

Kapitel 7: Varianzanalyse mit Messwiederholung

7.1 Einfaktorielle Varianzanalyse mit Messwiederholung	1
Berechnen der Teststärke a priori bzw. Stichprobenumfangsplanung	2
Teststärkebestimmung a posteriori	5
Berechnen der Effektgröße f aus empirischen Daten und Bestimmung der zugehörigen beobachteten Teststärke	6
7.2 Zweifaktorielle Varianzanalyse mit Messwiederholung auf einem Faktor	9
Berechnen der Teststärke a priori bzw. Stichprobenumfangsplanung	9
Teststärkebestimmung a posteriori	12
Berechnen der Effektgröße f aus empirischen Daten und Bestimmen der zugehörigen beobachteten Teststärke	13
7.3 Zweifaktorielle Varianzanalyse mit Messwiederholung auf beiden Faktoren	14
Berechnen der Teststärke a priori bzw. Stichprobenumfangsplanung	14
Teststärkebestimmung a posteriori	16
Berechnen der Effektgröße f aus empirischen Daten und bestimmen der zugehörigen beobachteten Teststärke	24
Literatur	25

7.1 Einfaktorielle Varianzanalyse mit Messwiederholung

Nach dem Starten von G*Power müssen Sie für einfaktorielle Varianzanalysen mit Messwiederholung zunächst unter Test family „F-Tests“ auswählen, unter Statistical test „ANOVA: Repeated measures, within factors“ und unter Type of power analysis „A priori“. Sie erhalten nun folgendes Eingabefenster:

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

G*Power-Ergänzungen (SPSS)

Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

The screenshot shows the G*Power software interface. At the top, 'Test family' is set to 'F tests' and 'Statistical test' is 'ANOVA: Repeated measures, within factors'. 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)' (0.95), 'Number of groups' (2), 'Repetitions' (4), 'Corr among rep measures' (0.5), and 'Nonsphericity correction epsilon' (1). The 'Output Parameters' section shows: 'Noncentrality parameter lambda', 'Critical F', 'Numerator df', 'Denominator df', 'Total sample size', and 'Actual power', all with question marks.

Im Fall der Messwiederholung ist es nicht möglich, einheitliche Konventionen für Effektstärken zu definieren, weil die Stärke der Abhängigkeit der wiederholten Messungen einen großen Einfluss auf die empirisch resultierende Effektstärke hat. In Bezug auf die Berechnung der Teststärke heißt das: Die Höhe der Korrelation zwischen den Stufen der Messwiederholung bestimmt, ob ein Populationseffekt mit einer hohen oder einer niedrigen Wahrscheinlichkeit gefunden werden kann. Zusätzlich hat auch das spezielle experimentelle Design (Anzahl der messwiederholten Faktoren, Anzahl der Stufen) einen Einfluss auf die Effektschätzung und die Teststärke. Dies führt dazu, dass Effektgrößen aus einer messwiederholten Studie nicht direkt mit Effekten aus nicht-messwiederholten Untersuchungen verglichen werden können. Auch der Vergleich zu anderen messwiederholten Studien ist nur dann möglich, wenn das experimentelle Design und die Höhe der Korrelationen zwischen den Stufen ähnlich sind.

Um dennoch den Rückgriff auf Konventionen zu ermöglichen, schlagen wir vor, sich an den Konventionen von Cohen (1988) zu orientieren, die für unabhängige, nicht-messwiederholte Daten formuliert wurden. G*Power verwendet als Effektstärkemaß das aus vorherigen Kapiteln bekannte f .

Die bekannten Konventionen für unabhängige Messungen lauten:

Kleiner Effekt:	$\Omega^2 = 0,01$	→	$f^2 = 0,01$	→	$f = 0,1$
Mittlerer Effekt:	$\Omega^2 = 0,06$	→	$f^2 = 0,0625$	→	$f = 0,25$
Großer Effekt:	$\Omega^2 = 0,14$	→	$f^2 = 0,16$	→	$f = 0,4^1$

Berechnen der Teststärke a priori bzw. Stichprobenumfangsplanung

Als Beispiel dient die dreifach wiederholte Durchführung eines motorischen Leistungstests. Der messwiederholte Faktor hat also drei Stufen ($p = 3$). Das Signifikanzniveau beträgt 5%. Da es keine

¹ Abweichungen bei der Umrechnung zwischen Omega und f entstehen durch Rundungsfehler.

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

G*Power-Ergänzungen (SPSS)

Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

Vorarbeiten gibt, orientiert sich der Studienleiter an den Konventionen für unabhängige Stichproben und legt einen mittleren Effekt von $f = 0,25$ ($\Omega^2 = 0,06$) als inhaltlich relevante Effektgröße fest. Diese soll in der Studie mit einer Wahrscheinlichkeit von mindestens 80% gefunden werden. Wie viele Versuchspersonen werden benötigt? Zur Beantwortung dieser Frage ist es zusätzlich notwendig, eine Abschätzung über die erwartete Höhe der Korrelation zwischen den einzelnen Messzeitpunkten vorzunehmen (siehe Kap. 7.1.8). Da es sich bei motorischen Leistungen um eine Fähigkeit handelt, die sich zwischen Versuchspersonen stark unterscheidet, innerhalb von Versuchspersonen aber relativ stabil ist, und weil die Messungen zeitlich relativ eng aufeinander folgen, nimmt der Forscher eine Korrelation von $r = 0,4$ an.

Zusätzlich zu diesen Werten verlangt G*Power eine Angabe über die Anzahl der Gruppen, die an der Studie teilnehmen. Das ist im Beispiel nur eine Gruppe (die drei Mal die Aufgabe wiederholt). Darüber hinaus ist es möglich, für eine Verletzung der Sphärizitätsannahme zu korrigieren. Diese Option ignorieren wir zunächst und belassen die Einstellung auf dem Standardwert 1.

Das Durchführen dieser Analyse liefert das Ergebnis, dass 33 Personen notwendig sind, um eine Teststärke von 80% zu erzielen. Die geringfügige Abweichung zwischen dem hier berechneten Stichprobenumfang $N = 33$ und dem in Kapitel 7.1.8 berechneten Stichprobenumfang von $N = 31$ ist zum einen auf geringfügige Veränderungen bei der Umrechnung von f und Ω^2 und zum anderen auf unterschiedliche Bestimmungsweisen des Nonzentralitätsparameters in den Tabellen im Buch und bei G*Power zurückzuführen.

The screenshot shows the G*Power software interface. The 'Test family' is set to 'F tests' and the 'Statistical test' is 'ANOVA: Repeated measures, within factors'. 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), 'Number of groups' (1), 'Number of measurements' (3), 'Corr among rep measures' (0.4), and 'Nonsphericity correction epsilon' (1). The 'Output Parameters' section includes: 'Noncentrality parameter lambda' (10.3125000), 'Critical F' (3.1404376), 'Numerator df' (2.0000000), 'Denominator df' (64.0000000), 'Total sample size' (33), and 'Actual power' (0.8086689).

Wie viele Versuchspersonen werden benötigt, wenn der Studienleiter anstatt einer Korrelation von $r = 0,4$ eine Korrelation von $r = 0,8$ annimmt?

