1. Einleitung

Im Rahmen des t-Tests wird untersucht, ob die Differenz der Mittelwerte von zwei Stichproben zufällig zustande gekommen ist, oder ob sie auch mit einer hohen Wahrscheinlichkeit in der Grundgesamtheit vorliegt, also im statistischen Sinne signifikant ist. Solche Mittelwertvergleiche bieten sich an, um Fallgruppen hinsichtlich bestimmter Eigenschaften auf Differenzen oder Ähnlichkeiten zu untersuchen.

Um Mittelwerte berechnen zu dürfen, muß die zu untersuchende abhängige Variable intervallskaliert sein (z.B. eine Likertskala). Die unabhängige Variable (oder auch Gruppenvariable) muß hingegen als Dichotomie vorliegen. Variablen, die mehr als zwei Kategorien aufweisen, aber dennoch als Gruppenvariable dienen sollen, müssen entsprechend dichotomisiert werden (z.B. am Median). Für nominalskalierte Variablen mit mehr als zwei Ausprägungen (z.B. Beruf) kann ein t-Test i.d.R. nicht durchgeführt werden, da es aus inhaltlichen Gründen häufig nicht möglich ist, verschiedene Kategorien einer nominalskalierten Variable zusammenzufassen. In solchen Fällen wird auf das Verfahren der Varianzanalyse zurückgegriffen (zur Varianzanalyse (ONEWAY, ANOVA, MANOVA) vgl. Brosius, 1995, S.417f).

Die Variable „Geschlecht" ist ein Beispiel für eine typische dichotome Gruppenvariable (zwei Ausprägungen). Durch einen Vergleich der Mittelwerte für Männer und Frauen in Bezug auf die jeweils interessierenden Merkmale oder Eigenschaften (z.B. Ausbildungsdauer, Einkommen oder Lebenserwartung als abhängige Variable) lassen sich geschlechtsspezifische Differenzen oder geschlechtsunabhängige Gemeinsamkeiten darlegen.

Eine weitere Anwendung des t-Tests ist die Berechnung der Trennschärfe für die Items einer Skala im Rahmen der Itemanalyse. Hierzu wird die Stichprobe anhand der Skalenwerte der vorläufigen Skala in eine obere und eine untere Gruppe unterteilt,[1] und die Gruppenmittelwerte für jedes einzelne Item werden verglichen. Unterscheiden sich die Gruppenmittelwerte eines Items signifikant, kann davon ausgegangen werden, daß das Item trennscharf ist.

Voraussetzungen zur Durchführung des t-Tests

1. In unseren Beispielen werden zwei unabhängige Zufallsstichproben vorausgesetzt. Dies bedeutet konkret, daß die Werte der Variablen in der einen Gruppe nicht abhängig sind von den Werten in der anderen Gruppe (z.B. wäre der Konsum von Techno-Musik von Jugendlichen unabhängig vom „Techno"-Konsum von Rentner/innen.) Der t-Test läßt sich jedoch auch für abhängige Zufallsstichproben durchführen (vgl. Brosius, 1995, S. 409f). In diesem Fall sind die Werte der Variablen in der zweiten Gruppe abhängig von den Werten in der ersten Gruppe

[1] Genauer gesagt: eine Gruppe besteht aus den 25% der Versuchspersonen mit den höchsten Skalenwerten, die andere Gruppe aus den 25% mit den niedrigsten Skalenwerten.

(z.B. die Beurteilung der Praxisrelevanz eines Faches vor dem Berufseinstieg der Befragten im Vergleich zu dieser Beurteilung durch die gleichen Befragten nach dem Berufseinstieg.)

2. Die abhängige Variable muß auf Intervallniveau gemessen worden sein und die Mittelwertdifferenz sollte normalverteilt sein. Bei hinreichend großen Stichproben kann man davon ausgehen, daß die Mittelwertdifferenz normalverteilt ist (siehe die folgende Erläuterung).

   Erläuterung:
   Für unterschiedliche Stichproben aus einer Grundgesamtheit können sich unterschiedliche Mittelwerte ergeben. Dabei streuen die Mittelwerte mehr oder weniger um den „wahren" Mittelwert der Grundgesamtheit. Ist die Grundgesamtheit normalverteilt, so sind es auch die Mittelwerte der aus dieser Grundgesamtheit gezogenen Stichproben. Nach dem zentralen Grenzwertsatz gilt auch für jede andere Verteilungsform in der Grundgesamtheit, daß sich mit wachsendem Stichprobenumfang die Verteilung der Stichprobenmittelwerte zunehmend der Normalverteilung annähert. Dabei wird die Normalverteilung mit wachsendem Stichprobenumfang um so schneller erreicht, je mehr die Verteilung der Grundgesamtheit einer Normalverteilung ähnelt. Konventionell geht man davon aus, daß die Verteilung der Stichprobenmittelwerte, unabhängig von der Verteilungsform in der Grundgesamtheit, bereits hinreichend normalverteilt ist, wenn N ³ 30 (vgl. Bortz, 1989, S. 163). Für unseren Fall, der zwei unabhängige Zufallstichproben voraussetzt, nimmt man an, daß die Verteilung der Mittelwertdifferenzen bei einem Stichprobenumfang von $N_1 + N_2$ ³ 50 annähernd normalverteilt ist (vgl. Bortz, 1989, S.166). Zusätzlich gilt, daß die Varianz der Stichprobenmittelwerte s_n^2 mit wachsender Stichprobengröße abnimmt. Ist s^2 die Varianz in der Grundgesamtheit, so gilt:

   $$s_n^2 = \frac{s^2}{N}$$

   Inhaltlich bedeutet dies, daß die „Genauigkeit" der Stichprobenschätzungen mit wachsendem Stichprobenumfang zunimmt.

3. Die unabhängige Variable (Gruppenvariable) ist dichotom (z.B. Männer versus Frauen, Hochgebildete versus niedrig Gebildete, Beschäftigte versus Arbeitslose)

4. In der einfachsten Form des t-Tests geht man davon aus, daß die Varianzen der zu vergleichenden Populationen gleich sind. Ist dies nicht der Fall, so wird eine modifizierte Formel für den t-Test verwendet.

2. t-Test bei zweiseitiger Fragestellung

Beim Vergleich der Mittelwerte von zwei voneinander unabhängigen Stichproben prüft der t-Test bei zweiseitiger Fragestellung die Gleichheit der Mittelwerte in der Grundgesamheit.[2] Letzteres wird auch als Nullhypothese (H_0) bezeichnet:

$$H_0: \mu_1 = \mu_2$$

Aus dieser Nullhypothese ergibt sich als Alternativhypothese (H_1), daß die Mittelwerte in der Grundgesamtheit verschieden sind:

$$H_1: \mu_1 \neq \mu_2$$

Bei der Durchführung des t-Tests wird immer nach folgender Logik vorgegangen:

1. **Bestimmung der Nullhypothese. Aus dieser ergibt sich logisch die Alternativhypothese.** Das Ziel ist es letztendlich, die Nullhypothese mit einer gewissen Sicherheit zu verwerfen und deshalb aus logischen Gründen die Alternativhypothese anzunehmen. I.d.R. soll also die Alternativhypothese der Hypothese entsprechen, die sich aus der theoretischen Fragestellung der Untersuchung ableitet (z.B. „Die Gruppe der Männer unterscheidet sich in ihrer normativen Geschlechtsrollenorientierung (NGRO) von der Gruppe der Frauen“). Deshalb muß die Nullhypothese so gewählt werden, daß die sich logisch ergebende Alternativhypothese genau der theoretisch postulierten Hypothese entspricht. Die theoretisch postulierte Hypothese (und somit auch die angestrebte Alternativhypothese) kann entweder einseitig oder zweiseitig formuliert sein. Bei der zweiseitigen Fragestellung wird lediglich ein Mittelwertunterschied postuliert (siehe obige Hypothese). Bei der einseitigen Fragestellung wird zusätzlich noch festgelegt, welche der beiden Gruppen den höheren Mittelwert aufweist (z.B.: „Die Gruppe der Frauen hat eine liberalere NGRO als die Gruppe der Männer“).[3]

---

[2] Es sei an dieser Stelle noch einmal drauf hingewiesen, daß es sich bei den Mittelwerten der Grundgesamtheit um feste, allerdings nicht beobachtbare Parameter handelt. Dies wird bei der Formulierung der Hypothesen durch die Verwendung von griechischen Buchstaben (Parameter) zum Ausdruck gebracht. Im Gegensatz dazu werden die beobachteten stichprobenabhängigen Statistiken (Mittelwerte und Varianzen der Stichproben), also die Ergebnisse (eine Realisation) des zugrundeliegenden Zufallsexperiments, mit lateinischen Buchstaben bezeichnet.

[3] Es liegt auf der Hand, daß es per se wahrscheinlicher ist, zufällig einen Mittelwertunterschied vorzufinden (zweiseitige Fragestellung), als zufällig zu beobachten, daß genau der Mittelwert der vorher festgelegten Gruppe größer ist, als der Mittelwert der anderen Gruppe (einseitige Fragestellung).

2. **Festlegung des Signifikanzniveaus Alpha (a).** Das Signifikanzniveau ist nichts anderes als die Wahrscheinlichkeit, H_0 irrtümlich zu verwerfen. α wird deshalb auch Irrtumswahrscheinlichkeit (oder Fehlerwahrscheinlichkeit) genannt.[4]

3. **Berechnung des t-Wertes ($\hat{t}$)**, der sich aufgrund der Stichprobe ergibt; $\hat{t}$ wird auch als Prüfgröße bezeichnet.

4. **Vergleich von $\hat{t}$ mit dem tabellierten t-Wert ($t_{a,df}$).** Unter Beachtung der Fragestellung (Einseitig oder zweiseitig?) hängt der tabellierte t-Wert, auch kritischer Wert genannt, von dem festgelegten Signifikanzniveau α und den Freiheitsgraden (*df*) ab. Der Vergleich von $\hat{t}$ mit $t_{a,df}$ gibt Auskunft darüber, ob die Differenz der Stichprobenmittelwerte statistisch signifikant ist.

2.1 Beispiel: Übungsaufgabe I.2

Das folgende hypothetische Beispiel bezieht sich auf den Vergleich von Stichprobenmittelwerten aus zwei unabhängigen Zufallsstichproben (vgl. Übungsaufgabe I.2; Skript 1995, S. A-2):

Besetzungszahlen für die unterschiedlichen Stichprobengrößen			Skalenwerte der Likert-Skala nGRO für Frauen und Männer	
N = 33	N = 330	N = 3	Frauen	Männer
8	80	1	26	24
17	170	1	32	30
8	80	1	38	36

In diesem Beispiel ergeben sich für Frauen und Männer jeweils unterschiedliche Stichprobenmittelwerte auf der nGRO-Skala (nGRO = normative Geschlechtsrollenorientierung) und zwar $\bar{x}_{\text{Frauen}} = 32$ und $\bar{x}_{\text{Männer}} = 30$.

Der t-Test wird hier für drei unterschiedlichen Stichprobengrößen durchgeführt. Dadurch soll veranschaulicht werden, daß das Ergebnis des Signifikanztest (Prüfung der Nullhypothese: Gleiche Mittelwerte in der Grundgesamtheit) von der Größe der Stichprobe abhängt.[5]

Die Beispiele haben folgende Stichprobengrößen für jede Gruppe (also Frauen und Männer):

---

[4] Eine 5%-ige Irrtumswahrscheinlichkeit bedeutet im übrigen das gleiche wie die 95%-ige Wahrscheinlichkeit, sich nicht zu irren (beim Zurückweisen der H_0). Deshalb wird hin und wieder auch von einer 95%-igen Sicherheitswahrscheinlichkeit gesprochen.

[5] Mit zunehmender Stichprobengröße nimmt die Varianz der Stichprobenmittelwerte (in dem hypothetischen Zufallsexperiment) ab (siehe oben).

Beispiel 1: $N_1 = 33$

Beispiel 2: $N_2 = 330$

Beispiel 3: $N_3 = 3$

Es soll nun geprüft werden, ob der Mittelwertunterschied zwischen den beiden Gruppen zufällig beobachtet wurde (H_0) oder ob dieser Unterschied auch in der Grundgesamtheit vorliegt (H_1). Da es hier nur um den Unterschied geht, liegt eine zweiseitige Fragestellung vor. Die logische Alternative zu einem Mittelwertunterschied ist die Gleichheit der Mittelwerte (H_0). Formal lauten deshalb Nullhypothese (H_0) und Alternativhypothese (H_1) wie folgt:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

Prüfung der Voraussetzungen

In der Übungsaufgabe gehen wir davon aus, daß es sich um unabhängige Stichproben und intervallskalierte Variablen handelt. Entsprechend der Stichprobengrößen der Beispiele können wir davon ausgehen, daß die abhängige Variable annähernd normalverteilt ist (mit Ausnahme des dritten Beispiels, die Fallzahl von N = 6 dient nur zu Demonstrationszwecken). Die unabhängige Variable Geschlecht ist dichotom und die Gleichheit der Varianzen wird im Rahmen der SPSS-Prozedur T-TEST zusätzlich geprüft (vgl. unten „Interpretation der Output-Datei“).

2.1.1 Beispiel 1 (N=33 pro Gruppe)

Zunächst müssen die Daten der Übungsaufgabe in ein **Daten**-Fenster für SPSS übertragen werden. Genauer gesagt, es werden zwei Variablen eingegeben, die Variable SEX (mit den Ausprägungen 0 für „weiblich“ und 1 für „männlich“) und die Variable NGRO (mit den Werten, die in der Übungsaufgabe stehen).[6]

Nach der Dateneingabe kann die Prozedur für den t-Test aufgerufen und ausgeführt werden (Statistik → Mittelwertvergleiche → T-Test bei unabhängigen Stichproben). Alternativ zu der „Menütechnik“ kann der t-Test auch über einen Befehl im **Syntax**-Fenster von SPSS aufgerufen werden. Dieser Befehl sieht wie folgt aus:

---

[6] Bei der Dateneingabe gibt es in diesem speziellen Fall zwei Alternativen: Entweder, es werden die Daten für 66 Versuchspersonen eingegeben (in diesem Fall erhalten 8 weibliche Befragte den Wert 26 für die NGRO, 17 weibliche Befragte den Wert 32, 8 weibliche Befr. den Wert 38, 8 männliche Befr. den Wert 24 ... usw.) oder es wird eine sog. Gewichtungsvariable gebildet, die als Wert jeweils die Häufigkeit der drei Antwortkategorien für die Männer und für die Frauen enthält (z.B.: SEX = 0, NGRO = 26, GEWICHT = 8; SEX = 1, NGRO = 24, GEWICHT = 8; ... usw.). Im letzeren Fall muß SPSS durch eine Anweisung mitgeteilt werden, daß die Fälle mit der Gewichtungsvariablen gewichtet werden sollen.

`T-TEST`	Dieses Schlüsselwort kennzeichnet die Prozedur
`GROUPS=sex(0 1)`	Dies ist die Anweisung, nach welcher Variable gruppiert werden soll, in diesem Fall nach SEX mit den Ausprägungen 0 und 1
`/MISSING=ANALYSIS`	Hier wird mitgeteilt, daß Personen mit fehlende Angaben (sog. Missings) für diesen Test ausgeschlossen werden sollen
`/VARIABLES=ngro`	Dies ist die abhängige Variable (in diesem Fall die NGRO), hier könnten ggf. auch noch weitere Variablen eingetragen werden
`/CRITERIA=CIN(.95).`	Durch diese Anweisung wird zusätzlich noch das 95%ige Konfidenzintervall für die Mittelwertdifferenz angefordert. (Am Ende des vollständigen Befehls muß ein Punkt stehen!!)

Unabängig davon, ob der Befehl „per Hand“ eingegeben wurde oder ob über das Menü die entsprechenden Variablen im T-Test-Dialogfenster eingegeben wurden, erhalten wird nach der Ausführung des Befehls im **Ausgabe**-Fenster folgenden Output:

```
t-tests for independent samples of  SEX   Geschlecht

                              Number
           Variable          of Cases    Mean        SD       SE of Mean
        -------------------------------------------------------------------
           NGRO  normative Geschlechtsrollenorientierung

          weiblich             33      32,0000      4,243         ,739
          männlich             33      30,0000      4,243         ,739
        -------------------------------------------------------------------

          Mean Difference = 2,0000
          Levene's Test for Equality of Variances: F= ,000   P= 1,000

       t-test for Equality of Means                                 95%
  Variances  t-value   df     2-Tail Sig    SE of Diff      CI for Diff
 ------------------------------------------------------------------------

  Equal      1,91     64          ,060         1,044       (-,087; 4,087)
  Unequal    1,91     64,00       ,060         1,044       (-,087; 4,087)
```

Interpretation der Output-Datei

Die erste Tabelle gibt für die beiden Ausprägungen (`weiblich`/`männlich`) der unabhängigen Variable Geschlecht die Anzahl der Fälle (`Number of Cases`), den Mittelwert (`Mean`), die Standardabweichung (`SD`) sowie den Standardfehler des Mittelwertes (`SE of Mean`) an. Aus dieser Tabelle läßt sich ersehen, daß sich für Frauen und Männer unterschiedliche Mittelwerte für „normative Geschlechtsrollenorientiertung“ ergeben. Männer (`Mean` = 30) haben einen niedrigeren Mittelwert als Frauen (`Mean` = 32). (Die Variable nGRO ist so skaliert, daß ein niedriger

Wert eine eher traditionale Geschlechtsrollenorientierung verkörpert, während hohe Werte mit einer eher liberalen Geschlechtsrollenorientierung korrespondieren).

Der `Levene´s Test for Equality of Variances` prüft, ob die Varianzen der Variablen (NGRO) für die beiden Fallgruppen in der Grundgesamtheit gleich sind (Voraussetzung 4.: $\sigma^2_{Männer} = \sigma^2_{Frauen}$). Als Prüfmaß wird hier ein F-Wert berechnet, für den es, wie für den t-Wert, eine entsprechende Zufallsverteilung gibt. Je höher die Wahrscheinlichkeit ist, die neben dem F-Wert angegeben wird, desto eher kann davon ausgegangen werden, daß in der Grundgesamtheit gleiche Varianzen vorliegen. In unserem Fall beträgt die Wahrscheinlichkeit `P = 1,000`, d.h. es liegt mit 100%iger Sicherheit Gleichheit der Varianzen in der Grundgesamtheit vor.[7]

Unter diesem Test folgt nun das Ergebnis des eigentlichen t-Tests (`t-test for Equality of Means`), genauer gesagt, es werden zwei Ergebnisse ausgegeben: Eins für den Fall gleicher Varianzen (`Equal`) und eins für den Fall ungleicher Varianzen (`Unequal`). In unserem Beispiel treffen die Ergebnisse des ersten Falles zu (da der F-Test die Gleichheit der Varianzen ergeben hat!).

Ausgegeben werden $\hat{t}$ (`t-value`), die Freiheitsgrade (`df`), das erreichte Signifikanzniveau (`2-Tail Sig`), der Standardfehler der Mittelwertdifferenz (`SE of Diff`) sowie das 95%ige Konfidenzintervall für die Differenz (`95% CI for Diff`). Um diese statistischen Maßzahlen vernünftig interpretieren zu können, werden wir an dieser Stelle die „Formel" für den t-Test genauer betrachten (diese ist im übrigen in nahezu jedem Lehrbuch für statistische Grundlagen nachzulesen):[8]

$$\hat{t} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

Im Zähler der Formel steht die Differenz der Gruppenmittelwerte (in diesem Fall ist nur der Betrag von Bedeutung, da es sich um die zweiseitige Fragestellung handelt). H_0 lautet (wie oben erläutert), daß diese Differenz in der Grundgesamtheit gleich Null sei (sich also die Mittelwerte <u>nicht</u> unterscheiden). Diese Differenz ist auch im SPSS-Output enthalten (`Mean Difference = ...`). Die Prüfgröße $\hat{t}$ wird nun berechnet, in dem die zu prüfende Größe (also die Differenz) durch ihren Standardfehler geteilt wird. Auch der Standardfehler der Differenz ist im SPSS-Output enthalten (`SE of Diff`). Berechnet wird dieser Standardfehler entsprechend dem Ausdruck, der im Nenner der Formel steht: s_i^2 ist die Varianz der Gruppe *i* (im SPSS-Output ist die Wurzel der

---

[7] Dieser sehr hohe Wert für `P` ist auf die Konstruktion unseres hypothetischen Beispieles zurückzuführen. In der Praxis wird kaum einmal dieser Wert erreicht werden. Von <u>gleichen</u> (im Sinne von nicht signifikant verschiedenen) Varianzen wird üblicherweise bei einem `P > 0,100` ausgegangen (Anpassungstest).

[8] Da es sich sowohl bei den Mittelwerten ($\bar{x}_1, \bar{x}_2$) als auch bei den Varianzen (s_1^2, s_2^2) um Stichprobenwerte handelt, werden die entsprechenden Statistiken mit lateinischen Buchstaben gekennzeichnet.

Varianz, also die Standardabweichung s_i angegeben [`SD`]), N_i ist die Fallzahl der Gruppe i (`Number of Cases`). Wenn wir die von SPSS ausgegebenen Maßzahlen in die Formel einsetzen, erhalten wir:

$$\hat{t} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} = \frac{|32 - 30|}{\sqrt{\frac{4{,}243^2}{33} + \frac{4{,}243^2}{33}}} = \frac{2}{\sqrt{\frac{18+18}{33}}} = \frac{2}{1{,}044} = 1{,}9148$$

Diese Prüfgröße $\hat{t}$ würde nun (wie oben erläutert) einem tabellierten t-Wert gegenübergestellt, dessen Größe von dem festgelegten Signifikanzniveau α sowie den Freiheitsgraden des T-tests abhängt (Tabellen für die t-Werte sind ebenfalls in nahezu jedem Statistik-Lehrbuch abgedruckt).

Die Freiheitsgrade (`df`) berechnen sich auf der Basis einer Stichprobe wie folgt:

$$\text{df} = N - 1$$

Unser Beispiel beruht jedoch auf zwei (unabhängigen) Stichproben, für die `df` wie folgt berechnet wird:

$$\text{df} = N_1 + N_2 - 2$$

Bei einem Signifikanzniveau von $\alpha = 0{,}05$ und 64 Freiheitsgraden beträgt der kritische t-Wert (für die zweiseitige Fragestellung):

$$t_{0,05;64} = 2{,}00$$

Dieser kritische Wert muß von unserer Prüfgröße überschritten werden, damit wir mit einer Irrtumswahrscheinlichkeit von 5% ($\alpha = 0{,}05$) H_0 ablehnen können. Wie oben berechnet, ist dies nicht der Fall. H_0 muß also beibehalten werden - der Mittelwertunterschied ist nicht signifikant.

Im Output von SPSS wird nicht der kritische Wert, sondern stattdessen das Signifikanzniveau (`2-Tail Sig`) ausgegeben, für das bei der zugrunde liegenden Anzahl von Freiheitsgraden (`64`) genau der Betrag der berechneten Prüfgröße (`1,91`) als t-Wert tabelliert wäre. Da wir vorher festgelegt hatten, daß diese Irrtumswahrscheinlichkeit höchsten 5% betragen darf, müßte dieser Wert unterschritten werden. In unserem Beispiel ist die `2-Tail Sig = 0,060`, also größer als die erlaubte Irrtumswahrscheinlichkeit. Die Schlußfolgerung ist natürlich dieselbe wie im letzten Absatz: H_0 muß beibehalten werden - der Mittelwertunterschied ist nicht signifikant.

Eine dritte Möglichkeit, zu beurteilen, ob die Mittelwertdifferenz in der Grundgesamtheit von Null verschieden ist, liefert uns das von SPSS ebenfalls ausgegebene Konfidenzintervall (`95% CI for Diff`). Dieses Intervall gibt an, in welchem Bereich die Differenz der Mittelwerte mit 95%iger Sicherheit in der Grundgesamtheit liegen wird. Wenn dieser Bereich den Wert Null mit einschließt (was in unserem Beispiel natürlich der Fall ist), dann ist der Mittelwertunterschied nicht signifikant.

Letztendlich ist es unerheblich, nach welcher der drei Alternativen die Signifikanz der Mittelwertdifferenz beurteilt wird, alle drei Methoden müssen die gleiche Schlußfolgerung implizieren.

2.1.2 Beispiel 2 (N=330 pro Gruppe)

Die Vorgehensweise ist hier prinzipiell dieselbe, der einzige Unterschied besteht in der größeren Fallzahl von N = 330 pro Gruppe. Deshalb wenden wir uns direkt dem SPSS-Output zu:

```
t-tests for independent samples of  SEX   Geschlecht

                            Number
           Variable        of Cases    Mean        SD       SE of Mean

            NGRO

          weiblich           330      32,0000     4,184        ,230
          männlich           330      30,0000     4,184        ,230


          Mean Difference = 2,0000
          Levene's Test for Equality of Variances: F= ,000   P= 1,000

       t-test for Equality of Means                             95%
  Variances  t-value   df    2-Tail Sig    SE of Diff      CI for Diff

  Equal      6,14    658         ,000         ,326       (1,360; 2,640)
  Unequal    6,14    658,00      ,000         ,326       (1,360; 2,640)
```

<u>Interpretation der Output-Datei</u>

Da dieselben Werte verwendet wurden, beträgt auch hier die Differenz der Mittelwerte 2 (`Mean Differenz`). Auch der F-Test ergibt wieder eine 100%ige Übereinstimmung der Varianzen in der Grundgesamtheit. Die Standardabweichung ist praktisch unverändert (`SD = 4,184`). Der entscheidende Unterschied liegt im Standardfehler der Differenz (`SE of Diff`). Wie in der Formel zu sehen, werden zur Berechnung des `SE` die Varianzen durch die Fallzahlen der Stichproben geteilt.[9] Da die Fallzahlen hier wesentlich größer sind, ist der Standardfehler der Differenz wesentlich kleiner als im vorherigen Beispiel. Die Prüfgröße $\hat{t}$ ergibt sich wiederum aus dem Quotienten der Mittelwertdifferenz und dem `SE`:

$$\hat{t} = \frac{2}{\sqrt{\frac{17{,}5 + 17{,}5}{330}}} = \frac{2}{0{,}326} = 6{,}14$$

Der kritische Wert $t_{0,05;658}$ beträgt nur 1,96 und ist somit wesentlich kleiner als die berechnete Prüfgröße. H_0 wird also abgelehnt - die Mittelwertdifferenz ist signifikant. Analog zu dieser Schlußfolgerung ist die von SPSS ausgegebene `2-Tail Sig` bei weitem kleiner als 0,05 und das 95%ige Konfidenzintervall enthält diesmal nicht den Wert Null.

---

[9] Als Faustregel gilt damit, daß eine Vervierfachung der Stichprobengröße gerade eine Halbierung des Standardfehlers der Stichprobe und damit eine doppelte „Genauigkeit“ ergibt. Generell gilt, daß die „Genauigkeit“ des T-tests mit der Wurzel der Stichprobengröße zunimmt (vgl.die Ausführungen zum zentralen Grenzwertsatz).

2.1.3 Beispiel 3 (N=3 pro Gruppe)

Auch im dritten Beispiel wurde lediglich die Fallzahl variiert. Die Stichprobengröße von insgesamt N = 6 ist jedoch nur zu Demonstrationszwecken so gewählt worden - normalerweise dürfte ein t-Test für so eine geringe Fallzahl nicht durchgeführt werden (es sei denn, man ist gewillt, für die nGRO Normalverteiltheit in der Grundgesamtheit vorauszusetzen; vgl. Erläuterungen zur Normalverteilungsvoraussetzung). SPSS hat folgenden Output ausgegeben:

```
t-tests for independent samples of  SEX   Geschlecht

                            Number
           Variable        of Cases    Mean         SD        SE of Mean
         -----------------------------------------------------------------
            NGRO

          weiblich             3      32,0000      6,000        3,464
          männlich             3      30,0000      6,000        3,464
         -----------------------------------------------------------------

          Mean Difference = 2,0000
          Levene's Test for Equality of Variances: F= ,000   P= 1,000

       t-test for Equality of Means                            95%
  Variances  t-value   df    2-Tail Sig    SE of Diff      CI for Diff
-----------------------------------------------------------------------------

  Equal        ,41      4          ,704        4,899      (-11,607; 15,607)
  Unequal      ,41      4,00       ,704        4,899      (-11,607; 15,607)
```

Interpretation der Output-Datei

Durch die niedrige Fallzahl erhalten wir bei gleicher Mittelwertdifferenz (`Mean Difference`) einen wesentlich höheren Standardfehler der Differenz (`SE of Diff = 4,899`). Die entsprechende Prüfgröße $\hat{t}$ (`t-value`) beträgt deshalb nur 0,41 und ist somit wesentlich kleiner als der kritische Wert von $t_{0,05;4}$ = 2,78. Die Wahrscheinlichkeit, daß das Verwerfen von H_0 falsch ist, beträgt bei einem $\hat{t}$ von 0,41 und 4 Freiheitsgraden (`df`) immerhin 70,4% (`2-Tail Sig`). Die Nullhypothese, in der Grundgesamtheit liege keine Mittelwertdifferenz vor, wird also beibehalten. Fazit: Die Signifikanz einer Mittelwertdifferenz hängt bei einem t-Test entscheidend davon ab, welchen Umfang die Stichprobe hat, denn die beobachtete Differenz (`2,00`) war in allen drei Stichproben gleich groß.

2.2 Beispiel für die zweiseitige Fragestellung: Geschlechtsspezifische materielle Gütergebundenheit

Im Rahmen der Kaufsuchtstudie von G. Scherhorn (1991) wurde unter anderem die materielle Gütergebundenheit der Befragten mit Hilfe einer Likert-Skala erhoben. Uns interessiert nun die Frage, ob zwischen Männern und Frauen ein Unterschied hinsichtlich der materiellen Gütergebundenheit besteht - also, ob die Männer auf der entsprechenden Skala einen anderen Mittelwert haben als die Frauen.

Das Geschlecht (in der Untersuchung die Variable `F109`) wird als dichotome Gruppierungsvariable verwendet, die abhängige Variable ist die materielle Gütergebundenheit (`GTGB`). Die `GTGB` wurde mit einer Likertskala gemessen, es liegt somit eine Messung auf Intervallniveau vor. In der Untersuchung wurden insgesamt 1523 Personen (`719` Männer und `804` Frauen) befragt, wir können also davon ausgehen, daß die Mittelwertdifferenz annähernd normalverteilt ist und einen t-Test für unabhängige Stichproben durchführen.

Der SPSS-Befehl im Syntax-Fenster lautet wie folgt:

```
T-TEST
  GROUPS=F109(1 2)
  /MISSING=ANALYSIS
  /VARIABLES=GTGB
  /CRITERIA=CIN(.95) .
```

Folgender Output wurde von SPSS im Ausgabe-Fenster erzeugt:

```
t-tests for independent samples of  F109   GESCHLECHT

                            Number
           Variable        of Cases    Mean        SD       SE of Mean
        ---------------------------------------------------------------
           GTGB  materielle Gütergebundenheit

          MAENNLICH          719      3,3861      ,924         ,034
          WEIBLICH           804      3,7081      ,948         ,033
        ---------------------------------------------------------------

          Mean Difference = -,3220
          Levene's Test for Equality of Variances: F= ,531   P= ,466

       t-test for Equality of Means                             95%
  Variances  t-value   df    2-Tail Sig    SE of Diff      CI for Diff
 -----------------------------------------------------------------------
  Equal     -6,70   1521          ,000         ,048       (-,416; -,228)
  Unequal   -6,71   1509,96       ,000         ,048       (-,416; -,228)
```

Zunächst einmal ist festzuhalten, daß in der Stichprobe eine Mittelwertdifferenz vorliegt: Männer haben hier einen niedrigeren Mittelwert als Frauen auf dieser Skala (`Mean Difference = -,3220`). Der F-Test zeigt keinen signifikanten Unterschied der Varianzen der `GTGB` in den beiden Stichproben an, `P = ,466` ist größer als der kritische Wert `,10` (Anpassungstest). Deshalb sind die Ergebnisse bei Gleicheit der Varianzen für uns relevant (Zeile `Equal`). SPSS berechnet folgende Prüfgröße:

$$\hat{t} = \frac{-,3220}{0,048} = -6,71 \quad \texttt{(Mean Difference/SE of Diff)}$$ [10]

Es fällt auf, daß SPSS <u>nicht</u> den Betrag der Differenz verwendet, obwohl dies bei der zweiseitigen Fragestellung eigentlich erforderlich wäre. Wir vergleichen folglich <u>den Betrag</u> von $\hat{t}$ mit dem tabellierten t-Wert. Dieser beträgt bei `1521 df` und einem $\alpha = 0,05$:

$$t_{0,05;1521} = 1,96.$$

`6,71` ist größer als 1,96 - die Nullhypothese kann also verworfen werden. Es besteht ein signifikanter Mittelwertunterschied zwischen Männern und Frauen bezüglich der materiellen Gütergebundenheit. Die Wahrscheinlichkeit, H_0 fälschlicherweise zu verwerfen, ist annähernd 0,0% (`2-Tail Sig`). Demzufolge schließt auch das 95%ige Konfidenzintervall (`95% CI of Diff`) den Wert Null nicht mit ein.

3. t-Test bei einseitiger Fragestellung

3.1 Die einseitige Fragestellung

Bisher haben wir den t-Test immer bei zweiseitiger Fragestellung durchgeführt, d.h. wir haben theoretisch postuliert, daß zwischen den beiden untersuchten Gruppen ein Mittelwertunterschied besteht, aber nichts über die Richtung dieses Unterschiedes ausgesagt. Entsprechend war unsere Alternativhypothese (die sich logisch aus der Ablehnung der Nullhypothese ergeben hat), daß überhaupt ein Unterschied zwischen beiden Gruppen besteht. Die einseitige Fragestellung bedeutet, daß nicht nur ein Mittelwertunterschied postuliert wird, sondern daß zusätzlich vorher (<u>aufgrund theoretischer Überlegungen</u>) festgelegt wird, welche der beiden Gruppen den höheren Mittelwert aufweisen sollte. Unsere theoretisch hergeleitete Hypothese (H_1) lautet folglich formal:

$$H_1: \quad \mu_1 > \mu_2.$$

Die zu testende Nullhypothese (aus deren Ablehnung sich die Alternativhypothese logisch ergeben muß) lautet deshalb formal:

$$H_0: \quad \mu_1 \leq \mu_2.$$

Für den t-Test ergibt sich aus der einseitigen Fragestellung nur eine einzige, jedoch entscheidende Änderung: Der kritische Wert (der tabellierte t-Wert) ist bei gleichem Signifikanzniveau α und gleicher Anzahl von Freiheitsgraden <u>niedriger</u> als bei der zweiseitigen Fragestellung. Die Chance, in der Stichprobe <u>zufällig</u> eine Mittelwert<u>differenz</u> zu beobachten (zweiseitige Fragestellung) ist nämlich <u>größer</u>, als die Wahrscheinlichkeit, daß in der Stichprobe <u>zufällig</u> eine Mittelwert<u>differenz</u> besteht <u>und zusätzlich zufällig</u> auch das <u>Vorzeichen</u> dieser Differenz der theoretisch postulierten Hypothese entspricht (einseitige Fragestellung). Infolgedessen reicht bereits ein geringerer

---

[10] Die Abweichung um 0,01 ist auf Rundungsfehler zurückzuführen.

Betrag für die Prüfgröße $\hat{t}$ aus, H_0 mit einer Irrtumswahrscheinlichkeit von α zu verwerfen. Außerdem muß das Vorzeichen der Prüfgröße $\hat{t}$ der in der Alternativhypothese postulierten Richtung der Mittelwertdifferenz entsprechen.

Da sich die Berechnung der Prüfgröße für die einseitige Fragestellung nicht ändert, können wir nun die Ergebnisse der Beispiele 1 bis 3 vor dem Hintergrund der einseitigen Fragestellung betrachten. Wir stellen also in unserem Beispiel bei einseitiger Fragestellung (theoriegeleitet!) die Hypothese auf, daß Frauen eine liberalere normative Geschlechtsrollenorientierung aufweisen (=einen höheren Mittelwert haben), als Männer.

N (pro Gruppe)	$\hat{t}$	df	$t_{0,05;df\ (zweiseitig)}$	sign.? (zweis.)	$t_{0,05;df\ (einseitig)}$	sign.? (eins.)
33	**1,91**	64	2,00		1,67	ja
330	**6,14**	654	1,96	ja	1,65	ja
3	**0,41**	4	2,78		2,13	

Erwartungsgemäß können wir H_0 für die Stichprobe mit je 330 Personen pro Gruppe auch bei einseitiger Fragestellung mit einer Irrtumswahrscheinlichkeit von maximal 5% zurückweisen. Aber auch bei 33 Personen pro Gruppe ist die Mittelwertdifferenz bei einseitiger Fragestellung bereits signifikant. $\hat{t}$ ist hier nämlich größer als der tabellierte t-Wert für die einseitige Fragestellung.

Fazit: Bei gleich großen beobachteten Mittelwertunterschieden ist die Signifikanz dieser Differenzen neben der Abhängigkeit von der Stichprobengröße auch noch von der Fragestellung (also der Art unserer theoretischen Annahmen) abhängig

3.2 Beispiel für die einseitige Fragestellung: Geschlechtsspezifische materielle Gütergebundenheit

Unter 2.2 wurde zweiseitig formuliert, Männer und Frauen hätten eine unterschiedliche materielle Gütergebundenheit. Man hätte auch (falls dies aufgrund theoretischer Vorüberlegungen plausibel erscheint) einseitig formulieren können:

Männer haben eine geringere materielle Gütergebundenheit als Frauen.

Der einzige Unterschied wäre (wie schon in 3.1 gesehen), daß sich der kritische Wert für t verringert hätte: $t_{0,05;\ 1521} = 1,65$. Alle anderen berechneten Größen bleiben gleich und somit kann (natürlich) auch bei einseitiger Fragestellung ein signifikanter Mittelwertunterschied festgestellt werden. Beachten sollte man jedoch folgendes: SPSS gibt als `t-value -6,70` aus. Das negative Vorzeichen kommt zustande, weil Männer als erste Gruppe und Frauen als zweite Gruppe angegeben wurden. Wie in der Formel für den t-Test (vgl. S. 7) zu sehen ist, wird die Differenz $x_1 - x_2$ zur Berechnung des t-Wertes gebildet. Bei der einseitigen Fragestellung sollte daher die

Gruppe als erste Gruppe eingegeben werden, bei welcher der höhere Mittelwert erwartet wird (dies wären hier die Frauen gewesen).

3.3 Beispiel für die einseitige Fragestellung: Trennschärfe von Items einer Likert-Skala zur Messung von Kaufsucht

Im Rahmen der Itemanalyse einer Likert-Skala wird unter anderem die Trennschärfe für jedes Item berechnet. Die Trennschärfe eines Items (zur Einstellungsmessung) kann interpretiert werden als das Ausmaß, in dem ein Item zwischen Personen mit einer negativen Einstellung und Personen mit einer positiven Einstellung zu unterscheiden vermag (vgl. z.B. Schnell/Hill/Esser 1994)

Um die Trennschärfe der Items einer Likert-Skala zu bestimmen, werden die Befragten in zwei Gruppen aufgeteilt:

1. Gruppe: Das Viertel der Befragten mit den höchsten Skalenwerten der vorläufigen Skala.

2. Gruppe: Das Viertel der Befragten mit den niedrigsten Skalenwerten der vorläufigen Skala.

Jedes gute Item sollte nun eine signifikante positive Mittelwertdifferenz für die beiden Gruppen aufweisen - dies wird über einen t-Test geprüft.[11] Die Fragestellung ist in diesem Fall einseitig, da klar ist, daß die Befragten mit den höchsten Skalenwerten auch bei jedem einzelnen Item den höheren Mittelwert aufweisen müssen (vorausgesetzt, die Items sind alle in der gleichen Richtung gepolt).[12]

Als Beispieldatensatz dient uns auch hier die Kaufsucht-Studie von G. Scherhorn (1991). Zur Messung des theoretischen Konstruktes „Kaufsucht“ wurden den Befragten 17 Items vorgelegt, zu denen sie sich auf einem 4-stufigen Rating zustimmend oder ablehnend äußern konnten. Im Rahmen der Itemanalyse soll nun die Trennschärfe dieser Items anhand eines t-Tests untersucht werden.

Dazu muß zunächst die vorläufige Skala durch das Aufsumieren der Items gebildet werden. Dies erfolgt in SPSS am besten in einem Syntax-Fenster mit der Anweisung `compute`:[13]

```
compute ksroh= v356+v357+v358+v359+v360+v361+v362+v363+
               v364+v365+v366+v367+v368+v369+v370+v371+v372.
execute.
```

Nach der Befehlsanweisung `compute` folgt zunächst der Name der neuen Variable (der vorläufigen Skala) `ksroh`.[14] Hinter dem Gleichheitszeichen folgt dann die Anweisung, wie `ksroh` zu

[11] Die Trennschärfe kann (und wird i.d.R.) auch als Korrelation des Items mit dem Rest der Skala berechnet. Dies ist im Rahmen der RELIABILITY-Prozedur von SPSS möglich.

[12] Die Polung muß jedoch schon vor der Bildung der vorläufigen Skala für alle Items gleich sein, ist also ggf. anzupassen!

[13] Keinesfalls sollte die SPSS-Funktion `SUM(v1, v2, ..., vn)` verwendet werden - sie rechnet falsch!

berechnen ist. In diesem Fall werden die Items der Kaufsuchtskala (`v356` bis `v372`) aufsumiert. Der Befehl `compute` endet (wie <u>jeder</u> Befehl) mit einem Punkt. Anschließend wird SPSS durch den Befehl `execute` veranlaßt, die soeben definierte Transformation auszuführen. Um diese beiden Anweisungen im Syntaxfenster auszuführen, müssen sie vorher <u>beide markiert</u> werden, dann kann die Schaltfläche **Ausführen** gewählte werden.

Nun wird über den Befehl `frequencies` (Statistik → deskriptive Statistik → Häufigkeiten) festgestellt, welche beiden Werte von `ksroh` von jeweils 25% der Befragten unter- bzw. überschritten werden (der folgende Output ist gekürzt):

KSROH

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent	
	17,00	92	6,0	6,4	6,4	
	18,00	86	5,6	5,9	12,3	
	19,00	77	5,0	5,3	17,6	
	20,00	82	5,4	5,7	23,3	
	21,00	**97**	**6,4**	**6,7**	**30,0**	¬
	22,00	77	5,0	5,3	35,3	
	...					
	30,00	57	3,7	3,9	70,2	
	31,00	**36**	**2,4**	**2,5**	**72,7**	¬
	32,00	45	2,9	3,1	75,8	
	33,00	36	2,4	2,5	78,3	
	...					
	60,00	1	,1	,1	99,8	
	61,00	1	,1	,1	99,9	
	62,00	2	,1	,1	100,0	
	,	79	5,2	Missing		
	Total	1527	100,0	100,0		

Befragte, die einen Wert < 21 auf der Skala erzielen, liegen im unteren Viertel der vorläufigen Skala, Befragte, die einen Wert > 31 erreichen, liegen im oberen Viertel der vorläufigen Skala. Aufgrund dieser Informationen ist eine dichotome Gruppierungsvariable zu bilden, nach der im t-Test gruppiert werden kann. Dies erfolgt wieder „per Hand“ im Syntax-Fenster:

```
if (ksroh<21) gruppe=2.
if (ksroh>31) gruppe=1.
execute.
```

Der Befehl `if` besteht aus einer bestimmten Bedingung (in Klammern), die erfüllt sein muß, damit die nachfolgende Rechenoperation ausgeführt wird. Die Rechenoperation wird genauso formuliert wie beim `compute`-Befehl. Es entsteht eine neue Variable `gruppe` mit den Ausprägungen `1` und

[14] Zu beachten ist, daß Variablennamen in SPSS maximal aus 8 Zeichen bestehen dürfen. Es sollten möglichst „normale“ Buchstaben und Zahlen verwendet werden, nicht zugelassen sind mathematische Zeichen (+,-,*, ...) oder Umlaute (ä,ö,ü,ß, ...).

