

Kapitel 6: Zweifaktorielle Varianzanalyse

Berechnen der Teststärke a priori bzw. Stichprobenumfangsplanung _____	1
Teststärkebestimmung a posteriori _____	4
Berechnen der Effektgröße f aus empirischen Daten und Bestimmung der beobachteten Teststärke _____	5
Literatur _____	8

Berechnen der Teststärke a priori bzw. Stichprobenumfangsplanung

Nach dem Starten von G*Power müssen Sie zunächst unter Test family „ANOVA: Fixed effects, special, main effects and interactions“ auswählen. Sie erhalten folgendes Eingabefenster:

Als Konventionen für einen kleinen, mittleren und großen Effekt gelten weiterhin $f = 0,10$, $f = 0,25$ und $f = 0,40$ (Cohen, 1988). Die Umrechnung in Ω^2 erfolgt nach der bekannten Formel:

$$\Phi^2 = \frac{\Omega^2}{1 - \Omega^2} \text{ bzw. } \Omega^2 = \frac{\Phi^2}{1 + \Phi^2}; \sqrt{\Phi^2} \text{ entspricht } f \text{ in G*Power}$$

Zur Veranschaulichung der Stichprobenumfangsplanung einer zweifaktoriellen Varianzanalyse ohne Messwiederholung soll das bekannte Beispiel des Einflusses der Verarbeitungsbedingung („strukturell“, „bildhaft“ und „emotional“) auf die Erinnerungsleistung dienen. Zusätzlich soll der Einfluss des Faktors „Geschlecht“ untersucht werden. Es liegt also eine 3 x 2 Varianzanalyse vor. Eine derartige Analyse unterscheidet drei Arten von Effekten: Den Haupteffekt der Verarbeitungsbedingung, den Haupteffekt des Geschlechts und die Wechselwirkung zwischen diesen beiden Faktoren. Für alle drei Effektarten müssen Sie eine separate Stichprobenumfangsplanung vornehmen, es sei denn ihre Hypothese bezieht sich nur auf eine bestimmte Effektart.

G*Power-Ergänzungen

Rasch, Friese, Hofmann & Naumann (2014). *Quantitative Methoden. Band 2* (4. Auflage). Heidelberg: Springer.

Wichtig für die Stichprobenumfangsplanung einer zweifaktoriellen Varianzanalyse sind die Freiheitsgrade der einzelnen Effektkarten. In der allgemeinen Schreibweise hat der Faktor A p Stufen, der Faktor B q Stufen. So ergeben sich folgende Freiheitsgrade:

Zählerfreiheitsgrade:

$$\text{Haupteffekt A: } df_A = p - 1 \quad \rightarrow \text{HE „Verarbeitungsbedingung“: } df = 3 - 1 = 2$$

$$\text{Haupteffekt B: } df_B = q - 1 \quad \rightarrow \text{HE „Geschlecht“: } df = 2 - 1 = 1$$

$$\text{Wechselwirkung A} \times \text{B: } df_{A \times B} = (p - 1) \cdot (q - 1) \quad \rightarrow \text{WW: } df = (3 - 1) \cdot (2 - 1) = 2$$

Nennerfreiheitsgrade:

$$df_{Res} = p \cdot q \cdot (n - 1)^1$$

Für die Stichprobenumfangsplanung legen wir einen mittleren Effekt als inhaltlich relevant fest ($f = 0,25$ bzw. $\Omega^2 = 0,06$). Das Signifikanzniveau beträgt 5%. Die Teststärke, einen mittleren Effekt zu entdecken falls er wirklich existiert, soll mindestens 80% betragen. Die Zählerfreiheitsgrade sind bei diesem Haupteffekt $df = 2$. Die Gruppenanzahl ist in diesem Fall 3 (Verarbeitungsbedingung) \times 2 (Geschlecht) = 6.

