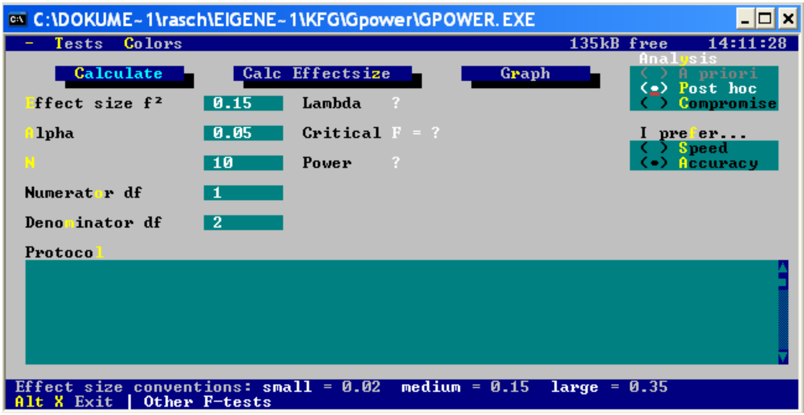
Kapitel 7: Varianzanalyse mit Messwiederholung

Nach dem Starten von GPower müssen Sie für Varianzanalysen mit Messwiederholung zunächst unter Tests „Other F-Tests" auswählen. Sie erhalten nun folgendes Eingabefenster:



Bitte ignorieren Sie in diesem Fenster die von GPower angegebenen Konventionen der Effektstärken. 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 bedeutet insgesamt, 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 Studien formuliert wurden. Für die Teststärkeberechnung müssen diese Konventionen zunächst unter Berücksichtigung der Korrelation zwischen den Faktorstufen und der Anzahl der Stufen des messwiederholten Faktors umgerechnet werden.

Die bekannten Konventionen für unabhängige Messungen lauten:

Kleiner Effekt:	$\Omega^2 = 0{,}01$	$\rightarrow$	$f^2 = 0{,}01$
Mittlerer Effekt:	$\Omega^2 = 0{,}06$	$\rightarrow$	$f^2 = 0{,}0625$
Großer Effekt:	$\Omega^2 = 0{,}14$	$\rightarrow$	$f^2 = 0{,}16$

7.1 Einfaktorielle Varianzanalyse mit Messwiederholung

Berechnen der Teststärke a priori bzw. Stichprobenumfangsplanung

Leider bietet GPower für die Varianzanalysen mit Messwiederholung keine a priori Berechnung der Teststärke an. Für die Berechnung des Stichprobenumfangs vor einer Untersuchung ist deshalb ein sukzessives Anpassen der Versuchspersonenzahl und der Freiheitsgrade notwendig, um die gewünschte Teststärke für einen inhaltlich relevanten Effekt zu erreichen.

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 Vorarbeiten gibt, orientiert sich der Studienleiter an den Konventionen für unabhängige Stichproben und legt einen mittleren Effekt von $\Omega^2 = 0{,}06$ 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.

Die Umrechung von $\Omega^2 = 0{,}06$ zu dem von GPower verlangen f'^2 erfolgt nach folgender Formel:

$$f'^2 = \frac{p}{1-\bar{r}} \cdot \Phi^2_{\text{unabhängig}} \text{ mit } \qquad \Phi^2_{\text{unabhängig}} = \frac{\Omega^2_{\text{unabhängig}}}{1-\Omega^2_{\text{unabhängig}}}$$

Beispiel:

$$f'^2 = \frac{3}{1-0{,}4} \cdot \frac{0{,}06}{1-0{,}06} = 5 \cdot 0{,}0638 = 0{,}319$$

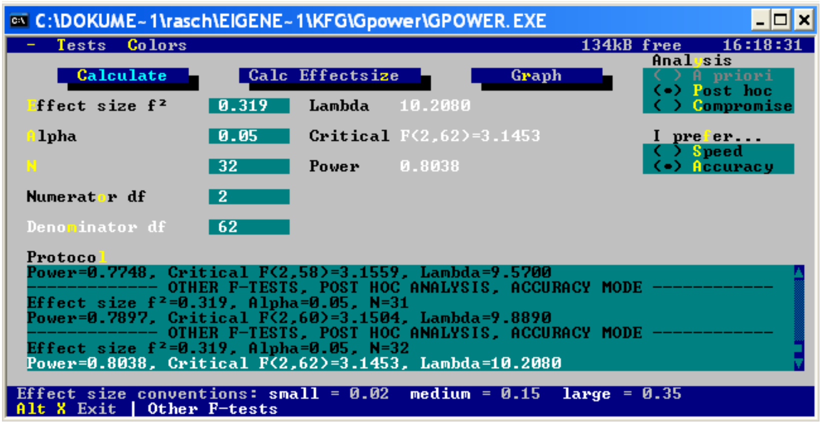
Diesen Wert können Sie nun in GPower eingeben.

Die Zähler- und Nennerfreiheitsgrade einer einfaktoriellen Varianzanalyse mit Messwiederholung berechnen sich zu (siehe Kap. 7.1.3):

$df_A = p - 1$ und $df_{A \times Vpn} = (n - 1) \cdot (p - 1)$

In unserem Beispiel liegen zwei Zählerfreiheitsgrade vor, die Nennerfreiheitsgrade müssen entsprechend der Versuchspersonenzahl angepasst werden. Wir beginnen mit einem Stichprobenumfang von $N = 20$ ($df_{A \times Vpn} = 38$). GPower berechnet für diese Versuchspersonenanzahl eine Teststärke von 57,59%.

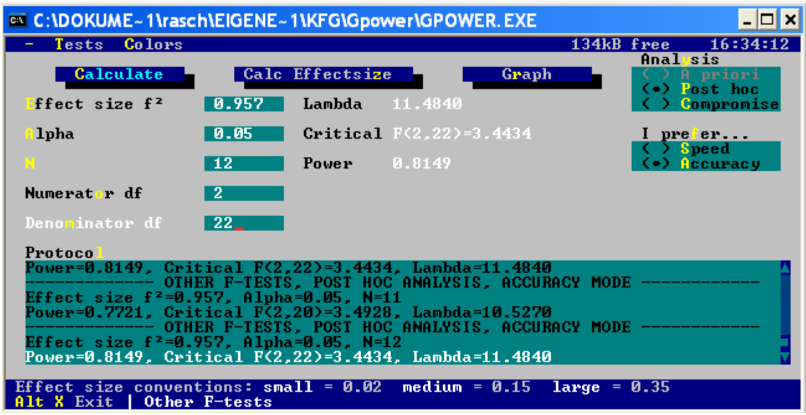
Nach sukzessivem Erhöhen der Personenzahl N und der Nennerfreiheitsgrade resultiert eine angemessene Teststärke von > 80% bei einem Stichprobenumfang von N = 32. Es werden also 32 Personen benötigt, um bei diesen Vorgaben einen mittleren Effekt von $\Omega^2 = 0{,}06$ bei einer angenommen Korrelation von r = 0,04 mit einer Wahrscheinlichkeit von 80% zu finden.



Die geringfügige Abweichung zwischen der Zahl N = 32 und dem in Kapitel 7.1.8 berechneten Stichprobenumfang von N = 31 liegt an der unterschiedlichen Art der Bestimmung der Wahrscheinlichkeit des Nonzentralitätsparameteres in den im Buch verwendeten Tabellen und GPower.

