# Kapitel 3: Der t-Test

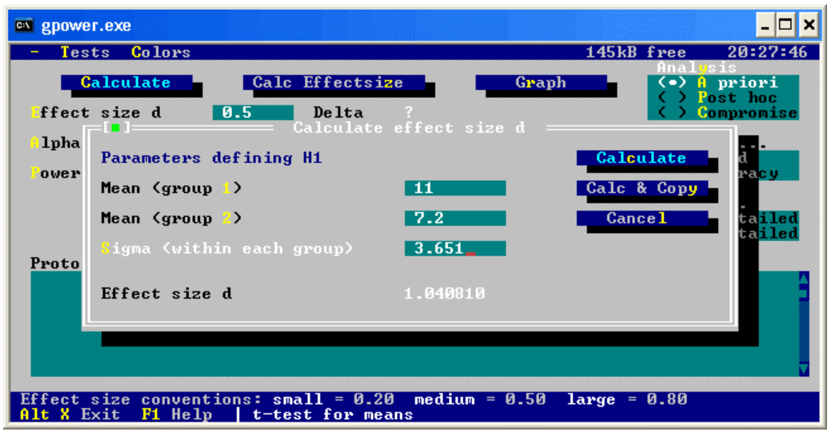
## t-Test für unabhängige Stichproben

### Berechnen der Effektgröße d

In Kapitel 3.3.1 haben Sie erfahren, wie sich die Effektgröße d aus empirischen Werten berechnen lässt. Dazu haben wir den Vergleich der Erinnerungsleistung der Verarbeitungsgruppen „strukturell“ und „bildhaft“ herangezogen. Wir wollen diese Berechnungen an dieser Stelle mit GPower nachvollziehen. Starten Sie dazu GPower, so dass Sie den folgenden Bildschirm vor sich sehen:


Ein Klick auf das Feld „Tests“ oben links verrät die Liste statistischer Verfahren, für die GPower Berechnungen durchführt. Die Standardeinstellung ist t-Test (means). Dies ist gleichbedeutend mit t-Test für unabhängige Stichproben. Wir können also von diesem Bildschirm aus mit unserer Analyse starten.

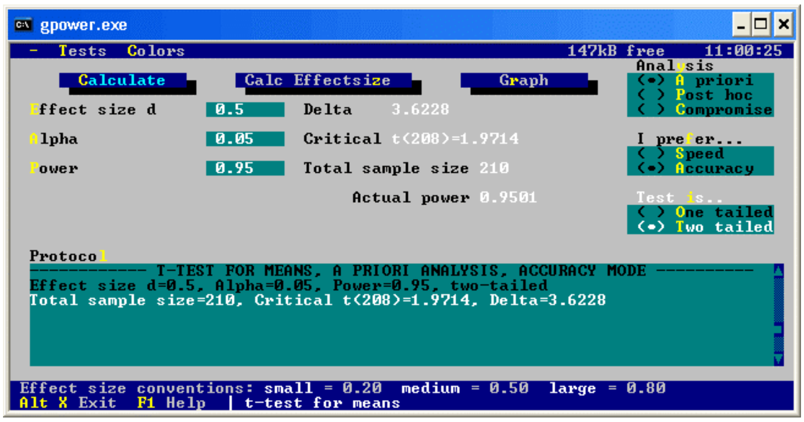
In der Mitte am oberen Bildschirmrand öffnen Sie das Feld Calc Effectsize. Dort geben Sie die Mittelwerte und die aus den Daten geschätzte Populationsstreuung an. Aus dem Datensatz bzw. aus Kapitel 3.3 können Sie diese Werte entnehmen: $(\bar{x}_{bildhaft} = 11;\ \bar{x}_{strukturell} = 7{,}2;\ \hat{\sigma}_x = 3{,}651)$. Beachten Sie, dass GPower ein Komma nicht akzeptiert und statt dessen einen Punkt erwartet. Sie erhalten den in Kapitel 3.3.1 berechneten Wert für d auch über die Berechnung mit GPower, d = 1,04. Dies ist den Konventionen von Cohen (1988) folgend ein großer Effekt. Die Konventionen sind in GPower am unteren Bildschirmrand eingeblendet.



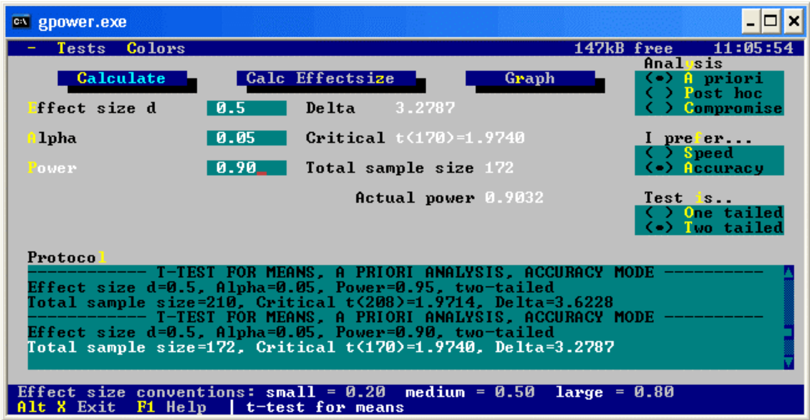

## Berechnen der Teststärke a priori bzw. Stichprobenumfangsplanung

GPower startet mit der Oberfläche für den t-Test für unabhängige Stichproben. Oben rechts sehen Sie die Optionen A Priori, Post hoc und Compromise. Die Standardeinstellung ist A Priori. Die Wahl zwischen Speed und Accuracy stammt aus einer Zeit, als die Computer noch wesentlich weniger leistungsfähig waren. Belassen Sie diese Einstellung immer auf Accuracy. Etwas weiter unten können Sie einstellen, ob Sie die Stichprobengröße für einen einseitigen oder zweiseitigen Test berechnen möchten.

