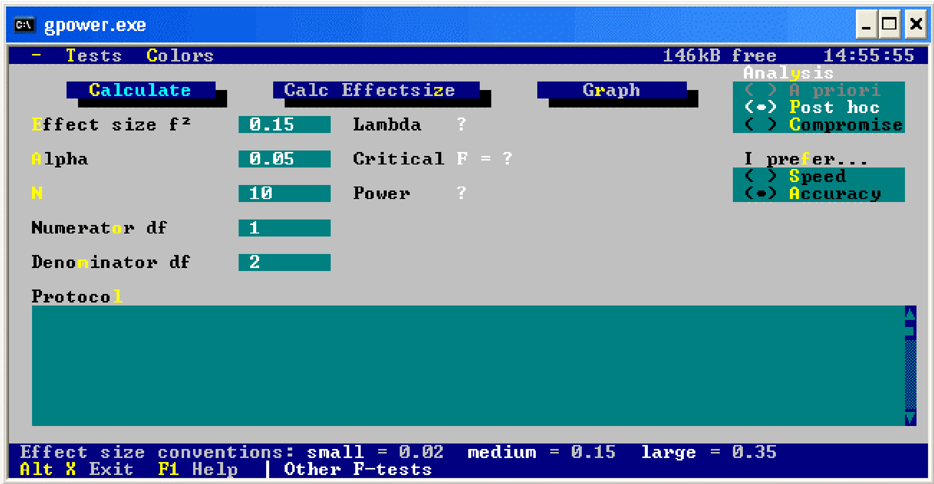
Kapitel 6: Zweifaktorielle Varianzanalyse

Berechnen der Teststärke a priori bzw. Stichprobenumfangsplanung

Nach dem Starten von GPower müssen Sie zunächst unter Tests „Other F-Tests“ auswählen. Sie erhalten nun folgendes Eingabefenster:



Leider bietet GPower für die zweifaktorielle Varianzanalyse 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.

Bitte ignorieren Sie in diesem Fenster die von GPower angegebenen Konventionen der Effektstärken. Wir verwenden hier weiterhin die folgenden Größen:

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

Umrechung[1]:

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

Zur veranschaulich der Stichprobenumfangsplanung einer zweifaktoriellen Varianzanalyse ohne Messwiederholung soll das bekannte Beispiel der Einfluss der Verarbeitungsbedingung („strukturell“, „bildhaft“ und „emotional“) auf die Erinnerungsleistung dienen. Zusätzlich soll der Einfluss des Faktors „Geschlecht“ untersucht werden. Es liegt also eine 3 x 2 Varianzanalyse vor. Eine derartige Analyse unterschiedet drei Arten von Effekten: Der Haupteffekt der Verarbeitungsbedingung, der Haupteffekt des Geschlechts und die Wechselwirkung zwischen diesen beiden Faktoren. Für alle drei Effektarten müssen Sie eine getrennte

[1] Anmerkung: Die Konventionen wurden von Cohen (1988) für f^2 definiert. Die Angaben für Ω^2 stellen daher gerundete Werte dar, deren Umrechnung nach Φ^2 zu Ungenauigkeiten führt.

Stichprobenumfangsplanung vornehmen, es sei denn ihre Hypothese bezieht sich nur auf eine bestimmte Effektart.

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

Zählerfreiheitsgrade:

Haupteffekt A: $df_A = p - 1$ → HE „Verarbeitungsbedingung“: $df = 3 - 1 = 2$

Haupteffekt B: $df_B = q - 1$ → HE „Geschlecht“: $df = 2 - 1 = 1$

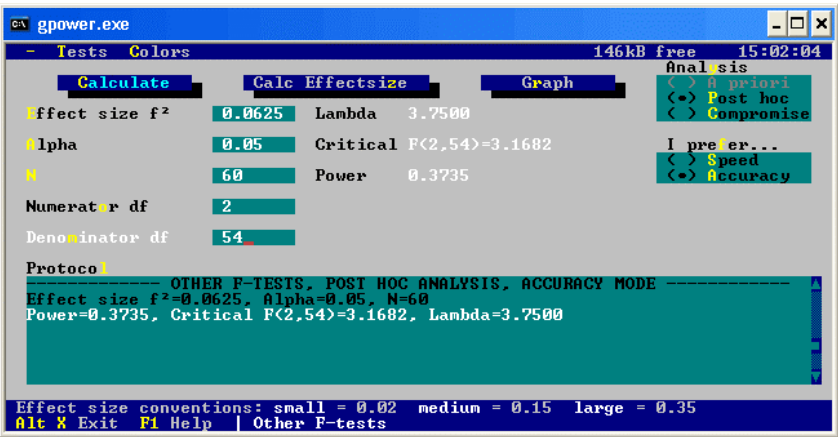
Wechselwirkung A×B: $df_{A \times B} = (p - 1) \cdot (q - 1)$ → WW: $df = (3 - 1) \cdot (2 - 1) = 2$

Nennerfreiheitsgrade:

$df_{Res} = p \cdot q \cdot (n - 1)$[2]

Für die Stichprobenumfangsplanung legen wir einen mittleren Effekt als inhaltlich relevant fest ($\Omega^2 = 0{,}06$). Das Signifikanzniveau beträgt 5%. Die Teststärke, einen mittleren Effekt zu entdecken falls er wirklich existiert, soll mindestens 80% betragen.

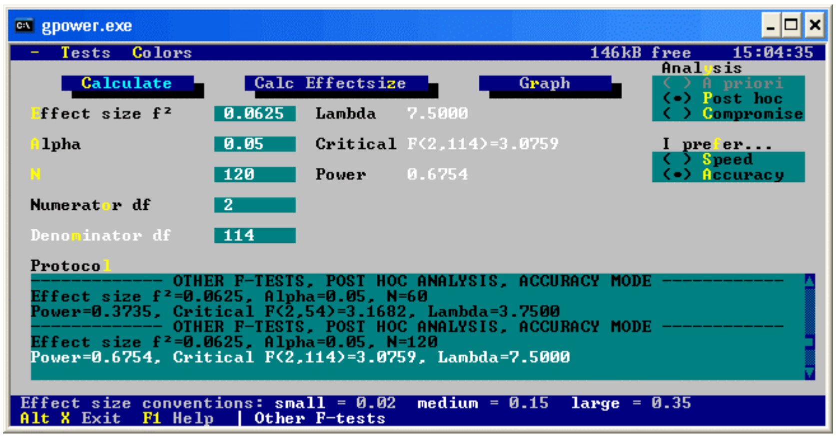
Für die Stichprobenumfangsplanung für den Haupteffekt „Verarbeitungsbedingung“ tragen Sie zunächst für f^2 den Wert 0,0625 ein. Das Signifikanzniveau Alpha ist bereits auf 5% voreingestellt. Für die Versuchspersonenzahl N können wir zunächst mit einer 60 beginnen (20 Personen in jeder der drei experimentellen Bedingungen, 10 davon männlich, 10 weiblich). Die Zählerfreiheitsgrade sind bei diesem Haupteffekt für den dreistufigen Faktor „Verarbeitungsbedingung“ df = 2. Die Nennerfreiheitsgrade ergeben sich nach der obigen Formel zu $df = 3 \cdot 2 \cdot (10 - 1) = 54$.



Die Teststärke ist mit $1-\beta = 0{,}3735$ zu niedrig. Sie müssen nun die Versuchspersonenzahl und die Nennerfreiheitsgrade sukzessive erhöhen, um die gewünschte Teststärke von 80% zu erreichen. Es empfiehlt sich, einen Stichprobenumfang zu wählen, der durch die Anzahl der Bedingungskombinationen teilbar ist, in dem Beispiel also durch Sechs. So lassen sich gleich große Gruppen bilden.