Wie viele Versuchspersonen werden benötigt, wenn der Studienleiter anstatt einer Korrelation von r = 0,04 eine Korrelation von r = 0,8 annimmt? In diesem Fall ergibt sich f'² zu:

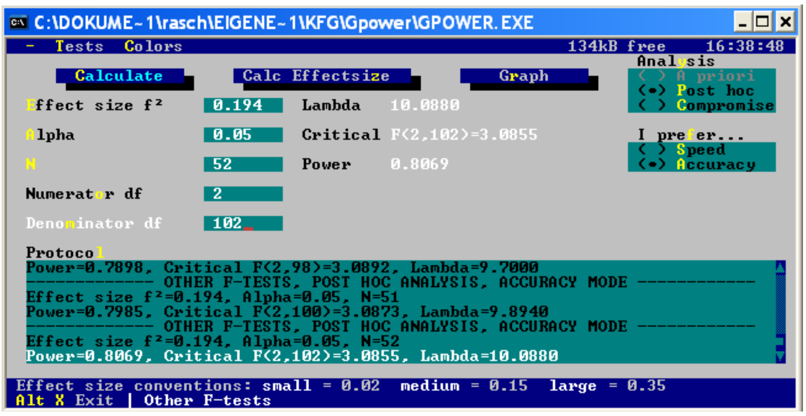
$$f'^2 = \frac{3}{1-0{,}8} \cdot \frac{0{,}06}{1-0{,}06} = 15 \cdot 0{,}0638 = 0{,}957$$



Wenn die Korrelation zwischen den messwiederholten Stufen r = 0,8 betragen würden, wären nur zwölf Versuchspersonen notwendig um einen mittleren Effekt von $\Omega^2 = 0{,}06$ mit einer Wahrscheinlichkeit von > 80% zu finden.

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

$$f'^2 = \frac{3}{1-0} \cdot \frac{0{,}06}{1-0{,}06} = 3 \cdot 0{,}0638 = 0{,}194$$



Im Fall einer Korrelation von r = 0 wären 52 Versuchspersonen notwendig gewesen, um einen mittleren Effekt von $\Omega^2 = 0{,}06$ mit einer Wahrscheinlichkeit von > 80% zu finden.

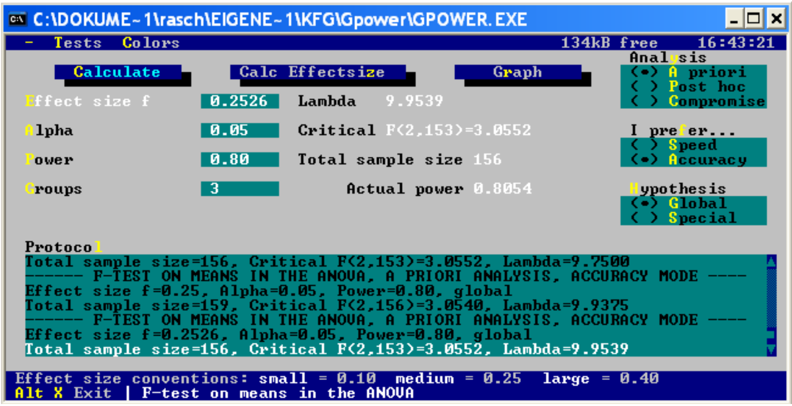
Ein identisches Ergebnis erhalten wir, wenn wir den Stichprobenumfang bei einer einfaktoriellen Varianzanalyse ohne Messwiederholung bestimmen. Gehen Sie unter „Tests“ zu „F-Test (ANOVA)“. Hier ist die Effektgröße f gefordert.

$$f = \sqrt{\Phi^2} = \sqrt{\frac{\Omega^2}{1-\Omega^2}} = \sqrt{\frac{0{,}06}{1-0{,}06}} = \sqrt{0{,}0638} = 0{,}2526$$ [1]

Unter der Angabe der Effektstärke, des α-Niveaus von 5%, der gewünschter Power von 80% und drei Gruppen errechnet GPower einen optimalen Stichprobenumfang von N = 156. Dies entspricht

[1] Abweichungen für die Angabe der Größe eines mittleren Effekts entstehen auf Grund von Rundungsfehlern bei der Umrechnung von Φ^2 zu Ω^2.

exakt 52 Versuchspersonen pro Gruppe. Die 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.



Achtung: Bei einer Verletzung der Sphärizitätsannahme reduziert sich die Teststärke in einer messwiederholten Varianzanalyse. Ist eine derartige Verletzung im Vorhinein absehbar, sollten mehr Versuchspersonen erhoben werden, als mit der Stichprobenumfangsplanung berechnet wurden.

Teststärkebestimmung a posteriori

Eine Teststärkebestimmung a posteriori ist nur dann sinnvoll, wenn vor einem Experiment keine Stichprobenumfangsplanung durchgeführt wurde und ein nicht signifikanten Ergebnis vorliegt. Trotzdem soll die a posteriori Teststärkenbestimmung nun an dem Beispiel der dreifach wiederholten Messung der motorischen Fertigkeiten aus Kapitel 7.1 durchgeführt werden, um das allgemeine Vorgehen zu veranschaulichen.

In dem Beispiel haben insgesamt 36 Versuchspersonen teilgenommen. Wie hoch war die Wahrscheinlichkeit, einen mittleren Effekt für unabhängige Stichproben von $\Omega^2 = 0,06$ in der Studie zu entdecken?

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

Korrelationen

		Messung1	Messung2	Messung3
Messung1	Korrelation nach Pearson	1	,661**	,649**
	Signifikanz (2-seitig)		,000	,000
	N	36	36	36
Messung2	Korrelation nach Pearson	,661**	1	,799**
	Signifikanz (2-seitig)	,000		,000
	N	36	36	36
Messung3	Korrelation nach Pearson	,649**	,799**	1
	Signifikanz (2-seitig)	,000	,000	
	N	36	36	36

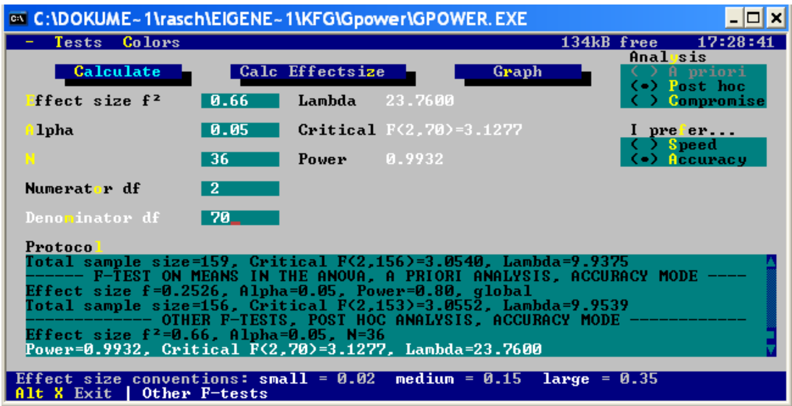
**. Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

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.

Nun können wir die von GPower geforderte Effektstärke f'² berechnen:

$$f'^2 = \frac{p}{1-\bar{r}} \cdot \frac{\Omega^2_{\text{unabhängig}}}{1-\Omega^2_{\text{unabhängig}}} = \frac{3}{1-0{,}71} \cdot \frac{0{,}06}{1-0{,}06} = 10{,}34 \cdot 0{,}0638 = 0{,}66$$

Die Anzahl der Zählerfreiheitsgrade ist $df_A = 2$, die Anzahl der Nennerfreiheitsgrade ist $df_{A\times Vpn} = 70$. GPower berechnet für diese Vorgaben eine Teststärke von > 99%. Die Wahrscheinlichkeit, in dieser Studie einen mittleren Effekt von $\Omega^2 = 0{,}06$ bei einer vorliegenden mittleren Korrelation zwischen den Faktorstufen von r = 0,71 zu finden, falls er tatsächlich existiert, war größer als 99%.