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

G*Power-Ergänzungen (SPSS)

Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

The screenshot shows the G*Power interface for an ANOVA test. The 'Test family' is 'F tests' and the 'Statistical test' is 'ANOVA: Repeated measures, within factors'. The 'Type of power analysis' is 'A priori: Compute required sample size - given alpha, power, and effect size'. The 'Input Parameters' are: Effect size f = 0.25, alpha err prob = 0.05, Power (1-beta err prob) = 0.80, Number of groups = 1, Number of measurements = 3, Corr among rep measures = 0.8, and Nonsphericity correction epsilon = 1. The 'Output Parameters' are: Noncentrality parameter lambda = 11.2500000, Critical F = 3.4433568, Numerator df = 2.0000000, Denominator df = 22.0000000, Total sample size = 12, and Actual power = 0.8064823.

Wenn die Korrelation zwischen den messwiederholten Stufen $r = 0,8$ betragen würden, wären nur zwölf Versuchspersonen notwendig um mit einer Wahrscheinlichkeit von mindestens 80% ein signifikantes Ergebnis zu erhalten, falls ein mittlerer Effekt von $f = 0,25$ existiert.

Und welcher Stichprobenumfang ergibt sich, wenn keine Korrelation ($r = 0$) zwischen den Faktorstufen besteht, also die Messungen statistisch voneinander unabhängig sind?

The screenshot shows the G*Power interface for an ANOVA test. The 'Test family' is 'F tests' and the 'Statistical test' is 'ANOVA: Repeated measures, within factors'. The 'Type of power analysis' is 'A priori: Compute required sample size - given alpha, power, and effect size'. The 'Input Parameters' are: Effect size f = 0.25, alpha err prob = 0.05, Power (1-beta err prob) = 0.80, Number of groups = 1, Number of measurements = 3, Corr among rep measures = 0, and Nonsphericity correction epsilon = 1. The 'Output Parameters' are: Noncentrality parameter lambda = 9.9375000, Critical F = 3.0837059, Numerator df = 2.0000000, Denominator df = 104, Total sample size = 53, and Actual power = 0.8008377.

Im Fall einer Korrelation von $r = 0$ wären 53 Versuchspersonen notwendig gewesen, um einen mittleren Effekt von $f = 0,25$ mit einer Wahrscheinlichkeit von mindestens 80% zu finden (d.h. ein signifikantes Ergebnis zu erhalten).

Da eine Varianzanalyse mit Messwiederholung, aber unkorrelierten Messwertereihen identisch ist mit einer Varianzanalyse für unabhängige Stichproben, erhalten wir ein identisches Ergebnis, wenn wir den Stichprobenumfang für diese berechnen. Gehen Sie unter „F tests“ zu „ANOVA: Fixed
Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

G*Power-Ergänzungen (SPSS)

Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

effects, omnibus, one-way“. Unter Angabe der Effektstärke, des α -Niveaus von 5%, der gewünschter Power von 80% und drei Gruppen errechnet G*Power einen optimalen Stichprobenumfang von $N = 159$. Dies entspricht exakt 53 Versuchspersonen pro Gruppe. Der einzige Unterschied ist also, dass bei einer Messwiederholung mit $r = 0$ jede Versuchsperson drei Werte abgibt, während ohne Messwiederholung jede Person nur einen Messwert erzeugt.

The screenshot shows the G*Power software interface. The 'Test family' is set to 'F tests' and the 'Statistical test' is 'ANOVA: Fixed effects, omnibus, one-way'. The 'Type of power analysis' is 'A priori: Compute required sample size - given α , power, and effect size'. The 'Input Parameters' section includes 'Determine =>', 'Effect size f' (0.25), ' α err prob' (0.05), 'Power (1- β err prob)' (0.80), and 'Number of groups' (3). The 'Output Parameters' section includes 'Noncentrality parameter λ ' (9.9375000), 'Critical F' (3.0540042), 'Numerator df' (2), 'Denominator df' (156), 'Total sample size' (159), and 'Actual power' (0.8048873).

Achtung: Bei einer Verletzung der Sphärizitätsannahme reduziert sich die Teststärke einer messwiederholten Varianzanalyse. Ist eine derartige Verletzung im Vorhinein absehbar, kann G*Power dafür mit Hilfe des Korrekturparameters ϵ („Nonsphericity correction“) kontrollieren und entsprechend mehr Versuchspersonen einplanen.

Teststärkebestimmung a posteriori

Eine Teststärkebestimmung a posteriori ist sinnvoll, wenn vor einem Experiment keine Stichprobenumfangsplanung durchgeführt wurde und ein nicht signifikantes Ergebnis vorliegt. Wir veranschaulichen das Vorgehen wieder für das Beispiel der dreifach wiederholten Messung der motorischen Fertigkeiten aus Kapitel 7.1. Stellen Sie dafür G*Power auf eine post hoc Analyse um.

The screenshot shows the G*Power software interface. The 'Test family' is set to 'F tests' and the 'Statistical test' is 'ANOVA: Repeated measures, within factors'. The 'Type of power analysis' is 'Post hoc: Compute achieved power - given α , sample size, and effect size'.

In dem Beispiel haben insgesamt 36 Versuchspersonen teilgenommen. Wie hoch war die Wahrscheinlichkeit, einen mittleren Effekt für unabhängige Stichproben von $f = 0,25$ zu entdecken? Zunächst ist es notwendig, die mittlere Korrelation zwischen den Faktorstufen zu bestimmen. Dazu verwenden wir das Programm SPSS. In dem Datensatz „Messwiederholung.sav“ finden Sie die Variablen der ersten, zweiten und dritten Messung der Studie unter dem Namen „Messung1“,
Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

G*Power-Ergänzungen (SPSS)

Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

„Messung2“ und „Messung3“ (vgl. SPSS Ergänzungen zu diesem Kapitel). Zur Berechnung der Korrelationen folgen Sie den Anleitungen aus den SPSS-Ergänzungen zu Kapitel 4. Sie erhalten folgenden SPSS Output:

Korrelationen

		Anzahl getippter Sequenzen bei Messung 1	Anzahl getippter Sequenzen bei Messung 2	Anzahl getippter Sequenzen bei Messung 3
Anzahl getippter Sequenzen bei Messung 1	Korrelation nach Pearson	1	,661 **	,649 **
	Signifikanz (2-seitig)		,000	,000
	N	36	36	36
Anzahl getippter Sequenzen bei Messung 2	Korrelation nach Pearson	,661 **	1	,799 **
	Signifikanz (2-seitig)	,000		,000
	N	36	36	36
Anzahl getippter Sequenzen bei Messung 3	Korrelation nach Pearson	,649 **	,799 **	1
	Signifikanz (2-seitig)	,000	,000	
	N	36	36	36

** . Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Die Errechnung des Mittelwerts der drei Korrelationen erfolgt über Fishers-Z Transformation und ist in Kapitel 7.1.7 beschrieben. Es ergibt sich eine mittlere Korrelation von $r = 0,71$. G*Power berechnet aus diesen Angaben eine Teststärke von mehr als 99%. Die Wahrscheinlichkeit, in dieser Studie einen mittleren Effekt von $f = 0,25$ oder größer bei einer vorliegenden mittleren Korrelation zwischen den Faktorstufen von $r = 0,71$ zu finden, falls er tatsächlich existiert, war größer als 99%.

Input Parameters		Output Parameters	
Determine =>	Effect size f	Noncentrality parameter λ	23.2758621
	α err prob	Critical F	3.1276756
	Total sample size	Numerator df	2.0000000
	Number of groups	Denominator df	70.0000000
	Number of measurements	Power (1- β err prob)	0.9922340
	Corr among rep measures		
	Nonsphericity correction ϵ		

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

Anmerkung: Wir beziehen uns hier auf die Stichprobenebene und lassen die komplizierteren Schätzungen für die Populationsebene außer Acht. SPSS und G*Power beschränken sich ebenfalls

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

G*Power-Ergänzungen (SPSS)

Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

auf Effektstärken auf der Stichprobengröße. Zu beachten ist allerdings, dass die Schätzung auf Stichprobenebene die Effekte auf Populationsebene teilweise deutlich überschätzen können.