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

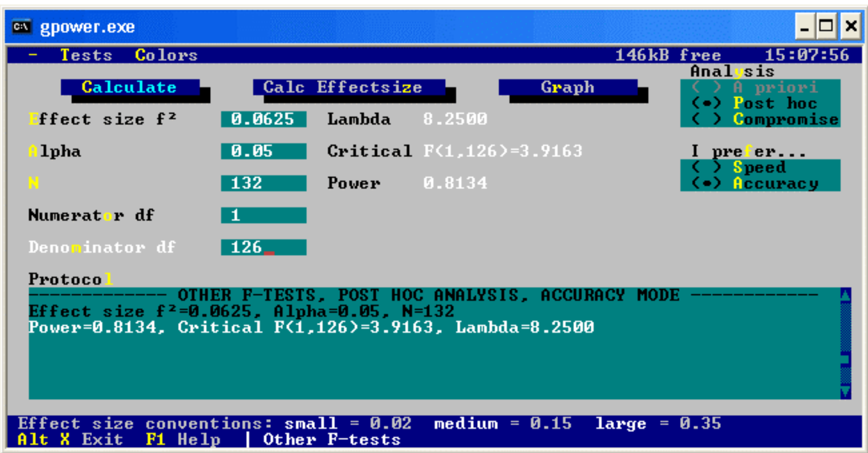
Probieren Sie z.B. die Versuchspersonenzahl N = 120.



Auch hier ist die Teststärke noch zu klein. Ein befriedigendes Ergebnis erhalten wir erst bei N = 162.

Um einen Effekt mittlerer Größe für den Haupteffekt „Verarbeitungsbedingung“ mit seinen drei Stufen mit 80%iger Wahrscheinlichkeit zu finden, falls er existiert, sind also mindestens 162 Versuchspersonen notwendig (27 pro Bedingungskombination).

Für den Haupteffekt „Geschlecht“ mit zwei Stufen ($df_{Zähler}$ = 1) ergibt sich ein benötigter Stichprobenumfang von N = 132.



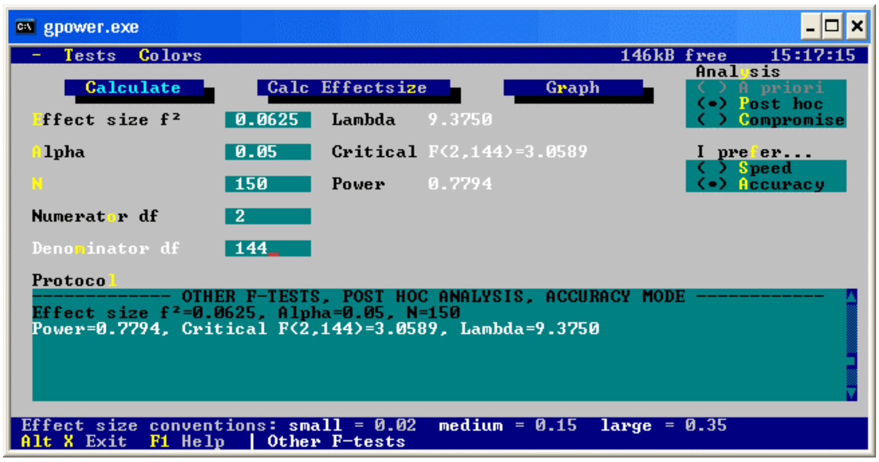
Die Berechnung des optimalen Stichprobenumfangs für die Wechselwirkung kommt zu dem gleichen Ergebnis wie die für den Haupteffekt „Verarbeitungsbedingung“, da sich die Zählerfreiheitsgrade der beiden Effektarten entsprechen.

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

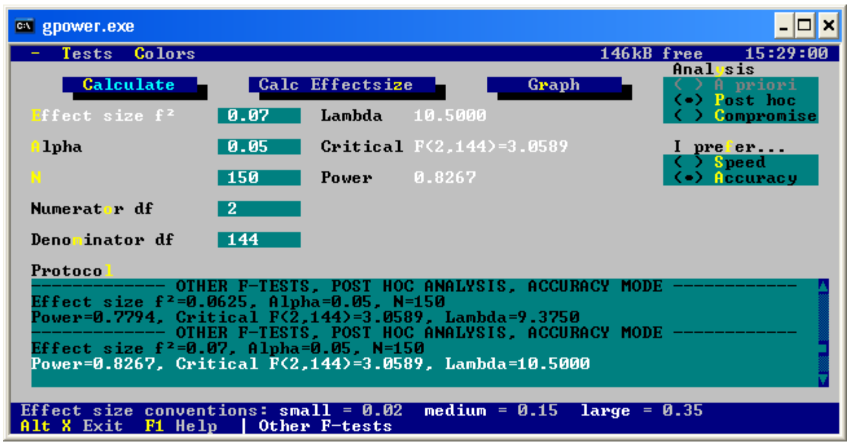
Teststärkebestimmung a posteriori

Ist in einer Untersuchung ohne vorangegangene Stichprobenumfangsplanung ein nicht signifikantes Ergebnis aufgetreten, so ist eine Teststärkenbestimmung a posteriori unerlässlich. In den SPSS Ergänzungen zu diesem Kapitel ergab sich für das Beispiel des Einflusses der Verarbeitungsbedingung und des Geschlechts eine nicht signifikante Wechselwirkung. Wie groß war die Teststärke für eine Wechselwirkung in dieser Untersuchung?

An der Untersuchung haben insgesamt 150 Versuchspersonen teilgenommen. Das Signifikanzniveau war 5%. Die Zählerfreiheitsgrade der Wechselwirkung sind bei einer 3×2 ANOVA df = 2. Ein mittlerer Effekt von $\Omega^2 = 0{,}06$ ($f^2 = 0{,}0625$) gelte als inhaltlich relevant. Wie groß war die Wahrscheinlichkeit, einen Effekt dieser Größe zu finden, falls er wirklich existiert?



Die Teststärke, einen mittleren Effekt der Wechselwirkung in dieser Untersuchung zu finden, betrug 77,94%. Dies ist noch unter der empfohlenen Grenze für die minimal akzeptable Teststärke von 80%. Die Entscheidung, die Existenz eines mittleren Effekts auf Grund des vorliegenden Ergebnisses auszuschließen, ist also mit einer Wahrscheinlichkeit von 22,06% falsch. Allerdings können Effektstärken von $f^2 = 0{,}07$ oder größer bereits mit einer Sicherheit von über 80% ausgeschlossen werden. Große Effekt von $f^2 = 0{,}16$ sind sogar mit über 99%iger Sicherheit nicht vorhanden.



Auf Grund der nicht signifikanten Wechselwirkung kann also mit über 80%iger Sicherheit die Aussage getroffen werden, dass Effekte der Größe $f^2 = 0{,}07$ oder größer nicht existieren.

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

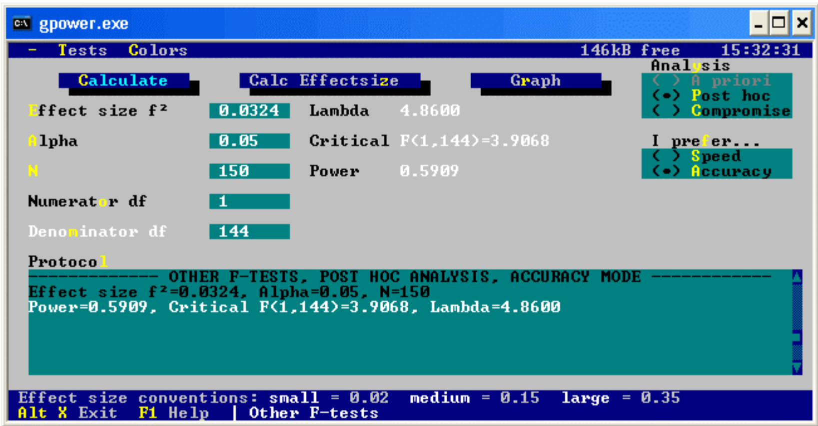
Berechnen der Effektgröße f² aus empirischen Daten und Bestimmung der beobachteten Teststärke