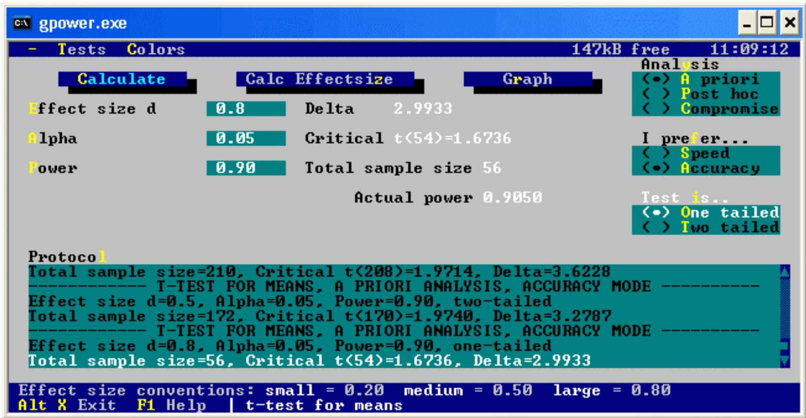
Die Stichprobenumfangsplanung kommt dann zum Zuge, wenn ein Forscher eine Untersuchung plant und wissen möchte, wie viele Personen er unter den gegebenen Annahmen rekrutieren muss, um auf jeden Fall ein interpretierbares Ergebnis zu erhalten. Der Forscher erwartet für seine ungerichtete Fragestellung einen mittleren Effekt von $d = 0{,}50$. Er setzt das Signifikanzniveau auf $\alpha = 0{,}05$ und möchte außerdem, dass der $\beta$-Fehler auch nicht größer ist. Dies würde in einer Teststärke von $1\text{-}\beta = 0{,}95$ resultieren. GPower errechnet für diese spezifische Konstellation einen Bedarf von 210 Versuchspersonen, also 105 Personen in jeder Gruppe.




Würde sich der Forscher mit einer Teststärke von 90% zufrieden geben, würde sich die benötigte Anzahl Versuchspersonen auf 172 reduzieren.




Ein anderer Forscher nimmt für seine Untersuchung einen großen Effekt zwischen den Gruppen an. Er verfolgt eine gerichtete Fragestellung und ist bereit, einen 10%igen β-Fehler zu akzeptieren. GPower berechnet einen benötigten Stichprobenumfang von 56 Personen.



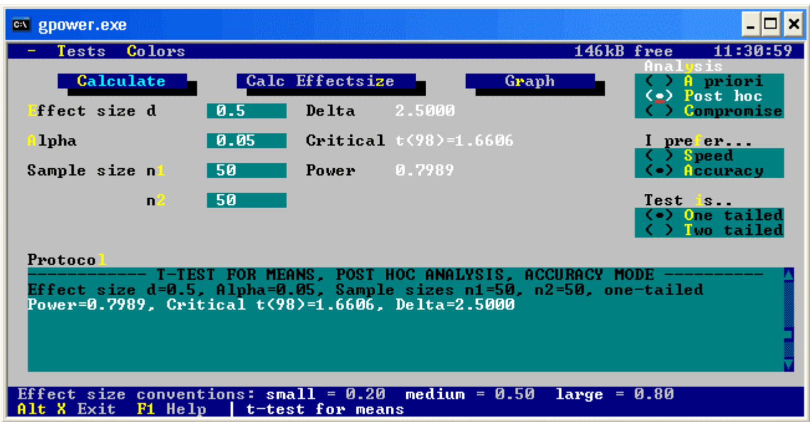

An diesen Beispielen können Sie sehr anschaulich nachvollziehen, wie sich die vier Determinanten eines statistischen Tests gegenseitig bedingen. Sind drei von ihnen festgelegt, ist auch die letzte eindeutig bestimmt. Wir möchten Sie ermutigen, selber einige Beispiele mit GPower zu rechnen, um zu sehen, wie die Veränderung einer Determinante den benötigten Stichprobenumfang beeinflusst: Ein großer angenommener Effekt verringert den Stichprobenumfang, während ein kleiner ihn erhöht. Eine geringere Teststärke erfordert weniger Versuchsteilnehmer als eine hohe Teststärke. Ein liberaleres α-Niveau verlangt ein kleineres N als ein strenges (siehe Kapitel 3.4.2).