Berechnen der Effektgröße f² aus empirischen Daten und Bestimmung der zugehörigen beobachteten Teststärke

In der messwiederholten Varianzanalyse gestaltet sich die empirische Schätzung der Effektgröße auf der *Ebene der Population* als schwierig. Wir orientieren uns deshalb an der von SPSS angebotenen Effektgrößenschätzung auf der *Stichprobenebene*, und verwenden das partielle Eta-Quadrat (η_p^2). Allerdings weisen wir an dieser Stelle noch einmal darauf hin, dass diese Effektgröße den Effekt auf der Ebene der Population teilweise deutlich überschätzt. Die Berechnungsvorschrift lautet wie folgt:

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

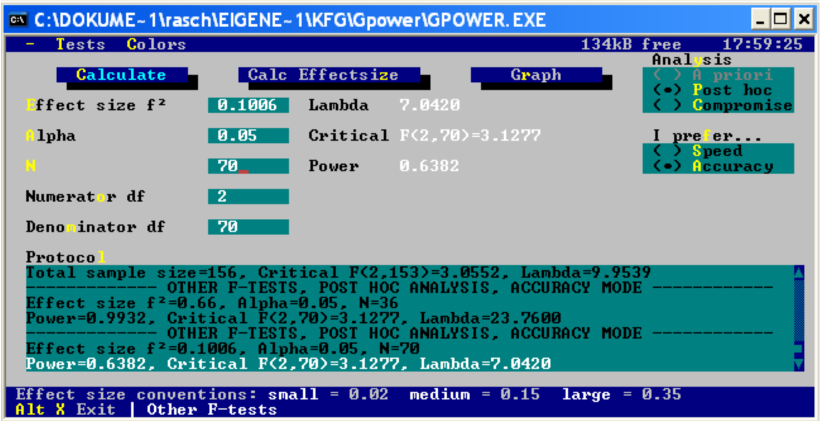
Zur Bestimmung der beobachteten Teststärke eines aus vorliegenden Daten geschätzten Effekts empfehlen wir die Ausgabe der „Beobachteten Schärfe" in SPSS. Insgesamt beinhaltet die beobachtete Teststärke allerdings wenig zusätzliche Information (vgl. SPSS-Ergänzungen zu den Kapiteln 5 und 6): Ist ein Ergebnis 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 vorherigen Abschnitt). Falls Sie trotzdem die von SPSS ausgegebenen Werte mit GPower nachvollziehen möchten, können Sie die Effektstärke f² nach der obigen Formel für $f^2_{S(\text{abhängig})}$ berechnen. Allerdings müssen Sie anstatt der Anzahl der Versuchspersonen N die Anzahl der Nennerfreiheitsgrade eingeben.

Achtung: Wenden Sie nicht die zu Beginn dieser GPower Ergänzungen angegebenen Formel auf die empirischen Effektstärkenschätzer auf der Stichprobenebene η_p^2 und $f^2_{S(\text{abhängig})}$ an. Diese Formeln beziehen sich ausschließlich auf die Umrechung der Konventionen für unabhängige Stichproben zu abhängigen Stichproben.

In dem Beispiel der Messwiederholung bei dem motorischen Leistungstest ergab sich ein F-Wert von $F_{(2;70)} = 3{,}52$. SPSS gibt für den Fall, dass keine Verletzung der Sphärizität vorliegt, eine beobachtete Teststärke von 63,8% an.

$$f^2_{S(\text{abhängig})} = \frac{F \cdot df_A}{df_{A\times Vpn}} = \frac{3{,}52 \cdot 2}{70} = 0{,}1006$$

Einsetzen in GPower und dabei für N die Anzahl der Nennerfreiheitsgrade $df_A = 70$ verwenden ergibt ebenfalls 63,8%.



Bei Verletzung der Sphärizität muss die eingegebene Effektstärke um den Faktor ε korrigiert werden. Für die Korrektur nach dem in SPSS mit „Greenhouse-Geisser“ bezeichneten Verfahren ergibt sich ein Korrekturfaktor von ε = 0,913 (Tabelle 7.4 in Kap. 7.1.5).

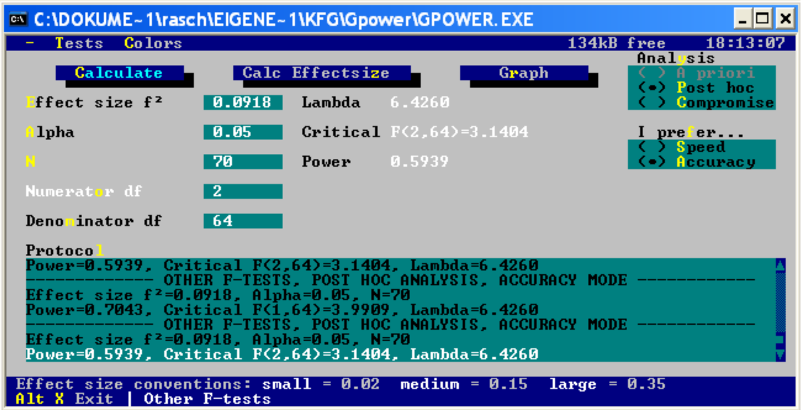
$$f'^2_{corr} = f^2_{S(\text{abhängig})} \cdot \varepsilon = 0{,}1006 \cdot 0{,}913 = 0{,}0918$$

Auch die Freiheitsgrade müssen korrigiert werden.

$$df_{A(corr)} = df_A \cdot \varepsilon = 2 \cdot 0{,}913 = 1{,}83$$

$$df_{A\times Vpn(corr)} = df_{A\times Vpn} \cdot \varepsilon = 70 \cdot 0{,}913 = 63{,}9$$

Leider erlaubt GPower nur die Eingabe von ganzen Zahlen bei den Freiheitsgraden. Dadurch kommt es zu geringfügigen Abweichungen der errechneten Teststärken von 59,39% zu dem von SPSS ausgegebenen 60,9%.


C:\DOKUME~1\rasch\EIGENE~1\KFG\Gpower\GPOWER.EXE
- Tests Colors
134kB free 18:13:07
Calculate
Calc Effectsize
Graph
Analysis
( ) A priori
(•) Post hoc
( ) Compromise
Effect size f²
0.0918
Lambda 6.4260
Alpha
0.05
Critical F(2,64)=3.1404
N
70
Power 0.5939
I prefer...
( ) Speed
(•) Accuracy
Numerator df
2
Denominator df
64
Protocol
Power=0.5939, Critical F(2,64)=3.1404, Lambda=6.4260
------------ OTHER F-TESTS, POST HOC ANALYSIS, ACCURACY MODE ------------
Effect size f²=0.0918, Alpha=0.05, N=70
Power=0.7043, Critical F(1,64)=3.9909, Lambda=6.4260
------------ OTHER F-TESTS, POST HOC ANALYSIS, ACCURACY MODE ------------
Effect size f²=0.0918, Alpha=0.05, N=70
Power=0.5939, Critical F(2,64)=3.1404, Lambda=6.4260
Effect size conventions: small = 0.02 medium = 0.15 large = 0.35
Alt X Exit | Other F-tests

7.2 Zweifaktorielle Varianzanalyse mit Messwiederholung auf einem Faktor

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

In dem Beispiel seien alle drei Effekte, der Haupteffekt A „Geschlecht“, der Haupteffekt B „Testwiederholung“ sowie ihre Wechselwirkung A×B inhaltlich relevant. Der Studienleiter legt vor der Studie fest, dass für alle drei Effektarten ein mittlerer Effekt (für unabhängige Stichproben) von $\Omega^2 = 0{,}06$ 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 des von GPower verlangen f’² für den nicht messwiederholten Faktor A „Geschlecht“ führt über folgende Formel. Die Anzahl der Stufen q bezieht sich auf den messwiederholten Faktor B (hier: q = 3)

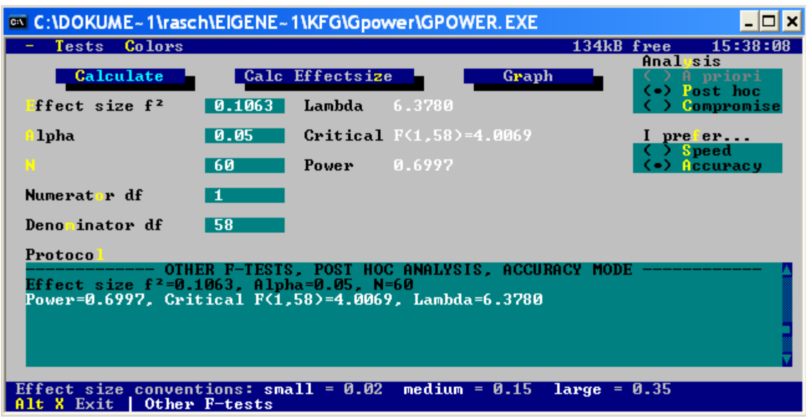
$$f'^2 = \frac{q}{1+(q-1)\cdot\bar{r}} \cdot \frac{\Omega^2_{\text{unabhängig}}}{1-\Omega^2_{\text{unabhängig}}} = \frac{3}{1+(3-1)\cdot 0{,}4} \cdot \frac{0{,}06}{1-0{,}06} = \frac{3}{1{,}8} \cdot 0{,}0638 = 0{,}1063$$

