

DATENAUSWERTUNG MIT SPSS

Claudia Krüger, Lars Borgmann, Tobias Antonik, Ann-Kathrin Meyer

Lehrstuhl für Personalentwicklung und Veränderungsmanagement

Zentrum für HochschulBildung

TU Dortmund

Version 1.3 (Sommersemester 2012)

Inhaltsverzeichnis

1 Grundlegendes zur Software SPSS	4
1.1 Ziel und Beschaffung	4
1.2 Programmstruktur	4
1.2.1 Exkurs Syntax-Editor	5
1.3 Unterstützung.....	5
2 Dateneingabe	7
2.1 Variablenansicht	7
2.1.1 Variablenname und -label	7
2.1.2 Typ.....	8
2.1.3 Spaltenformat und Dezimalstellen	8
2.1.4 Wertelabels	8
2.1.5 Fehlende Werte	9
2.1.6 Messniveau	9
2.1.6.1 Exkurs Messniveau	9
2.2 Datenansicht.....	10
2.3 Datenimport aus Excel.....	12
3 Datensichtung.....	13
3.1 Häufigkeiten	13
3.2 Deskriptive Statistik	14
3.2.1 Exkurs Maße der zentralen Tendenz (Lagemaße)	16
3.2.2 Exkurs Streuungsmaße	17
3.3 Diagramme.....	17
3.3.1 Kreisdiagramm	17
3.3.2 Säulendiagramm	18
3.3.3 Histogramm.....	19
3.3.4 Boxplot	19
3.5 Überprüfung auf Normalverteilung	20
3.5.1 Exkurs Normalverteilung.....	23

3.6 Kreuztabelle.....	24
4 Datenstrukturierung	26
4.1 Fälle auswählen	26
4.2 Datei aufteilen	32
5 Datenaufbereitung.....	35
5.1 Umkodieren.....	35
5.2 Visuelles Klassieren	36
5.3 z-Standardisierung	38
6.1 Reliabilitätsanalyse	39
6.1.1 Exkurs Reliabilität	41
6.2 Variable berechnen	42
7 Prüfung von Zusammenhangshypothesen.....	43
7.1 Bivariate Korrelation	43
7.1.1 Exkurs Pearson-Korrelation	46
7.1.2 Streudiagramm	47
7.2 Partielle Korrelation.....	48
7.3 Lineare Regression	49
7.3.1 Exkurs Regression.....	52
7.3.2 Hierarchische Regression	54
7.3.3 Kontrollvariablen.....	55
7.3.4 Dummy-Variablen	58
7.4 Mediatoreffekt	60
7.5 Moderatoreffekt.....	63
7.5.1 Kontinuierlicher Moderator.....	63
7.5.2 Dichotomer Moderator	65
7.5.2.1 Exkurs Vergleich von Korrelationskoeffizienten	68

1 GRUNDLEGENDES ZUR SOFTWARE SPSS

1.1 Ziel und Beschaffung

SPSS steht für ‚Statistical Package for the Social Sciences‘. Es handelt sich um ein modular aufgebautes Programmpaket zur statistischen Datenanalyse. Die erste Version wurde 1968 von Norman H. Nie und seinen Kommilitonen an der Stanford Universität noch für Lochkarten entwickelt und gab gleichzeitig dem Unternehmen (SPSS Inc.) seinen Namen, das zur Verbreitung des entstandenen „Pakets“ gegründet wurde.

2009 wurde das Unternehmen SPSS von IBM übernommen. Die Software selbst hieß zwischenzeitlich PASW (Predictive Analysis SoftWare). Das angebotene Programmpaket wurde im Laufe der Jahrzehnte erheblich erweitert, so dass der Name SPSS heute für das ursprüngliche Kernprodukt verwendet wird.

IBM SPSS Statistics kann auf den Rechnerplattformen Windows, Apple Macintosh (nicht auf PowerPC basierten Systemen) und Linux verwendet werden. IBM Deutschland bietet für die Software „IBM SPSS Statistics“ Studentenversionen an (http://www.spss.com/de/vertical_markets/academia.htm), auch die TU Dortmund hält entsprechende Angebote bereit (<http://www.itmc.tu-dortmund.de/>).

1.2 Programmstruktur

IBM SPSS Statistics umfasst als Software vier Programmebenen: Daten, Syntax, Ausgabe, Skript. Diese Ebenen werden jeweils als separate Dateien genutzt. Um alle Bearbeitungsschritte zu sichern, müssen daher ggf. alle vier Ebenen einzeln gespeichert werden.

- Daten (.sav): Die Datenebene dient der Eingabe und Ansicht der Daten [s. [Dateneingabe](#)]
- Syntax (.sps): Im Syntax-Editor können Befehle zur Durchführung von Berechnungen eingeben, angepasst, kommentiert und ausgeführt werden [s. [Exkurs Syntax-Editor](#)].
- Ausgabe (.spv): Die Ausgabe umfasst die Ergebnisse der ausgeführten Berechnungen, soweit sie sich nicht ausschließlich auf Transformationen der Daten beziehen. Auf Wunsch (inzwischen ist das die Standardeinstellung nach der Installation) wird die jeweils zugrunde liegende Befehlssyntax hier ebenfalls in einem Log berichtet.
- Skript (.www, .sbs): Skripts sind kleine Programme, die bestimmte Funktionen/Berechnungen automatisieren können. SPSS liefert einige Skripte mit, weitere können selbst programmiert werden.

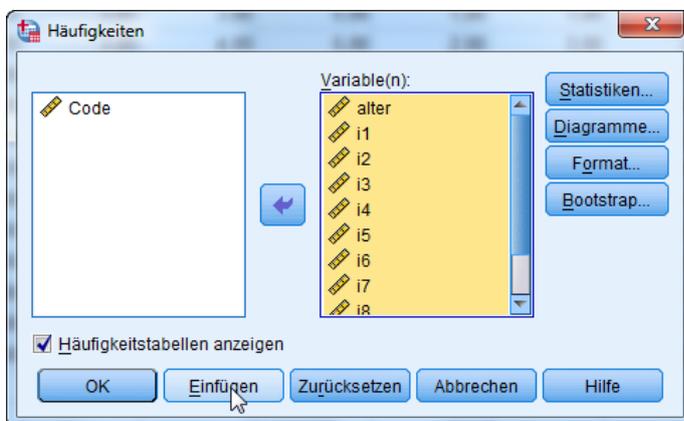
Jede dieser Ebenen zeigt eine typische Menüleiste am oberen Rand. Unter *Bearbeiten/Optionen* können zahlreiche Anpassungen vorgenommen werden, u.a. zur Darstellung und Nutzung der verschiedenen Programmebenen.

1.2.1 Exkurs Syntax-Editor

Die Nutzung der Syntax-Ebene ist sehr zu empfehlen, um eigene Berechnungen zu dokumentieren, auf Knopfdruck wiederholen und bei Bedarf für ähnliche Berechnungen modifizieren zu können.

Der Syntax-Editor ist wie ein Textprogramm zu benutzen, d. h. Textteile können u. a. ausgeschnitten, kopiert, eingefügt, gesucht und ersetzt werden.

Der Syntax-Editor wird über *Datei/Neu/Syntax* aufgerufen. Die Befehle können direkt manuell eingegeben werden (Hilfe hierzu im Menü *Hilfe/Befehlssyntax-Referenz*) oder automatisch erzeugt werden, indem im Anschluss an die Einstellungen in einem Dialogfenster die Option „Einfügen“ statt „OK“ gewählt wird.



Die auszuführenden Befehle werden markiert und mittels Kontextmenü, Tastenkombination (Strg + R), Symbolleiste (grünes Dreieck) oder Menü (Ausführen) gestartet.

1.3 Unterstützung

Im Hilfe-Menü finden sich folgende Optionen, die inzwischen alle (außer der Befehlssyntax-Referenz, s. u.) auf Hilfeseiten im Internet verweisen.

- **Themen:** Die Registerkarten "Inhalt", "Index" und "Suchen" ermöglichen die Suche nach bestimmten Hilfethemen. Über die Schaltfläche „Hilfe“ in den Dialogfenstern können ebenfalls die jeweiligen Themen aufgerufen werden.
- Das **Lernprogramm** bietet eine Einführung ins Programm.
- Die **Fallstudien** liefern praktische Beispiele, die veranschaulichen, wie verschiedene statistische Analysen durchgeführt und die Ergebnisse interpretiert werden. Die in den Beispielen verwendeten Datendateien werden bereitgestellt. Die Beispiele können somit durchgearbeitet werden, um zu verfolgen, wie die Ergebnisse zustande kommen. Innerhalb der Ausgabe können in Pivot-Tabellen über das Kontextmenü weitere Informationen aufgerufen werden.

- Der **Statistik-Assistent** unterstützt die Auswahl einer geeigneten Prozedur per Entscheidungsfragen.
- Bei der **Befehlssyntax-Referenz** handelt es sich um ein pdf-Dokument, das detaillierte Informationen zur Befehlssyntax liefert.

Der technische Support (FAQs, Patches) ist unter <http://support.spss.com/> zu finden. IBM bietet unter dem Namen „SPSS Consulting“ Beratungsleistungen in beliebigem Umfang für Organisationen an.

An der TU Dortmund bietet das „Statistische Beratungs- und Analyse Zentrum“ (SBAZ) der Fakultät Statistik Unterstützung zu SPSS an (<http://www.statistik.uni-dortmund.de/sbaz/de/index.html>).