## Teststärkebestimmung a posteriori

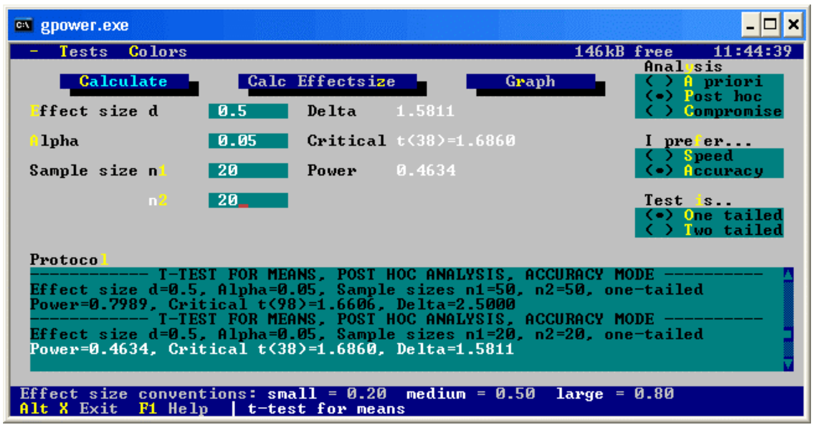
In der Forschungspraxis ist eine Teststärkebestimmung a priori bzw. eine Stichprobenumfangsplanung bedauerlicher Weise noch kein Standard. Häufig wünschen sich Wissenschaftler aber nach einem nicht signifikanten Ergebnis in einer Untersuchung zumindest eine Antwort auf die Frage, wie groß denn die Chance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

Auch ein weiterer Fall verhilft dieser Analyse zur häufigen Anwendung: In der Realität ist es häufig so, dass Forscher schon vor einer Untersuchung wissen, wie viele Versuchspersonen sie für die Studie erheben können. Gründe dafür sind z.B. ein begrenzter Zugang zu finanziellen Mitteln oder Räumen zur Datenerhebung. In diesem Fall bietet die Teststärkebestimmung a posteriori trotz ihres Namens schon vor der Durchführung eine Entscheidungshilfe für die Frage, ob sich die Datenerhebung überhaupt lohnt.

Ein Forscher hat eine Untersuchung mit je 50 Versuchspersonen in zwei Gruppen durchgeführt. Er hatte eine gerichtete Hypothese und vermutete einen mittleren Effekt hinter dem untersuchten Phänomen. Das Ergebnis war allerdings auf dem 5%-Niveau nicht signifikant. Mit Hilfe von GPower ermittelt er eine empirische Teststärke von knapp 80%. Der β-Fehler lag also bei 20%. Sollte er seine Hypothese auf Grund dieser Daten verwerfen und einen Nullunterschied zwischen den Gruppen annehmen, würde er mit 20%iger Wahrscheinlichkeit einen Fehler machen. Eine Power von 80% gilt als gerade noch akzeptabel.




Ein anderer Forscher weiß, dass er nur 20 Versuchspersonen pro Bedingung erheben kann. Er nimmt ebenfalls einen mittleren Effekt an und setzt das α-Niveau auf 5% für seine gerichtete Fragestellung. Wenn er diese Untersuchung durchführen möchte, muss er mit einer Teststärke von weniger als 50% vorlieb nehmen. Er könnte also ebenso gut eine Münze werfen.




Auch an diesen Beispielen sehen Sie, wie sich die vier Determinanten des t-Tests gegenseitig bedingen. Probieren Sie ein wenig aus, welche Auswirkungen es auf die Teststärke hat, wenn Sie die Effektgröße, das α-Niveau und/oder die Stichprobengröße verändern!

## t-Test für abhängige Stichproben

### Berechnen der Effektgröße $d_z$

Auch für abhängige Stichproben gibt es eine Effektgröße d. Um deutlich zu machen, dass sie die Effektgröße für abhängige Stichproben ist, heißt sie $d_z$. Die Berechnung von $d_z$ ist in GPower nicht möglich. Allerdings ist diese Berechnung leicht per Hand auszuführen.

Es gilt: $d_z = \frac{\bar{x}_d}{\sigma_d}$,

wobei $\bar{x}_d$ der Mittelwert der Differenzen ist und $\sigma_d$ die Streuung dieser Differenzen (Cohen, 1988). Sowohl der Mittelwert der Differenzen als auch die Streuung der Differenzen sind dem SPSS-Output eines t-Tests für abhängige Stichproben zu entnehmen.

So wie die Berechnung des t-Werts zur Prüfung auf Signifikanz ist auch die Berechnung des Effektstärkenmaßes $d_z$ beeinflusst durch die Abhängigkeit der beiden Messzeitpunkte bei einem t-Test für abhängige Stichproben. Diese muss also bei der Berechnung des Maßes berücksichtigt werden. Im Fall von $d_z$ ist die relevante Information über die Abhängigkeit der Daten in der Streuung der Differenzen verwoben.

Mathematisch lässt sich zeigen, dass $d_z$ eng verwandt ist mit der in Kapitel 3.5 diskutierten Effektgröße $f^2_{s(\text{abhängig})}$ ist. Es gilt die Beziehung $d_z = \sqrt{f^2_{s(\text{abhängig})}} = f_{s(\text{abhängig})}$.

