

SPSS-Beispiel zu Kapitel 6: Methoden der Reliabilitätsbestimmung¹

Karin Schermelleh-Engel & Christina S. Werner

Inhaltsverzeichnis

1	Empirischer Datensatz.....	1
2	Interne Konsistenz.....	2
3	Split-Half-Reliabilität.....	4
4	Retest- Reliabilität.....	6
5	Literatur.....	11

1 Empirischer Datensatz

Die verschiedenen Methoden der Reliabilitätsschätzung werden demonstriert am Beispiel der Skala "Aufgeregtheit" des TAI-G (Test Anxiety Inventory - German; Hodapp, 1991, 1996). Diese Skala misst emotionale und physiologische Aspekte der Prüfungsangst anhand der folgenden acht Items:

Item 1: „Ich spüre ein komisches Gefühl im Magen.“

Item 2: „Ich bin am ganzen Körper verkrampft.“

Item 3: „Ich fühle mich unbehaglich.“

Item 4: „Das Herz schlägt mir bis zum Hals.“

Item 5: „Ich fühle mich ängstlich.“

Item 6: „Ich zittere vor Aufregung.“

Item 7: „Ich habe ein beklemmendes Gefühl.“

Item 8: „Ich bin aufgeregt.“

¹ „IBM SPSS Statistics“ hieß früher „PASW Statistics“. - Alle Abbildungen in diesem Beitrag stammen aus dem Programm IBM SPSS Statistics, Release 18.0 2009. Chicago: SPSS, an IBM Company. Wir danken IBM für die freundliche Genehmigung zur Verwendung der Screenshots.

Die Antwortskala ist eine vierstufige Likert-Skala mit den Abstufungen "1 = fast nie", "2 = manchmal", "3 = oft" und "4 = fast immer".

Die Daten entstammen einer umfangreichen Längsschnittstudie (Keith, Hodapp, Schermelleh-Engel & Moosbrugger, 2003; Schermelleh-Engel, Keith, Moosbrugger & Hodapp, 2004). Für die folgenden Analysen wurde per Zufall eine Teilstichprobe von $N = 30$ Personen ausgewählt, die den Test zu zwei Messzeitpunkten bearbeitet hatten.

Wird das Programm SPSS gestartet und die Datei "Pruefungsangst.sav" geladen, so zeigt der Daten-Editor die Rohwerte der 30 Personen an, die in 30 Zeilen (für die Personen) und 16 Spalten (für die acht Items zu den zwei Messzeitpunkten) angeordnet sind (Abbildung 1). Die ersten acht Spalten beinhalten die Antworten auf die acht Items der Skala "Aufgeregtheit" zum ersten Messzeitpunkt, die weiteren acht Spalten die Antworten auf die acht Items zum zweiten Messzeitpunkt.

	Item1_1	Item2_1	Item3_1	Item4_1	Item5_1	Item6_1	Item7_1	Item8_1	Item1_2	Item2_2	Item3_2	Item4_2	Item5_2	Item6_2	Item7_2	Item8_2
1	2	1	2	1	2	1	2	1	2	1	1	1	1	1	1	2
2	4	2	3	1	3	1	3	4	3	1	4	1	4	1	4	4
3	1	2	3	3	1	3	4	4	2	1	4	2	1	3	4	4
4	4	1	4	3	4	1	2	4	3	2	4	2	4	2	2	4
5	1	2	3	3	2	2	2	3	1	2	2	2	2	2	2	3
6	4	2	4	3	2	3	3	4	3	2	3	2	2	4	3	4
7	4	3	4	3	3	2	3	4	4	2	4	3	3	1	4	4
8	2	1	2	1	2	1	2	2	2	1	2	2	2	1	1	2
9	2	1	2	2	2	2	2	4	3	2	2	2	3	2	2	4
10	4	3	4	3	3	3	4	4	3	2	3	2	3	2	2	3
11	4	2	4	2	3	1	3	3	4	1	4	2	3	1	4	4
12	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
13	2	1	2	1	1	1	2	3	1	1	2	1	1	1	1	2
14	4	3	4	4	4	3	4	4	4	3	4	4	4	4	2	4
15	4	3	4	4	4	3	4	4	4	3	4	4	4	4	4	4
16	2	1	2	2	1	1	2	1	2	1	2	2	2	1	1	2
17	4	2	4	4	4	4	4	4	3	2	3	3	3	3	2	3
18	3	2	2	2	2	2	3	3	2	1	2	2	2	2	2	2
19	4	2	4	3	3	2	3	3	3	2	3	2	3	2	3	3
20	4	1	4	3	2	2	2	4	3	2	3	3	2	2	2	3
21	4	2	2	4	2	4	2	4	4	3	3	4	3	4	4	4
22	4	3	4	3	4	1	3	4	4	3	3	3	3	1	4	3
23	4	3	4	4	4	3	4	4	4	2	3	4	3	2	3	3
24	3	1	3	1	1	1	2	3	3	1	2	1	1	1	2	2
25	2	2	3	1	3	1	2	4	2	1	3	1	3	1	2	3
26	3	1	2	3	1	2	2	3	3	1	2	3	1	2	2	3
27	3	2	2	2	2	2	2	3	2	2	3	2	3	1	3	4
28	2	2	3	2	3	2	2	3	2	3	2	2	2	2	2	3
29	3	1	2	1	2	1	2	3	3	1	2	1	2	1	3	2

Abbildung 1. SPSS-Daten-Editor mit den Rohdaten der Datei "Pruefungsangst.sav".