Die Freiheitsgrade berechnen sich zu:

$df_A = p - 1$ und $df_{\text{Vpn in S}} = p \cdot (n-1)$

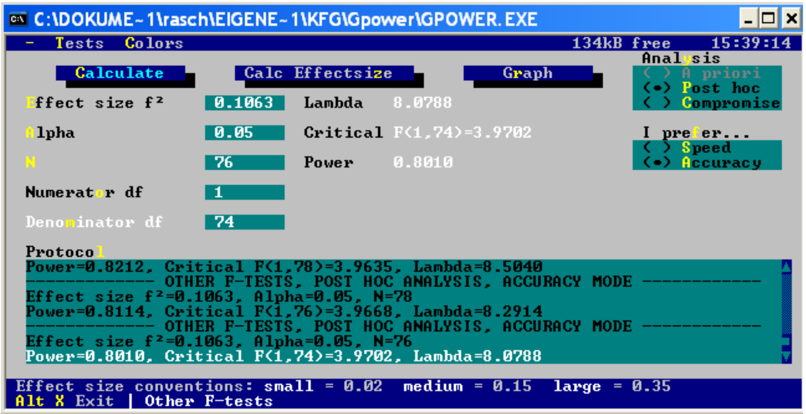
Das kleine n bezieht sich auf die Anzahl der Probanden pro Gruppe des nicht messwiederholten Faktors. Der Zusammenhang zwischen n und N ist hier N = n · p, wenn wir davon ausgehen, dass in jeder Stufe des nicht messwiederholten Faktors gleich viele Probanden teilnehmen .

Wir beginnen zunächst mit einer Gesamtversuchspersonenzahl von N = 60 (30 Männer, 30 Frauen). Die Zählerfreiheitsgradzahl für den zweistufigen Faktor „Geschlecht“ (p = 2) ist $df_A = 1$, die Nennerfreiheitsgrade sind $df_{\text{Vpn in S}} = 58$.



Quelle: http://www.quantitative-methoden.de

Die Teststärke ist mit 69,97% zu niedrig. Erst bei 76 Personen (38 Männer, 38 Frauen) ergibt sich eine ausreichende Power von > 80% unter den Vorgaben.



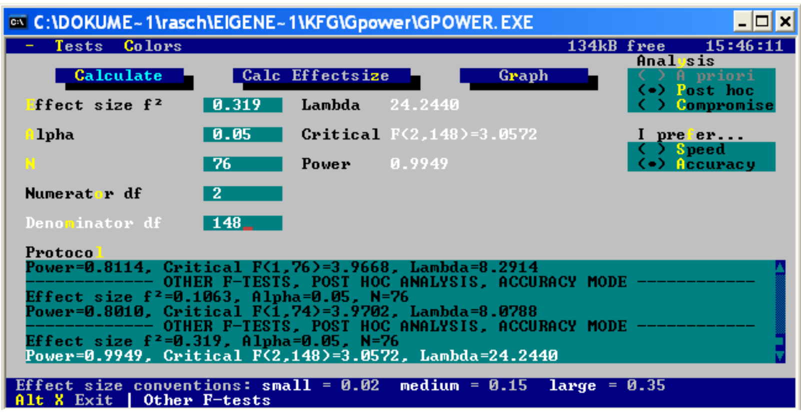
Die Stichprobenumfangsplanung für den Faktor B „Testwiederholung“ wurde bereits im vorherigen Abschnitt durchgeführt:

$$f'^2 = \frac{p}{1-\bar{r}} \cdot \frac{\Omega^2_{\text{unabhängig}}}{1-\Omega^2_{\text{unabhängig}}} = \frac{3}{1-0{,}4} \cdot \frac{0{,}06}{1-0{,}06} = 5 \cdot 0{,}0638 = 0{,}319$$

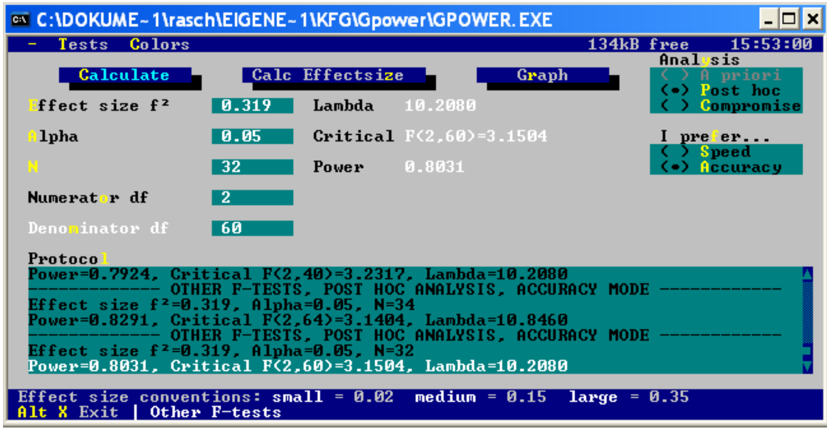
Allerdings verändert sich die Berechnung der Freiheitsgrade:

$df_B = q - 1$ und $df_{B \times Vpn} = p \cdot (q-1) \cdot (n-1)$.

Auf Grund der Berechnung für den nicht messwiederholten Faktor A „Geschlecht“ können wir die Teststärkenberechnung mit N = 76 beginnen. Die Zählfreiheitsgrade für den Faktor B „Testwiederholung“ (q = 3) sind $df_B = 2$, die Nennerfreiheitsgrade bei 38 Personen pro Gruppe $df_{B \times Vpn} = 148$.



Die Teststärke, einen mittleren Effekte (für unabhängige Stichproben) in dem messwiederholten Faktor B bei einer geschätzten Korrelation von r = 0,4 zu finden, ist mehr als ausreichend (1-β > 99%). Eigentlich wären für diesen Faktor nur N = 32 Personen notwendig gewesen (16 Männer, 16 Frauen), um eine Power von > 80% zu erreichen. Diese Analyse entspricht mit Ausnahme der Anzahl der Nennerfreiheitsgrade der errechneten optimalen Versuchspersonenzahl in der einfaktoriellen Varianzanalyse mit Messwiederholung, ohne den zusätzlichen Faktor Geschlecht (siehe oben).



Die Berechnung der Effektstärke f'² für die Wechselwirkung zwischen einem messwiederholten und einem nicht-messwiederholten Faktor ist identisch mit dem f'² für den messwiederholten Faktor. Da auch für die Wechselwirkung ein mittlerer Effekt (für unabhängige Stichproben) als inhaltlich relevant erachtet wurde, ergibt sich derselbe Wert.

$$f'^2 = \frac{p}{1-\bar{r}} \cdot \frac{\Omega^2_{\text{unabhängig}}}{1-\Omega^2_{\text{unabhängig}}} = \frac{3}{1-0{,}4} \cdot \frac{0{,}06}{1-0{,}06} = 5 \cdot 0{,}0638 = 0{,}319$$

Hier verändert sich nur die Berechnung der Freiheitsgrade des Zählers, die Anzahl der Nennerfreiheitsgrade entspricht den Nennerfreiheitsgraden des messwiederholten Faktors.

$df_{A\times B(mw)} = (p-1)\cdot(q-1)$ und $df_{B\times Vpn} = p\cdot(q-1)\cdot(n-1)$.

Allerdings hat in dem verwendeten Beispiel der nicht messwiederholte Faktor nur zwei Stufen. In diesem besonderen Fall ist die Stichprobenumfangsplanung für die Wechselwirkung und für den messwiederholten Faktor identisch. Es ergibt sich also auch eine optimale Versuchspersonenzahl von N = 32, einen mittleren Effekt (für unabhängige Stichproben) bei einem Signifikanzniveau von 5% und einer geschätzten mittleren Korrelation zwischen den Stufen von r = 0,4 mit einer Wahrscheinlichkeit von 80% zu finden.