### Berechnen der Teststärke a priori bzw. Stichprobenumfangsplanung

Die a priori Bestimmung des Stichprobenumfangs ist in GPower für den t-Test für abhängige Stichproben nicht möglich. Es gibt aber unter Einschränkungen die Möglichkeit, sich durch ein iteratives Verfahren im Post hoc Modus der gewünschten Teststärke zu nähern (siehe unten).

### Teststärkebestimmung a posteriori

Um die Teststärke für einen t-Test für abhängige Stichproben zu bestimmen, ist „Other t-Tests“ die richtige Option. Als Effektgröße ist f aufgeführt mit entsprechenden Konventionen nach Cohen (1988). Für unabhängige Stichproben gilt $d = 2 \cdot f$, deshalb sind die angegebenen Konventionen hier halb so groß im Vergleich zu der Option „t-Test Means“. Allerdings ist es wichtig zu beachten, *dass diese Konventionen nur für unabhängige Stichproben zutreffen*. Bei abhängigen Stichproben muss die Größe der Korrelation in den Effekt einberechnet werden, um die Teststärke für einen kleinen, mittleren oder großen Effekt bestimmen zu können. Leider gibt es keine Möglichkeit, in GPower die Korrelation direkt einzugeben. Wir behelfen uns, indem wir in das Programm eine veränderte Effektstärke $d_z$’ oder f’ eingeben, in die wir den Einfluss der Korrelation per Hand einrechnen. Dies geschieht beim t-Test für abhängige Stichproben nach folgender Formel:

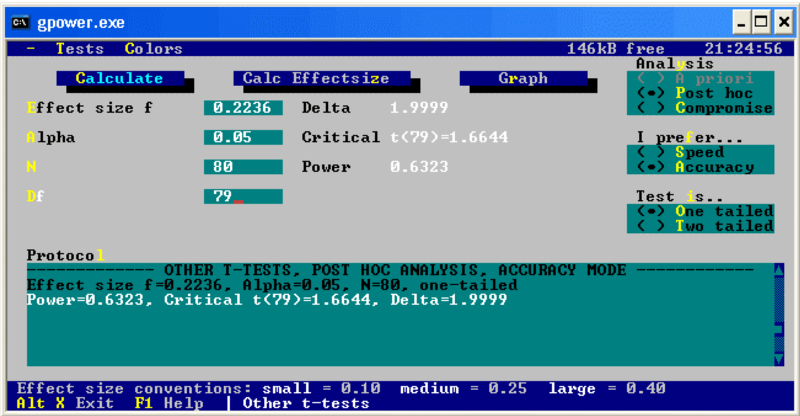
$$d_z' = f' = \sqrt{\frac{2}{1-r}} \cdot f_{\text{unabhängig}} = \sqrt{\frac{2}{1-r}} \cdot \frac{d_{\text{unabhängig}}}{2} \quad \text{(siehe Kapitel 3.5)}$$

Eine Untersuchung mit einer gerichteten Fragestellung und 80 Versuchspersonen liefert bei einem α-Niveau von 5% ein nicht signifikantes Ergebnis. Wie groß war die Teststärke, einen kleinen Effekt von $d_{\text{unabhängig}} = 0{,}2$ in dieser Studie zu entdecken? Für die Berechnung der Teststärke ist die Korrelation zwischen den Stichproben mit entscheidend. Je höher die Korrelation, desto größer die Teststärke für einen bestimmten Effekt. In der vorliegenden Untersuchung betrug die Korrelation zwischen den Stichproben r = 0,6.

Zunächst müssen wir das $d_z$’ bzw. f’ ausrechnen, dass wir in GPower eingeben müssen, um dem Programm eine Effektstärke unter Berücksichtigung der Korrelation zwischen den Stichproben mitzuteilen.

$$d_z' = f' = \sqrt{\frac{2}{1-r}} \cdot f_{\text{unabhängig}} = \sqrt{\frac{2}{1-r}} \cdot \frac{d_{\text{unabhängig}}}{2} = \sqrt{\frac{2}{1-0{,}6}} \cdot \frac{0{,}2}{2} = 0{,}2236$$

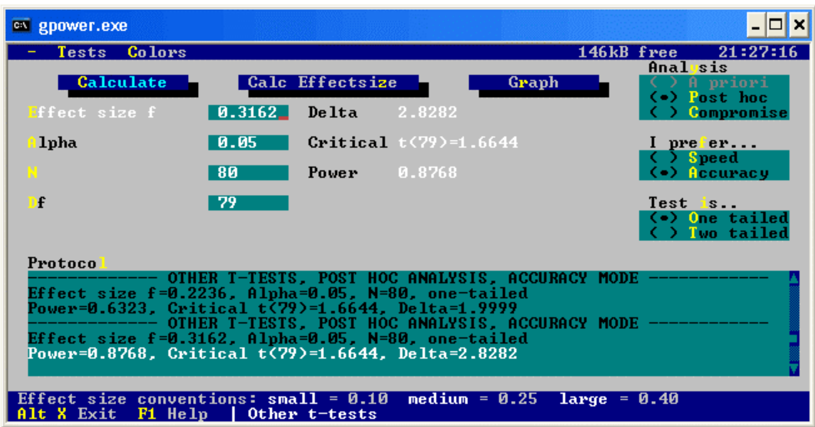
Diesen Wert für $d_z$’ geben wir nun in GPower ein, zusammen mit einem alpha-Niveau von 5%, einer Versuchspersonenzahl von 80 und 79 Freiheitsgraden.