G*Power arbeitet auch bei einfaktoriellen Varianzanalysen mit Messwiederholung mit der Effektgröße f . Um eine empirische Effektgröße zu bestimmen, haben Sie – analog zur Berechnung für Varianzanalysen ohne Messwiederholung – die Möglichkeit, dies über Varianzen oder über die von SPSS gelieferte Effektgröße partielles Eta-Quadrat zu tun (siehe G*Power-Erläuterungen zu Kapitel 6).

Wir verwenden als Beispiel wie bereits in den SPSS-Erläuterungen und oben die wiederholte Messung eines motorischen Tests. Der folgenden Tabelle des SPSS-Outputs können Sie die relevanten Angaben entnehmen.

Tests der Innersubjekteffekte

Maß: MEASURE_1

Quelle		Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat	Dezent. Parameter	Beobachtete Schärfe ^a
Faktor1	Sphärizität angenommen	52,056	2	26,028	3,518	,035	,091	7,035	,638
	Greenhouse-Geisser	52,056	1,826	28,512	3,518	,039	,091	6,422	,609
	Huynh-Feldt	52,056	1,921	27,099	3,518	,037	,091	6,757	,625
	Untergrenze	52,056	1,000	52,056	3,518	,069	,091	3,518	,446
Fehler(Faktor1)	Sphärizität angenommen	517,944	70	7,399					
	Greenhouse-Geisser	517,944	63,901	8,105					
	Huynh-Feldt	517,944	67,234	7,704					
	Untergrenze	517,944	35,000	14,798					

a. Unter Verwendung von Alpha = ,05 berechnet

Um den empirischen Effekt zu bestimmen, klicken Sie in G*Power auf „Determine“ und geben die Varianz des Faktors ein (Quadratsumme vom Typ III) sowie die Fehlervarianz. G*Power ermittelt einen Effekt von $f = 0,32$. Zum gleichen Ergebnis (abgesehen von Rundungsungenauigkeiten) kommen Sie, wenn Sie bei G*Power unter der Möglichkeit „Direct“ das partielle Eta Quadrat des Faktors angeben (0,091).

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

G*Power-Ergänzungen (SPSS)

Rasch, Frieze, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

Zur Bestimmung der beobachteten Teststärke eines aus vorliegenden Daten geschätzten Effekts empfehlen wir die Ausgabe der „Beobachteten Trennschärfe“ in SPSS. Eine a posteriori Berechnung der Teststärke für einen empirischen Effekt einer messwiederholten Varianzanalyse ist in G*Power nicht möglich, da in der empirischen Effektstärke die Korrelation zwischen den Messzeitpunkten bereits enthalten ist.

Achtung: Obwohl bei G*Power die Eingabeoption für das partielle Eta Quadrat besteht, ist es nicht zulässig, hier die empirische Effektstärke einer messwiederholten Varianzanalyse einzutragen und die Teststärke zu berechnen. G*Power geht davon aus, dass das partielle Eta-Quadrat keine Abhängigkeit der Daten enthält, und kommt somit zu einem falschen Ergebnis für die Teststärke. Sie können es gerne einmal ausprobieren.

Insgesamt beinhaltet die beobachtete Teststärke allerdings wenig zusätzliche Information (vgl. SPSS-Ergänzungen zu den Kapiteln 5 und 6): Ist ein Ergebnis signifikant, so ist es unerheblich, mit welcher Wahrscheinlichkeit der empirisch geschätzte Effekt gefunden werden konnte, denn er wurde ja soeben „entdeckt“. Im Fall der Nichtsignifikanz wird die beobachtete Teststärke für den empirisch geschätzten Effekt immer unbefriedigende Werte annehmen. Weit aussagekräftiger ist es, eine a posteriori Teststärkenanalyse mit einem als inhaltlich relevant erachteten Effekt durchzuführen (siehe vorheriger Abschnitt). In diesem Sinne erlaubt G*Power nur die Eingabe von Effektstärken, die keine Abhängigkeit der Daten enthalten, d.h. Effektstärkenmaße für unabhängige Stichproben wie z.B. die Konventionen von Cohen (1988).

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Frieze, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

7.2 Zweifaktorielle Varianzanalyse mit Messwiederholung auf einem Faktor

In mehrfaktoriellen Varianzanalysen ist es notwendig, für alle inhaltlich relevanten Effekte eine eigene Stichprobenumfangsplanung vorzunehmen. Der größte berechnete Stichprobenumfang bestimmt die notwendige Anzahl Versuchspersonen.

Auch in diesem Abschnitt greifen wir auf das Beispiel der dreifachen Durchführung eines motorischen Leistungstests zurück. Zusätzlich soll der Einfluss des nicht-messwiederholten Faktors „Geschlechts“ untersucht werden.

Berechnen der Teststärke a priori bzw. Stichprobenumfangsplanung

In dem Beispiel seien alle drei Effekte, der Haupteffekt A „Geschlecht“, der Haupteffekt B „Testwiederholung“ sowie ihre Wechselwirkung $A \times B$ inhaltlich relevant. Der Studienleiter legt vor der Studie fest, dass für alle drei Effektarten ein mittlerer Effekt (für unabhängige Stichproben) von $f = 0,25$ bei einem Signifikanzniveau von 5% mit einer Wahrscheinlichkeit von mindestens 80% gefunden werden soll. Die mittlere Korrelation zwischen den drei wiederholten Messzeitpunkten schätzt er auf $r = 0,4$. Wie viele Versuchspersonen muss er erheben?

Die Berechnung für den nicht messwiederholten Faktor Geschlecht führt über die Option „ANOVA: Repeated measures, between factors“. Dort machen Sie die relevanten Eingaben. G*Power errechnet einen Bedarf von 78 Versuchspersonen (38 Frauen, 38 Männer).

Test family		Statistical test	
F tests		ANOVA: Repeated measures, between factors	
Type of power analysis			
A priori: Compute required sample size - given α , power, and effect size			
Input Parameters		Output Parameters	
Determine =>		Noncentrality parameter λ	8.1250000
Effect size f	0.25	Critical F	3.9667598
α err prob	0.05	Numerator df	1.0000000
Power ($1 - \beta$ err prob)	0.80	Denominator df	76.0000000
Number of groups	2	Total sample size	78
Number of measurements	3	Actual power	0.8035267
Corr among rep measures	.4		

Für die Stichprobenumfangsplanung für den Faktor „Testwiederholung“ wechseln wir zum Menüpunkt „ANOVA: Repeated measures, within factors“ (siehe vorheriger Abschnitt). Im Gegensatz zum vorherigen Abschnitt müssen wir nun vermerken, dass es insgesamt zwei Gruppen gibt (Frauen und Männer). G*Power ermittelt einen Bedarf von lediglich 34 Personen, deutlich weniger als für den nicht messwiederholten Faktor Geschlecht.