2 DATENEINGABE

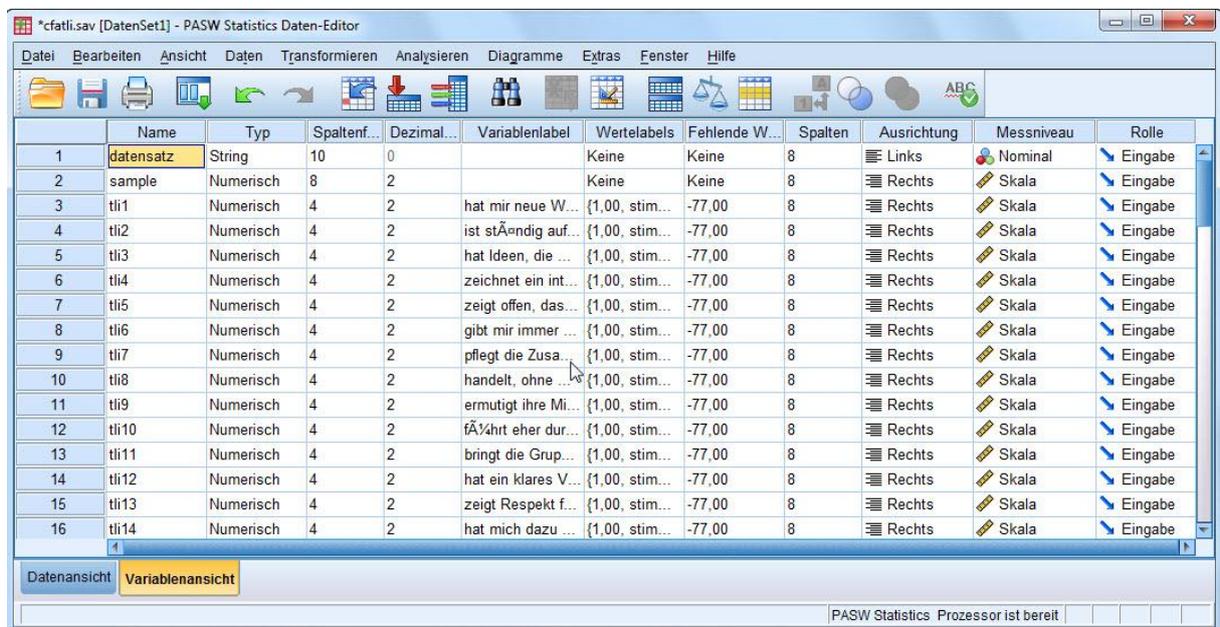
Es werden bei SPSS zwei Ansichten unterschieden: Datenansicht und Variablenansicht.

In der Datenansicht werden die untersuchten Objekte (z. B. Personen) in Zeilen, sog. Fällen, mit ihren jeweiligen Werten in den spaltenweise angeordneten Variablen dargestellt. Diese Ansicht entspricht am ehesten einer Datentabelle wie bspw. bei Excel.

In der Variablenansicht werden die Variablen unabhängig von den einzelnen Fällen und ihren Werten beschrieben. Bevor die Werte eines Falls eingetragen werden können, müssen die Variablen definiert werden. Hierzu dient die Variablenansicht.

2.1 Variablenansicht

Hier kann jede Variable hinsichtlich Name, Typ, Spaltenformat, Dezimalstellen, Variablenlabel, Wertelabel, fehlenden Werte, Spalten, Ausrichtung und Messniveau definiert werden.



	Name	Typ	Spaltenf...	Dezimal...	Variablenlabel	Wertelabels	Fehlende W...	Spalten	Ausrichtung	Messniveau	Rolle
1	datensatz	String	10	0		Keine	Keine	8	Links	Nominal	Eingabe
2	sample	Numerisch	8	2		Keine	Keine	8	Rechts	Skala	Eingabe
3	tli1	Numerisch	4	2	hat mir neue W...	{1,00, stim...	-77,00	8	Rechts	Skala	Eingabe
4	tli2	Numerisch	4	2	ist stÄndig auf...	{1,00, stim...	-77,00	8	Rechts	Skala	Eingabe
5	tli3	Numerisch	4	2	hat Ideen, die ...	{1,00, stim...	-77,00	8	Rechts	Skala	Eingabe
6	tli4	Numerisch	4	2	zeichnet ein int...	{1,00, stim...	-77,00	8	Rechts	Skala	Eingabe
7	tli5	Numerisch	4	2	zeigt offen, das...	{1,00, stim...	-77,00	8	Rechts	Skala	Eingabe
8	tli6	Numerisch	4	2	gibt mir immer ...	{1,00, stim...	-77,00	8	Rechts	Skala	Eingabe
9	tli7	Numerisch	4	2	pfl egt die Zusa...	{1,00, stim...	-77,00	8	Rechts	Skala	Eingabe
10	tli8	Numerisch	4	2	handelt, ohne ...	{1,00, stim...	-77,00	8	Rechts	Skala	Eingabe
11	tli9	Numerisch	4	2	ermutigt ihre Mi...	{1,00, stim...	-77,00	8	Rechts	Skala	Eingabe
12	tli10	Numerisch	4	2	fÄhrt eher dur...	{1,00, stim...	-77,00	8	Rechts	Skala	Eingabe
13	tli11	Numerisch	4	2	bringt die Grup...	{1,00, stim...	-77,00	8	Rechts	Skala	Eingabe
14	tli12	Numerisch	4	2	hat ein klares V...	{1,00, stim...	-77,00	8	Rechts	Skala	Eingabe
15	tli13	Numerisch	4	2	zeigt Respekt f...	{1,00, stim...	-77,00	8	Rechts	Skala	Eingabe
16	tli14	Numerisch	4	2	hat mich dazu ...	{1,00, stim...	-77,00	8	Rechts	Skala	Eingabe

2.1.1 Variablenname und -label

Unter *Name* kann jede Variable benannt werden. Es bietet sich an, diese Zeichenkombinationen nach einem einheitlichen Prinzip zu gestalten. Meist wird ein Kurzname wie bspw. „tli1“ für das 1. Item des TLI-Fragebogens vergeben.

Unter *Variablenlabel* wird häufig ein ausführlicher Name oder der konkrete Itemtext eingegeben. Hier kann eine bis zu 256 Zeichen lange Beschreibung der Variablen vorgenommen werden. Da die Labels aber auch in den Ergebnisausgaben erscheinen, wo sie durchaus nützlich sind, sollten sie knapp gehalten werden. Dadurch kann ein unnötiges Aufblähen der Ausgaben vermieden werden.

2.1.2 Typ

Hier kann aus verschiedenen Optionen der *Typ* der Variable ausgewählt werden. Da SPSS mit „Zahlen“ rechnet, ist hier, soweit möglich „numerisch“ als Typ auszuwählen (s. Wertelabels). Für freie Texteingänge bietet sich der Typ „String“ an.

2.1.3 Spaltenformat und Dezimalstellen

Das *Spaltenformat* gibt an, wie viele Zeichen die Werte dieser Variablen maximal umfassen.

Das Variablenmerkmal *Dezimalstellen* definiert lediglich die Zahl der in der Datenansicht und Ausgabe berichteten Dezimalstellen. Gespeichert werden so viele Stellen wie es das Spaltenformat zulässt.

2.1.4 Wertelabels

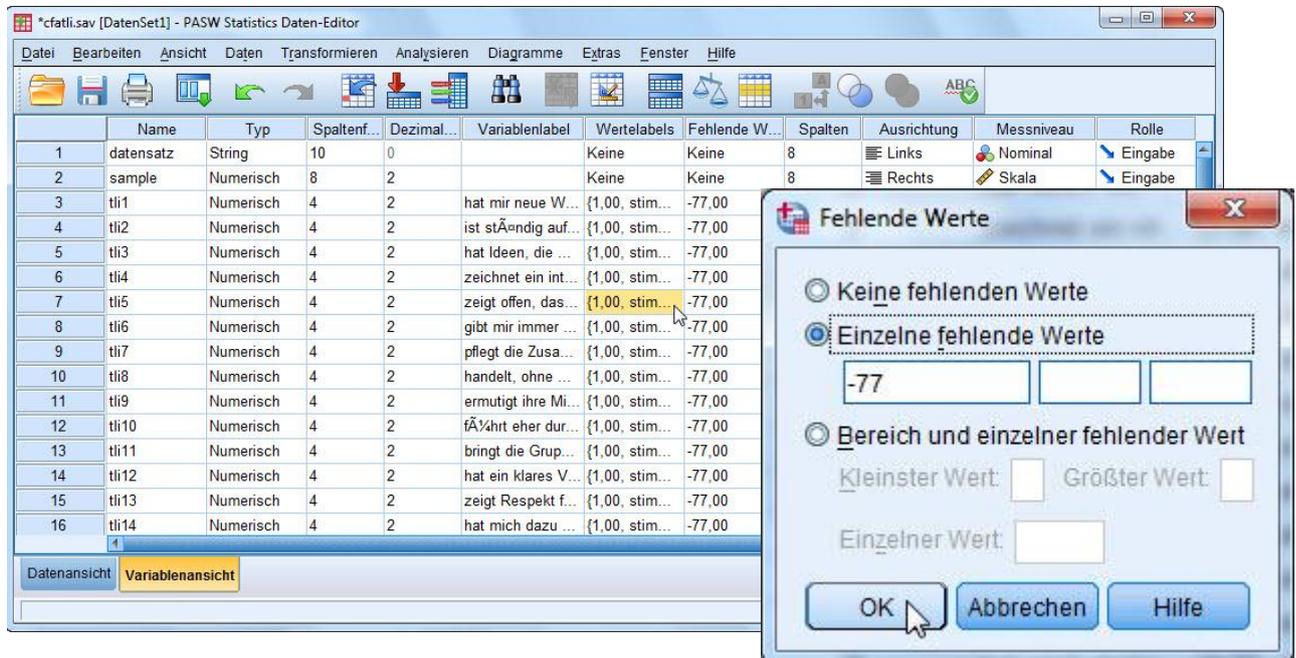
Mit dieser Funktion können einer Variablen numerische Kodierungen zugewiesen werden. Das heißt es wird ersichtlich, welche inhaltliche Bedeutung ein bestimmter Zahlenwert hat. Bei Antwortskalen reicht es häufig, die höchste und die niedrigste Stufe anhand von Wertelabels zu kodieren, um die inhaltliche Richtung der Variable festzuhalten.

Zum Definieren oder Ändern der Wertelabels wird auf das kleine Kästchen ([...]) geklickt. Anschließend wird im Dialogfenster jeweils der numerische Wert, unter „Beschriftung“ die inhaltliche Bedeutung eingetragen und auf „Hinzufügen“ geklickt.

The screenshot shows the SPSS Statistics interface with the 'Wertelabels' dialog box open. The background window displays a list of variables with columns for Name, Typ, Spaltenf..., Dezimal..., Variablenlabel, Wertelabels, Fehlende W..., Spalten, Ausrichtung, Messniveau, and Rolle. The 'Wertelabels' dialog box has a 'Wert' field containing '5' and a 'Beschriftung' field. Below the 'Beschriftung' field, there is a list of existing labels: '1,00 = "stimme gar nicht zu"', '2,00 = "stimme eher nicht zu"', '3,00 = "stimme teilweise zu"', and '4,00 = "stimme eher zu"'. The dialog box includes buttons for 'Hinzufügen', 'Ändern', 'Entfernen', 'Rechtschreibung...', 'OK', 'Abbrechen', and 'Hilfe'.