Um in der Beispieluntersuchung einen mittleren Effekt auch in dem Faktor „Geschlecht" mit ausreichender Wahrscheinlichkeit zu finden, sollten N = 76 Personen untersucht werden. Sind die Geschlechtsunterschiede nicht inhaltlich relevant, genügen 32 Versuchspersonen.

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 in dem 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 $\Omega = 0{,}06$ wird als inhaltlich relevant festgelegt.

$$f'^2 = \frac{p}{1-\bar{r}} \cdot \frac{\Omega^2_{\text{unabhängig}}}{1-\Omega^2_{\text{unabhängig}}} = \frac{3}{1-0{,}71} \cdot \frac{0{,}06}{1-0{,}06} = 10{,}34 \cdot 0{,}0638 = 0{,}66$$

Als einzige Veränderung müssen Sie die entsprechende Berechnung der Zähler- und Nennerfreiheitsgrade beachten:

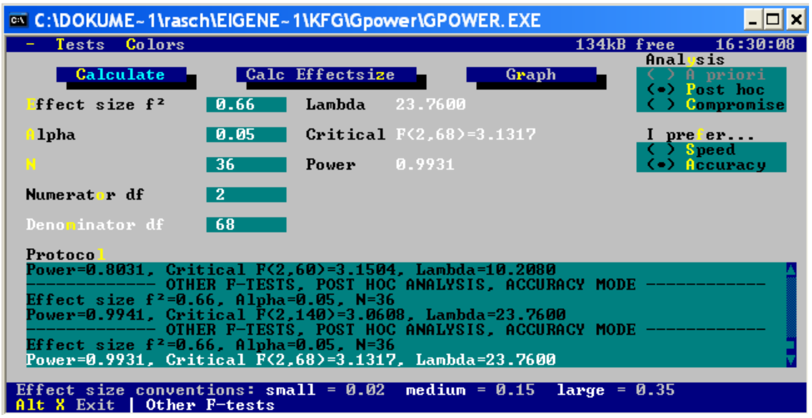
Messwiederholter Faktor B:

$df_B = q - 1$ und $df_{B \times Vpn} = p \cdot (q-1) \cdot (n-1)$

Wechselwirkung:

$df_{A \times B(mw)} = (p - 1) \cdot (q - 1)$ und $df_{B \times Vpn} = p \cdot (q-1) \cdot (n-1)$

Auf Grund der hohen Korrelation von r = 0,71 errechnet GPower für beide Effektarten für die untersuchten 36 Versuchspersonen (18 Männer, 18 Frauen) eine sehr hohe Teststärke von > 99%.

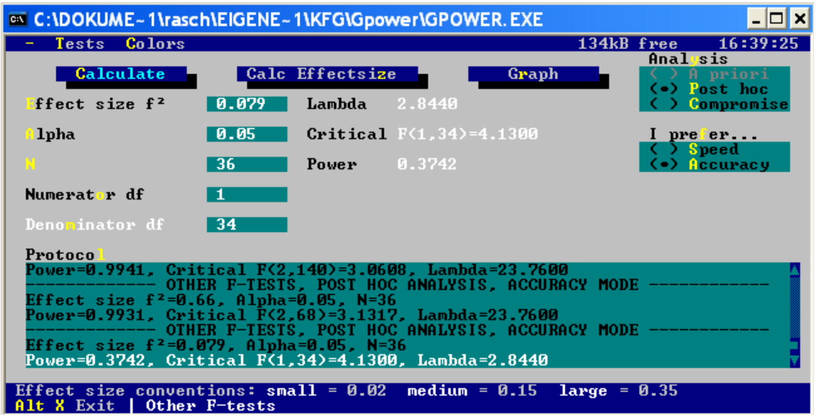


Wie hoch war die Teststärke für den nicht messwiederholten Faktor „Geschlecht", einen mittleren Effekt mit insgesamt 36 Personen zu finden? Auch hier geht die mittlere Korrelation zwischen den Stufen des messwiederholten Faktors in die Teststärkeberechnung mit ein. Allerdings *verkleinert* eine hohe Korrelation die Teststärke des nicht messwiederholten Faktors.

$$f'^2 = \frac{q}{1+(q-1)\cdot\bar{r}} \cdot \frac{\Omega^2_{\text{unabhängig}}}{1-\Omega^2_{\text{unabhängig}}} = \frac{3}{1+(3-1)\cdot 0{,}71} \cdot \frac{0{,}06}{1-0{,}06} = \frac{3}{2{,}42} \cdot 0{,}0638 = 0{,}079$$

$df_A = p - 1 = 2 - 1 = 1$ und $df_{Vpn\ in\ S} = p \cdot (n - 1) = 2 \cdot (18 - 1) = 34$

GPower errechnet für den Beispieldatensatz nur eine Teststärke von 37,42%, um einen mittleren Effekt bei einer gegebenen Korrelation von r = 0,71 mit insgesamt N = 36 Versuchspersonen zu finden.



Berechnen der Effektgröße f² 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:

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

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

7.3 Zweifaktorielle Varianzanalyse mit Messwiederholung auf beiden Faktoren

In diesem Abschnitt stellen wir die Teststärkeberechnung für die vollständig messwiederholte Varianzanalyse mit zwei Faktoren vor. 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 $\Omega^2 = 0{,}06$ 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. $r = 0{,}3$.

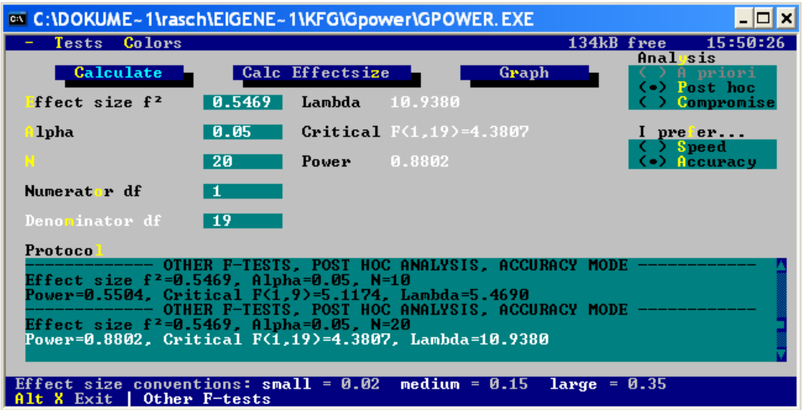
In die Berechnung des von GPower verlangten f'^2 geht die Anzahl der Stufen beider messwiederholten Faktoren ein:

$$f'^2 = \frac{p \cdot q}{1-\bar{r}} \cdot \frac{\Omega^2_{\text{unabhängig}}}{1-\Omega^2_{\text{unabhängig}}} = \frac{2 \cdot 3}{1-0{,}3} \cdot \frac{0{,}06}{1-0{,}06} = 8{,}571 \cdot 0{,}0638 = 0{,}5469$$

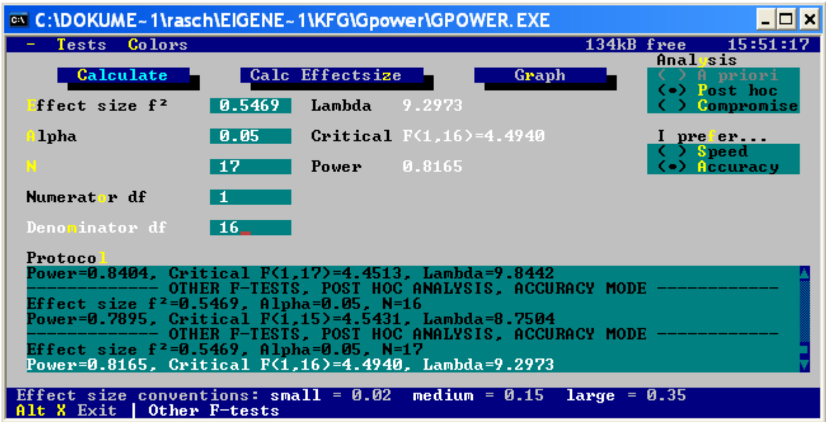
Bei der Berechnung der Freiheitsgrade ist es wichtig, die Nennerfreiheitsgrade der entsprechenden Prüfvarianz des F-Werts zu verwenden (vgl. Kap. 7.3.1):

$df_A = p - 1$ und $df_{A\times Vpn} = (p-1)\cdot(n-1)$.