The screenshot shows the G*Power software interface. The 'Test family' is set to 'F tests' and the 'Statistical test' is 'ANOVA: Fixed effects, special, main effects and interactions'. The 'Type of power analysis' is 'A priori: Compute required sample size - given alpha, power, and effect size'. The 'Input Parameters' section includes: Effect size f (0.25), alpha err prob (0.05), Power (1-beta err prob) (0.80), Numerator df (2), and Number of groups (6). The 'Output Parameters' section includes: Noncentrality parameter lambda (9.8750000), Critical F (3.0555579), Denominator df (152), Total sample size (158), and Actual power (0.8020390).

Um einen Effekt mittlerer Größe für den Haupteffekt Verarbeitungsbedingung mit seinen drei Stufen mit 80%iger Wahrscheinlichkeit zu finden, falls er existiert, sind mindestens 158 Versuchspersonen notwendig. Da diese Zahl nicht durch Sechs teilbar ist (der Anzahl Zellen im Versuchsplan), würde man in diesem Fall insgesamt mindestens 162 Personen erheben (27 pro Bedingungskombination).

Für den Haupteffekt Geschlecht mit zwei Stufen ($df_{\text{Zähler}} = 1$) ergibt sich ein benötigter Stichprobenumfang von $N = 128$ bzw. 132, um gleich viele Personen in jeder Zelle des Versuchsplans zu haben.

¹ Anmerkung: Das kleine n bezeichnet hier die Anzahl der Versuchspersonen pro Bedingungskombination (auch „Zelle des Versuchsplans“ genannt).

G*Power-Ergänzungen

Rasch, Frieze, Hofmann & Naumann (2014). *Quantitative Methoden. Band 2* (4. Auflage). Heidelberg: Springer.

Input Parameters		Output Parameters	
<input type="button" value="Determine =>"/>	Effect size f	Noncentrality parameter λ	8.0000000
	α err prob	Critical F	3.9188157
	Power (1- β err prob)	Denominator df	122
	Numerator df	Total sample size	128
	Number of groups	Actual power	0.8012613

Die Berechnung des optimalen Stichprobenumfangs für die Wechselwirkung kommt zu dem gleichen Ergebnis wie die für den Haupteffekt Verarbeitungsbedingung, da sich die Zählerfreiheitsgrade der beiden Effektarten entsprechen (siehe oben).

Sind alle drei Effektarten (Haupteffekt A, Haupteffekt B und Wechselwirkung A x B) für die inhaltliche Hypothese relevant, so sollte der größte errechnete Stichprobenumfang gewählt werden (hier: $N = 158$ bzw. 162). Ist allein der Geschlechtsunterschied von inhaltlichem Interesse, ist die für den Faktor Geschlecht errechnete Versuchspersonenzahl ausreichend.

Teststärkebestimmung a posteriori

Ist in einer Untersuchung ohne vorangegangene Stichprobenumfangsplanung ein nicht signifikantes Ergebnis aufgetreten, so ist eine Teststärkebestimmung a posteriori unerlässlich. Dafür wählen Sie in G*Power unter „Type of power analysis“ post hoc aus.

In den SPSS Ergänzungen zu diesem Kapitel ergab sich für die Wechselwirkung der Faktoren Verarbeitungsbedingung und Geschlecht ein nicht signifikantes Ergebnis. Wie groß war die Teststärke für eine Wechselwirkung in dieser Untersuchung? An der Untersuchung haben insgesamt 150 Versuchspersonen teilgenommen. Das Signifikanzniveau war 5%. Die Zählerfreiheitsgrade der Wechselwirkung sind bei einer 3×2 ANOVA $df = 2$. Ein mittlerer Effekt von $f = 0,25$ ($\Omega^2 = 0,06$) gelte als inhaltlich relevant. Wie groß war die Wahrscheinlichkeit, einen Effekt dieser Größe zu finden, falls er wirklich existiert?