2 Interne Konsistenz

Die am häufigsten verwendete Methode der Reliabilitätschätzung einer Skala ist die Bestimmung der internen Konsistenz über Cronbachs α . Zur Berechnung der internen Konsistenz muss zunächst das Menü *Analysieren* → *Skalierung* → *Reliabilitätsanalyse* aufgerufen werden (Abbildung 2).

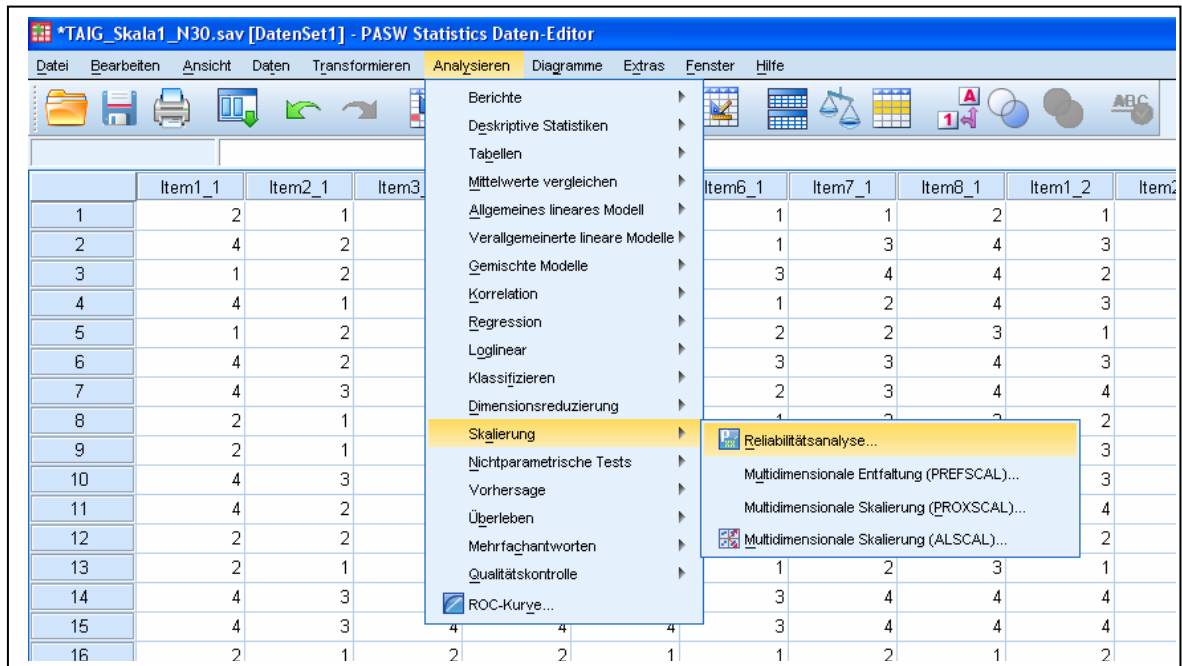


Abbildung 2. Auswahl des Untermenüs "Reliabilitätsanalyse".

Im Menüfenster sind nun links die Variablen untereinander in einem Fenster angeordnet. Um die interne Konsistenz der Skala zum ersten Messzeitpunkt zu bestimmen, werden die ersten acht Items markiert und durch Anklicken des Pfeils in das rechte Fenster mit der Bezeichnung "Items" übertragen (s. Abbildung 3).

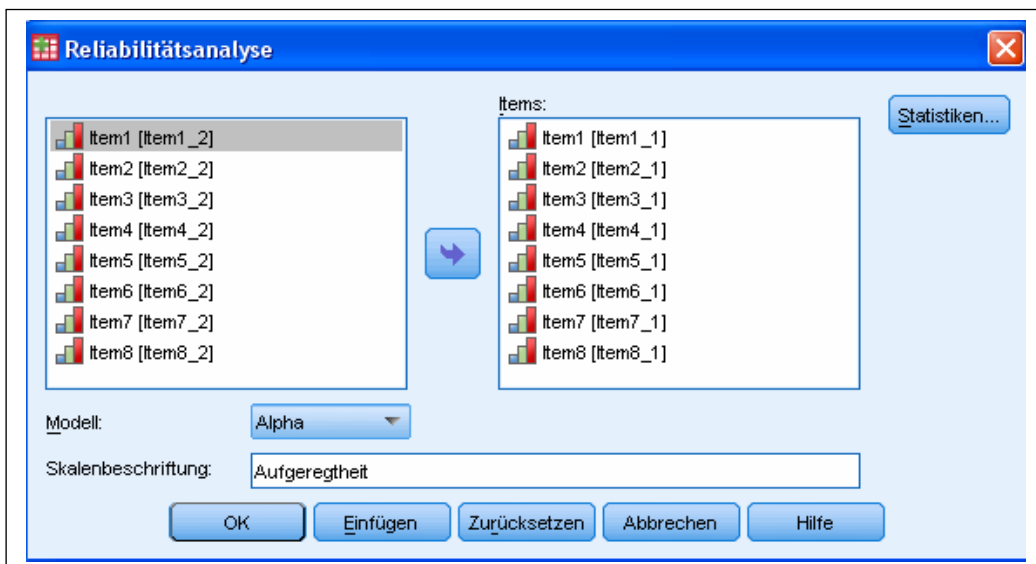


Abbildung 3. Auswahl der Items für die Reliabilitätsanalyse.