Die Berechnung der Effektgröße f² ist bereits aus Kapitel 5 bekannt:

$$f^2 = \frac{(F_{(df_{Zähler};df_{Nenner})} - 1) \cdot df_{Zähler}}{N}$$

Durch Einsetzen dieses Wertes zusammen mit dem Signifikanzniveau α, der Versuchspersonenzahl N und den Zähler- und Nennerfreiheitsgraden ergibt sich die beobachtete Teststärke für einen aus vorliegenden Daten geschätzten Effekt. So ergibt sich in dem Beispiel folgende beobachtete Teststärke für den Faktor Geschlecht ($F_{(1;144)} = 5{,}861$):

$$f^2 = \frac{(F_{(df_{Zähler};df_{Nenner})} - 1) \cdot df_{Zähler}}{N} = \frac{(5{,}861 - 1) \cdot 1}{150} = 0{,}0324$$



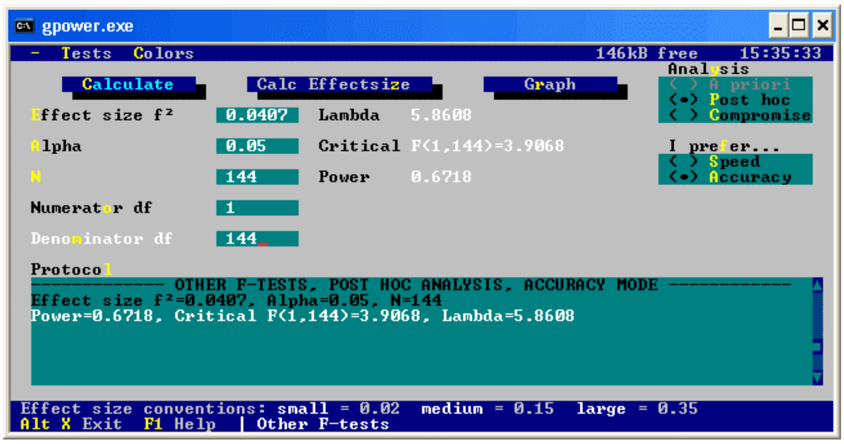
Der Wert der beobachteten Teststärke von 59,09% für den empirisch geschätzten Effekt f² = 0,0324 ist kleiner als der von SPSS angegebene Wert für die „Beobachtete Schärfe" (siehe SPSS Ergänzungen zu Kapitel 6). Dies hat zwei Gründe:

1. SPSS verwendet Effektgrößenschätzer auf der Stichprobenebene. Sie fallen größer als Effektgrößenschätzungen auf der Ebene der Population. Die Formel für die Effektschätzung auf der Stichprobenebene lautet:

$$f_S^2 = \frac{F_{(df_{Zähler};df_{Nenner})} \cdot df_{Zähler}}{df_{Nenner}} = \frac{5{,}861 \cdot 1}{144} = 0{,}0407 \qquad \rightarrow \qquad \eta_p = \frac{f_S^2}{1 + f_S^2} = \frac{0{,}0407}{1 + 0{,}0407} = 0{,}0391$$

2. SPSS berechnet die Teststärke anhand der Freiheitsgrade, und nicht über die Versuchspersonenzahl N.

Um den Wert für die „Beobachtete Schärfe" für den Haupteffekt „Geschlecht" von 67,2% in GPower zu reproduzieren (siehe SPSS Ergänzungen zu Kapitel 6), müssen wir also die Effektgröße f_S^2 verwenden, und anstatt N die Anzahl der Nennerfreiheitsgrade eingeben.


gpower.exe
- Tests Colors
146kB free 15:35:33
Analysis
Calculate
Calc Effectsize
Graph
( ) A priori
(•) Post hoc
( ) Compromise
Effect size f²
0.0407
Lambda 5.8608
Alpha
0.05
Critical F(1,144)=3.9068
I prefer...
( ) Speed
(•) Accuracy
N
144
Power 0.6718
Numerator df
1
Denominator df
144
Protocol
------------- OTHER F-TESTS, POST HOC ANALYSIS, ACCURACY MODE ------------
Effect size f²=0.0407, Alpha=0.05, N=144
Power=0.6718, Critical F(1,144)=3.9068, Lambda=5.8608
Effect size conventions: small = 0.02 medium = 0.15 large = 0.35
Alt X Exit F1 Help | Other F-tests

Literatur

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