Wir beginnen mit einer Gesamtversuchspersonenzahl N = 20, und setzen die entsprechenden Werte für die Freiheitsgrade in GPower ein. Es ergibt sich eine Teststärke von 88,02%.



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.

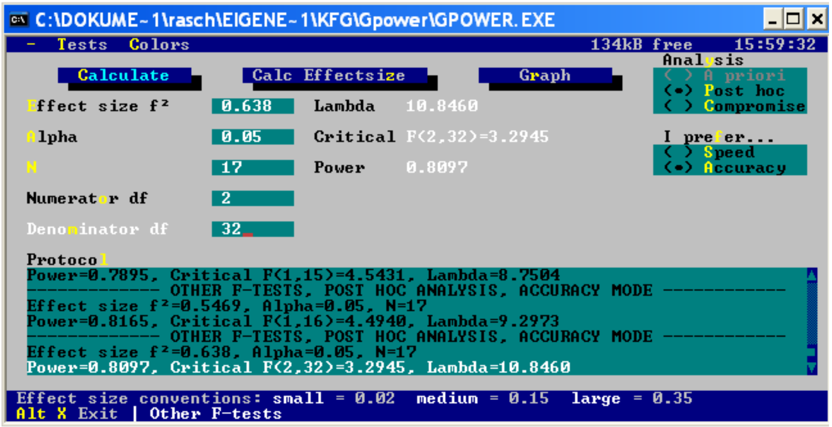


Wie viele Personen sind für den messwiederholten Faktor B „Testwiederholung“ unter den gleichen Voraussetzungen notwendig? Da die drei wiederholten Messungen zeitlich direkt aufeinander folgen, nehmen wir hier eine etwas höhere Korrelation zwischen den Faktorstufen des Faktors B von r = 0,4 an. Die Berechnung des f’² ist identisch mit der Berechnung des von GPower geforderten Effekts für den messwiederholten Faktor A, mit Ausnahme der unterschiedlichen Höhe der angenommenen Korrelation zwischen den drei Stufen. Auch die Anzahl der Freiheitsgrade unterscheidet sich (q = 3).

$$f'^2 = \frac{p \cdot q}{1-\bar{r}} \cdot \frac{\Omega^2_{\text{unabhängig}}}{1-\Omega^2_{\text{unabhängig}}} = \frac{2 \cdot 3}{1-0{,}4} \cdot \frac{0{,}06}{1-0{,}06} = 10 \cdot 0{,}0638 = 0{,}638$$

$df_B = q - 1$ und $df_{B \times Vpn} = (q-1) \cdot (n-1)$

Auf Grund der Berechnung für den Faktor A beginnen wir die Berechnung mit einer Gesamtzahl von Personen N = 17.

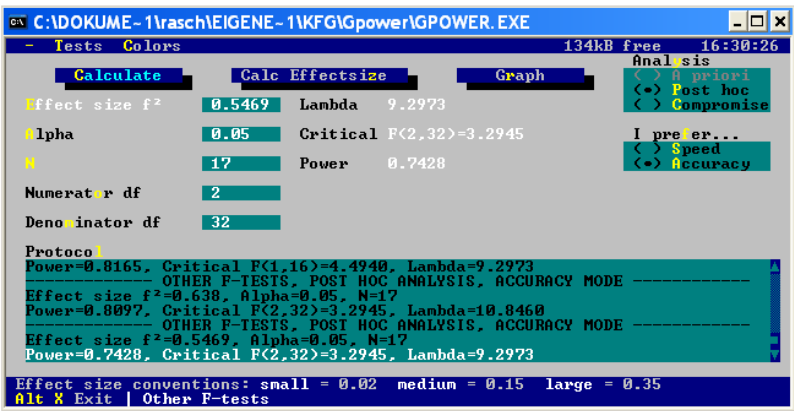


Die Anzahl N = 17 ist ausreichend, um in dem Haupteffekt B „Testwiederholung" einen mittleren Effekt (für unabhängige Stichproben) bei einem Signifikanzniveau von 5% und einer angenommen Korrelation zwischen den Faktorstufen von r = 0,4 mit einer Power von > 80% zu finden.

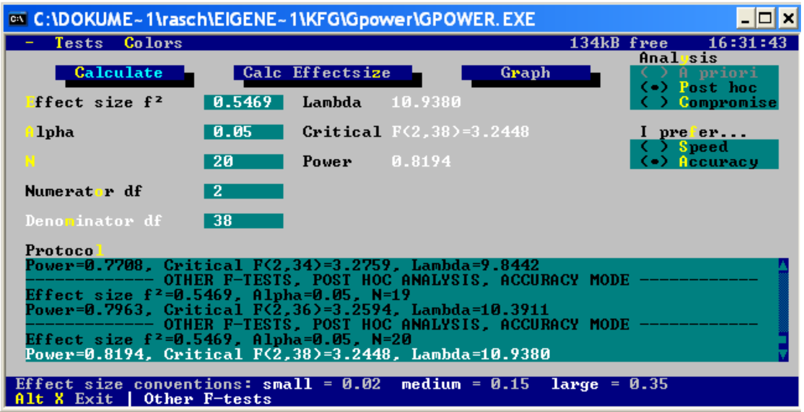
Reichen 17 Personen auch aus, um einen mittleren Effekt der Wechselwirkung mit mehr als 80%iger Sicherheit zu finden? Auch hier ist es angemessen, eine geringere Korrelation von r = 0,3 anzunehmen, da für die Wechselwirkung die mittlere Korrelation zwischen allen sechs möglichen Faktorstufenkombinationen ausschlaggebend ist. Da auch hier ein größerer zeitlicher Abstand zwischen den Messzeitpunkten besteht, z.B. bei der Korrelation zwischen Messung 1 unter Placebo mit der Messung 3 unter dem Medikament, sollte die mittlere Korrelation für die Stichprobenumfangsplanung nicht zu hoch eingeschätzt werden.

$$f'^2 = \frac{p \cdot q}{1-\bar{r}} \cdot \frac{\Omega^2_{\text{unabhängig}}}{1-\Omega^2_{\text{unabhängig}}} = \frac{2 \cdot 3}{1-0{,}3} \cdot \frac{0{,}06}{1-0{,}06} = 8{,}571 \cdot 0{,}0638 = 0{,}5469$$

$$df_{A \times B} = (p-1) \cdot (q-1) \quad \text{und} \qquad df_{A \times B \times Vpn} = (p-1) \cdot (q-1) \cdot (n-1)$$



Für die Wechselwirkung ergibt sich für N = 17 nicht die gewünschte Teststärke von > 80%. Sie wird erst bei N = 20 Personen erreicht.

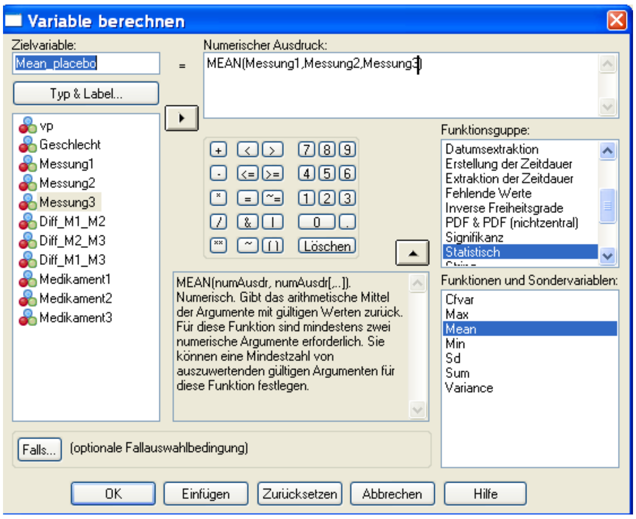


In dieser Studie sollten also insgesamt N = 20 Versuchspersonen untersucht werden, um einen mittleren Effekt (für unabhängige Stichproben) mit einer Wahrscheinlichkeit von mindestens 80% für die beiden Haupteffekte und die Wechselwirkung unter den angenommenen Korrelationen zwischen den entsprechenden Faktorstufen zu finden. 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 GPower Ergänzungen zu Kapitel 3).