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

G*Power-Ergänzungen (SPSS)

Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

The screenshot shows the G*Power software interface for an ANOVA: Repeated measures, within factors test. The 'Test family' is set to 'F tests' and the 'Statistical test' is 'ANOVA: Repeated measures, within factors'. The 'Type of power analysis' is 'A priori: Compute required sample size - given alpha, power, and effect size'. The 'Input Parameters' section includes: 'Determine =>' (selected), 'Effect size f' (0.25), 'alpha err prob' (0.05), 'Power (1-beta err prob)' (0.80), 'Number of groups' (2), 'Number of measurements' (3), 'Corr among rep measures' (0.4), and 'Nonsphericity correction epsilon' (1). The 'Output Parameters' section includes: 'Noncentrality parameter lambda' (10.6250000), 'Critical F' (3.1404376), 'Numerator df' (2.0000000), 'Denominator df' (64.0000000), 'Total sample size' (34), and 'Actual power' (0.8208620).

Schließlich erfolgt die Stichprobenumfangsplanung für die Wechselwirkung der Faktoren Geschlecht und Testwiederholung. Dafür wechseln wir in den Menüpunkt „ANOVA: Repeated measures, within-between interaction“. Auch hier machen wir die relevanten Angaben und erzielen ein identisches Ergebnis wie für den messwiederholten Faktor Testwiederholung. Dies muss aber nicht so sein. In Fällen, in denen der nicht messwiederholte Faktor mehr als zwei Gruppen hat, erhöht sich der Bedarf an Versuchspersonen im Vergleich zum messwiederholten Faktor (probieren Sie es aus).

The screenshot shows the G*Power software interface for an ANOVA: Repeated measures, within-between interaction test. The 'Test family' is set to 'F tests' and the 'Statistical test' is 'ANOVA: Repeated measures, within-between interaction'. The 'Type of power analysis' is 'A priori: Compute required sample size - given alpha, power, and effect size'. The 'Input Parameters' section includes: 'Determine =>' (selected), 'Effect size f' (0.25), 'alpha err prob' (0.05), 'Power (1-beta err prob)' (0.80), 'Number of groups' (2), 'Number of measurements' (3), 'Corr among rep measures' (0.4), and 'Nonsphericity correction epsilon' (1). The 'Output Parameters' section includes: 'Noncentrality parameter lambda' (10.6250000), 'Critical F' (3.1404376), 'Numerator df' (2.0000000), 'Denominator df' (64.0000000), 'Total sample size' (34), and 'Actual power' (0.8208620).

Um in der Beispieluntersuchung einen mittleren Effekt zu finden, reichen für den Faktor Testwiederholung und die Interaktion zwischen Testwiederholung und Geschlecht 34 Versuchspersonen. Um allerdings auch für den Faktor Geschlecht einen mittleren Effekt mit ausreichender Wahrscheinlichkeit zu finden, sollten mindestens 78 Personen untersucht werden.

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

G*Power-Ergänzungen (SPSS)

Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

Teststärkebestimmung a posteriori

Für die Teststärkebestimmung a posteriori in einer zweifaktoriellen Varianzanalyse mit Messwiederholung auf einem Faktor ist die mittlere Korrelation zwischen den Stufen des messwiederholten Faktors von Bedeutung. Ihre Abschätzung wurde bereits im Abschnitt zur einfaktoriellen Varianzanalyse mit Messwiederholung vorgestellt. Für den Beispieldatensatz beläuft sie sich auf $r = 0,71$. Diese Korrelation verändert sich nicht, wenn zusätzlich Faktoren ohne Messwiederholung aufgenommen werden.

Für die a posteriori Berechnung der Teststärke für den messwiederholten Faktor und die Wechselwirkung können Sie sich an den Beschreibungen in dem Abschnitt der einfaktoriellen Varianzanalyse orientieren (siehe oben). Ein mittlerer Effekt für unabhängige Stichproben von $f = 0,25$ wird als inhaltlich relevant festgelegt. Für sowohl den messwiederholten Faktor als auch die Interaktion ermittelt G*Power eine sehr hohe Teststärke von über 99% bei der vorhandenen Stichprobengröße von 36 Personen (18 Frauen, 18 Männer).

Input Parameters		Output Parameters	
Determine =>	Effect size f	Noncentrality parameter λ	23.2758621
	α err prob	Critical F	3.1316720
	Total sample size	Numerator df	2.0000000
	Number of groups	Denominator df	68.0000000
	Number of measurements	Power (1- β err prob)	0.9921694
	Corr among rep measures		
	Nonsphericity correction ϵ		

Zuletzt die Teststärkeberechnung für den nicht messwiederholten Faktor Geschlecht. Auch hier geht die mittlere Korrelation zwischen den Stufen des messwiederholten Faktors in die Teststärkeberechnung mit ein. Allerdings *verkleinert* eine hohe Korrelation die Teststärke des nicht messwiederholten Faktors (probieren Sie es aus!). G*Power errechnet für den Faktor Geschlecht nur eine Teststärke von etwa 37%, um einen mittleren Effekt bei einer gegebenen Korrelation von $r = 0,71$ mit insgesamt $N = 36$ Versuchspersonen zu finden.

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Frieze, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

G*Power-Ergänzungen (SPSS)

Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

The screenshot shows the G*Power software interface. The 'Test family' is set to 'F tests' and the 'Statistical test' is 'ANOVA: Repeated measures, between factors'. 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), 'Effect size f' (0.25), 'alpha err prob' (0.05), 'Total sample size' (36), 'Number of groups' (2), 'Number of measurements' (3), and 'Corr among rep measures' (.71). The 'Output Parameters' section includes: 'Noncentrality parameter lambda' (2.7892562), 'Critical F' (4.1300177), 'Numerator df' (1.0000000), 'Denominator df' (34.0000000), and 'Power (1-beta err prob)' (0.3682317).

Berechnen der Effektgröße f aus empirischen Daten und Bestimmen der zugehörigen beobachteten Teststärke

Die Formeln für die von SPSS verwendete Schätzung der Effektstärke aus den Daten wurden bereits im Abschnitt der einfaktoriellen ANOVA mit Messwiederholung angegeben:

$$f_{S(\text{abhängig})}^2 = \frac{F \cdot df_A}{df_{A \times Vpn}} \quad \rightarrow \quad \eta_p^2 = \frac{f_S^2}{1 + f_S^2}$$

Für die Berechnung der beobachteten Teststärke für einen empirisch gefundenen Effekt verweisen wir auf die Option „Beobachtete Trennschärfe“ in SPSS.

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

7.3 Zweifaktorielle Varianzanalyse mit Messwiederholung auf beiden Faktoren

Die Berechnung der Teststärke für Varianzanalysen mit Messwiederholung auf mehreren Stufen ist in G*Power Version 3 leider noch nicht möglich. Die Autoren der Software planen diese Option in einer neueren Version zu integrieren, der Erscheinungstermin ist allerdings noch unklar.

Wir stellen hier übergangsweise eine alternative Berechnungsart vor, mit der die Rechnung trotzdem mit der aktuellen Version von G*Power durchgeführt werden kann. Dies geschieht über einen „Trick“: Wir passen das Effektstärkenmass f vor der Eingabe in G*Power so an, dass die bestehenden Eingabeoptionen in der Software ausreichen, um das korrekte Ergebnis zu erhalten. Zur Unterscheidung bezeichnen wir das einzugebende, angepasste Effektstärkenmass als f' .

Für alle folgenden Berechnungen verwenden wir den Menüpunkt „ANOVA: Repeated measures, within factors“. Der Eintrag „Number of Groups“ steht auf 1, da kein nicht-messwiederholter Gruppenfaktor vorliegt.