Der Test hatte nur eine Teststärke von etwa 63%.

An diesem Beispiel lässt sich sehr gut der Einfluss der Höhe der Korrelation auf die Teststärke verdeutlichen. So hätte sich in derselben Studie ein kleiner Effekt von $d_{\text{unabhängig}} = 0{,}2$ mit 80 Versuchspersonen mit einer eher akzeptablen Teststärke entdecken lassen, wenn die Korrelation z.B. r = 0,8 gewesen wäre.

$$d_z' = f' = \sqrt{\frac{2}{1-r}} \cdot f_{\text{unabhängig}} = \sqrt{\frac{2}{1-r}} \cdot \frac{d_{\text{unabhängig}}}{2} = \sqrt{\frac{2}{1-0{,}8}} \cdot \frac{0{,}2}{2} = 0{,}3162$$



Die Teststärke beträgt nun ca. 88%. Zu beachten ist bei diesen Überlegungen natürlich, dass die empirisch auftretende Korrelation vor einer Untersuchung nie bekannt ist. Sie kann lediglich auf Basis von vorherigen Studien geschätzt werden.

Liegt keine Korrelation (r = 0) zwischen den Stichproben vor, so ergibt sich mit dz' = 0,1414 eine Teststärke von nur 35% (bitte nachrechnen). Dieser Wert für die Teststärke bei r = 0 entspricht der Teststärke eines t-Tests für unabhängige Stichproben für einen kleinen Effekt $d_{\text{unabhängig}} = 0{,}2$ bei einer identischen Anzahl von Messwerten. Im Fall abhängiger Stichproben geben die 80 Versuchspersonen jeweils zwei Messwerte ab. Die Teststärke entspricht also der eines t-Tests für unabhängige Stichproben mit 160 Personen, also 80 Personen pro Gruppe.

## Berechnung des Stichprobenumfangs vor einer Untersuchung

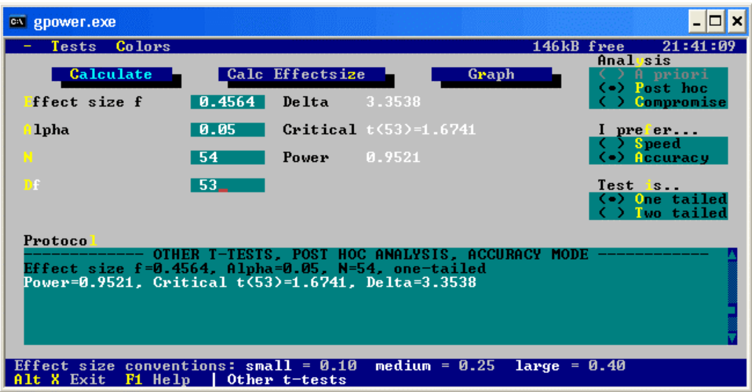
Die Methode A priori ist für den t-Test für abhängige Stichproben nicht verfügbar. Allerdings lässt sich über die Post hoc Methode in einem iterativen Verfahren und unter Annahme einer Korrelation zwischen den Messwertreihen der benötigte Stichprobenumfang ermitteln. Dieser Ansatz hat einen großen Nachteil: Vor einer Untersuchung ist die Korrelation zwischen den Messwertreihen nie bekannt. Sie lässt sich bestenfalls auf Basis von anderen veröffentlichten Studien in dem Bereich oder eigenen Vorstudien abschätzen. Dieser Umstand hat zur Folge, dass die auf diese Weise ermittelten a priori Teststärken nur grobe Schätzungen sind und keine exakten Werte.

Ein Forscher hält für seine gerichtete Fragestellung einen mittleren Effekt von $d_{\text{unabhängig}} = 0{,}5$ für inhaltlich relevant. Auf Grund von Vorstudien geht er von einer Korrelation von r = 0,40 zwischen den Messwertreihen aus. Das Signifikanzniveau setzt er auf 5% fest. Wie viele Versuchspersonen braucht er, um eine Power von 95% zu erreichen?

Um diese Frage zu beantworten müssen wir zunächst $d_z$' oder f' berechnen.

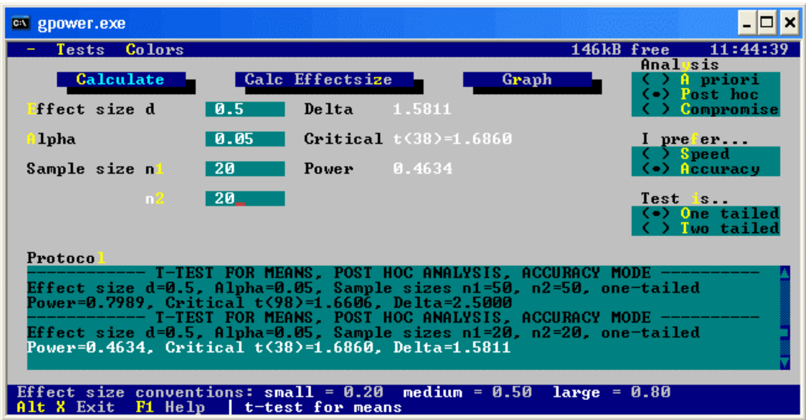
$$d_z' = f' = \sqrt{\frac{2}{1-r}} \cdot f_{\text{unabhängig}} = \sqrt{\frac{2}{1-r}} \cdot \frac{d_{\text{unabhängig}}}{2} = \sqrt{\frac{2}{1-0{,}4}} \cdot \frac{0{,}5}{2} = 0{,}4564$$