Als Modell ist bereits als Voreinstellung "Alpha" eingestellt, mit der Cronbachs Alpha berechnet wird. Bei "Skalenbeschriftung" wird hier die Bezeichnung "Aufgeregtheit" eingegeben, da es sich um die Items dieser Skala handelt. Mit "OK" wird der Befehl abgeschlossen und die Analyse durchgeführt.

Das Ergebnis wird im Ausgabe-Editor "Statistics Viewer" angezeigt (Abbildung 4). Zunächst werden Angaben zur Stichprobengröße gemacht ($N = 30$) und mitgeteilt, ob Fälle aus der Analyse ausgeschlossen wurden (was ist hier nicht erfolgt ist); danach wird das Ergebnis der Reliabilitätsschätzung ausgegeben. Der Wert von $\alpha = .904$ zeigt eine sehr hohe Messgenauigkeit für die acht Items an, die zur Skala "Aufgeregtheit" zusammengefasst wurden.

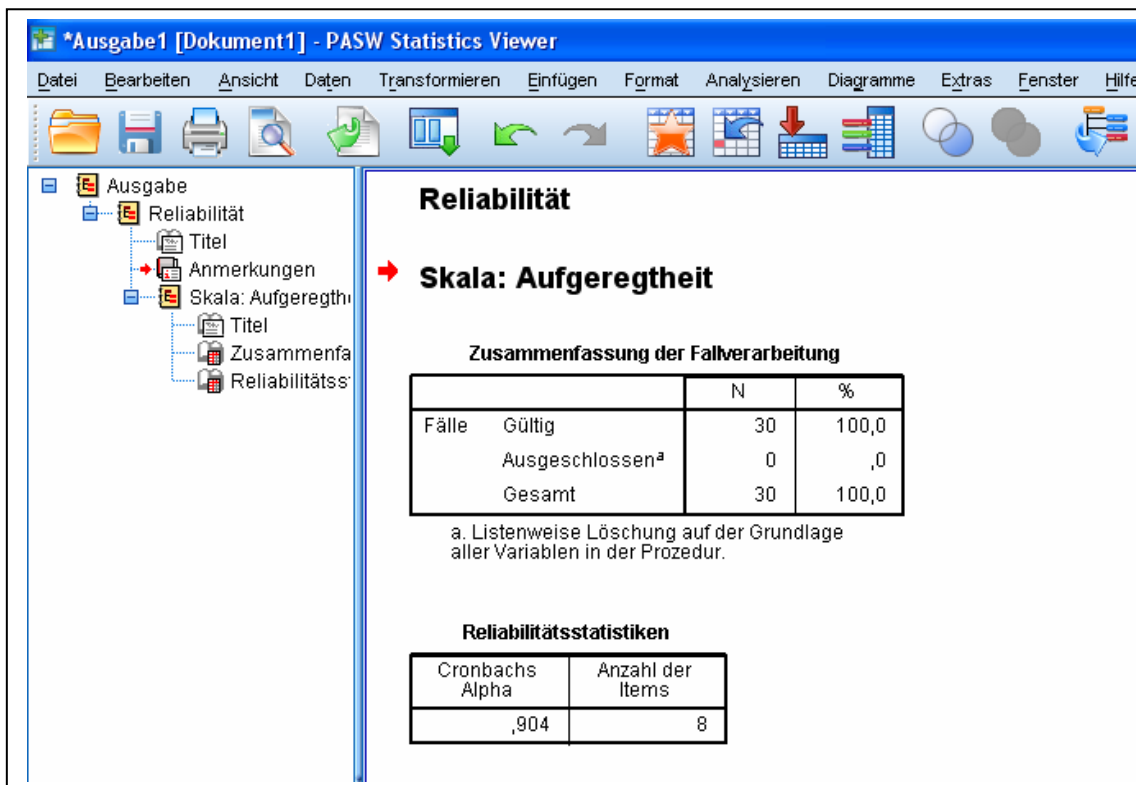


Abbildung 4. Ergebnisse der Berechnung der Internen Konsistenz (Cronbachs Alpha).

3 Split-Half-Reliabilität

Besteht ein Testverfahren aus einer größeren Anzahl von Items, so kann die Reliabilität auch über die Splithalf-(Testhalbierungs-)Reliabilität geschätzt werden. Hierzu werden die Items dieses Tests in zwei möglichst parallele Testhälften aufgeteilt und die

Korrelation der beiden Testhälften bestimmt. Dies soll am vorliegenden Datensatz demonstriert werden.

Zur Testhalbierung gibt es verschiedene Methoden. Häufig werden die Items so aufgeteilt, dass alle geradzahigen Items in die eine Hälfte und alle ungeradzahigen Items in die andere Hälfte kommen (Odd-Even-Methode). Bei heterogenen Items wird die Aufteilung üblicherweise nach der Trennschärfe und der Schwierigkeit der Items vorgenommen (Methode der Itemzwillinge). SPSS nimmt dagegen die Aufteilung standardmäßig so vor, dass die erste Hälfte der Items eines Tests (bzw. einer Skala) die erste Testhälfte bilden und die zweite Hälfte der Items die zweite Testhälfte, was bei homogenen Items vertretbar ist. Hierzu muss im Menü *Reliabilitätsanalyse* als Modell "Split-Half" gewählt werden (Abbildung 5).

