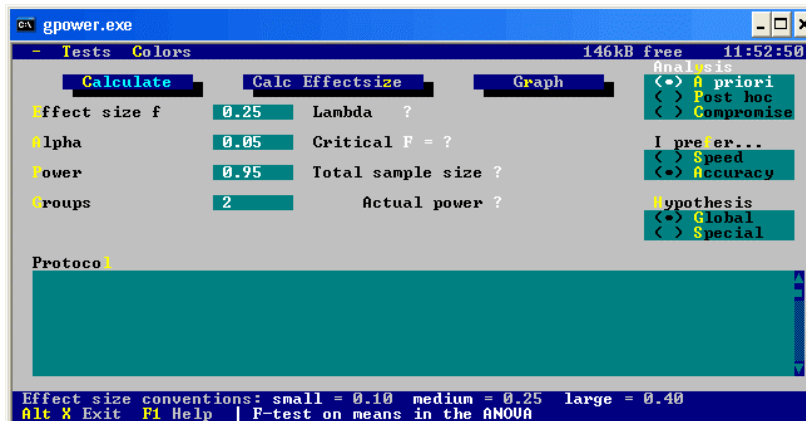


Kapitel 5: Einfaktorielle Varianzanalyse

Berechnen der Teststärke a priori bzw. Stichprobenumfangsplanung

Nach dem Starten von GPower müssen Sie für dieses Kapitel zunächst unter Tests „F-Test (ANOVA)“ auswählen. Sie erhalten nun folgendes Eingabefenster:

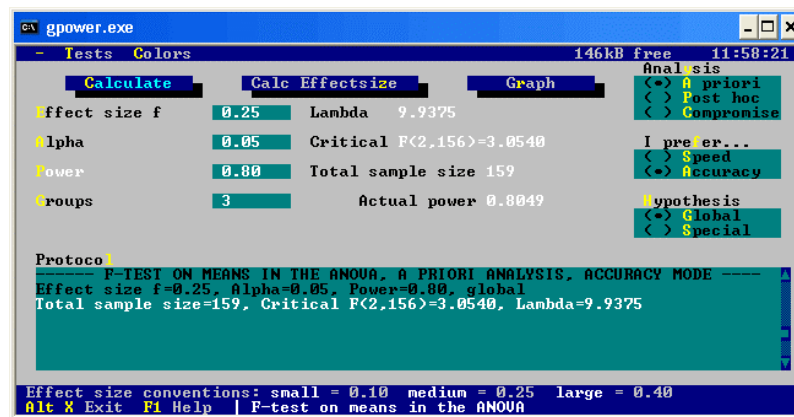


Für die Berechnung der Teststärke a priori sind die Optionen bereits voreingestellt. Bei einer einfaktoriellen Varianzanalyse erfolgt eine globale Signifikanztestung, deshalb sollte auch bei „Hypothesis“ die Option „Global“ ausgewählt bleiben. Ist keine Effektgröße aus Vorstudien oder anderen Untersuchungen aus der Literatur bekannt, so empfiehlt sich eine Orientierung an den von Cohen (1988) vorgeschlagenen Konventionen. Allerdings verwendet das Programm GPower die Effektstärke f im Gegensatz zu Ω^2 . Die in GPower angegebenen Konventionen für f entsprechen aber den in den Kapitel 5.3 berichteten Konventionen für Ω^2 . Die Maße lassen sich nach folgender Formel ineinander umrechnen:

$$f = \sqrt{\Phi^2} = \sqrt{\frac{\Omega^2}{1 - \Omega^2}}$$

Als Beispiel für die Stichprobenumfangsplanung einer einfaktoriellen Varianzanalyse ohne Messwiederholung soll der Vergleich von drei Gruppen dienen. Das Signifikanzniveau ist $\alpha = 0,05$. Wie viele Versuchspersonen werden benötigt, um in diesem experimentellen Design einen mittleren Effekt von $\Omega^2 = 0,06$ bzw. $f = 0,25$ mit einer Teststärke von 80% zu finden, falls er wirklich existiert?