Als Beispiel dient wiederum die Frage, wie sich eine dreifach wiederholte Durchführung eines motorischen Tests auf die Performanz auswirkt. Zusätzlich möchte der Studienleiter untersuchen, ob ein bestimmtes Medikament Auswirkungen auf die motorische Leistung hat. Die Versuchspersonen kommen zweimal in das Labor, an einem Tag bekommen sie das Medikament und führen den motorischen Test drei Mal durch, an einem anderen Tag bekommen sie ein Scheinmedikament (Placebo), und führen den Test ebenfalls drei Mal durch. Es ergibt sich ein messwiederholter Faktor A „Medikament“ mit zwei Stufen („Medikament“ und „Placebo“), sowie der bekannte messwiederholte Faktor B „Testwiederholung“ mit drei Stufen.

Berechnen der Teststärke a priori bzw. Stichprobenumfangsplanung

Wie bei allen mehrfaktoriellen Varianzanalysen muss der optimale Stichprobenumfang getrennt für alle inhaltlich relevanten Effektarten bestimmt werden. Im Beispiel sind also sowohl die beiden Haupteffekte als auch die Wechselwirkung relevant. Die Teststärke, einen mittleren Effekt (für unabhängige Stichproben) von $f = 0,25$ bei einem Signifikanzniveau von 5% zu finden, falls dieser existiert, soll mindestens 80% betragen.

Wir berechnen zunächst den optimalen Stichprobenumfang für den messwiederholten Faktor A „Medikament“. Er hat zwei Stufen ($p = 2$). Entscheidend für die Teststärke dieses Faktors ist die Korrelation zwischen den mittleren Werten in dem Leistungstest am Tag des Medikaments und am Tag der Placebogabe. Da diese Messungen zeitlich um mehrere Tage (oder Wochen) auseinander liegen, ist es ratsam, eine etwas geringere Korrelation zwischen diesen beiden Messungen für die Stichprobenumfangsplanung anzunehmen, z.B. $r = 0,3$.

Die fehlende Eingabeoption besteht in den Stufen des zweiten messwiederholten Faktors „Testwiederholung“. Diesen rechnen wir nach folgender Formel in das Effektstärkenmass mit ein:

$$f' = \sqrt{q} \cdot f = \sqrt{3} \cdot 0,25 = 0,4330$$

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

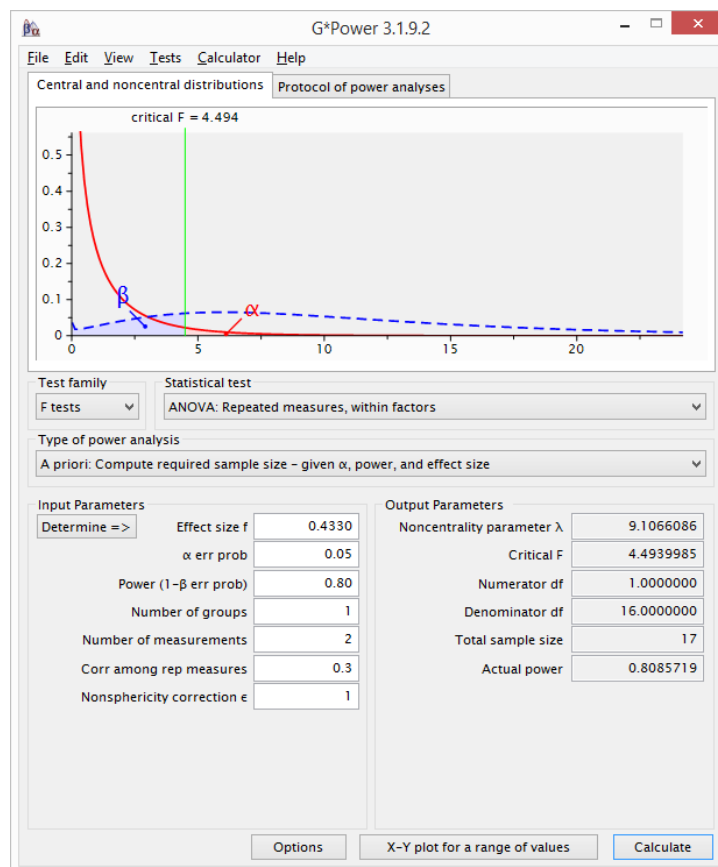
Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

G*Power-Ergänzungen (SPSS)

Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

Wir geben nun das angepasste Effektstärkenmass f' in G*Power an, zusammen mit den anderen Angaben.



Um für den Haupteffekt A („Medikament“) einen mittleren Effekt bei einer geschätzten Korrelation zwischen den Stufen von $r = 0,3$ mit $> 80\%$ iger Wahrscheinlichkeit zu finden, sind nur insgesamt $N = 17$ Personen notwendig. Diese niedrige Zahl ist durch die zwei messwiederholten Faktoren zu erklären. Jede dieser Personen gibt sechs Messwerte ab, drei unter Einfluss des Medikaments, drei unter Placebo.

Wie viele Personen sind für den messwiederholten Faktor B „Testwiederholung“ unter den gleichen Voraussetzungen notwendig? Für die Anpassung des Effektstärkemaßes f' beziehen wir nun die Anzahl Stufen des Faktors „Medikament“ mit ein.

$$f' = \sqrt{p} \cdot f = \sqrt{2} \cdot 0.25 = 0.3536$$

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

G*Power-Ergänzungen (SPSS)

Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

The screenshot shows the G*Power software interface. The 'Test family' is set to 'F tests' and the 'Statistical test' is 'ANOVA: Repeated measures, within factors'. 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), 'Effect size f' (0.3536), 'alpha err prob' (0.05), 'Power (1-beta err prob)' (0.80), 'Number of groups' (1), 'Number of measurements' (3), 'Corr among rep measures' (0.3), and 'Nonsphericity correction epsilon' (1). The 'Output Parameters' section includes: 'Noncentrality parameter lambda' (10.7171109), 'Critical F' (3.2448184), 'Numerator df' (2.0000000), 'Denominator df' (38.0000000), 'Total sample size' (20), and 'Actual power' (0.8110323).

Die Anzahl $N = 20$ ist ausreichend, um für den Haupteffekt B „Testwiederholung“ einen mittleren Effekt (unabhängige Stichproben) bei einem Signifikanzniveau von 5% und einer angenommenen Korrelation zwischen den Faktorstufen von $r = 0,3$ mit einer Power von $> 80\%$ zu finden.

In unserem Beispiel entspricht die Berechnung des Stichprobenumfangs für die Wechselwirkung der vorgestellten Art der Berechnung des Haupteffekts B „Testwiederholung“ (aber siehe Anmerkung unten). Es sollten also in unserem Beispiel mindestens 20 Versuchspersonen untersucht werden, um eine ausreichende Teststärke für beide Haupteffekte sowie die Wechselwirkung sicherzustellen. Die wahre Teststärke hängt natürlich – wie in allen messwiederholten Analysen – von den tatsächlichen Korrelationen zwischen den Faktorstufen und einer etwaigen Verletzung der Zirkularitätsannahme ab. Insofern sind diese Ergebnisse mit größerer Vorsicht zu interpretieren als die Berechnungen für Analysen ohne Messwiederholung (siehe oben und G*Power Ergänzungen zu Kapitel 3).