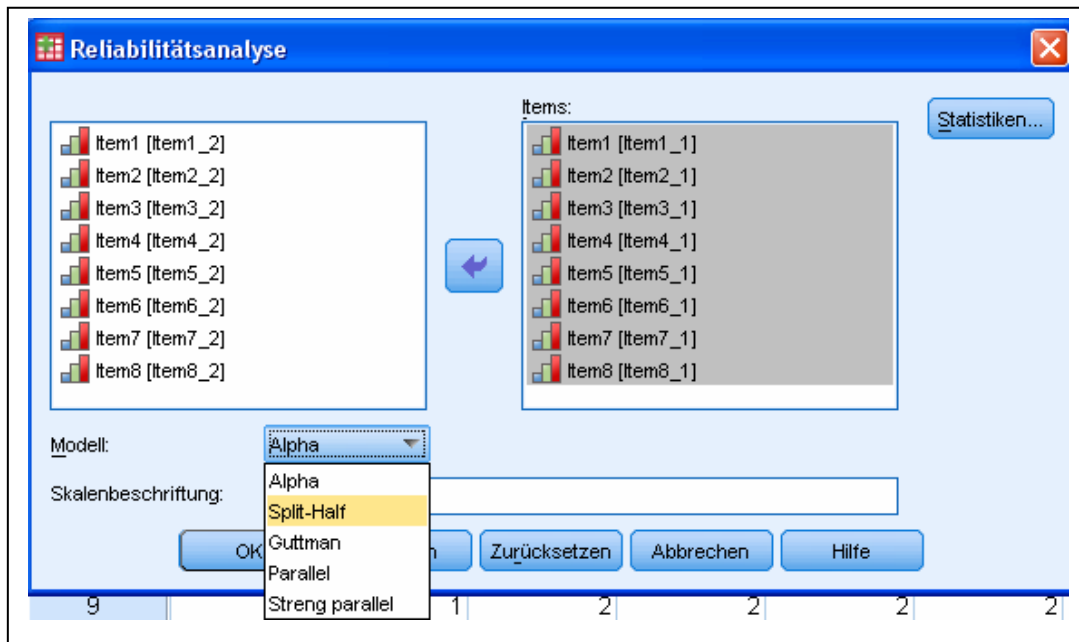


Abbildung 5. Wahl der Split-Half-Methode für die Reliabilitätsschätzung.

Mit der Eingabe "OK" wird der Befehl abgeschlossen und die Ergebnisse werden im "Statistics Viewer" (Abbildung 6) aufgelistet. Die Ergebnisse zeigen, dass beide Testhälften jeweils bereits zufriedenstellende Reliabilitäten (Cronbachs Alpha) aufweisen, dass die beiden Testhälften aber nicht exakt parallel sind: Die erste Testhälfte weist eine höhere interne Konsistenz (.837) auf als die zweite Testhälfte (.773). Die Korrelation zwischen den Testhälften (ohne Spearman-Brown-Korrektur) ist bereits recht hoch (.886). Die Schätzung der Reliabilität der Gesamtskala unter Verwendung der Spearman-Brown-Korrektur, die eine Aufwertung der Halbttest-Reliabilität auf die Reliabilität des Gesamttests vornimmt, zeigt mit einem Wert von

.939 einen etwas höheren Wert an als die Schätzung der Reliabilität der Gesamtskala über Cronbachs Alpha.