Nachdem wir diesen Wert ermittelt und bei GPower eingetragen haben, müssen wir lediglich die Zahl der Versuchspersonen von einem frei wählbaren Ausgangswert sukzessive erhöhen, bis die Power beim angestrebten Wert liegt. Im Beispiel ist dies bei 54 Versuchspersonen der Fall.



## Vergleich von t-Test für unabhängige und abhängige Stichproben sowie Vertiefung des Konzeptes „Abhängigkeit“

Im Abschnitt über den t-Test für unabhängige Stichproben haben wir ein Beispiel betrachtet, in dem ein Forscher wusste, dass er nur 40 Versuchspersonen zur Verfügung hatte. Für seine Studie mit einem angenommenen Effekt mittlerer Größe und einem Signifikanzniveau von 5% ergab sich bei einseitiger Fragestellung eine Teststärke von 46% (siehe Graphik). Würde er die Studie in dieser Form durchführen, wäre das mit einem großen Risiko verbunden, am Ende ohne interpretierbares Ergebnis dazustehen. Gibt es eine Alternative für den Forscher?




Bisher haben wir abhängige Daten als solche bezeichnet, die auf der selben abhängigen Variablen an zwei unterschiedlichen Messzeitpunkten von derselben Person produziert wurden. Das Konzept der Abhängigkeit von Daten greift aber noch weiter. Die Messwiederholung ist nur einer von vielen möglichen Fällen abhängiger Daten. Denken Sie an unser Gedächtnisexperiment. Dort ging es darum, positive, negative und neutrale Wörter zu erinnern. Eine Person, die besonders viele positive Wörter erinnert, wird in aller Regel auch viele negative Wörter erinnern. Die dahinter stehende Eigenschaft „gutes Gedächtnis von Person X“ wirkt sich auf alle drei Wortarten aus. Die Werte für positive, negative und neutrale Wörter kommen also nicht unabhängig voneinander zu Stande, sondern werden alle von der Fähigkeit derselben Person beeinflusst. Sie sind abhängig voneinander. (Beachten Sie, dass diese Ausführungen nichts mit der Einteilung in bildhafte, emotionale und strukturelle Verarbeitung zu tun hat, die wir bislang thematisiert haben.)

Wenn das Ziel darin besteht, herauszufinden, ob es Unterschiede in der Erinnerungsfähigkeit positiver und negativer Adjektive gibt, hat ein Forscher mehrere Möglichkeiten, dieses Ziel zu verfolgen. Zum einen kann er zwei Gruppen bilden, die entweder positive oder negative Adjektive präsentiert bekommen und später abrufen sollen. Die adäquate Auswertungsstrategie für diesen Versuchsaufbau wäre ein t-Test für unabhängige Stichproben. Eine von mehreren anderen Möglichkeiten wäre aber, allen Personen beide Arten von Adjektiven zu präsentieren und später die Daten mit einem t-Test für abhängige Stichproben auszuwerten.

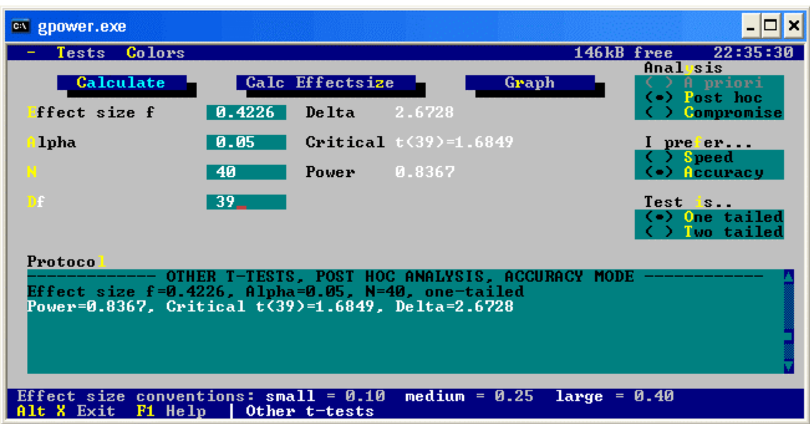
Betrachten Sie ein anderes Beispiel abhängiger Daten: Ein Sozialpsychologe möchte die Einstellung gegenüber der SPD und den Grünen erfassen. Dafür hat er mehrere Möglichkeiten. Zum einen könnte er eine Personengruppe zu ihrer Einstellung gegenüber der SPD befragen und eine andere Gruppe zu ihrer Einstellung gegenüber den Grünen. Von diesen Gruppen könnte der Forscher mit einem t-Test für unabhängige Stichproben die Mittelwerte vergleichen und somit überprüfen, ob er einen Unterschied in den Einstellungen gegenüber beiden Parteien feststellen kann. Eine andere Möglichkeit bestünde darin, *alle* Versuchspersonen zu *beiden* Parteien zu befragen. Allerdings ist es plausibel anzunehmen, dass die Einstellungen einer Person zu den beiden Parteien nicht unabhängig voneinander sind, denn sie sind politisch zwar verschieden, aber verwandt. Eine Person, welche die eine Partei positiv bewertet, hat vermutlich auch eine zumindest ähnliche Einstellung gegenüber der anderen Partei. Die dahinter liegende Eigenschaft „politische Einstellung" würde sich also auf beide Einstellungsangaben dieser Person positiv auswirken. Eine andere Person hingegen ist möglicherweise deutlich konservativer eingestellt und gibt deshalb bei beiden Gruppen wenig positive Einstellungen an. Auch hier kämen also die Daten zu beiden Einstellungsmaßen nicht unabhängig voneinander zu Stande. Mit anderen Worten: Sie sind korreliert (siehe Kapitel 4).