Geben Sie die entsprechenden Angaben in GPower ein, und drücken Sie „Calculate“. Es ergibt sich folgendes Ergebnis:



Um einen mittleren Effekt mit einer Wahrscheinlichkeit von mindestens 80% zu finden, falls er existiert, sollten insgesamt 159 Versuchspersonen untersucht werden, 53 pro Gruppe. GPower gibt auch den Nonzentralitätsparameter λ und den kritischen F-Wert an. Berechnen Sie selber einmal, wie viele Versuchspersonen benötigt werden, um mit demselben experimentellen Design einen kleinen bzw. einen großen Effekt mit einer Wahrscheinlichkeit von 80% zu entdecken! Dies hilft Ihnen, um ein Gefühl für die Abhängigkeit zwischen Teststärke und Effektgröße zu entwickeln.

Teststärkebestimmung a posteriori

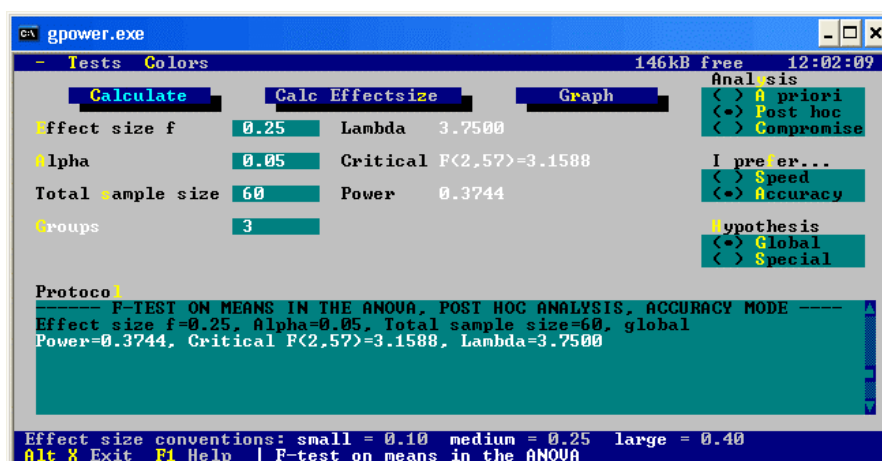
In der Forschungspraxis ist eine Teststärkebestimmung a priori bzw. eine Stichprobenumfangsplanung bedauerlicher Weise noch kein Standard. Häufig wünschen sich Wissenschaftler aber nach einem nicht-signifikanten Ergebnis in einer Untersuchung zumindest eine Antwort auf die Frage, wie groß denn die Chance überhaupt war, den vermuteten Effekt zu finden. Die Teststärkebestimmung a posteriori beantwortet diese Frage.

Stellen Sie sich vor, ein Forscher hat eine Untersuchung mit drei Bedingungen durchgeführt. In jeder dieser experimentellen Bedingungen waren 20 Versuchspersonen. Es haben also 60 Personen an der Studie teilgenommen. Der Forscher hat die Unterschiede zwischen den Gruppen mit Hilfe einer einfaktoriellem Varianzanalyse ausgewertet und findet ein nicht signifikantes Ergebnis. Das Signifikanzniveau war $\alpha = 0,05$. Kann der Forscher auf Grund seines Ergebnisses mit großer Sicherheit ausschließen, dass kein Effekt mittlerer Größe existiert?

Zur Beantwortung dieser Frage schalten Sie GPower für den F-Test um auf eine a posteriori Analyse. Geben Sie die entsprechenden Angaben ein ($f = 0,25$, $\alpha = 0,05$, Total Sample Size = 60, Groups = 3) und drücken Sie „Calculate“. Sie erhalten folgendes Ergebnis:

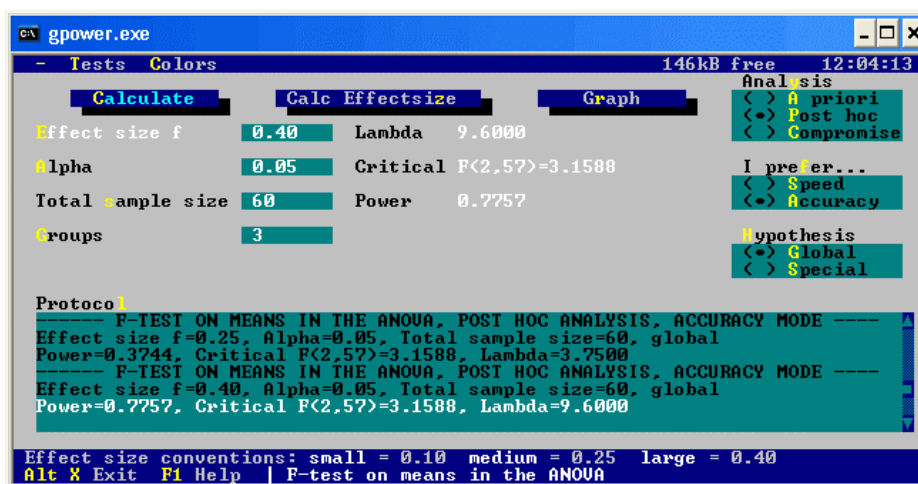
GPower-Ergänzungen

Rasch, Frieze, Hofmann & Naumann (2006). *Quantitative Methoden. Band 2* (2. Auflage). Heidelberg: Springer.



Die Wahrscheinlichkeit, einen mittleren Effekt der Größe $f = 0,25$ ($\Omega^2 = 0,06$) mit einer Versuchspersonenanzahl von $N = 60$ zu finden, falls er existiert, war $1 - \beta = 37,44\%$. Der Forscher konnte also einen Effekt mittlerer Größe mit seiner Anzahl an Versuchspersonen nur mit einer sehr geringen Wahrscheinlichkeit entdecken. Anders ausgedrückt: Die Entscheidung anzunehmen, dass kein Effekt mittlerer Größe existiert, ist mit sehr hoher Wahrscheinlichkeit falsch ($\beta = 62,56\%$). Der Forscher kann also auf Grund seines Ergebnisses nicht behaupten, dass ein Effekt mittlerer Größe nicht existiert, obwohl er ein nicht signifikantes Ergebnis erhalten hat.

Er kann allerdings mit einer höheren Wahrscheinlichkeit ausschließen, dass ein großer Effekt existiert, denn den hätte er eher mit 60 Versuchspersonen finden sollen.



Die Teststärke der Studie mit drei Bedingungen und 60 Versuchspersonen betrug $77,57\%$, um einen großen Effekt von $f = 0,4$ ($\Omega^2 = 0,14$) zu entdecken, falls dieser wirklich existiert. Der Forscher kann also auf Grund seines nicht signifikanten Ergebnisses mit einer Sicherheit von $77,57\%$ ausschließen, dass ein großer Effekt von $f = 0,4$ (oder größer) existiert. Kleinere Effekte kann er dagegen nicht mit einer ausreichenden Sicherheit ausschließen.

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

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

In Kapitel 5.3.2 haben Sie erfahren, wie sich die Effektgröße f^2 aus dem empirischen F-Wert berechnen lässt. Die Daten beruhen auf dem Vergleich zwischen der Erinnerungsleistung von Adjektiven in den Verarbeitungsbedingungen „strukturell“, „bildhaft“ und „emotional“. An dem Versuch haben 150 Personen teilgenommen (50 pro Bedingung), die Unterschiede zwischen den drei Gruppen waren insgesamt signifikant. Für dieses Beispiel wurde folgender Effekt geschätzt:

$$f^2 = \frac{(F_{(df_{\text{Zähler}}, df_{\text{Nenner}} - 1)} - 1) \cdot df_{\text{Zähler}}}{N} = \frac{(21,59 - 1) \cdot 2}{150} = 0,2745$$

Die Effektstärke f ergibt sich als Wurzel aus f^2 :

$$f = \sqrt{f^2} = \sqrt{0,2745} = 0,5239$$

GPower bietet eine weitere Möglichkeit, die Effektstärke f zu schätzen. Nach dem Drücken der Taste „Calc Effectsize“ haben Sie die Möglichkeit, die Mittelwerte der Gruppen und die Streuung der Werte einzutragen (siehe folgende Tabelle).