Input Parameters		Output Parameters	
Effect size f	0.25	Noncentrality parameter λ	9.3750000
α err prob	0.05	Critical F	3.0589280
Total sample size	150	Denominator df	144
Numerator df	2	Power ($1 - \beta$ err prob)	0.7793948
Number of groups	6		

Die Teststärke, einen mittleren Effekt der Wechselwirkung in dieser Untersuchung zu finden, betrug etwa 78%. Dies ist noch unter der empfohlenen Grenze für die minimal akzeptable Teststärke von 80%. Die Entscheidung, die Existenz eines mittleren Effekts auf Grund des vorliegenden Ergebnisses auszuschließen, ist also mit einer Wahrscheinlichkeit von etwa 22% falsch. Allerdings können Effektstärken von $f = 0,30$ oder größer bereits mit einer Sicherheit von über 90% ausgeschlossen werden. Große Effekt von $f = 0,40$ sind sogar mit über 99%iger Sicherheit nicht vorhanden (Bitte ausprobieren). Auf Grund der nicht signifikanten Wechselwirkung kann also mit über 90%iger Sicherheit die Aussage getroffen werden, dass Effekte der Größe $f = 0,30$ oder größer nicht existieren.

Berechnen der Effektgröße f aus empirischen Daten und Bestimmung der beobachteten Teststärke

Für die Berechnung der Effektgröße f mit Hilfe von G*Power gibt es zwei Möglichkeiten. Die erste Möglichkeit nutzt empirische Varianzen, die zweite basiert auf dem von SPSS gelieferten Effektstärkenmaß Eta-Quadrat bzw. „partielles Eta hoch zwei“ (Englisch: partial η^2 bzw. eta squared).

Für beide Wege ist es sinnvoll, sich den Output der Varianzanalyse in SPSS noch einmal zu verdeutlichen. Unter „Optionen“ haben wir zuvor eingestellt, dass das Programm sowohl die Effektstärke als auch die beobachtete Trennschärfe anzeigen soll.

Tests der Zwischensubjekteffekte

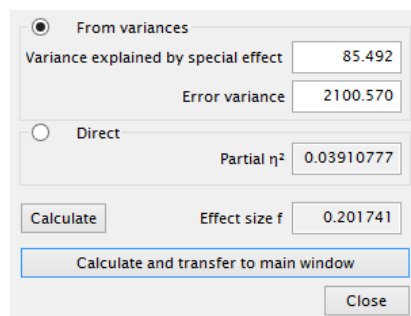
Abhängige Variable: Gesamtzahl erinnertes Adjektive

Quelle	Typ III Quadratsumme	df	Quadratischer Mittelwert	F	Sig.	Partielles Eta hoch zwei	Dezent. Parameter	Beobachtete Trennschärfe ^b
Korrigiertes Modell	741,623 ^a	5	148,325	10,168	,000	,261	50,840	1,000
Konstanter Term	12976,548	1	12976,548	889,579	,000	,861	889,579	1,000
bed	583,122	2	291,561	19,987	,000	,217	39,975	1,000
sex	85,492	1	85,492	5,861	,017	,039	5,861	,672
bed * sex	7,514	2	3,757	,258	,773	,004	,515	,090
Fehler	2100,570	144	14,587					
Gesamtsumme	18063,000	150						
Korrigierter Gesamtwert	2842,193	149						

a. R-Quadrat = ,261 (Angepasstes R-Quadrat = ,235)

b. Berechnet mit alpha = ,05

In G*Power öffnet sich durch Klicken auf „Determine“ das Fenster, in dem man f sehr einfach berechnen kann. Standardmäßig ist die Option „From variances“ aktiviert. Als Effektvarianz trägt man den Wert ein, der bei dem entsprechenden Effekt im Output in der Spalte „Type III Quadratsumme“ steht. Für den Faktor Geschlecht ist dies der Wert 85,492. Weiterhin verlangt G*Power die Fehlervarianz, die kurz darunter im Output zu finden ist (2100,570). Nun können Sie auf „Calculate and transfer to main window“ klicken. Sie sehen, dass G*Power ein f von 0,20 ermittelt, was einem kleinen bis mittleren Effekt entspricht.