2.1.5 Fehlende Werte

Hier können fehlende Werte definiert werden. Wurden z. B. einzelne Items nicht beantwortet, können diese als -77 oder -99 kodiert werden. Diese Zahlen werden gewählt, da sie üblicherweise nicht vorkommen oder durch Tippfehler entstehen.



Im Dialogfenster können solche einzelnen fehlenden Werte eingetragen werden. SPSS schließt nun bei sämtlichen Analysen die fehlenden Werte aus. Wären diese nicht als fehlende Werte definiert, würden sie als Wert -77 in die Analyse eingehen.

2.1.6 Messniveau

Abhängig von der Art der Variable wird ein bestimmtes Messniveau angegeben. Es ist wichtig, das Messniveau zu spezifizieren, da die Wahl der statistischen Analyse vom Messniveau bestimmt wird. In SPSS stehen „Skala“ (meint metrisches bzw. Intervallskalenniveau), „Ordinal“ und „Nominal“ als Skalierung zur Auswahl.

2.1.6.1 Exkurs Messniveau

Das Messniveau oder die Skalendignität ist in der Statistik und Empirie eine wichtige Eigenschaft von Merkmalen bzw. von Variablen. Je nach der Art eines Merkmals bzw. je nachdem, welche Vorschriften bei seiner Messung eingehalten werden können, lassen sich verschiedene Stufen der Skalierbarkeit unterscheiden:

Nominalskala: Niedrigstes Skalenniveau. Für verschiedene Objekte oder Erscheinungen wird mithilfe eines Vergleichs lediglich eine Entscheidung über Gleichheit oder Ungleichheit der Merkmalsausprägung getroffen. Es handelt sich also nur um qualitative Merkmale (z. B. Blutgruppen oder Geschlecht). Es gilt die Gleichheitsrelation, also kann entschieden werden, ob zwei Ausprägungen gleich oder ungleich sind. Die Werte können aber nicht der Größe nach sortiert

werden, im Sinne von "ist größer als" oder "besser als". Ein Beispiel ist die Rückennummer beim Fußball. Spieler können unterschieden werden, die Nummer 10 ist aber nicht schlechter als die Nummer 8.

Ordinalskala: Für ein ordinal skalierbares Merkmal bestehen Rangordnungen der Art "größer", "kleiner", "mehr", "weniger", "stärker", "schwächer" zwischen je zwei unterschiedlichen Merkmalswerten. Über die Abstände zwischen diesen benachbarten Urteilklassen ist jedoch nichts ausgesagt. Meist handelt es sich um qualitative Merkmale, wie z. B. der in der Frage gesuchte "höchste erreichbare Bildungsabschluss". Ein weiteres Beispiel ist die Platzierung bei einem Sportwettkampf: Platz 1 ist besser als Platz 2, es gibt aber keine Auskunft darüber, ob der Unterschied zwischen Platz 1 und 2 gleich groß ist wie der zwischen Platz 3 und Platz 4.

Intervallskala: Die Reihenfolge der Merkmalswerte ist festgelegt, und die Größe des Abstandes zwischen zwei Werten lässt sich sachlich begründen. Die Ausprägungen dieses Skalenniveaus lassen sich quantitativ mittels Zahlen darstellen. Es ist für die metrische Skala charakteristisch, dass die Abstände zwischen den Zahlen auch in der Realität gleich groß sind. Aus diesem Grund ist es auf diesem Skalenniveau sinnvoll, bspw. einen Mittelwert zu berechnen. Ein Beispiel für metrische Daten ist die Körpergröße in Metern oder das Gewicht in Kilogramm. Bei Fragebogenitems, die anhand einer mehrstufigen (mindestens vierstufigen) Skala beantwortet werden, wird ebenfalls Intervallskalenniveau angenommen.

2.2 Datenansicht

Die Datenansicht eignet sich zur Eingabe der Werte. Eine Zeile stellt dabei einen Fall (z. B. eine Person) und seine zugehörige Werte dar. Im Screenshot findet sich in Zeile 316 beispielsweise eine Person aus dem Datensatz „consulting“ und dem Sample 1. Das erste TLI-Item (tli1) wurde mit einer 1 (stimme gar nicht zu) beantwortet, das zweite TLI-Item (tli2) wurde mit einer 4 (stimme eher zu) beantwortet.

cfatli.sav [DatenSet1] - IBM SPSS Statistics Daten-Editor

316 : datensatz consulting Sichtbar: 46 von 46 Variablen

	datensatz	sample	tli1	tli2	tli3	tli4	tli5	tli6	tli7
310	consulting	1,00	3,00	4,00	4,00	4,00	5,00	3,00	3,00
311	consulting	1,00	3,00	3,00	2,00	4,00	4,00	2,00	2,00
312	consulting	1,00	3,00	2,00	1,00	1,00	2,00	3,00	3,00
313	consulting	1,00	3,00	2,00	2,00	2,00	3,00	3,00	4,00
314	consulting	1,00	3,00	3,00	3,00	2,00	3,00	3,00	3,00
315	consulting	1,00	3,00	3,00	4,00	5,00	5,00	5,00	4,00
316	consulting	1,00	3,00	4,00	3,00	2,00	2,00	4,00	3,00
317	consulting	1,00	2,00	2,00	3,00	2,00	4,00	3,00	3,00
318	consulting	1,00	3,00	4,00	3,00	2,00	3,00	2,00	2,00
319	consulting	1,00	3,00	3,00	4,00	4,00	4,00	4,00	4,00
320	consulting	1,00	3,00	4,00	4,00	4,00	4,00	3,00	4,00
321	consulting	1,00	4,00	4,00	4,00	3,00	5,00	5,00	3,00
322	consulting	1,00	1,00	1,00	2,00	1,00	1,00	2,00	2,00
323	consulting	1,00	3,00	2,00	2,00	1,00	4,00	3,00	3,00
324	consulting	1,00	4,00	3,00	2,00	3,00	5,00	4,00	4,00

Datenansicht Variablenansicht

IBM SPSS Statistics Prozessor ist bereit

Die Fälle lassen sich entsprechend ihrer Werte in einer Variablen ordnen, indem mit der rechten Maustaste auf den Variablenamen geklickt und „Aufsteigend sortieren“ ausgewählt wird.

cfatli.sav [DatenSet1] - IBM SPSS Statistics Daten-Editor

1 : tli2 1,00 Sichtbar: 46 von 46 Variablen

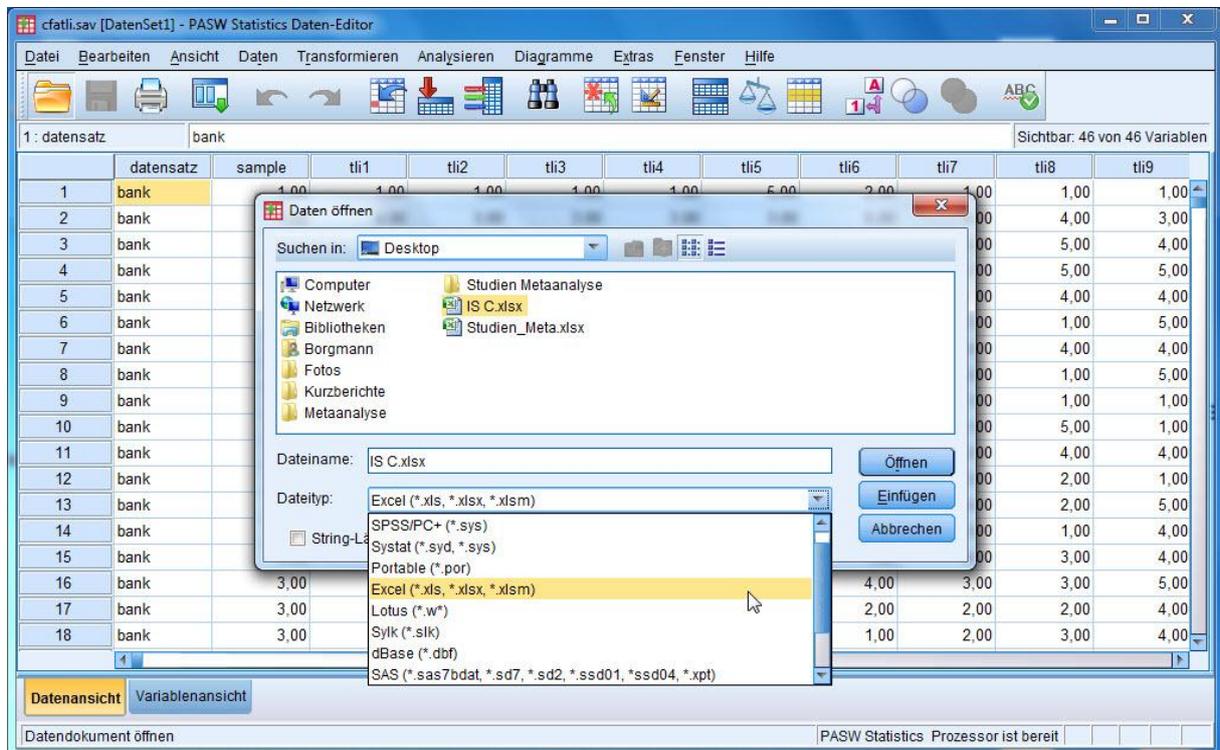
	datensatz	sample	tli1	tli2	tli3	tli4	tli5	tli6	tli7
1	bank	1,00	1,00	1			5,00	2,00	1,00
2	bank	1,00	4,00	3			3,00	5,00	5,00
3	bank	1,00	4,00	3			5,00	5,00	5,00
4	bank	1,00	4,00	4			4,00	4,00	5,00
5	bank	1,00	5,00	4			4,00	5,00	4,00
6	bank	1,00	5,00	4			5,00	2,00	3,00
7	bank	2,00	4,00	4			5,00	4,00	4,00
8	bank	2,00	5,00	5			5,00	5,00	5,00
9	bank	2,00	1,00	1			1,00	1,00	1,00
10	bank	2,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
11	bank	2,00	3,00	2,00	3,00	2,00	4,00	1,00	2,00
12	bank	2,00	2,00	1,00	1,00	1,00	5,00	1,00	3,00
13	bank	2,00	4,00	4,00	4,00	4,00	4,00	4,00	4,00
14	bank	2,00	3,00	3,00	2,00	3,00	5,00	2,00	3,00
15	bank	3,00	3,00	3,00	3,00	3,00	4,00	1,00	4,00

Datenansicht Variablenansicht

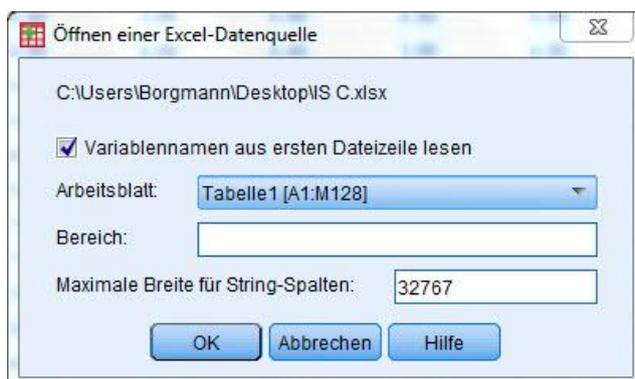
IBM SPSS Statistics Prozessor ist bereit

2.3 Datenimport aus Excel

SPSS ist in der Lage, auch fremde Datenformate zu lesen. Um z. B. Excel-Dateien zu öffnen, kann unter *Datei/Öffnen* als Dateityp Excel ausgewählt werden.