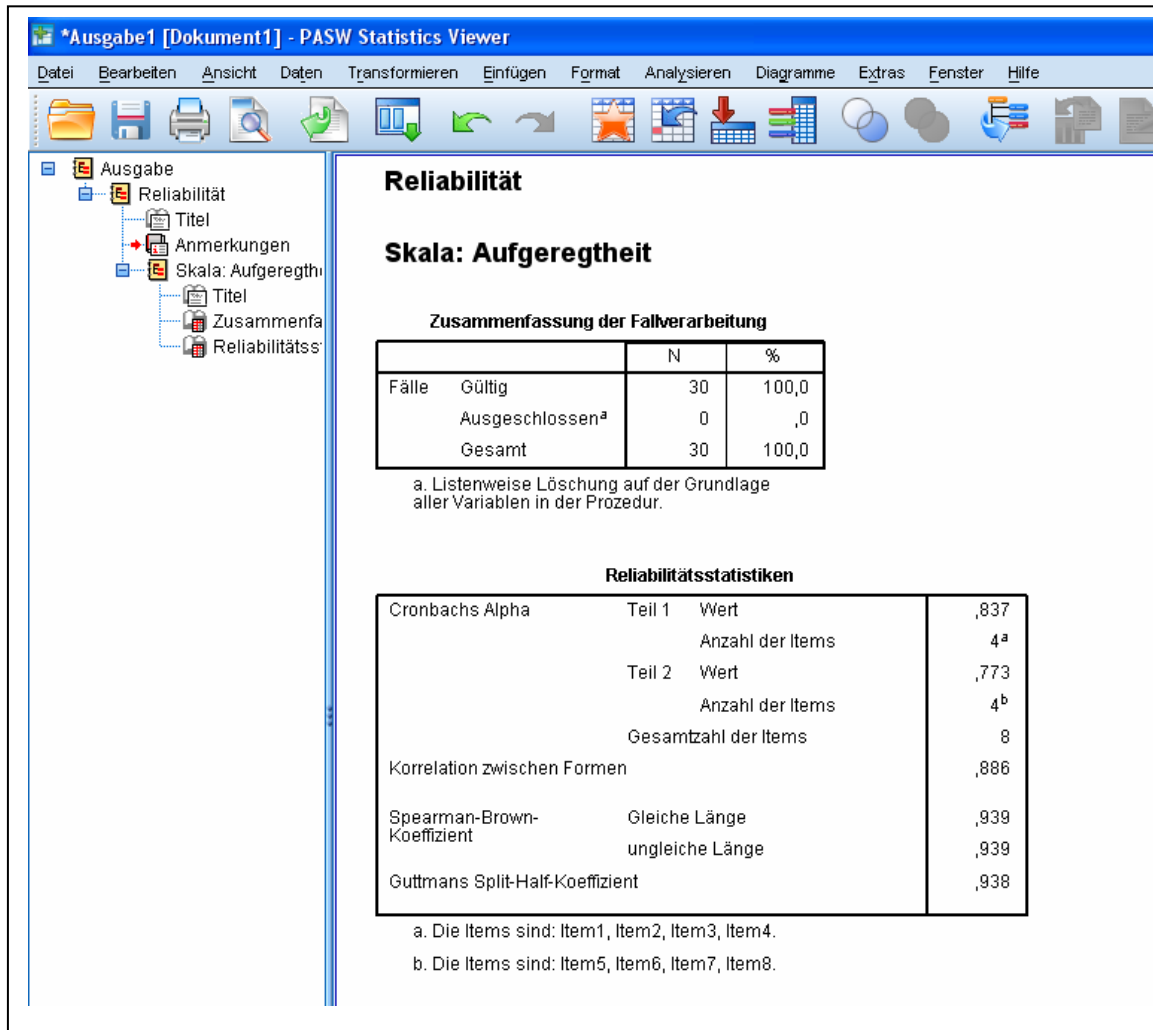


Abbildung 6. Ergebnisse der Berechnung der Split-Half-Reliabilität.

4 Retest-Reliabilität

Zur Berechnung der Retest-Reliabilität könnte man zunächst überprüfen, ob die interne Konsistenz der Skala "Aufgeregtheit" zum zweiten Messzeitpunkt ebenfalls hoch ist. Die Berechnung erfolgt dann analog zur Berechnung der internen Konsistenz der Skala zum ersten Messzeitpunkt, nun mit den Items des zweiten Messzeitpunktes. Die in Abbildung 7 dargestellten Ergebnisse zeigen, dass die Messung zum zweiten Messzeitpunkt ebenfalls hoch konsistent ist (Abbildung 7) mit einem Wert von $\alpha = .915$.