Ein Klick auf „Calculate“ berechnet Ihnen die empirische Teststärke von nahezu 69%. Dies entspricht nahezu dem von SPSS ermittelten Wert für die beobachtete Trennschärfe von 67%. Die kleine Abweichung zwischen den Werten kommt dadurch zu Stande, dass SPSS die Teststärke anhand der Freiheitsgrade berechnet und nicht wie G*Power an Hand der Versuchspersonenzahl. Dies können Sie leicht überprüfen, in dem Sie bei G*Power in das Feld „Total sample size“ den Wert der Fehlerfreiheitsgrade für den Effekt eintragen (144). Nun liefert

G*Power-Ergänzungen

Rasch, Frieze, Hofmann & Naumann (2014). *Quantitative Methoden. Band 2* (4. Auflage). Heidelberg: Springer.

auch G*Power eine empirische Teststärke von gerundet 67,2% (siehe SPSS Ergänzungen zu Kapitel 6).

Auf der Basis der Anzahl der Versuchspersonen (regulärer G*Power-Weg):

The screenshot shows the G*Power software interface. The 'Test family' is set to 'F tests' and the 'Statistical test' is 'ANOVA: Fixed effects, special, main effects and interactions'. The 'Type of power analysis' is 'Post hoc: Compute achieved power – given α , sample size, and effect size'. The 'Input Parameters' section includes: Effect size $f = 0.2017410$, α err prob = 0.05, Total sample size = 150, Numerator df = 1, and Number of groups = 6. The 'Output Parameters' section includes: Noncentrality parameter $\lambda = 6.1049147$, Critical F = 3.9068490, Denominator df = 144, and Power ($1 - \beta$ err prob) = 0.6894726.

Auf der Basis der Fehlerfreiheitsgrade (SPSS-Weg für beobachtete Trennschärfe):

The screenshot shows the G*Power software interface. The 'Test family' is set to 'F tests' and the 'Statistical test' is 'ANOVA: Fixed effects, special, main effects and interactions'. The 'Type of power analysis' is 'Post hoc: Compute achieved power – given α , sample size, and effect size'. The 'Input Parameters' section includes: Effect size $f = 0.2017410$, α err prob = 0.05, Total sample size = 144, Numerator df = 1, and Number of groups = 6. The 'Output Parameters' section includes: Noncentrality parameter $\lambda = 5.8607181$, Critical F = 3.9097293, Denominator df = 138, and Power ($1 - \beta$ err prob) = 0.6715040.

Um f mit Hilfe des partiellen Eta-Quadrats zu berechnen, entnehmen wir den entsprechenden Wert dem SPSS Output und tragen ihn bei G*Power ein. Für den Faktor Geschlecht ist dies im Beispiel 0,039, was zu einer fast identischen Schätzung von f führt wie der Weg über die Varianzen.

G*Power-Ergänzungen

Rasch, Frieze, Hofmann & Naumann (2014). *Quantitative Methoden. Band 2* (4. Auflage). Heidelberg: Springer.

From variances
Variance explained by special effect: 0
Error variance: 0

Direct
Partial η^2 : 0.039

Effect size f: 0.2014515

Buttons: Calculate, Calculate and transfer to main window, Close

Diesen Wert kopieren wir in das Hauptfenster und berechnen die empirische Teststärke. Auch diese Berechnung führt natürlich zu einem nahezu identischen Ergebnis einer Teststärke von 69% wie beim zuerst gezeigten Weg über Varianzen.

Test family: F tests
Statistical test: ANOVA: Fixed effects, special, main effects and interactions

Type of power analysis: Post hoc: Compute achieved power - given α , sample size, and effect size

Input Parameters
Determine =>
Effect size f: 0.2014515
 α err prob: 0.05
Total sample size: 150
Numerator df: 1
Number of groups: 6

Output Parameters
Noncentrality parameter λ : 6.0874060
Critical F: 3.9068490
Denominator df: 144
Power ($1-\beta$ err prob): 0.6882282

Literatur

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