Im anschließenden Dialogfenster kann, falls in der Excel-Datei die Variablenamen in der ersten Zeile stehen, das Häkchen bei "Variablenamen aus der ersten Dateizeile lesen" gesetzt bleiben. Sind in der Excel-Datei keine Variablenamen definiert, wird das Häkchen weg geklickt und die Variablenamen müssen in der Variablenansicht manuell eingetragen werden.



3 DATENSICHTUNG

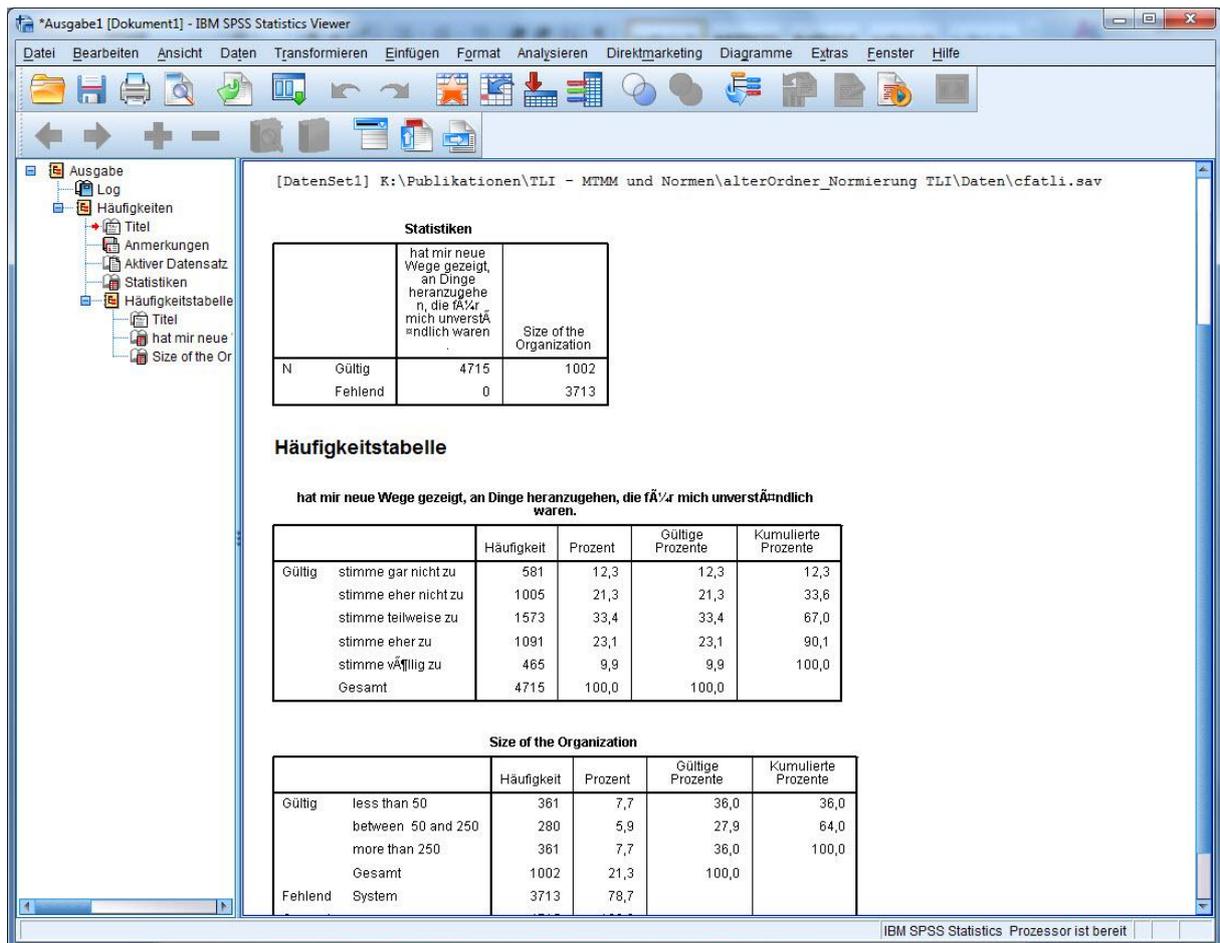
3.1 Häufigkeiten

Einer der ersten Schritte beim Verwenden von neuen Datensätzen ist zunächst eine deskriptive Statistik über die Daten anzufertigen, um einen Überblick über die Daten zu erhalten.

Hierfür kann zum einen der Befehl Häufigkeiten genutzt werden, der sich vor allem bei kategorialen Variablen (Nominal-/Ordinalskalenniveai) anbietet, wie Geschlecht, Schul- und Berufsbildung. Die Analyse der Häufigkeiten kann auch genutzt werden, um mögliche Tippfehler zu entdecken, da zu jedem Wert, der vorkommt, die Häufigkeit angezeigt wird.

Hierzu sind unter *Analysieren/Deskriptive Statistiken/Häufigkeiten* die gewünschten Variablen auszuwählen.

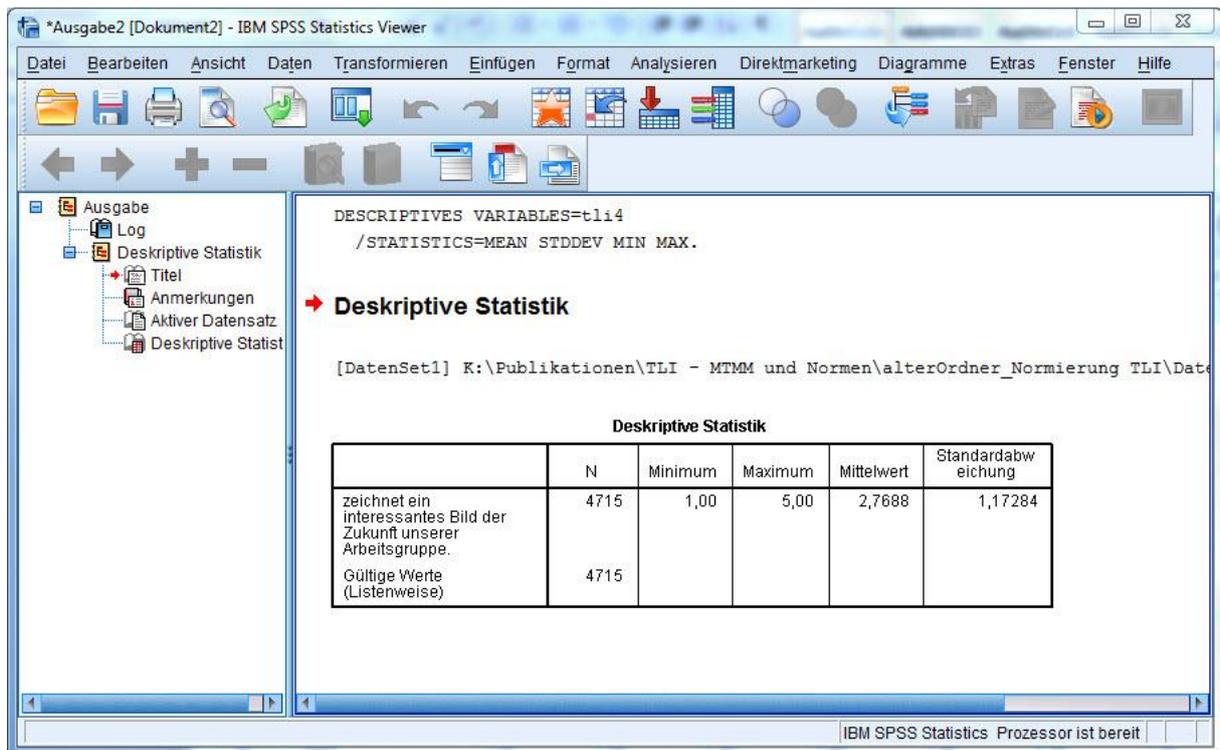




Innerhalb der Ausgabedatei findet sich für jede Variable, die ausgewählt wurde, eine Häufigkeitstabelle. In der oberen Häufigkeitstabelle im Screenshot ist ersichtlich, dass „stimme gar nicht zu“ von 581 Personen angekreuzt wurde bzw. „stimme gar nicht zu“ eine (absolute) Häufigkeit von 581 hat. Bezogen auf die Gesamtstichprobe von $N = 4715$ entspricht dies einer relativen Häufigkeit von 12,3%. Die kumulierten Prozente geben an, bei wie viel Prozent der Fälle der Wert kleiner oder gleich ist als der betrachtete Wert. Im Screenshot haben beispielsweise 67% „stimme teilweise zu“, „stimme eher nicht zu“ oder „stimme gar nicht zu“ angekreuzt.

3.2 Deskriptive Statistik

Der Befehl *Analysieren/Deskriptive Statistiken/Deskriptive Statistiken* dient dazu, statistische Kennwerte (wie Lage- oder Streuungsmaße) für metrische Variablen (wie z. B. Alter) zu bestimmen.



In der Ausgabe erscheint eine Tabelle mit den deskriptiven Statistiken zu den gewählten Variablen. Die Gesamtstichprobe umfasst $N = 4715$ Fälle. Das Minimum, also der kleinste angekreuzte Wert, liegt bei 1. Das Maximum, also der größte angekreuzte Wert, liegt bei 5. Der Mittelwert der Stichprobe ist $M = 2.77$ (es werden in wissenschaftlichen Arbeiten üblicherweise zwei Nachkommastellen angegeben). Die Standardabweichung liegt bei $SD = 1.17$. Mittelwert und Standardabweichung werden im Folgenden erläutert.

3.2.1 Exkurs Maße der zentralen Tendenz (Lagemaße)

Mittelwert: Der allgemein bekannte Durchschnitt ist in der Statistik das arithmetische Mittel. Er errechnet sich als Summe der Werte geteilt durch ihre Anzahl:

$$m_x = \frac{\sum_{i=1}^n x_i}{n}$$

Voraussetzung für die sinnvolle Interpretation des Mittelwertes ist das Intervallskalenniveau der Werte.

Median: Der Median ist der Wert, der in der Mitte liegt. Wenn etwa die durchschnittliche Körpergröße von fünf Jungen aus der achten Klasse gesucht wird, zeigt das folgende Beispiel die Berechnung des Median. Die Größe der Jungen in Zentimeter: 156, 146, 136, 167 und 177. Werden die Zahlen sortiert und genau die gewählt, die in der Mitte liegt, ergibt das: 156. Diese Zahl hat links genauso viele Nachbarn wie rechts. Im Unterschied zum Mittelwert kann der Median bereits ab Ordinalskalenniveau bestimmt und interpretiert werden. Zudem ist er robust gegen Ausreißer.

Angenommen statt des größten Mitschülers stellt sich der größte Mensch der Welt zu den Achtklässlern, so dass sich als Größen ergäbe: 157, 146, 136, 167 und 257 Zentimeter (statt 177). Der Median bliebe von dem Ausreißer unberührt, er wäre immer noch 157. Das arithmetische Mittel läge hier hingegen bei 172.6, obwohl nur zwei Personen über 1.70 Meter groß sind.

3.2.2 Exkurs Streuungsmaße

Standardabweichung: Um die Mittelwerte einer Verteilung herum streuen die einzelnen Werte. Das bekannteste Maß, um das Ausmaß dieser Streuung zu beschreiben, ist die Standardabweichung. Sie berechnet sich als Wurzel aus der Varianz bzw. als

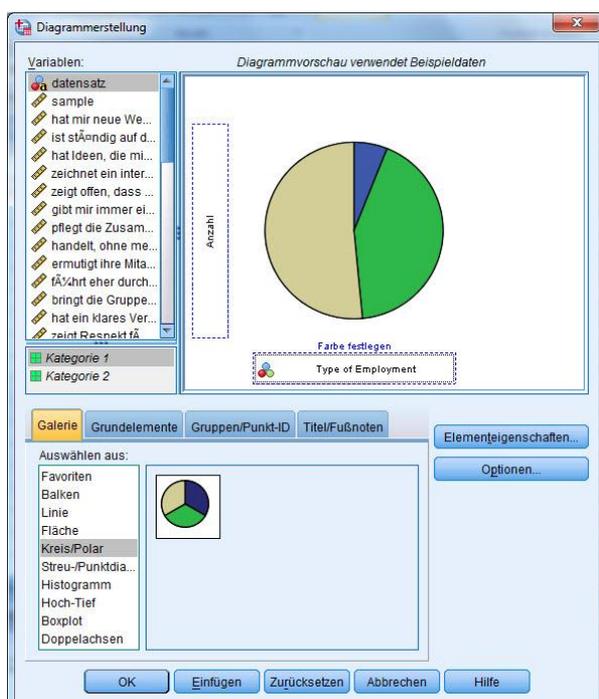
$$s_x = \sqrt{\frac{\sum_{i=1}^n (x_i - m_x)^2}{n}}$$

3.3 Diagramme

Häufig bietet es sich an, Werte und ihre Verteilung in Diagrammen darzustellen. Einzelne Diagrammtypen stehen innerhalb bestimmter Analysieren-Befehle zur Verfügung. Sämtliche Diagramme können zudem über das Menü *Diagramme/Diagrammerstellung* erzeugt werden.

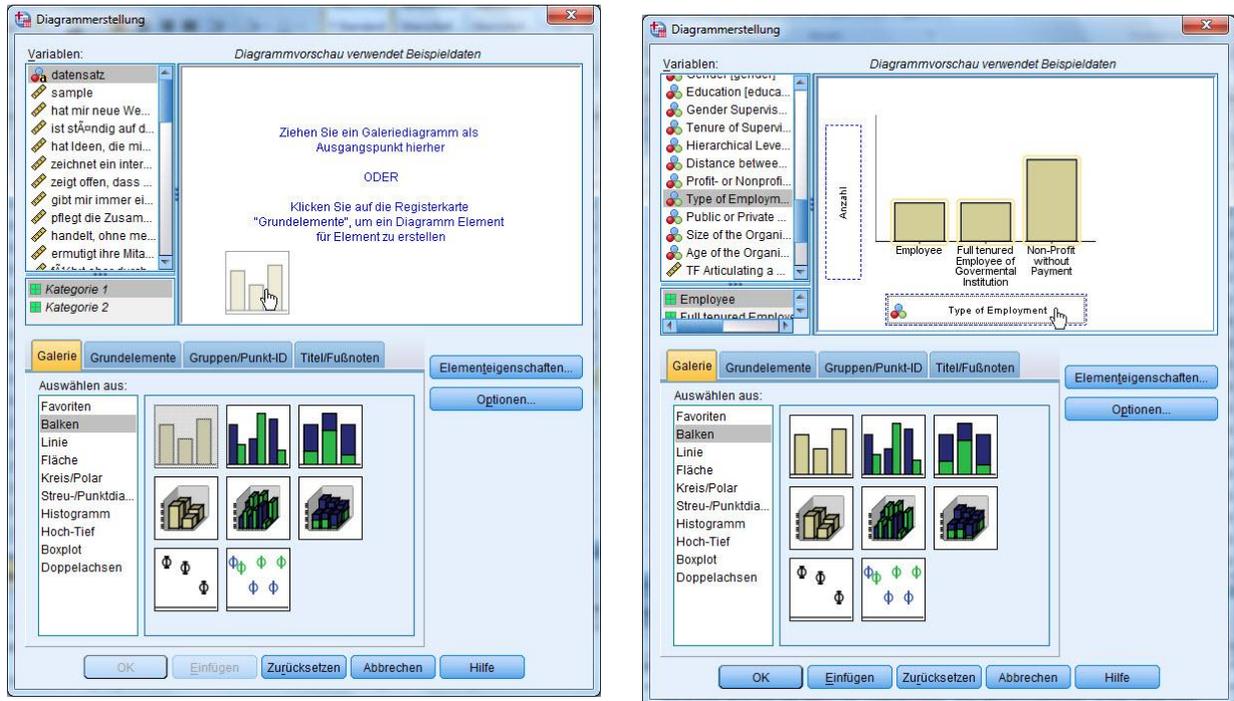
3.3.1 Kreisdiagramm

Ein Kreisdiagramm bietet sich zur Darstellung von Häufigkeiten an, wenn die Häufigkeiten aller dargestellten Werte zusammen, die gesamte Stichprobe darstellen, so wie die im Kreisdiagramm dargestellten Kreissektoren den Kreis bilden. Das Anfordern eines Kreisdiagramms erfolgt analog zum Säulendiagramm, statt „Balken“ wird lediglich „Kreis/Polar“ gewählt.

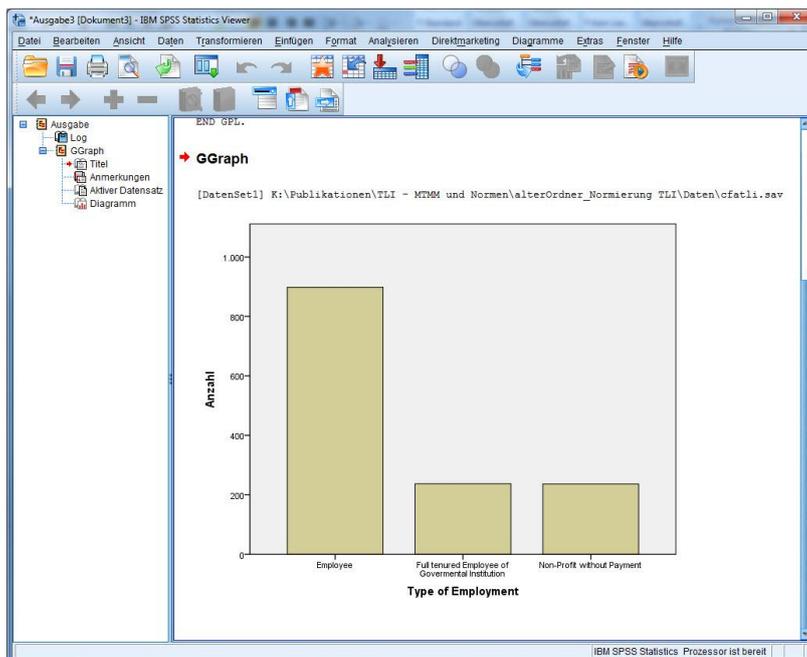


3.3.2 Säulendiagramm

Das Säulendiagramm ist ein Diagramm, das durch auf der x-Achse senkrecht stehende Säulen die Häufigkeitsverteilung einer diskreten Variablen veranschaulicht. Das Säulendiagramm eignet sich besonders, um die Häufigkeit weniger verschiedener Werte (bis ca. 15) zu veranschaulichen. Im Dialogfenster zur Diagrammerstellung wird mit der Maus der Diagrammtyp, hier Säulendiagramm, in das obige Feld gezogen.



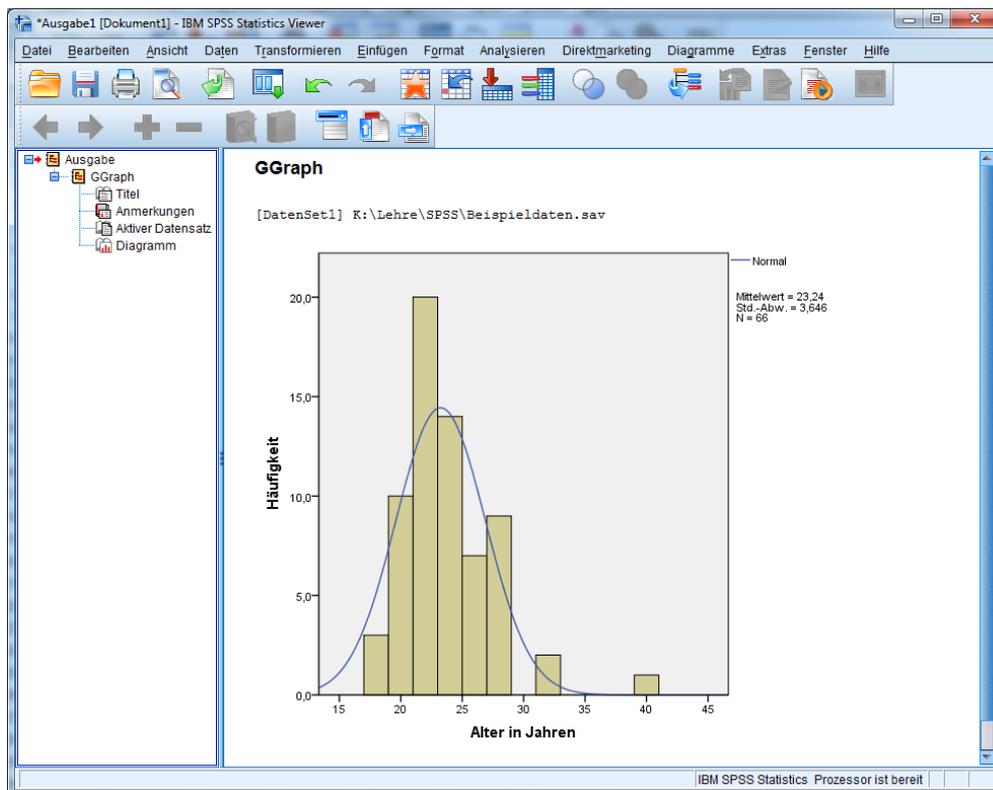
Anschließend wird die Variable, für die ein Diagramm angefordert werden soll, in das Feld „X-Achse?“ gezogen.



3.3.3 Histogramm

Das Histogramm zeigt im Wesentlichen ebenfalls Häufigkeiten in Form von Säulen an, nutzt hierzu im Unterschied zum Säulendiagramm auf der X-Achse jedoch nicht die beobachteten Werte der Variable, sondern die metrisch Skala der Variable. Nur das Histogramm ist daher geeignet die empirische Verteilung mit einer [Normalverteilung](#) graphisch zu vergleichen.

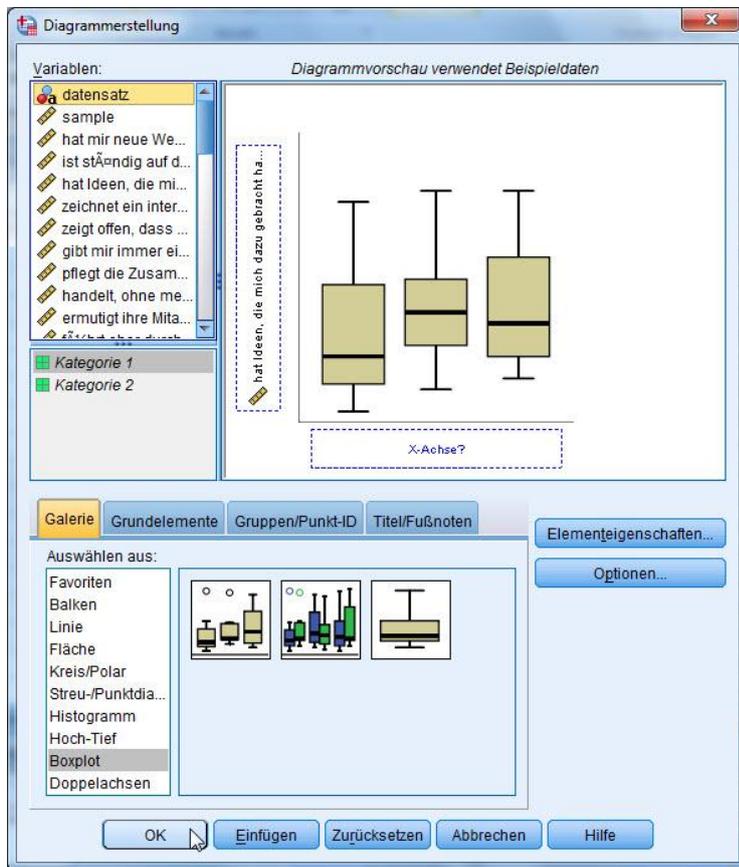
Es kann über das Dialogfenster zur [Diagrammerstellung](#) oder über den Befehl [Häufigkeiten](#) erzeugt werden. In beiden Fällen kann eine Normalverteilungskurve zum Vergleich angefordert werden.



3.3.4 Boxplot

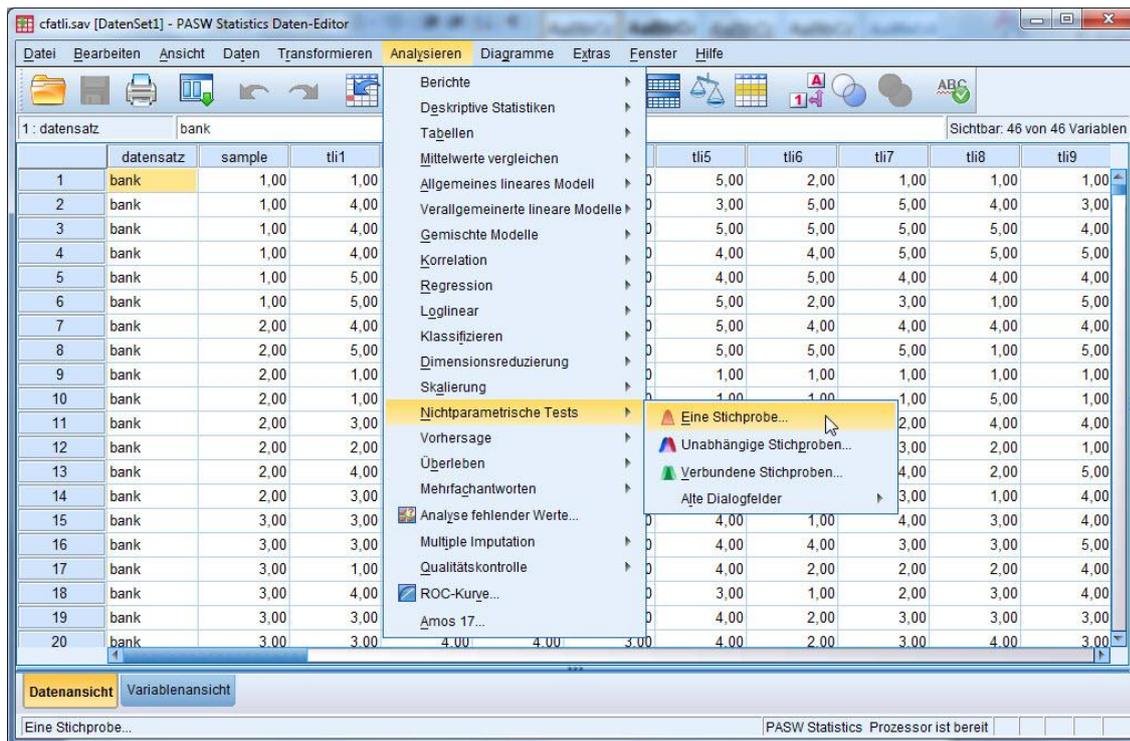
Der Boxplot ist ein Diagramm, das zur grafischen Darstellung der Verteilung statistischer Daten verwendet wird. Es fasst dabei verschiedene robuste Streuungs- und Lagemaße in einer Darstellung zusammen. Ein Boxplot soll schnell einen Eindruck darüber vermitteln, in welchem Bereich die Daten liegen und wie sie sich über diesen Bereich verteilen. Hierzu werden alle Werte der sogenannten Fünf-Punkte-Zusammenfassung dargestellt: Minimum, erstes Quartil, Median, drittes Quartil und Maximum. Die Quartile teilen die Fälle in vier Gruppen zu jeweils 25%. Das erste Quartil trennt also die unteren 25% von den oberen 75%. Das 50%-Quartil entspricht dem Median.

Bei der Diagrammerstellung im Dialogfenster wird als erstes mit der Maus der Diagrammtyp, hier Boxplot, in das obige Feld gezogen. Anschließend wird die Variable, für die ein Diagramm angefordert werden soll, in das Feld „Y-Achse?“ gezogen.

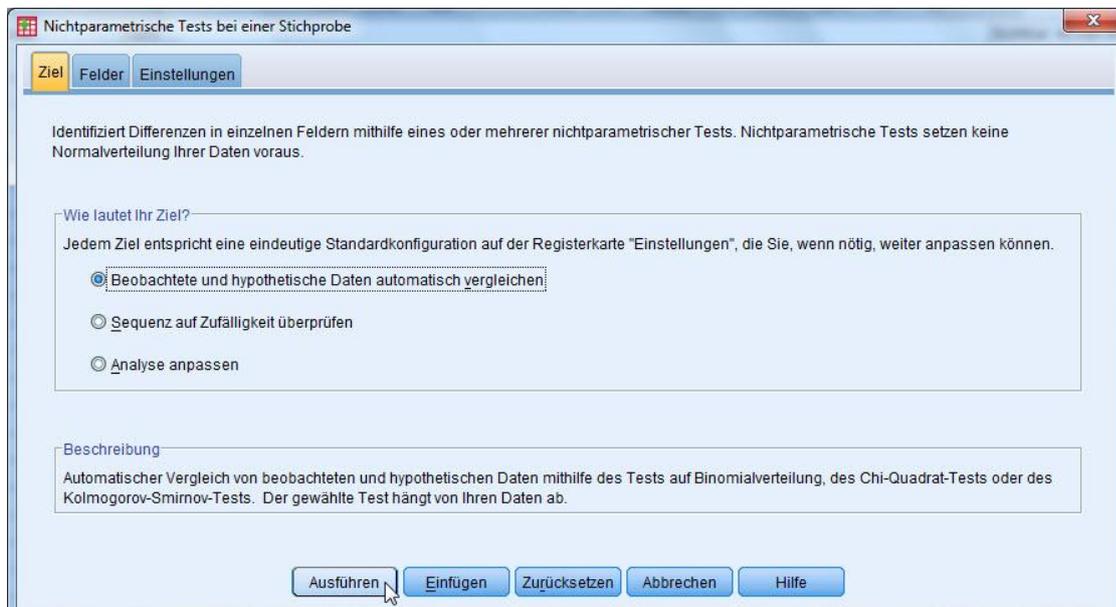


3.5 Überprüfung auf Normalverteilung

Für die Auswahl geeigneter statistischer Verfahren ist es wichtig, die Verteilungsform der Daten in den einzelnen Variablen zu kennen. Für eine Vielzahl parametrischer Verfahren ist die Normalverteilung eine wesentliche Voraussetzung. Bei einigen Maßen, wie zum Beispiel Intelligenz, kann eine Normalverteilung vorausgesetzt werden, auch wenn diese sich nicht in der Stichprobe findet. Mit dem Kolmogorov-Smirnov-Test bei einer Stichprobe (Anpassungstest) wird die beobachtete kumulative Verteilungsfunktion für eine Variable mit einer festgelegten theoretischen Verteilung verglichen (meist mit der Normalverteilung).



Unter *Analysieren/Nichtparametrische Tests/Eine Stichprobe* ist dieser Befehl („Beobachtete und hypothetische Daten automatisch vergleichen“) voreingestellt, so dass direkt „Ausführen“ gewählt werden kann. Falls nicht alle in der Datendatei enthaltenen Variablen geprüft werden sollen, können sie unter „Felder“ von rechts nach links verschoben und somit ausgeschlossen werden.



NPTESTS

```
/ONESAMPLE TEST (datensatz sample tli1 (...) isn cr)
KOLMOGOROV_SMIRNOV(NORMAL=SAMPLE)
/MISSING SCOPE=ANALYSIS USERMISSING=EXCLUDE
/CRITERIA ALPHA=0.05 CILEVEL=95.
EXECUTE.
```

SPSS vergleicht nun anhand des Kolmogorov-Smirnov-Tests bei den gewählten Variablen die vorliegende Verteilung der Werte mit einer Normalverteilung. Der Signifikanzwert gibt an, ob sich die Verteilung der Daten signifikant von einer Normalverteilung unterscheidet. Ist dieser Wert größer als .05 unterscheidet sich die Verteilung der Daten nicht von der angenommenen Normalverteilung. Ist eine Normalverteilung gegeben, können parametrische Verfahren verwendet werden. Ist der Wert kleiner als .05, dann sollten non-parametrische Verfahren verwendet werden, da sich die Verteilung signifikant von einer Normalverteilung unterscheidet.

The screenshot shows the IBM SPSS Statistics Viewer interface. The main window displays the command window with the following text:

```
GET
FILE='K:\Exchange\Für Tobias\Projekte\SPSS\Beispieldaten_TLI_bearbeitet.sav'.
DATASET NAME DataSet1 WINDOW=FRONT.
SET TVars=Names TNumbers=Labels ONumbers=L
*Nonparametric Tests: One Sample.
NPTESTS
/ONESAMPLE TEST (cr tf)
/MISSING SCOPE=ANALYSIS USERMISSING=EXCL
/CRITERIA ALPHA=0.05 CILEVEL=95.
```

Below the command window, the output for 'Nichtparametrische Tests' is shown. It includes a table titled 'Übersicht über Hypothesentest' which is also shown in a callout box:

	Nullhypothese	Test	Sig.	Entscheidung
1	Die Verteilung von TA Contingend Reward ist normal mit Mittelwert 3.33 und Standardabweichung 1.10.	Kolmogorov-Smirnov-Test einer Stichprobe	,106	Nullhypothese behalten.
2	Die Verteilung von Transformational Leadership ist normal mit Mittelwert 3.35 und Standardabweichung 0.75.	Kolmogorov-Smirnov-Test einer Stichprobe	,402	Nullhypothese behalten.

Asymptotische Signifikanzn werden angezeigt. Das Signifikanzniveau ist .05.

Bei sehr großen Stichproben führt der K-S-Test auch dann zu signifikanten Ergebnissen, wenn die Werte de facto normal verteilt sind. In diesen Fällen sind ggf. deskriptive Verteilungsmaße (Schiefe, Exzess) und ein Histogramm zu nutzen, um die Normalverteilung zu zeigen.

Die Prüfung der Normalverteilung ist vorgeschrieben, wenn parametrische Berechnungen durchgeführt werden sollen. Hierzu zählen u. a. die Verfahren des Allgemeinen Linearen Modells (ALM) wie Varianz-, Kovarianz- und Regressionsanalysen. Bei Berechnungen mit parametrischen

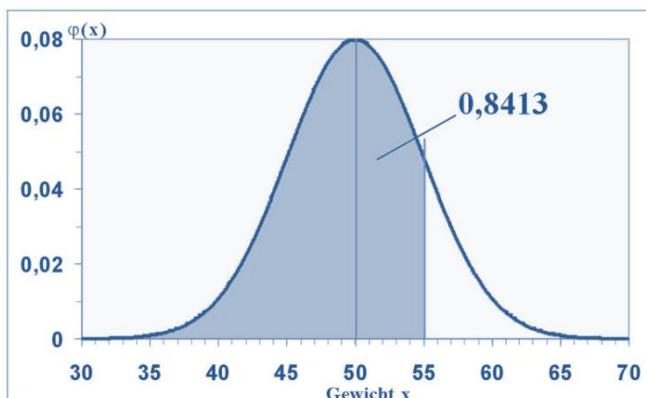
Verfahren, müssen **alle** Variablen normalverteilt sein. Sobald eine Variable nicht normal verteilt ist, müssen non-parametrische Verfahren genutzt werden, auch wenn alle anderen Variablen normalverteilt sind.

Bei diesem Beispiel sind beide Werte nicht signifikant ($.106 > .050$ und $.402 > .050$), folglich wird die Nullhypothese nicht abgelehnt. Die Verteilung der Daten unterscheidet sich also nicht von der angenommenen Normalverteilung. Da hier beide Variablen normalverteilt sind, können anschließend parametrische Analyseverfahren angewendet werden.

3.5.1 Exkurs Normalverteilung

Die Normal- oder Gauß-Verteilung (nach Carl Friedrich Gauß) ist ein wichtiger Typ stetiger Wahrscheinlichkeitsverteilungen. Sie ist definiert für Werte von minus Unendlich bis plus Unendlich. Die besondere Bedeutung der Normalverteilung beruht unter anderem auf dem zentralen Grenzwertsatz, der besagt, dass eine Summe von n unabhängigen, identisch verteilten Zufallsvariablen im Grenzwert $n \rightarrow \infty$ normalverteilt ist.

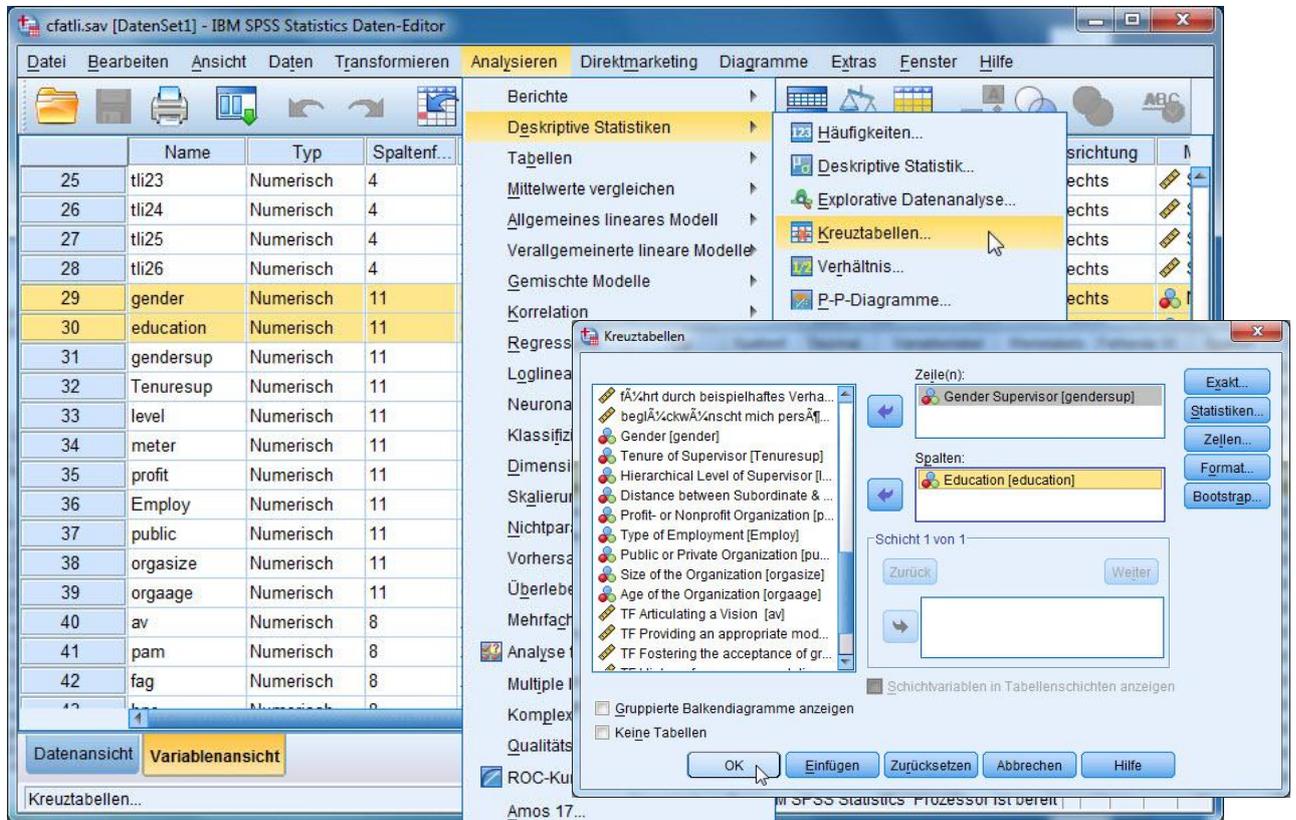
Das heißt zum Beispiel: Auf einer Hühnerfarm mit sehr vielen Hühnern werden eine Woche lang die einzelnen Eier gewogen. Es stellt sich heraus, dass ein Ei im Durchschnitt 50g wiegt. Der Mittelwert ist daher 50. Die Verteilung des Gewichts kann annähernd wie in der Grafik dargestellt werden.