Achtung: Die hier vorgestellte Berechnung des Stichprobenumfangs für die Wechselwirkung ist nur exakt für zwei messwiederholte Faktoren, von denen einer nur zwei Stufen hat. Es ist dabei sehr wichtig, die Stufenzahl des Faktors mit mehr als zwei Stufen in G*Power einzugeben, und die Effektstärke mit $\sqrt{2}$ anzupassen (siehe Vorgehen Faktor B). Haben beide Faktoren mehr als zwei Stufen, geht G*Power von anderen Freiheitsgraden aus. Dies führt zu einer leichten Überschätzung der Teststärke und damit zu einer Unterschätzung der benötigten Versuchspersonenanzahl.

Teststärkebestimmung a posteriori

Die mittlere Korrelation zwischen den messwiederholten Faktorstufen (zusammen mit einer potentiellen Verletzung der Zirkularitätsannahme) sind maßgeblich für die tatsächlich in einer Studie mit Messwiederholung erreichte Teststärke. Die mittlere Korrelation muss hier allerdings für die drei in einer zweifaktoriellen Varianzanalyse möglichen Effektarten (zwei Haupteffekte, eine Wechselwirkung) separat bestimmt werden. Hierzu verwenden wir das Programm SPSS, den Datensatz finden Sie in der Datei „Messwiederholung.sav“.

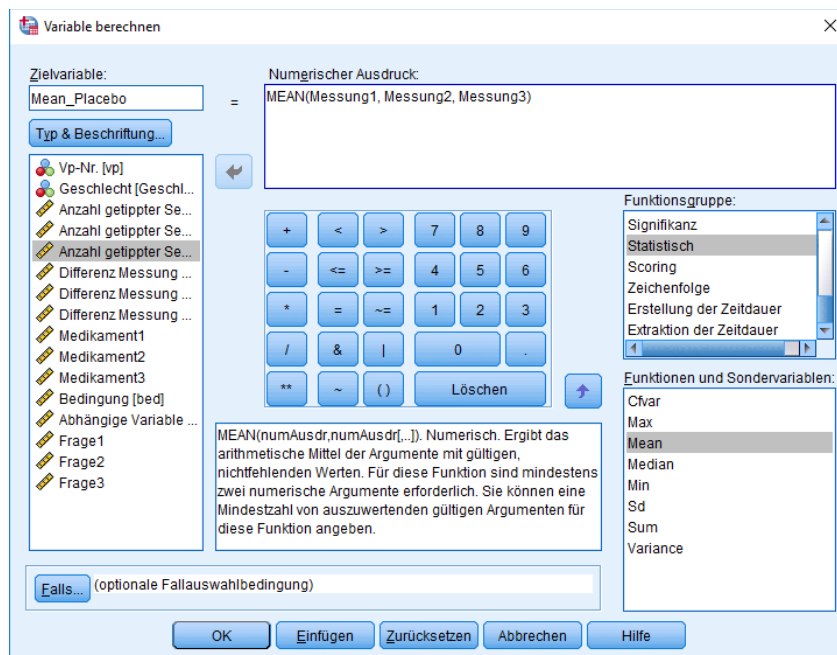
Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

G*Power-Ergänzungen (SPSS)

Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

Für den Faktor A „Medikament“ („Medikament“ vs. „Placebo“) ist die Korrelation zwischen den über die drei Messwiederholungen des motorischen Tests gemittelten Werten unter Placebo und unter dem Medikament entscheidend. Um diese zu berechnen, gehen Sie in SPSS auf „Transformieren“ → „Variable berechnen...“. Geben Sie einen neuen Namen für die Zielvariable in das Feld oben links ein, z.B. „Mean_Placebo“. Unter „Typ & Beschriftung...“ können Sie einen aussagekräftigeren Namen definieren, z.B. „Mittelwert der drei Messungen unter Placebo“. Suchen Sie unter der Funktionsgruppe „Statistisch“ die Funktion „Mean“ und bewegen Sie sie in das Feld „numerischer Ausdruck“. Hier geben Sie die Namen der drei Messungen unter Placebo durch Kommata getrennt ein. Drücken Sie anschließend auf OK. In dem SPSS Datenfenster sollte nun eine neue Variable hinzugekommen sein, die „Mean_Placebo“ heißt.



Führen Sie dieselbe Berechnung für die drei Messungen unter dem Einfluss des Medikaments durch und erstellen Sie eine neue Variable „Mean_Medikament“.

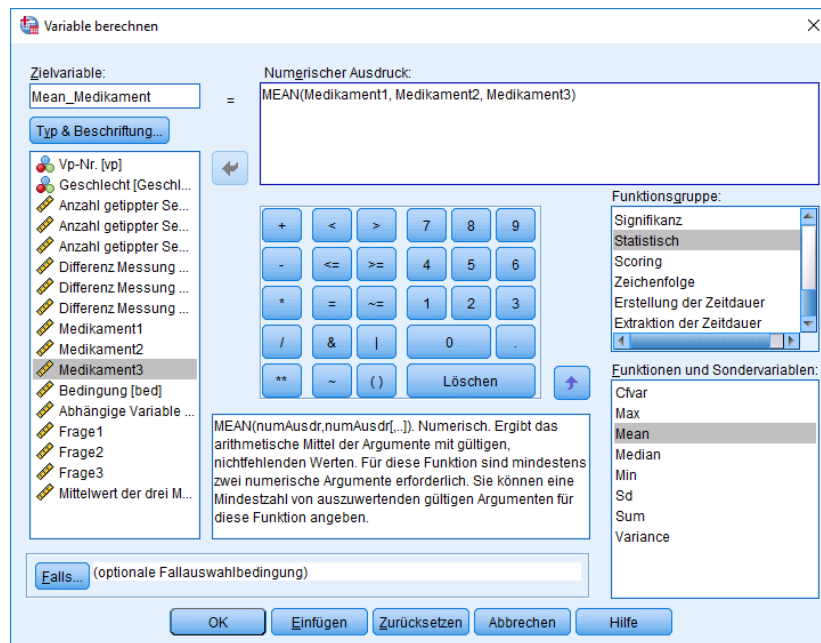
Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

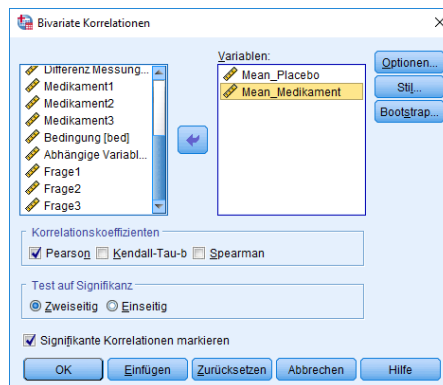
© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

G*Power-Ergänzungen (SPSS)

Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.



Berechnen Sie nun die Korrelation zwischen den beiden neu erstellten Variablen.



Korrelationen

		Mean_Placebo	Mean_Medikament
Mean_Placebo	Korrelation nach Pearson	1	,718**
	Signifikanz (2-seitig)		,000
	N	36	36
Mean_Medikament	Korrelation nach Pearson	,718**	1
	Signifikanz (2-seitig)	,000	
	N	36	36

** . Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Die Korrelation zwischen den mittleren Testwerten unter dem Einfluss des Medikaments und dem mittleren Testwert unter Placebo ist $r = 0,718$. Diese Stärke der Abhängigkeit zwischen den Stufen

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

G*Power-Ergänzungen (SPSS)

Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

des messwiederholten Faktors A „Medikament“ müssen Sie nun in die Berechnung der Teststärke mit einbeziehen. Die weiteren Vorgaben für die a posteriori Teststärkenberechnung sind: mittlerer Effekt ($f = 0,25$; Anpassung $f' = \sqrt{q} \cdot f = \sqrt{3} \cdot 0,25 = 0,4330$, siehe vorheriger Abschnitt), Signifikanzniveau 5%, $p = 2$ Stufen, Anzahl der Versuchspersonen $N = 36$.