Deskriptive Statistiken

Abhängige Variable: Gesamtzahl erinnertes Adjektive

Verarbeitungsbedingung	Mittelwert	Standardabweichung	N
strukturell	7,20	3,162	50
bildhaft	11,00	4,140	50
emotional	12,02	4,206	50
Gesamt	10,07	4,368	150

Geben Sie in das Feld „Number of Groups“ eine Drei ein. In dem nächsten Feld wird die Standardabweichung innerhalb der Gruppen verlangt. Sie sollte auf Grund der Annahme der Varianzhomogenität in allen verglichen experimentellen Gruppen gleich sein. Ist dies nicht genau der Fall, muss ein mittlerer Wert angenommen werden, der allerdings bei großen Unterschieden zwischen den Varianzen innerhalb der Gruppen zu Verzerrungen führen kann. Die Mittelung bei drei Versuchsbedingungen sollte nach folgender Formel erfolgen:

$$\bar{\sigma} = \sqrt{\frac{\hat{\sigma}_1^2 + \hat{\sigma}_2^2 + \hat{\sigma}_3^2}{3}} = \sqrt{\frac{3,162^2 + 4,14^2 + 4,206^2}{3}} = 3,8656$$

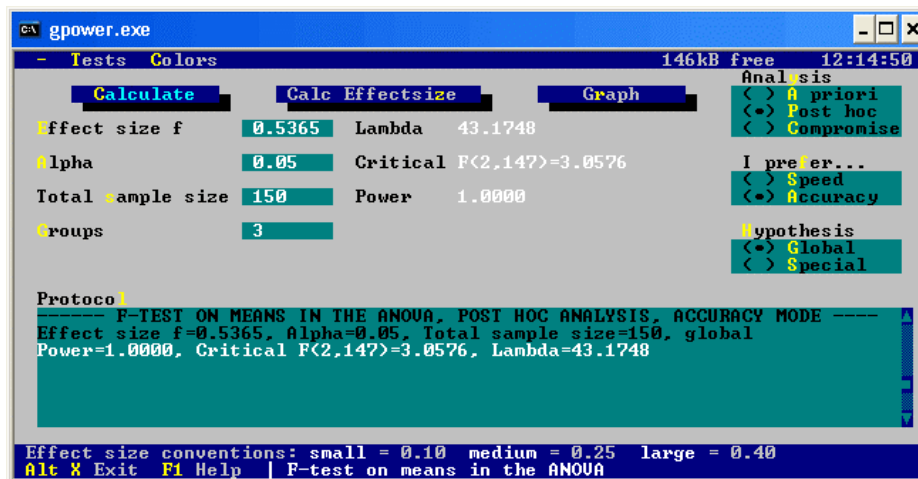
Geben Sie diesen mittleren Wert für die Streuung bei „Sigma (within each group)“ ein (Achtung: Punkt statt Komma!). Führen Sie nun einen Doppelklick auf die Zeile der Gruppe 1 in den unteren Feldern („List“) aus, so dass die Zahlen im Eingabefeld direkt unter den Bezeichnungen „Group“, „Mean“ und „Size n“ stehen. Die entsprechende Zeile ist nun weiß markiert. Geben Sie die Werte für den Mittelwert der Gruppe 1 und die Anzahl an Versuchspersonen in der Gruppe ein. Ein weiterer Doppelklick auf die weiß markierte Zeile in der Liste (!) überträgt die Zahlen vom Eingabefeld in die Liste. Ein Doppelklick auf die Eingabefelder hat dagegen keine Wirkung. Wiederholen Sie den Vorgang, bis Sie folgendes Bild erhalten:

GPower-Ergänzungen

Rasch, Frieze, Hofmann & Naumann (2006). *Quantitative Methoden. Band 2* (2. Auflage). Heidelberg: Springer.



Betätigen Sie nun „Calculate“. Es ergibt sich eine Effektstärke f von 0,5365, der nur geringfügig von dem oben berechneten Wert abweicht. Über die Option Calc & Copy können Sie diesen Wert direkt in das vorherige Fenster für die Teststärkenberechnung einfügen. Die a posteriori Powerberechnung für den aus den empirischen Daten geschätzten F-Wert führt zu folgendem Ergebnis:



Die Wahrscheinlichkeit, einen Effekt der Größe $f = 0,5365$ bei drei Gruppen à 50 Versuchspersonen und einem Signifikanzniveau von 5% zu finden, falls er existiert, war größer als 99%. GPower gibt hier den asymptotischen Wert von 1 an. Diese a posteriori Berechnung der Teststärke für einen aus vorliegenden Daten geschätzten Effekt entspricht der Angabe der „Beobachteten Schärfe“ in SPSS (siehe SPSS Ergänzungen für Kapitel 5).

Literatur

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