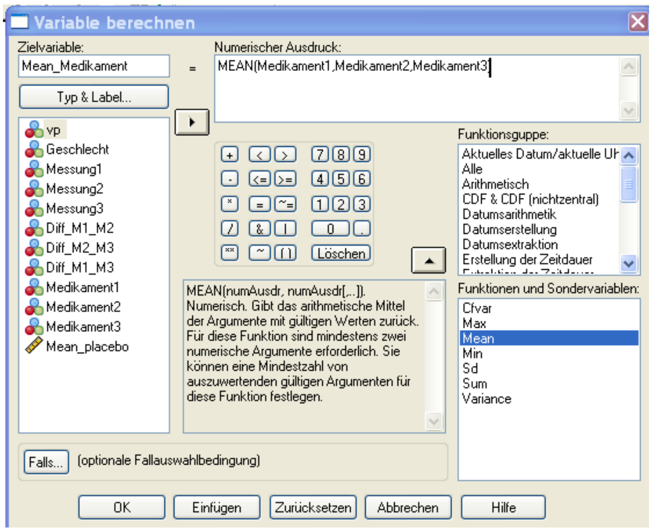
Teststärkebestimmung a posteriori

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 „Datensatz_Messwiederholung.sav“.

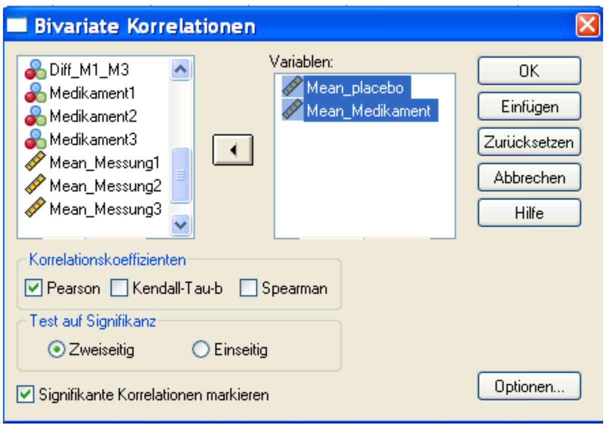
Für den Faktor A „Medikament“ („Medikament“ vs. „Placebo“) ist die Korrelation zwischen den über die drei Messwiederholungen des motorischen Tests gemittelten Werten unter Placebo und unter dem Medikament entscheidend. Um diese zu berechnen, gehen Sie in SPSS auf „Transformieren“ → „Berechnen“. Geben Sie einen neuen Namen für die Zielvariable in das Feld oben links ein, z.B. „Mean_placebo“. Unter „Typ & Label“ können Sie einen aussagekräftigeren Namen definieren, z.B. „Mittelwert der drei Messungen unter Placebo“. Suchen Sie unter der Funktionsgruppen „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“.



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



Quelle: http://www.quantitative-methoden.de

Korrelationen

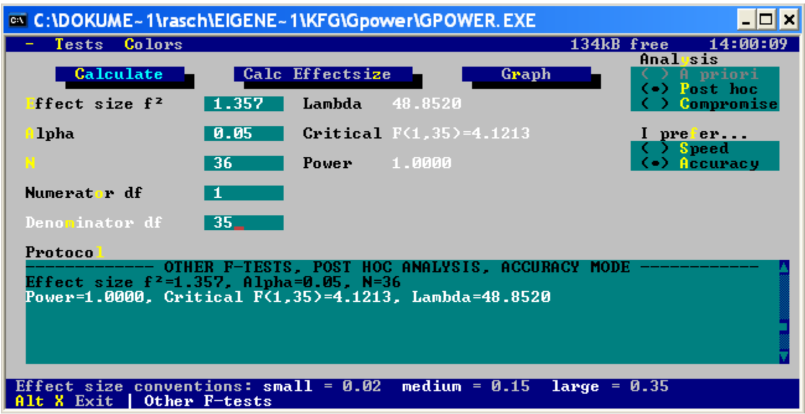
		Mean_ placebo	Mean_ Medikament
Mean_placebo	Korrelation nach Pearson	1	,718**
	Signifikanz (2-seitig)		,000
	N	36	36
Mean_Medikament	Korrelation nach Pearson	,718**	1
	Signifikanz (2-seitig)	,000	
	N	36	36

**. 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 des messwiederholten Faktors A „Medikament" müssen Sie nun in die Berechnung des von GPower geforderten Effekts f'² mit einbeziehen. Die weiteren Vorgaben für die a posteriori Teststärkenberechnung sind: mittlerer Effekt (Ω^2 = 0,06), Signifikanzniveau 5%, zwei messwiederholte Faktoren mit p = 2 und q = 3 Stufen, Anzahl der Versuchspersonen N = 36.

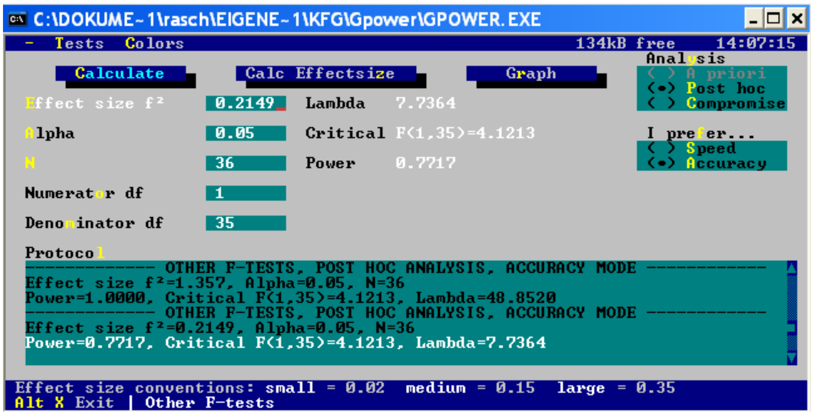
$$f'^2 = \frac{p \cdot q}{1-\bar{r}} \cdot \frac{\Omega^2_{\text{unabhängig}}}{1-\Omega^2_{\text{unabhängig}}} = \frac{2 \cdot 3}{1-0{,}718} \cdot \frac{0{,}06}{1-0{,}06} = 21{,}277 \cdot 0{,}0638 = 1{,}357$$

$df_A = p - 1 = (2 - 1) = 1$ und $df_{A \times Vpn} = (p-1) \cdot (n-1) = (2-1) \cdot (36-1) = 35$



Die Teststärke, einen mittleren Effekt (für unabhängige Stichproben) in der vorliegenden Untersuchung zu finden, war > 99%. Für das Aufspüren eines kleinen Effekts von Ω^2 = 0,01 ist die erreichte Teststärke von 77,17% knapp nicht ausreichend.

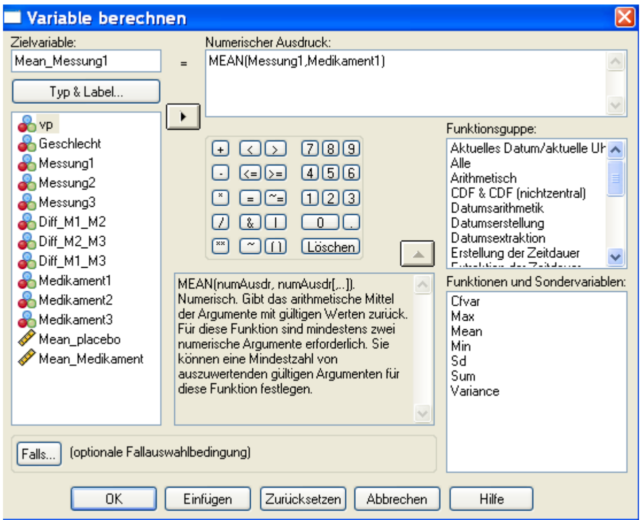
$$f'^2 = \frac{p \cdot q}{1-\bar{r}} \cdot \frac{\Omega^2_{\text{unabhängig}}}{1-\Omega^2_{\text{unabhängig}}} = \frac{2 \cdot 3}{1-0{,}718} \cdot \frac{0{,}01}{1-0{,}01} = 21{,}277 \cdot 0{,}0101 = 0{,}2149$$