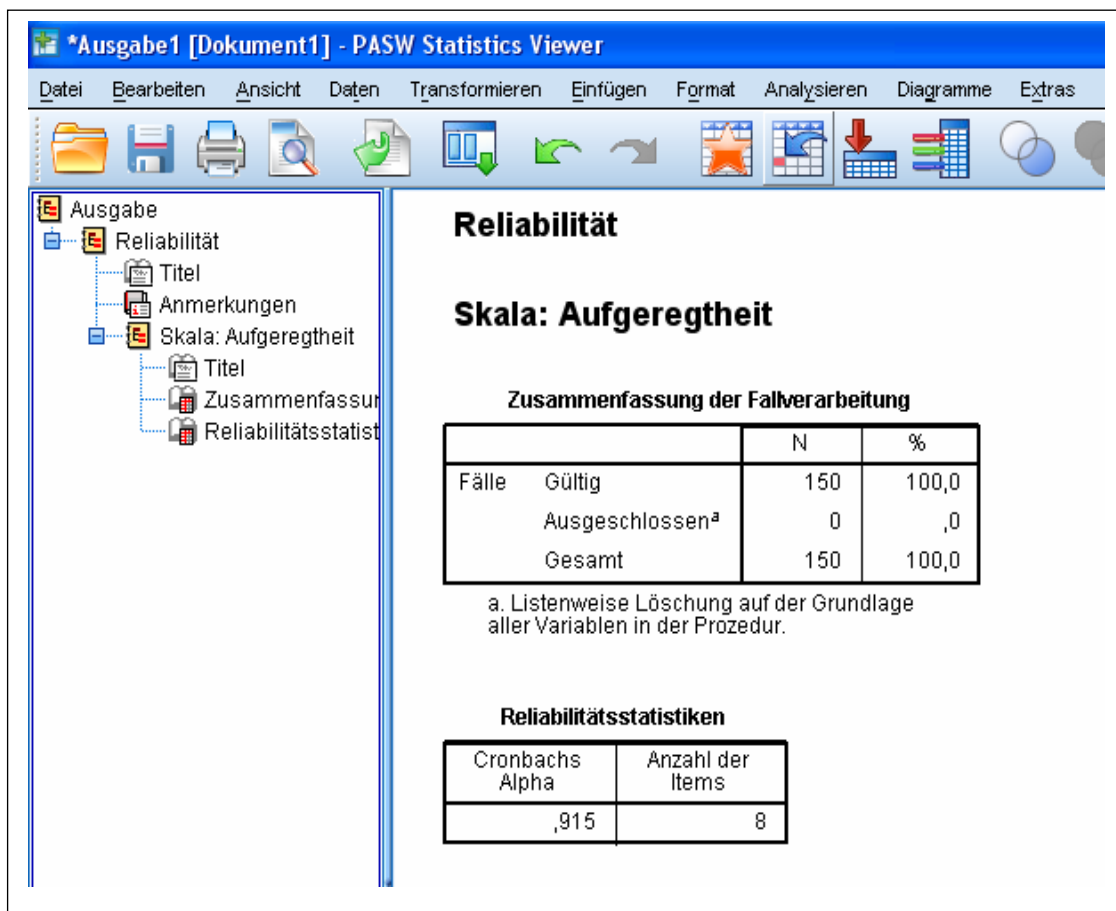


Abbildung 7. Ergebnisse der Berechnung der Internen Konsistenz für die Skala "Aufgeregtheit" zum zweiten Messzeitpunkt.

Zur Berechnung der Retest-Reliabilität müssen nun Skalenwerte aus den acht Items zum ersten Messzeitpunkt sowie aus den acht Items zum zweiten Messzeitpunkt gebildet werden, damit diese Skalenwerte dann miteinander korreliert werden können. Die Reliabilitätsschätzung als Korrelation zwischen den beiden Messzeitpunkten beruht auf der Annahme, dass sich die wahren Werte der Personen zwischen den beiden Messzeitpunkten nicht verändert haben und auch die Messfehlereinflüsse gleich geblieben sind, so dass das Ausmaß, in dem die Korrelation von ihrem Idealwert eins abweicht, eine Abschätzung der Größe des beteiligten Messfehlers erlaubt..

Zur Bildung der Skalenwerte zu den beiden Messzeitpunkten wird im Daten-Editor das Menü *Transformieren -> Variable berechnen* aufgerufen (Abbildung 8).

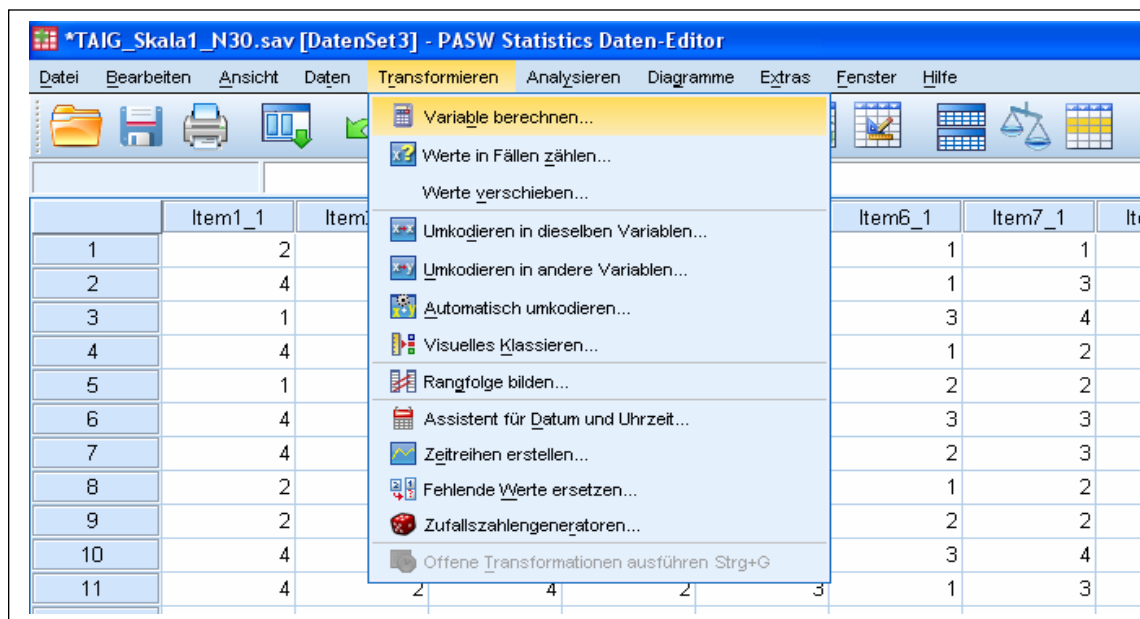


Abbildung 8. Auswahl des Untermenüs "Variable berechnen".

Es öffnet sich das Menü *Variable berechnen*, das in Abbildung 9 dargestellt ist.

