenzen bei Messung 1	16,56	36	4,931	,822
Anzahl getippter Sequenzen bei Messung 3	18,25	36	4,789	,798

Korrelationen für Stichproben mit paarigen Werten

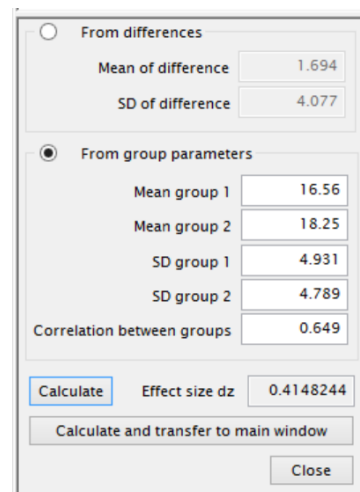
	H	Korrelation	Sig.
Paar 1 Anzahl getippter Sequenzen bei Messung 1 & Anzahl getippter Sequenzen bei Messung 3	36	,649	,000

Test für Stichproben mit paarigen Werten

	Paarige Differenzen					t	df	Sig. (2-seitig)
	Mittelwert	Standardabweichung	Standardfehler Mittelwert	95% Konfidenzintervall der Differenz				
				Unterer	Oberer			
Paar 1 Anzahl getippter Sequenzen bei Messung 1 - Anzahl getippter Sequenzen bei Messung 3	-1,694	4,077	,679	-3,074	-,315	-2,494	35	,018

Der Unterschied von durchschnittlich 1,694 mehr geschafften Sequenzen am dritten Messzeitpunkt ist signifikant.

- b) Mit G\*Power lässt sich  $d_z$  auf zwei verschiedene Weisen berechnen, über die Differenzen zwischen den Messwertereihen oder über Mittelwerte und Streuungen der beiden Messzeitpunkte. Beide Wege führen zum Ergebnis von  $d_z = 0,41$ .



- c) Die Teststärke lag bei 83%.

# Aufgaben mit SPSS & G\*Power

Rasch, Friese, Hofmann & Naumann (2014). *Quantitative Methoden. Band 1* (4. Auflage). Heidelberg: Springer.

Test family		Statistical test	
t tests		Means: Difference between two dependent means (matched pairs)	
Type of power analysis			
Post hoc: Compute achieved power - given alpha, sample size, and effect size			
Input Parameters		Output Parameters	
Tail(s) Two		Noncentrality parameter $\delta$	3.000000
Determine =>	Effect size dz 0.5	Critical t	2.0301079
	$\alpha$ err prob 0.05	Df	35
	Total sample size 36	Power (1- $\beta$ err prob)	0.8306656

- d) Um einen Effekt von  $d_{unabhängig} = 0,5$  mit einer Power von 95% bei  $\alpha = 0,05$  und einer angenommenen Korrelation zwischen den abhängigen Messwertpaaren von  $r = 0,30$  zu finden, wären 75 Personen nötig gewesen.

$$d_z = \sqrt{\frac{2}{1-r}} \cdot f_{unabhängig} = \sqrt{\frac{2}{1-0,3}} \cdot \frac{d_{unabhängig}}{2} = \sqrt{\frac{2}{1-0,3}} \cdot \frac{0,5}{2} = 0,4226$$

Test family		Statistical test	
t tests		Means: Difference between two dependent means (matched pairs)	
Type of power analysis			
A priori: Compute required sample size - given alpha, power, and effect size			
Input Parameters		Output Parameters	
Tail(s) Two		Noncentrality parameter $\delta$	3.6598234
Determine =>	Effect size dz 0.4226	Critical t	1.9925435
	$\alpha$ err prob 0.05	Df	74
	Power (1- $\beta$ err prob) 0.95	Total sample size	75
		Actual power	0.9507288

## Aufgaben mit SPSS & G\*Power

Rasch, Friese, Hofmann & Naumann (2014). *Quantitative Methoden. Band 1* (4. Auflage). Heidelberg: Springer.

- e) Wäre diese Fragestellung mit unabhängigen Stichproben untersucht worden, hätte sich der Bedarf an Versuchspersonen auf 210 Personen im Gegensatz zu 75 Personen in Aufgabe d) mehr als verdoppelt. Dieses Beispiel macht die große Überlegenheit des  $t$ -Tests für abhängige Stichproben in Fragen der Teststärke deutlich. Ohne den Einfluss der Korrelation (also bei  $r = 0$ ) wären im  $t$ -Test für abhängige Stichproben ungefähr 105 Versuchspersonen notwendig gewesen (bitte nachrechnen). Dies entspricht der Hälfte der hier ermittelten Anzahl von 210 Personen. Bei der Messwiederholung gibt jede Person zwei Messwerte ab. Im Fall einer Nullkorrelation zwischen abhängigen Stichproben führt eine identische Zahl von Messwerten also zu identischen Ergebnissen bei  $t$ -Tests für unabhängige und abhängige Stichproben.

The screenshot shows the G\*Power software interface for a t-test. The 'Test family' is set to 't tests' and the 'Statistical test' is 'Means: Difference between two independent means (two groups)'. The 'Type of power analysis' is 'A priori: Compute required sample size - given  $\alpha$ , power, and effect size'. The 'Input Parameters' section includes: Tail(s) set to 'Two', Effect size d set to 0.5,  $\alpha$  err prob set to 0.05, Power ( $1 - \beta$  err prob) set to 0.95, and Allocation ratio N2/N1 set to 1. The 'Output Parameters' section includes: Noncentrality parameter  $\delta$  set to 3.6228442, Critical t set to 1.9714347, Df set to 208, Sample size group 1 set to 105, Sample size group 2 set to 105, Total sample size set to 210, and Actual power set to 0.9501287.

Input Parameters		Output Parameters	
Tail(s)	Two	Noncentrality parameter $\delta$	3.6228442
Effect size d	0.5	Critical t	1.9714347
$\alpha$ err prob	0.05	Df	208
Power ( $1 - \beta$ err prob)	0.95	Sample size group 1	105
Allocation ratio N2/N1	1	Sample size group 2	105
		Total sample size	210
		Actual power	0.9501287

## Literatur

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NY: Erlbaum.