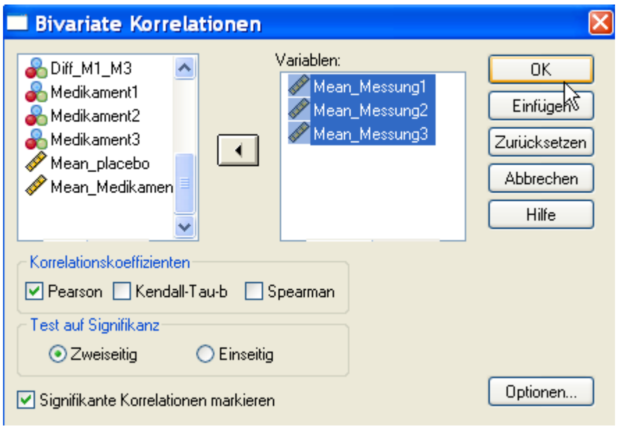
Für die a posteriori Berechnung der Teststärke des messwiederholten Faktors B „Testwiederholung“ ist eine andere Korrelation entscheidend: Die Korrelation zwischen den gemittelten Werten der drei aufeinander folgenden Messungen des motorischen Leistungstests. Um diese Korrelation abzuschätzen, müssen 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.



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.

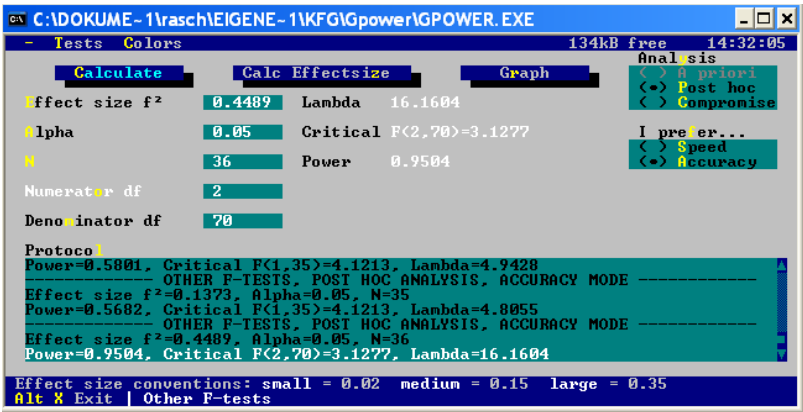
Die mittlere Korrelation zwischen den Stufen des Faktors lässt sich über die Fishers-Z Transformation ermitteln. Benutzen Sie dafür Tabelle D in Band I. Es ergibt sich folgender mittlerer Z-Wert:

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

Die mittlere Korrelation von r = 0,865 setzen Sie nun in die Berechnung des von GPower geforderten f'²-Werts ein. Wie groß ist die Teststärke der Studie, bei N = 36 Personen einen kleinen Effekt für unabhängige Stichproben von $\Omega^2 = 0{,}01$ bei einem Signifikanzniveau von 5% zu finden, falls dieser existiert (q = 3)?

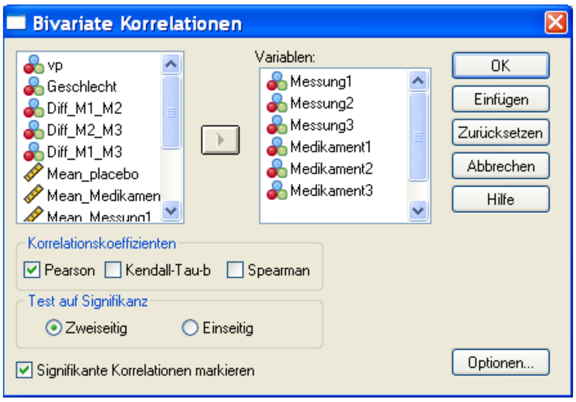
$$f'^2 = \frac{p \cdot q}{1 - \bar{r}} \cdot \frac{\Omega^2_{\text{unabhängig}}}{1 - \Omega^2_{\text{unabhängig}}} = \frac{2 \cdot 3}{1 - 0{,}865} \cdot \frac{0{,}01}{1 - 0{,}01} = 44{,}44 \cdot 0{,}0101 = 0{,}4489$$

$df_B = q - 1 = 3 - 1 = 2$ und $df_{B \times Vpn} = (q-1) \cdot (n-1) = (3-1) \cdot (36-1) = 70$



Die Teststärke, einen kleinen Effekt für unabhängige Stichproben in der vorliegenden Studie in dem messwiederholten Faktor B „wiederholte Testdurchführung" zu finden, war 95,04%.

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 diesem Kapitel). Können wir bei dem vorliegenden Ergebnis mit ausreichender Sicherheit ausschließen, dass ein mittlerer Effekt für unabhängige Stichproben 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.



Korrelationen

		Messung1	Messung2	Messung3	Medikament1	Medikament2	Medikament3
Messung1	Korrelation nach Pearson	1	,661**	,649**	,705**	,795**	,638**
	Signifikanz (2-seitig)		,000	,000	,000	,000	,000
	N	36	36	36	36	36	36
Messung2	Korrelation nach Pearson	,661**	1	,799**	,548**	,568**	,490**
	Signifikanz (2-seitig)	,000		,000	,001	,000	,002
	N	36	36	36	36	36	36
Messung3	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.

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

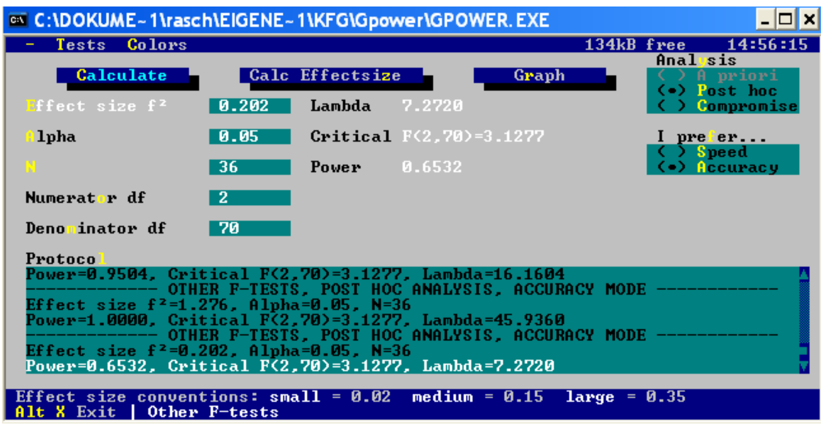
$$\overline{Z} = 0{,}8711 \qquad \rightarrow \qquad \bar{r} = 0{,}7$$

$$f'^2 = \frac{p \cdot q}{1 - \bar{r}} \cdot \frac{\Omega^2_{\text{unabhängig}}}{1 - \Omega^2_{\text{unabhängig}}} = \frac{2 \cdot 3}{1 - 0{,}7} \cdot \frac{0{,}06}{1 - 0{,}06} = 20 \cdot 0{,}0638 = 1{,}276$$

$$df_{A \times B} = (p-1) \cdot (q-1) = 2 \qquad \text{und} \qquad df_{A \times B \times Vpn} = (p-1) \cdot (q-1) \cdot (n-1) = 70$$

Einsetzen in GPower 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, das ein mittlerer Effekt (für unabhängige Stichproben) im Fall der Wechselwirkung existiert. Kleine Effekte können wir allerdings nur mit 65,32%iger Sicherheit ausschließen.

$$f'^2 = \frac{p \cdot q}{1 - \bar{r}} \cdot \frac{\Omega^2_{\text{unabhängig}}}{1 - \Omega^2_{\text{unabhängig}}} = \frac{2 \cdot 3}{1 - 0{,}7} \cdot \frac{0{,}01}{1 - 0{,}01} = 20 \cdot 0{,}0101 = 0{,}202$$



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 Effekt mit einer ausreichenden Sicherheit ausschließen zu können. Effekte mittlerer Größe existieren aber mit hoher Wahrscheinlichkeit nicht.

Berechnen der Effektgröße f^2 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 Schärfe“ in SPSS.

Literatur

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