2. Der Wert 2 kennzeichnet die Zugehörigkeit des/der Befragten zur unteren Gruppe, der Wert 1 steht für die Zugehörigkeit zur oberen Gruppe. Alle Befragten, die in keine der beiden Gruppen fallen, gelten in der Variable `gruppe` als `Missing`.

Jetzt folgen die t-Tests mit den Items als abhängigen Variablen. Aus Platzgründen betrachten wir nur den Output des ersten t-Tests, die weiteren T-tests werden analog interpretiert:

```
t-tests for independent samples of  GRUPPE   Extremgruppen von KSROH

                              Number
           Variable          of Cases    Mean        SD       SE of Mean
          ---------------------------------------------------------------
          V356  Geld muß ich ausgeben

          obere Gruppe        395       2,3873      ,830        ,042
          untere Gruppe       337       1,1128      ,335        ,018
          ---------------------------------------------------------------

          Mean Difference = 1,2746
          Levene's Test for Equality of Variances: F=387,365 P= ,000

       t-test for Equality of Means                                95%
  Variances  t-value   df    2-Tail Sig    SE of Diff       CI for Diff
------------------------------------------------------------------------
  Equal     26,40    730         ,000         ,048        (1,180; 1,369)
  Unequal   27,95    535,79      ,000         ,046        (1,185; 1,364)
```

Zunächst ist festzuhalten, daß die obere Gruppe (wie erwartet) einen höheren Mittelwert beim Item `V356` erzielt als die untere Gruppe. Die Mittelwertdifferenz ist also positiv (`Mean Difference`). Der `Levene's Test` auf Gleichheit der Varianzen ist signifikant (Irrtumswahrscheinlichkeit `P = 0,000`), d.h. die Varianzen der beiden Gruppen sind unterschiedlich. Deshalb sind die Ergebnisse der Zeile `Unequal` für uns relevant. Für die Prüfgröße $\hat{t}$ (`t-value`) erhalten wir `27,95`. Dieser Wert ist um ein Vielfaches größer als der kritische Wert für t bei einseitiger Fragestellung, `536` Freiheitsgraden (`df`) und 5%iger Irrtumswahrscheinlichkeit ($t_{0,05;536}$ = 1,65). Wir können deshalb H_0 zurückweisen, der Mittelwertunterschied ist signifikant.

Die Ergebnisse aller t-Tests sind in der folgenden Tabelle zusammengefaßt:

Ergebnisse der Itemanalyse der Kaufsuchtskala t-Werte der `t-tests` und Trennschärfekoeffizienten der `Reliability`-Prozedur			
Item	$\hat{t}$	corrected Item-Scale-Correlation	Kurzbeschreibung des Items
V356	27,95	,6048	Geld muß ich ausgeben
V357	40,25	,6970	Im Kaufhaus - Verlangen zu kaufen
V358	37,04	,6959	Drang, loszugehen und etwas zu kaufen
V359	39,55	,6748	Impuls, etwas zu kaufen
V360	39,76	,6859	Gefühl etwas haben zu müssen
V361	31,25	,5211	Wichtigkeit von Einkäufen ?
V362	20,92	**,4430**	kaufen, weil etwas billig ist
V363	31,16	,6926	Lust etwas zu kaufen
V364	21,46	,5474	Bestellung über Werbebriefe
V365	25,71	,5803	nicht benutzen von Gekauftem
V366	27,48	,6734	kaufen, obwohl ich es mir nicht leisten kann
V367	22,32	,6205	Ich bin verschwenderisch
V368	27,96	,6057	Kaufen ist wie Entspannen
V369	29,55	,7080	Antrieb, einkaufen zu gehen
V370	26,35	,6210	manchmal schlechtes Gewissen nach Kauf
V371	18,49	,5665	gekaufte Sachen anderen nicht zeigen
V372	**16,06**	**,2763**	Einkaufen mit Kreditkarten ist praktisch

Für alle t-Tests gilt als kritischer Wert ein $t_{0,05;df}$ von 1,65. Kein einziges $\hat{t}$ unterschreitet diesen Wert; im Gegenteil, die Mittelwertdifferenzen aller Items sind hochsignifikant. Kann daraus geschlossen werden, daß alle Items „sehr gut" sind? Nein! Wie schon weiter oben dargestellt, ist die Größe der Prüfgröße $\hat{t}$ neben der tatsächlichen Differenz sehr stark abhängig von der Größe der beiden Fallgruppen und damit von der Zahl der Freiheitsgrade. Letztere betragen in unserem Beispiel (korrigiert) `df = 536`. Bei dieser hohen Zahl von Freiheitsgraden sind bereits minimale Mittelwertdifferenzen signifikant.

Bei großen Stichproben ist es deshalb immer ratsam, anstelle des t-Tests zu Bestimmung der Trennschärfe auf die korrigierte Item-Skalen-Korrelation der Reliabilitätsanalyse (`RELIABILITY`) **zurückzugreifen.** Hier gilt die Faustregel, daß der Trennschärfekoeffizient > 0,5 sein sollte, um von einem trennscharfen Item zu sprechen. Wie in der Tabelle zu sehen, ist das Item `V372` alles andere als trennscharf, es korreliert nur sehr gering mit der Summe der übrigen Items der Skala (0,2763). Dieses Item erzielte auch das geringste $\hat{t}$ (16,06). Die <u>endgültige</u> Skala würde ohne dieses Item gebildet werden.[15]

[15] Ferner würde man möglicherweise auch noch das Item `V362` eliminieren, sofern dadurch die Reliabilität der Skala (Cronbachs α) noch zu steigern wäre. Auch eine Faktorenanalyse zur Überprüfung der Eindimensionalität der Skala ist empfehlenswert.

Inhalt