Das Konzept der Abhängigkeit von Daten kann sogar noch weiter gefasst werden. Stellen Sie sich eine Untersuchung mit Zwillingen vor, die im selben Elternhaus aufgewachsen sind. Auch wenn diese sich natürlich voneinander unterscheiden, ist es doch plausibel anzunehmen, dass Zwillingspaare häufig ähnliche Werthaltungen und Ansichten teilen. In diesem Fall lassen sich sogar die Daten von zwei *unterschiedlichen* Personen als abhängig betrachten. Noch einmal im Kontrast dazu der Fall von unabhängigen Stichproben: Hier geht man davon aus, dass sich in den zwei Gruppen unterschiedliche Personen befinden, die in keinem besonderen Verhältnis zueinander stehen. Ihre Daten sind unkorreliert, denn keine zwei Datenpunkte sind von dem selben dahinter stehenden Konstrukt beeinflusst, wie z.B. der Intelligenz einer Person, den motorischen Fähigkeiten einer Person oder auch nur dem gemeinsamen Elternhaus mit ähnlicher Erziehung etc.

Es gibt also wissenschaftliche Fragestellungen, die sowohl mit unabhängigen als auch mit abhängigen Stichproben untersucht werden können. Welche Auswirkungen hat die Entscheidung für die eine oder andere Vorgehensweise auf die Teststärke? Betrachten wir das obige Beispiel noch einmal, in dem ein Forscher eine schwache Teststärke von 46% mit den ihm zur Verfügung stehenden Mitteln erzielen konnte. Welche Teststärke würde erzielt, wenn die Untersuchung an abhängigen Stichproben durchgeführt würde, die zu $r = 0{,}30$ miteinander korrelieren? (t-Test für abhängige Stichproben: $N = 40$, $\alpha = 5\%$, angenommenes $d_{\text{unabhängig}} = 0{,}5$, einseitige Fragestellung.)

Zunächst müssen wir $d_z$' bzw. f' ermitteln.

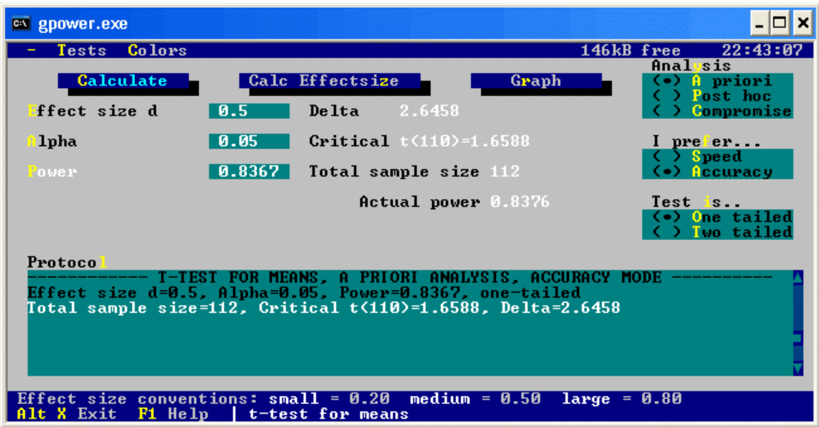
$$d_z' = f' = \sqrt{\frac{2}{1-r}} \cdot f_{\text{unabhängig}} = \sqrt{\frac{2}{1-r}} \cdot \frac{d_{\text{unabhängig}}}{2} = \sqrt{\frac{2}{1-0{,}30}} \cdot \frac{0{,}5}{2} = 0{,}4226$$