The screenshot shows the G*Power software interface. The 'Test family' is set to 'F tests' and the 'Statistical test' is 'ANOVA: Repeated measures, within factors'. The 'Type of power analysis' is 'Post hoc: Compute achieved power - given alpha, sample size, and effect size'. The 'Input Parameters' section includes: Effect size f (0.4330), alpha err prob (0.05), Total sample size (36), Number of groups (1), Number of measurements (2), Corr among rep measures (0.718), and Nonsphericity correction epsilon (1). The 'Output Parameters' section includes: Noncentrality parameter lambda (47.8695319), Critical F (4.1213382), Numerator df (1.0000000), Denominator df (35.0000000), and Power (1-beta err prob) (0.9999990).

Die Teststärke, einen mittleren Effekt (für unabhängige Stichproben) in der vorliegenden Untersuchung zu finden, war $> 99\%$. Für das Aufspüren eines kleinen Effekts von $f = 0,1$ (Anpassung $f' = \sqrt{q} \cdot f = \sqrt{3} \cdot 0,1 = 0,1732$) ist die erreichte Teststärke von 76,76% knapp nicht ausreichend.

The screenshot shows the G*Power software interface with a smaller effect size. The 'Test family' is 'F tests' and the 'Statistical test' is 'ANOVA: Repeated measures, within factors'. The 'Type of power analysis' is 'Post hoc: Compute achieved power - given alpha, sample size, and effect size'. The 'Input Parameters' section includes: Effect size f (0.1732), alpha err prob (0.05), Total sample size (36), Number of groups (1), Number of measurements (2), Corr among rep measures (0.718), and Nonsphericity correction epsilon (1). The 'Output Parameters' section includes: Noncentrality parameter lambda (7.6591251), Critical F (4.1213382), Numerator df (1.0000000), Denominator df (35.0000000), and Power (1-beta err prob) (0.7675876).

Für die a posteriori Berechnung der Teststärke des messwiederholten Faktors B „Testwiederholung“ ist eine andere Korrelation entscheidend: Die Korrelation zwischen den

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

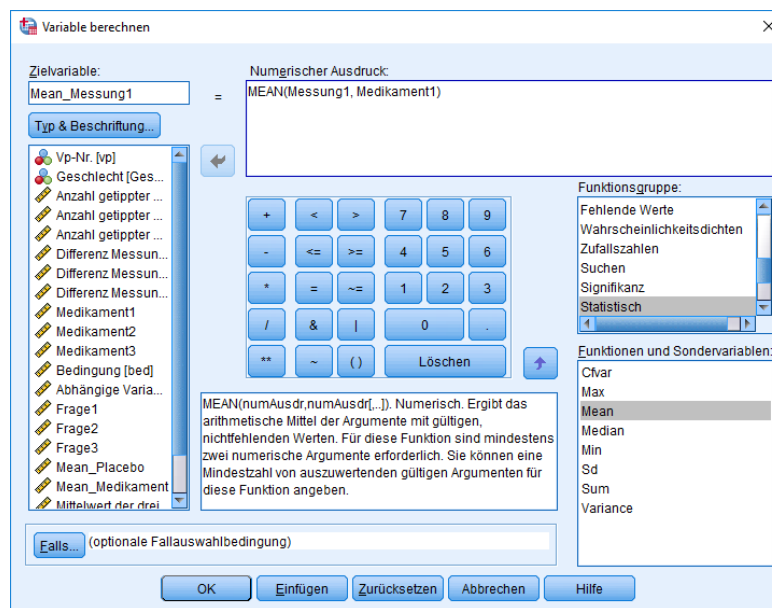
© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

G*Power-Ergänzungen (SPSS)

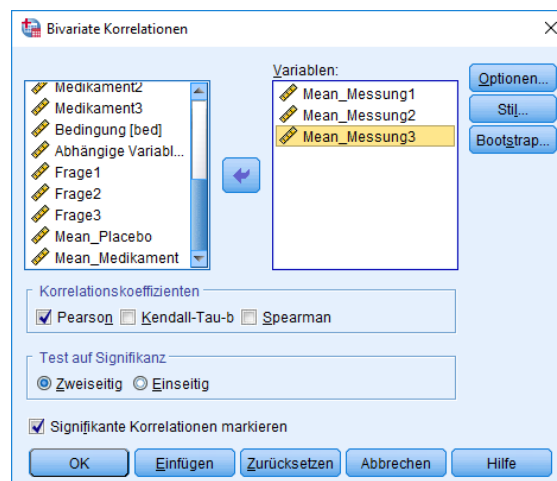
Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

gemittelten Werten der drei aufeinander folgenden Messungen des motorischen Leistungstests. Um diese Korrelation abzuschätzen, müssen Sie zunächst diese Werte mit Hilfe von SPSS berechnen.

Analog zu der Vorgehensweise für den Faktor A, erstellen Sie nun drei neue Variablen, z.B. mit den Namen „Mean_Messung1“, „Mean_Messung2“ und „Mean_Messung3“, in denen Sie SPSS den Mittelwert der jeweiligen Messung unter Placebo und Medikament berechnen lassen. Die Abbildung zeigt das Fenster für die Berechnung der Variable „Mean_Messung1“.



Nachdem Sie alle drei neuen Variablen erstellt haben, berechnen Sie die Korrelationen zwischen diesen Variablen.



Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

G*Power-Ergänzungen (SPSS)

Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

Korrelationen

		Mean_Messung1	Mean_Messung2	Mean_Messung3
Mean_Messung1	Korrelation nach Pearson	1	,891 **	,838 **
	Signifikanz (2-seitig)		,000	,000
	N	36	36	36
Mean_Messung2	Korrelation nach Pearson	,891 **	1	,858 **
	Signifikanz (2-seitig)	,000		,000
	N	36	36	36
Mean_Messung3	Korrelation nach Pearson	,838 **	,858 **	1
	Signifikanz (2-seitig)	,000	,000	
	N	36	36	36

** . Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Die mittlere Korrelation zwischen den Stufen des Faktors lässt sich über die Fishers-Z Transformation ermitteln. Benutzen Sie dafür Tabelle D in Band 1. Es ergibt sich folgender mittlerer Z-Wert:

$$\bar{Z} = \frac{1,422 + 1,221 + 1,293}{3} = 1,312 \quad \rightarrow \quad \bar{r} = 0,865$$

Die mittlere Korrelation von $r = 0,865$ setzen Sie nun in die Berechnung der Teststärke ein. Wie groß ist die Teststärke der Studie, bei $N = 36$ Personen einen kleinen Effekt für unabhängige Stichproben $f = 0,1$ (Anpassung $f' = \sqrt{p} \cdot f = \sqrt{2} \cdot 0,1 = 0,1414$) bei einem Signifikanzniveau von 5% zu finden, falls dieser existiert ($q = 3$)?

Die Teststärke, einen kleinen Effekt für unabhängige Stichproben in der vorliegenden Studie in dem messwiederholten Faktor B „wiederholte Testdurchführung“ zu finden, war 94,84%.