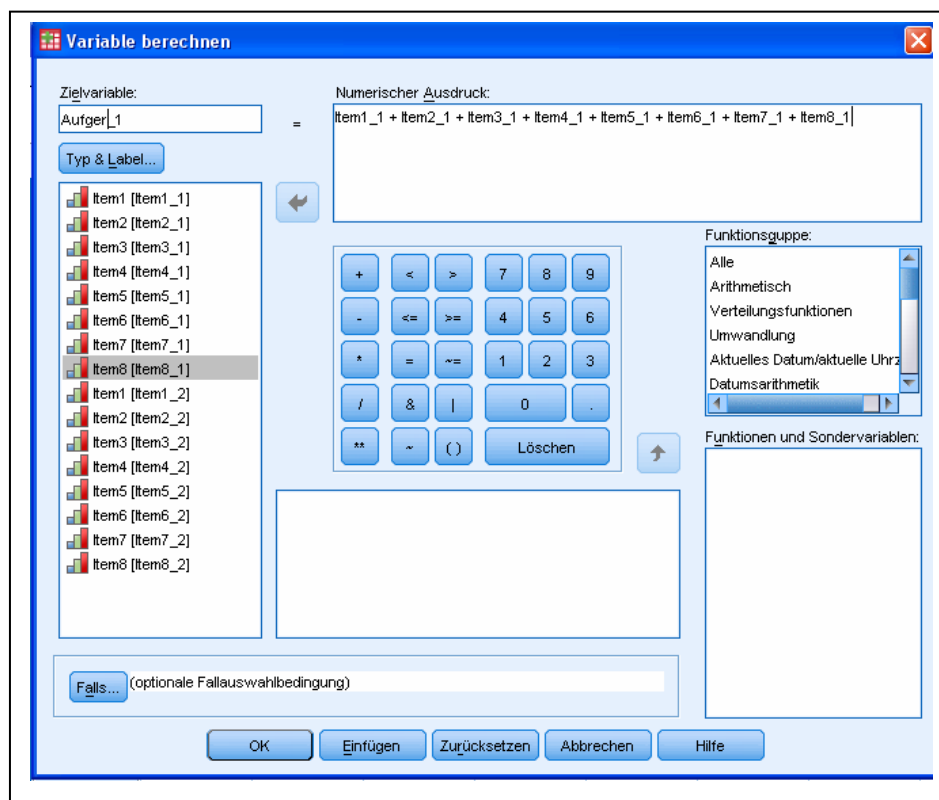


Abbildung 9. Bildung der Skala "Aufgeregtheit" zum ersten Messzeitpunkt (Aufger_1) durch Aufaddieren der Werte der acht Items, die zum ersten Messzeitpunkt erhoben wurden.

Unter "Zielvariable" des Menüs *Variable berechnen* wird zunächst ein sinnvoller Name bzw. eine Abkürzung eingetragen, hier "Aufger_1" (für "Aufgeregtheit" zum Messzeitpunkt 1). Unter "Numerischer Ausdruck" werden die Items eingetragen und durch ein "+" verbunden, die zu einem Skalenwert aufaddiert werden. Dies geschieht konkret, indem zunächst das erste Item (Item1) markiert und durch Anklicken des Pfeils nach rechts ins Fenster "Numerischer Ausdruck" übertragen wird. Danach wird auf dem Rechner das Symbol "+" angeklickt und dieses Symbol ins Fenster "Numerischer Ausdruck" übertragen, danach das zweite Item und schließlich die weiteren sechs Items hinzugefügt. Alternativ kann die Anweisung zur Summenbildung auch direkt in das Textfeld "Numerischer Ausdruck" geschrieben werden. Wenn alle acht Items eingegeben sind, wird die Berechnung durch Anklicken von "OK" ausgeführt.

Die Bildung der Skala "Aufgeregtheit" zum zweiten Messzeitpunkt erfolgt analog, indem alle Items zum zweiten Messzeitpunkt (Item1_2 bis Item8_2) aufaddiert werden. Der SPSS-Daten-Editor enthält nun zwei zusätzliche Variablen, "Aufger_1" und "Aufger_2".

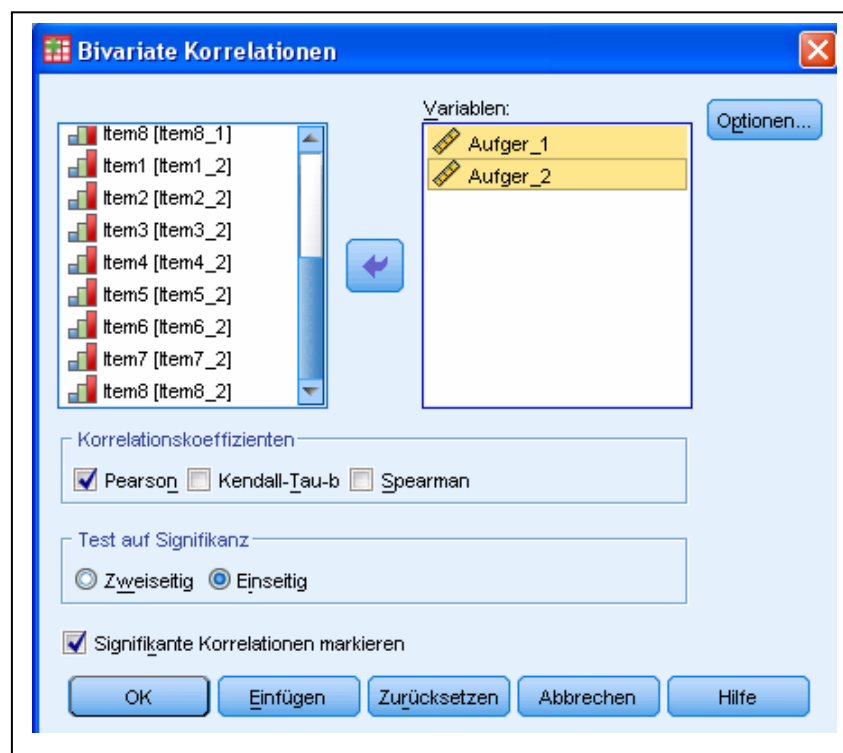


Abbildung 10. Auswahl der beiden Skalen im Untermenü "Bivariate Korrelationen".

Zur Bestimmung der Retest-Reliabilität wird die Korrelation zwischen den beiden Skalen "Aufger_1" und "Aufger_2" bestimmt. Dazu wird das Menü *Analysieren* ->

Korrelation -> Bivariat... aufgerufen. Einige Voreinstellungen im Menü *Bivariate Korrelationen* (Abbildung 10) werden beibehalten: Unter "Korrelationskoeffizienten" die Option "Pearson" sowie - ganz unten - "Signifikante Korrelationen markieren". Unter "Test auf Signifikanz" kann die Voreinstellung "zweiseitig" durch "einseitig" ersetzt werden, da hier ja eine positive Korrelation erwartet wird. Mit der Eingabe "OK" wird der Befehl abgeschlossen.

Die Ergebnisse der Korrelationsanalyse werden im "Statistics Viewer" (Abbildung 11) angezeigt.

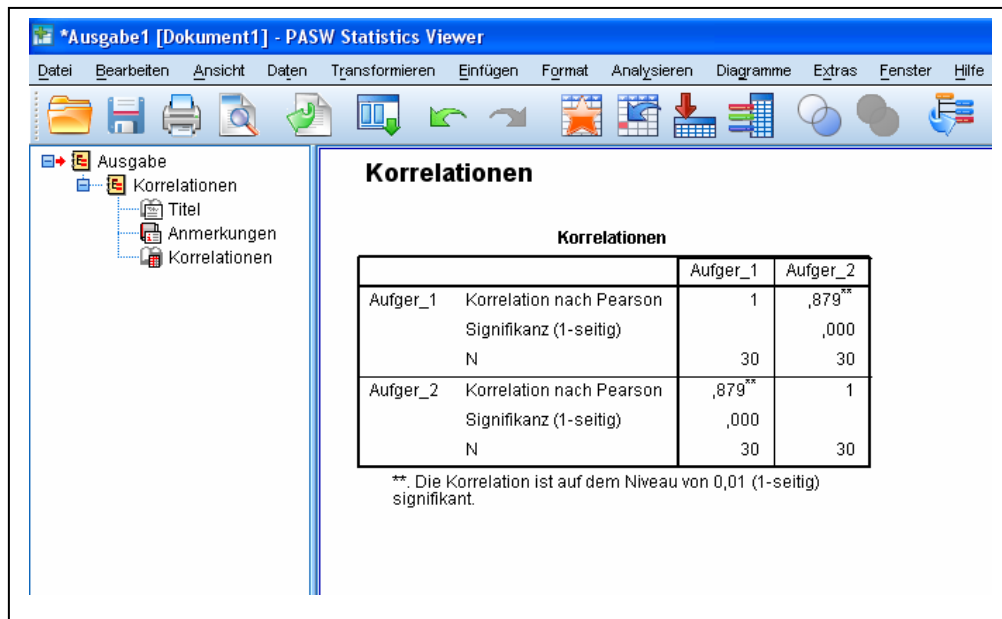


Abbildung 11. Ergebnisse der Berechnung der Retest-Reliabilität über die Korrelation der Skalenwerte von "Aufgeregtheit" zu zwei Messzeitpunkten.

Das Ergebnis der Korrelationsanalyse zeigt eine sehr hohe Retest-Reliabilität mit einer Korrelation von .879 an.

Hätte man anstelle der zwei Messzeitpunkte, bei denen die selben Items erhoben wurden, zwei unterschiedliche, parallele Testformen eingesetzt, so könnte statt der Retest-Reliabilität die Paralleltest-Reliabilität bestimmt werden. Analog zum Vorgehen bei der Retest-Reliabilität werden dann für beide Parallelformen Skalenwerte gebildet, in dem für jede Parallelform die Skalenwerte (die Summen der Items) berechnet werden. Die Korrelation dieser Skalenwerte der Parallelformen ergibt dann die Paralleltest-Reliabilität.

5 Literatur

- Hodapp, V. (1991). Das Prüfungsängstlichkeitsinventar TAI-G: Eine erweiterte und modifizierte Version mit vier Komponenten. *Zeitschrift für Pädagogische Psychologie*, 5, 121-130.
- Hodapp, V. (1996). The TAI-G: A multidimensional approach to the assessment of test anxiety. In C. Schwarzer & M. Zeidner (Eds.), *Stress, anxiety, and coping in academic settings* (pp. 95-130). Tübingen: Francke.
- Keith, N., Hodapp, V., Schermelleh-Engel, & Moosbrugger (2003). Cross-sectional and longitudinal confirmatory factor models for the German Test Anxiety Inventory: A construct validation. *Anxiety, Stress & Coping*, 16, 251 - 270.
- Schermelleh-Engel, K., Keith, N., Moosbrugger, H. & Hodapp, V. (2004). Decomposing person and occasion-specific effects: An extension of latent state-trait theory to hierarchical models. *Psychological Methods*, 9(2), 198-219.