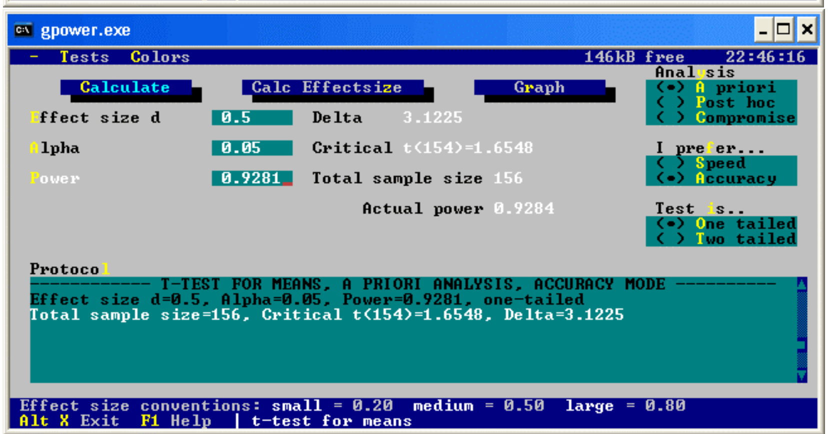


Während der Forscher bei einem Untersuchungsdesign mit unabhängigen Stichproben eine Teststärke von 46% erzielt hat, liegt die Teststärke bei abhängigen Stichproben für r = 0,30 bei wesentlich höheren 84%!

Zwei Gründe führen zu dieser hohen Teststärke: Zum einen gehen in den Test für abhängige Stichproben bei gleicher Gesamtanzahl Probanden doppelt so viele Messwerte ein, da jede Person zwei Werte abgibt, während bei dem Vergleich von unabhängigen Gruppen jede Person nur einen Messwert liefert. In der Formel für f’ bzw. $d_z$’ zeigt sich dieser Einfluss an der Zahl 2 im Zähler unter der Wurzel, die f’ im Vergleich zu f um den Faktor $\sqrt{2}$ erhöht, und so die Teststärke vergrößert. Zum anderen bewirkt die positive Korrelation einen Anstieg der Teststärke. Dies lässt sich ebenso an der Formel veranschaulichen: je höher die Korrelation, desto kleiner die Zahl im Nenner unter der Wurzel, desto größer wird $d_z$’ und damit die Teststärke.

Der Unterschied in dem Beispiel zwischen einer Teststärke von 46% und 84% bei gleicher Versuchspersonenzahl ist beachtlich, wenn man bedenkt, wie viele Kosten unterschiedlicher Natur mit der Rekrutierung und Datenerhebung von Versuchspersonen in der Regel verbunden sind. Läge die Korrelation zwischen den Messwertreihen bei r = 0,50 (ein durchaus realistischer Wert für viele Fragestellungen), würde sich die Power sogar auf nahezu 93% erhöhen! (Rechnen Sie dieses Beispiel nach!) Würde der Forscher auf eine ähnlich hohe Teststärke im Fall unabhängiger Stichproben abzielen, bräuchte er 112 Versuchspersonen für knapp 84% Power und sogar 156 für nahezu 93% Power! (siehe folgende Graphiken)



Der t-Test für abhängige Stichproben ist also bei gleicher Anzahl von Personen teststärker als der t-Test für unabhängige Stichproben, da jede Versuchsperson zwei Werte abgibt, und weil eine (positive) empirische Korrelation zwischen den Messwertreihen der abhängigen Stichproben die Teststärke zusätzlich erhöht[1]. Die genaue Höhe der Teststärke hängt letztendlich von der Größe der Korrelation ab. Je größer diese ist, desto größer ist der Vorteil der abhängigen Stichproben gegenüber den unabhängigen. Abgesehen von den weiter oben erwähnten eher mathematischen Gründen, warum sich die Korrelation zwischen abhängigen Daten positiv auf die Teststärke auswirkt, gibt es auch noch andere, eher inhaltlich fassbare. Diese Überlegungen sind stark mit dem Konzept der Varianz verbunden. Mehr darüber erfahren Sie in Kapitel 7, Band II.

Wenn die inhaltliche Fragestellung es zulässt, ein Untersuchungsdesign zu wählen, das ohne Einbußen in der Aussagekraft mit abhängigen Stichproben arbeitet, so hat dies also Vorteile für die Teststärke und damit für die Effizienz der Forschung! In diesen Fällen bestimmt die Planung einer Untersuchung, ob am Ende abhängige oder unabhängige Stichproben vorliegen. Einschränkend sei allerdings gesagt, dass in vielen Fällen unveränderbare Umstände die Frage bestimmen, ob man seine Daten an zwei unabhängigen oder abhängigen Gruppen untersucht. Wenn z.B. Geschlechterunterschiede im Fokus einer Untersuchung stehen, gibt es keine Möglichkeit, die Daten beider Ausprägungen des Merkmals Geschlecht von ein und derselben Person zu erhalten. Etliche andere versuchplanerische Erwägungen schließen die Erhebung mehrer Messwerte pro Person bei bestimmten Fragestellungen ebenfalls aus (Reihenfolgeeffekte, Übungseffekte, Ermüdungseffekte etc.). In diesen Fällen sind unabhängige Stichproben erforderlich.

Anmerkungen: Zum Vergleich von einem t-Test für unabhängige und abhängige Stichproben in SPSS siehe Datei „Kapitel_3_SPSS_Ergaenzungen.pdf".

[1] Anmerkung: Interessanter Weise würde eine negative Korrelation zwischen abhängigen Stichproben zu einer *Verringerung* der Teststärke führen, wie Sie leicht an der Formel für f' bzw. $d_z$' nachvollziehen können. Allerdings tritt dieser Fall in der Praxis nur sehr selten auf.

# Literatur

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