Besonders relevant ist eine a posteriori Teststärkenberechnung für das Datenbeispiel im Fall der Wechselwirkung, da hier die Varianzanalyse keine Signifikanz ergab (siehe SPSS Ergänzungen zu

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

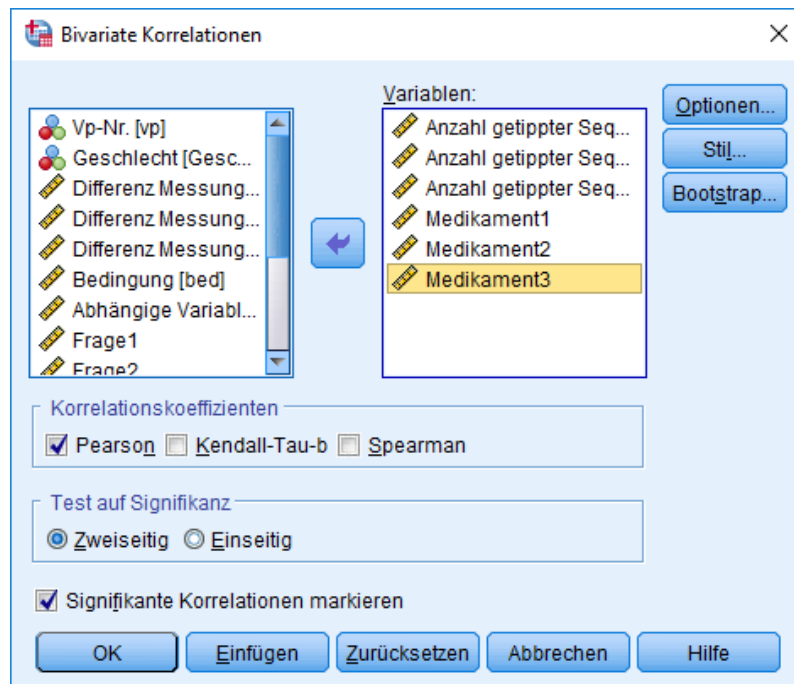
Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

G*Power-Ergänzungen (SPSS)

Rasch, Frieze, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

diesem Kapitel). Können wir bei dem vorliegenden Ergebnis mit ausreichender Sicherheit ausschließen, dass ein mittlerer Effekt für unabhängige Stichproben ($f = 0.25$) für die Wechselwirkung existiert? Zur Beantwortung dieser Frage ist die mittlere Korrelation zwischen den einzelnen Kombinationen der Faktorstufen entscheidend. Die einzelnen Korrelationen können wir direkt mit dem Programm SPSS berechnen, die Bildung des Mittelwerts der Korrelationen erfolgt wiederum mit Hilfe der Fishers-Z Transformation.



Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Frieze, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

G*Power-Ergänzungen (SPSS)

Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

Korrelationen

		Anzahl getippter Sequenzen bei Messung 1	Anzahl getippter Sequenzen bei Messung 2	Anzahl getippter Sequenzen bei Messung 3	Medikament1	Medikament2	Medikament3
Anzahl getippter Sequenzen bei Messung 1	Korrelation nach Pearson	1	,661 **	,649 **	,705 **	,795 **	,638 **
	Signifikanz (2-seitig)		,000	,000	,000	,000	,000
	N	36	36	36	36	36	36
Anzahl getippter Sequenzen bei Messung 2	Korrelation nach Pearson	,661 **	1	,799 **	,548 **	,568 **	,490 **
	Signifikanz (2-seitig)	,000		,000	,001	,000	,002
	N	36	36	36	36	36	36
Anzahl getippter Sequenzen bei Messung 3	Korrelation nach Pearson	,649 **	,799 **	1	,614 **	,588 **	,561 **
	Signifikanz (2-seitig)	,000	,000		,000	,000	,000
	N	36	36	36	36	36	36
Medikament1	Korrelation nach Pearson	,705 **	,548 **	,614 **	1	,914 **	,816 **
	Signifikanz (2-seitig)	,000	,001	,000		,000	,000
	N	36	36	36	36	36	36
Medikament2	Korrelation nach Pearson	,795 **	,568 **	,588 **	,914 **	1	,817 **
	Signifikanz (2-seitig)	,000	,000	,000	,000		,000
	N	36	36	36	36	36	36
Medikament3	Korrelation nach Pearson	,638 **	,490 **	,561 **	,816 **	,817 **	1
	Signifikanz (2-seitig)	,000	,002	,000	,000	,000	
	N	36	36	36	36	36	36

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

$$\bar{Z} =$$

$$\frac{0,802 + 0,775 + 0,877 + 1,085 + 0,758 + 1,099 + 0,618 + 0,648 + 0,536 + 0,717 + 0,678 + 0,633 + 1,557 + 1,142 + 1,142}{15}$$

$$\bar{Z} = 0,8711 \quad \rightarrow \quad \bar{r} = 0,7$$

Das Einsetzen eines mittleren Effekts ($f = 0,25$; Anpassung $f' = \sqrt{q} \cdot f = \sqrt{2} \cdot 0,25 = 0,3536$) ergibt eine sehr hohe Teststärke von mehr als 99%. Auf Grund des nicht signifikanten Ergebnisses in der Beispielstudie können wir also mit über 99%iger Sicherheit ausschließen, dass ein mittlerer Effekt (für unabhängige Stichproben) im Fall der Wechselwirkung existiert. Kleine Effekte ($f = 0,1$; Anpassung $f' = \sqrt{q} \cdot f = \sqrt{2} \cdot 0,1 = 0,1414$) können wir allerdings nur mit 64,84%iger Sicherheit ausschließen.

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

G*Power-Ergänzungen (SPSS)

Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

The image displays two screenshots of the SPSS G*Power dialog box, showing the configuration for a power analysis. Both screenshots use the same test family (F tests), statistical test (ANOVA: Repeated measures, within factors), and type of power analysis (Post hoc: Compute achieved power - given α , sample size, and effect size).

Top Screenshot:

Input Parameters	Value
Determine =>	
Effect size f	0.3536
α err prob	0.05
Total sample size	36
Number of groups	1
Number of measurements	3
Corr among rep measures	0.7
Nonsphericity correction ϵ	1

Output Parameters	Value
Noncentrality parameter λ	45.0118656
Critical F	3.1276756
Numerator df	2.0000000
Denominator df	70.0000000
Power (1- β err prob)	0.9999887

Bottom Screenshot:

Input Parameters	Value
Determine =>	
Effect size f	0.1414
α err prob	0.05
Total sample size	36
Number of groups	1
Number of measurements	3
Corr among rep measures	0.7
Nonsphericity correction ϵ	1

Output Parameters	Value
Noncentrality parameter λ	7.1978256
Critical F	3.1276756
Numerator df	2.0000000
Denominator df	70.0000000
Power (1- β err prob)	0.6483695

Es ist also durchaus möglich, dass trotz des nicht signifikanten Ergebnisses kleine Effekte der Wechselwirkung zwischen der wiederholten Durchführung des motorischen Tests und dem Faktor „Treatment“ existieren. Die Teststärke in der vorliegenden Studie war nicht ausreichend, um kleine Effekte mit einer ausreichenden Sicherheit ausschließen zu können. Effekte mittlerer Größe existieren aber mit hoher Wahrscheinlichkeit nicht.

Berechnen der Effektgröße f aus empirischen Daten und bestimmen der zugehörigen beobachteten Teststärke

Die Formeln für die von SPSS verwendete Schätzung der Effektstärke aus den Daten wurden bereits in den vorherigen Abschnitten angegeben. Für die Berechnung der beobachteten Teststärke für einen empirisch gefundenen Effekt verwenden Sie die Option „Beobachtete Trennschärfe“ in SPSS.

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

G*Power-Ergänzungen (SPSS)

Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

Literatur

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

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature