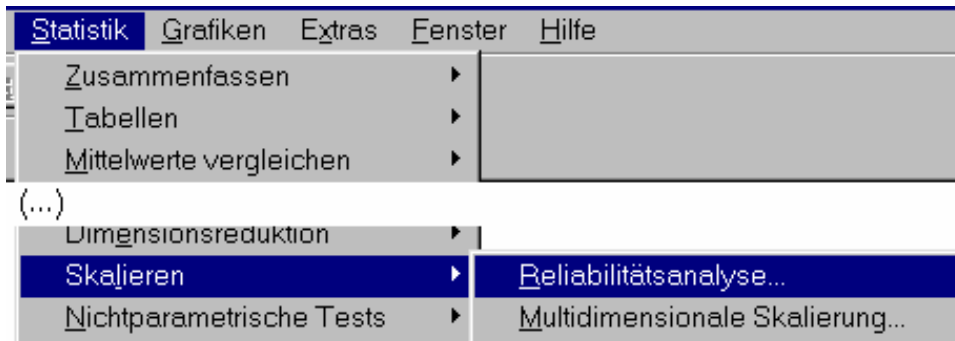


# Reliabilitätsanalyse mit SPSS

Die meisten statistischen Funktionen, die bei der (Reliabilitäts-) Analyse eines auf der klassischen Testtheorie aufbauenden Tests geläufig sind, finden sich unter „Skalieren“-„Reliabilitätsanalyse“:



Die wesentlichen Schritte bei der Reliabilitätsanalyse sind die Wahl der Items und die Wahl eines Reliabilitätskriteriums:

The image shows the 'Reliabilitätsanalyse' dialog box in SPSS. It has a list of items on the left (tr011 to tr020) and a list of items on the right (tr001 to tr009). The 'Modell' dropdown is set to 'Alpha'. There are buttons for 'OK', 'Einfügen', 'Zurücksetzen', 'Abbrechen', and 'Hilfe'. A 'Statistik...' button is also present. Several callout boxes provide additional information:

- Items der zu analysierenden Skala:** Setzt sich ein Test aus mehreren Skalen zusammen (z.B. Extraversion und Neurotizismus) muss die Reliabilitätsbestimmung selbstverständlich für jede Skala gesondert erfolgen, d.h. immer nur für Items, die zusammen auch nur eine Dimension erfassen sollen.
- Reliabilitätsmaß (s.u.):** Die Standard-einstellung „Alpha“ (=Cronbachs Alpha) ist das verbreitetste Maß für die interne Konsistenz einer Skala.
- Statistiken s.u.**
- Zeigt zu Beginn der Ausgabe eine Liste mit den Variablennamen und -labels der Items in der Analyse (i.A. ziemlich überflüssig).**

## Wahl des Reliabilitätskriteriums

Die zur Auswahl stehenden Kriterien sind im weiteren Sinne alles Tests zur Prüfung der internen Konsistenz. Cronbachs Alpha („Alpha“) ist die Standardeinstellung und zugleich das gängigste Reliabilitätskriterium, wenn nur Daten von einem Messzeitpunkt vorliegen. Hier Kurzdefinitionen der verfügbaren Methoden:

- **Alpha (Cronbach):** Dies ist ein Modell der inneren Konsistenz, welches auf der durchschnittlichen Korrelation zwischen den Items beruht.
- **Split-Half:** Bei diesem Modell wird die Skala in zwei Hälften geteilt und die Korrelation zwischen den beiden Hälften berechnet. **Anmerkung:** SPSS teilt die Items einfach nach ihrer Reihenfolge in der Variablenliste in eine erste und zweite Hälfte. Dies entspricht *nicht* dem üblicherweise empfohlenen Vorgehen, Testhälften auf Basis von Itemschwierigkeit und -trennschärfe zu bilden! Hierzu muss man die Zuordnung der Items zu den Testhälften „per Hand“ (d.h. nach Ansicht der Itemstatistiken, s.u.)

vornehmen und die Items dann in der entsprechenden Reihenfolge in der Variablenliste anordnen.

- *Guttman*: Bei diesem Modell werden Guttmans untere Grenzen für die wahre Zuverlässigkeit berechnet.
- *Parallel*: Bei diesem Modell wird angenommen, dass alle Items gleiche Varianzen und gleiche Fehlervarianzen für mehrere Wiederholungen aufweisen.
- *Streng parallel*: Bei diesem Modell werden die Annahmen des parallelen Modells übernommen und zusätzlich die Gleichheit der Mittelwerte der Items angenommen.

### Zusätzliche Statistiken

Wählt man nur Items und eine Methode zur Reliabilitätsbestimmung, besteht die Ausgabe auch lediglich aus den Informationen zur Reliabilität. Über das Dialogfeld „Statistik“ können eine Reihe mehr oder weniger nützlicher Zusatzinformationen angefordert werden. Die in den meisten Fällen wichtigsten Informationen, nämlich die Itemgütekriterien Varianz, Schwierigkeit und Trennschärfe, finden sich hierbei unter „Deskriptive Statistik...“ (s.u. unter 1).

#### 1) Deskriptive Statistiken für:

- *Item*: Zeigt Mittelwerte und Standardabweichungen aller Items an. Es gibt keine Option zur Berechnung der Itemschwierigkeit, hierfür muss der Itemmittelwert herangezogen werden und die Berechnung der Schwierigkeit dann von Hand erfolgen.
- *Skala*: Zeigt Skalenmittelwert, -varianz und -standardabweichung an, d.h. Deskriptivstatistiken für den Summenwert aus allen Items.
- *Skala wenn Item gelöscht*: Zeigt eine Auswertungsstatistik, bei der jedes Item mit dem Testwert verglichen wird, welche *ohne* das jeweilige Item gebildet wird. Die Statistik enthält den Mittelwert, Varianz und Cronbachs Alpha der Skala ohne das jeweilige Item, außerdem die Korrelation zwischen dem Item und der Skala ohne das jeweilige Item („corrected item-total-correlation“) und die multiple Korrelation des Items mit allen anderen Items als Prädiktoren. Die „corrected item-total-correlation“ ist die *Itemtrennschärfe!*

- 2) **Zwischen Items Korrelationen und Kovarianzen:** Kovarianz- und Korrelationsmatrizen aller Items untereinander.
- 3) **Auswertung** (Mittelwert, Varianzen, Kovarianzen, Korrelationen): Gibt deskriptive Statistiken für Itemparameter an. Insbesondere interessant, weil bei Testkonstruktionsstudien häufig berichtet, sind die hierüber erhältlichen Minima, Maxima und Mittel der Iteminterkorrelationen. *Allerdings:* SPSS bildet den Mittelwert der Interkorrelationen untransformiert. Dies ist eigentlich nicht zulässig, die Koeffizienten müssten vorher einer Transformation in Fisher-Z-Werte unterzogen werden. Dies wiederum geht nur „per Hand“.
- 4) **ANOVA-Tabelle:** Hier wird eine Varianzzerlegung in die Varianz zwischen und innerhalb der Personen sowie zwischen Items und Residualvarianz vorgenommen. Der F-Test hierbei entspricht einer Varianzanalyse mit Messwiederholungen mit den Items als „Messzeitpunkten“. Er sagt nichts über die Reliabilität der Skala und ist eigentlich überflüssig, interessanter ist die „between people“-Varianz, diese sollte im Vergleich zur „within people“-Varianz möglichst hoch sein. Die drei Methoden unterscheiden sich im Skalenniveau der Daten (F-Test: Intervallskaliert; Friedmanns Chi-Quadrat: Ordinal; Cochrans Chi-Quadrat: Dichotome Daten).
- 5) **T-Quadrat nach Hotelling:** Ein multivariater Test für die Nullhypothese, dass alle Items einer Skala den gleichen Mittelwert besitzen.
- 6) **Tukey-Additivitäts-Test:** Testet die Annahme, dass keine multiplikative Wechselwirkung zwischen Items vorliegt. (Schätzt die Potenz, mit der eine Skala potenziert werden muss, um Additivität zu erhalten.)

### **Sonstige Gütekriterien**

Das Modul „Reliabilitätsanalyse“ ist – wie der Name sagt – nicht zur Prüfung von *Validitätskriterien* gedacht oder geeignet. Als Kriterium der Konstruktvalidität wird in der Literatur z.T. die „faktorielle Validität“ angeführt, d.h. ob die Items einer als eindimensional konzipierten Skala auf einem gemeinsamen Faktor laden oder sich theoretisch angenommene Dimensionen in einer Faktorenanalyse als Faktoren wiederfinden lassen. Hierfür ist die Faktorenanalyse und das zugehörige SPSS-Modul das Verfahren der Wahl – oder aber eine Analyse auf Basis linearer Strukturgleichungsmodelle, z.B. mit LISREL.

Zur Ermittlung von Maßen der Kriteriumsvalidität oder der diskriminanten oder konvergenten Validität wird der Summenscore der Skala als eigenständige Variable benötigt. Diese muss über „Transformieren“-„Berechnen“ gebildet werden. Für einen Testwert mit dem Variablennamen „score“, der die Summe aus Items namens „a01“ bis „a12“ darstellt, würde das z.B. aussehen wie auf der nächsten Seite dargestellt. Das gleiche Ziel würde über folgende Syntax erreicht:

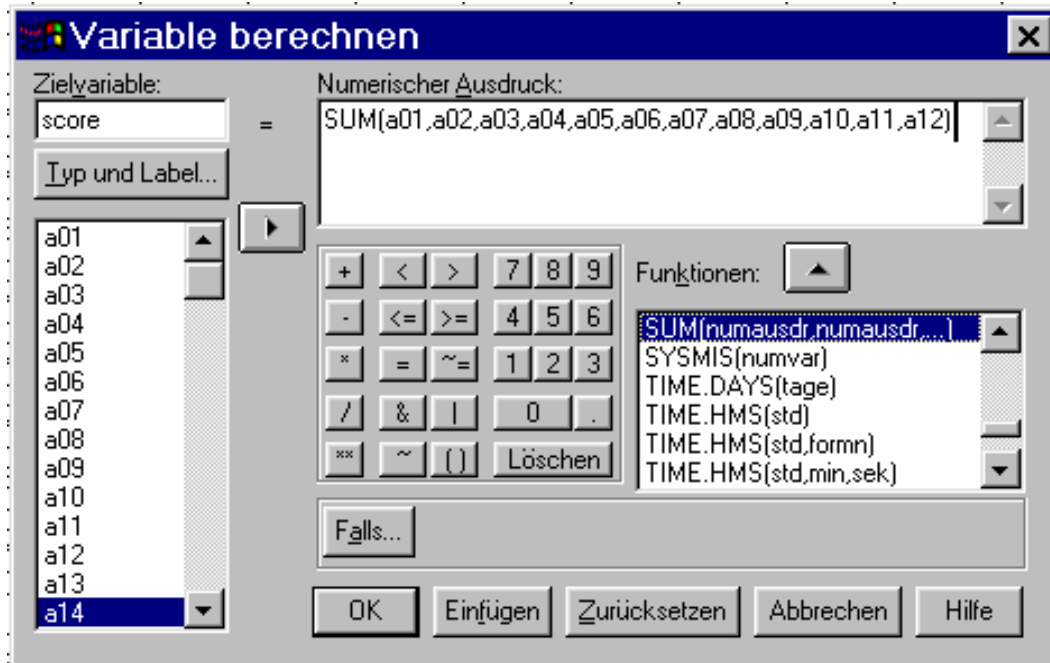
```
COMPUTE score = SUM(a01, a02, a03, a04, a05, a06, a07, a08, a09,
a10, a11, a12) .
EXECUTE .
```

Falls die zu summierenden Variablen in der Datendatei am Stück in aufeinanderfolgenden Spalten stehen (d.h. ohne andere Variablen dazwischen!) ist auch folgende Kurzschreibweise möglich:

```
COMPUTE score = SUM(a01 TO a12) .
EXECUTE .
```

Bei der Berechnung mehrerer (Sub-) Skalenwerte stellt das Vorgehen über die Syntax den bequemeren Weg dar.

Ist eine Testwertvariable gebildet, lässt sich z.B. ein Maß für die Kriteriumsvalidität als einfache Korrelation dieser Variablen mit einem (quantifizierten) Außenkriterium berechnen (z.B. Testwert Konzentrationsleistung mit der Fehlerzahl bei der Montage von Fernsehgeräten).



*Berechnung eines Summenwertes über das Dialogfeld „Transformieren“/„Berechnen“*

### **Retest-Reliabilität**

Mit den Daten einer Stichprobe zu einem Zeitpunkt, die mit dem Modul „Reliabilitätsanalyse“ typischerweise ausgewertet werden, lässt sich natürlich auch keine Retest-Reliabilität bestimmen. Auch hierfür ist – selbstverständlich neben wiederholter Datenerhebung – die Bildung von Summenwerten und die Berechnung der Korrelation zwischen Testwert zum Zeitpunkt eins und zum Zeitpunkt zwei nötig.

### **Allgemeine Anmerkung zur Ausgabe**

Die Ausgabe des Moduls „Reliabilitätsanalyse“ besteht im Ausgabe-Navigator auch bei SPSS 7.x bis 10.x aus „Textausgabe“-Feldern. In diesen Feldern steht derselbe Schreibmaschinen-Schrift-Output, den ältere SPSS-Versionen ausgeben. Entsprechend ist auch die Beschriftung der Ausgabe ausschließlich englisch.