Die X-Achse beschreibt das Gewicht der Eier, die Y-Achse beschreibt die relative Häufigkeit. Es zeigt sich, dass sich die meisten Eier in der Nähe des Mittelwerts 50 befinden und dass die Wahrscheinlichkeit, sehr kleine oder sehr große Eier zu erhalten, sehr klein wird. Dies sind die zentralen Merkmale einer Normalverteilung. Sie ist typisch für Zufallsvariablen, die sich aus sehr vielen verschiedenen Einflüssen zusammensetzen, wie im Beispiel u. a. Gewicht des Huhns, Alter, Gesundheit, Standort, Vererbung.

3.6 Kreuztabelle

Kreuztabellen sind Tabellen, die die absoluten oder relativen Häufigkeiten von Kombinationen bestimmter Merkmalsausprägungen enthalten. In Kreuztabellen wird das gemeinsame Auftreten von zwei Merkmalen dargestellt, bspw. von Geschlecht und Bildungsabschluss.



```
CROSSTABS
/TABLES=gendersup BY education
/FORMAT=AVALUE TABLES
/CELLS=COUNT
/COUNT ROUND CELL.
EXECUTE.
```

*Ausgabe2 [Dokument2] - IBM SPSS Statistics Viewer

Datei Bearbeiten Ansicht Daten Transformieren Einfügen Format Analysieren Direktmarketing Diagramme Extras Fenster Hilfe

Ausgabe
 Log
 Kreuztabellen
 Titel
 Anmerkungen
 Aktiver Datensatz
 Verarbeitete Fälle
 Gender Superviso

```

CROSSTABS
  /TABLES=gendersup BY education
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ROUND CELL.
  
```

→ Kreuztabellen

[DatenSet1] K:\Publikationen\TLI - MTMM und Normen\alterOrdner_Normierung TLI\Daten\cfatli.sav

Verarbeitete Fälle

	Fälle					
	Gültig		Fehlend		Gesamt	
	N	Prozent	N	Prozent	N	Prozent
Gender Supervisor * Education	2081	44,1%	2634	55,9%	4715	100,0%

Gender Supervisor * Education Kreuztabelle

Anzahl

		Education				Gesamt
		high school	Advanced high school	Higher education entrance qualification	University degree	
Gender Supervisor	female	144	281	115	184	724
	male	155	478	272	452	1357
Gesamt		299	759	387	636	2081

IBM SPSS Statistics Prozessor ist bereit

In der Ausgabedatei ergibt sich eine Kreuztabelle für die Variablen Education und Gender Supervisor. Die nominalskalierte Variable "Education" hat die vier Ausprägungsstufen „high school“, „Advanced Highschool“, „Higher education entrance qualification“ und „University degree“. Die Variable "Gender Supervisor" weist die Ausprägungen „female“ und „male“ auf. In der Kreuztabelle ist ersichtlich, dass es insgesamt 636 Mitarbeiter mit einem "University degree" gibt. 184 davon haben einen weiblichen Vorgesetzten und 452 einen männlichen.

4 DATENSTRUKTURIERUNG

Wenn Berechnungen nur für einen Teil der vorliegenden Daten ausgeführt oder Statistiken unterschiedlicher Teilgruppen verglichen werden sollen, können in SPSS [Fälle ausgewählt](#) oder die [Datei aufgeteilt](#) werden.

4.1 Fälle auswählen

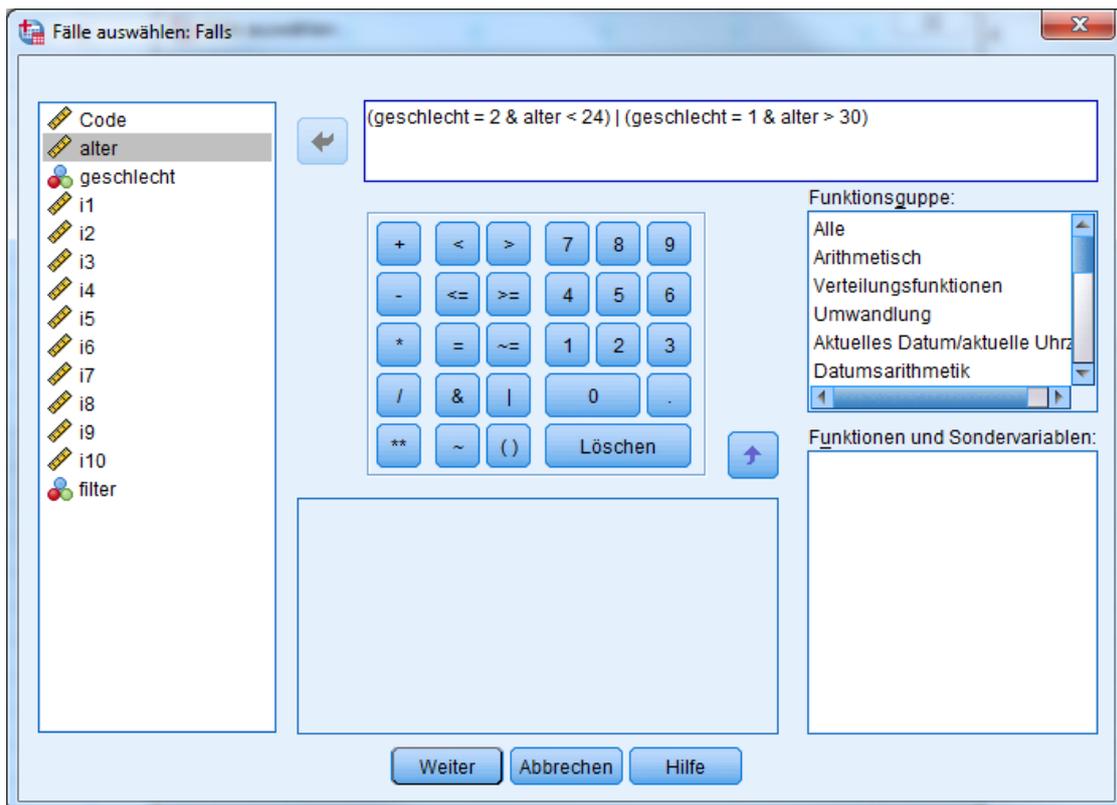
Wichtig: Ausgewählt werden stets jene Fälle, für die anschließend Berechnungen durchgeführt werden sollen. Nicht ausgewählte Fälle werden nicht berücksichtigt.

Unter *Daten/Fälle auswählen* öffnet sich ein Dialogfenster, das verschiedene Möglichkeiten bereithält, um Fälle auszuwählen. Diese Möglichkeiten sind untereinander exklusiv, d.h. es kann immer nur eine aktiv sein. Um in der Syntax Überlagerungen verschiedener Filter-Befehle zu vermeiden, beginnt jeder Filter-Befehl dort mit der Aufhebung aller evtl. zuvor bestehenden Filter-Befehle („FILTER OFF.“ oder „USE ALL.“).

Folgende Optionen stehen zur Verfügung:

- **Falls Bedingung zutrifft**

Diese Option wird am häufigsten gewählt. Im anschließenden Dialogfenster kann mit Hilfe der bestehenden Variablen und der verfügbaren Funktionen (vgl. [Variable berechnen](#)) eine Bedingungsgleichung aufgestellt werden. Alle Fälle, für die diese Gleichung zutrifft, werden ausgewählt. Sollen mehrere Bedingungen miteinander verknüpft werden, so sind sie mittels „und“ (&) oder „oder“ (|) zu verbinden. Im Beispiel (s. Screenshot) werden alle Fälle ausgewählt, die entweder zugleich weiblich *und* jünger als 24 Jahre alt sind *oder* die zugleich männlich *und* älter als 30 Jahre sind. Es werden also junge Frauen *und* [sic!] alte Männer ausgewählt. Alte Frauen und junge Männer werden *nicht* ausgewählt.

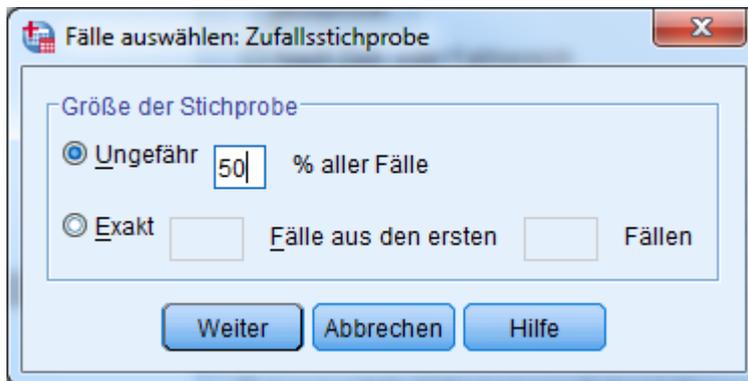


SPSS erstellt hierzu eine neue Variable (mit dem Namen „filter_\$“), die für die ausgewählten Fälle eine 1 und für die nicht ausgewählten eine 0 enthält; die Bedingung wird zudem als Variablenlabel hinterlegt. In der Syntax kann die Variable per Hand umbenannt werden, damit z. B. der Name aussagekräftiger wird („Filter_männlich“) oder damit die Filtervariable nicht bei der nächsten Auswahl überschrieben wird.

```
USE ALL.
COMPUTE filter_$=((geschlecht = 2 & alter < 24) | (geschlecht = 1 & alter > 30)).
VARIABLE LABELS filter_$ '(geschlecht = 2 & alter < 24) | (geschlecht = 1 & alter > 30)
(FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.
EXECUTE.
```

- **Zufallsstichprobe**

Zur Auswahl einer Zufallsstichprobe ist lediglich die gewünschte Größe zu spezifizieren, und zwar entweder als Prozentsatz aller Fälle (s. Screenshot) oder als Absolutzahl.

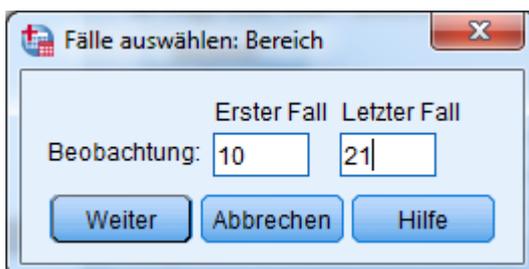


Diese Funktion wird genutzt, um zum Beispiel die Stabilität von Ergebnissen zu prüfen. Sofern die Gesamtstichprobe ausreichend groß ist, können die Berechnungen an zufällig ausgewählten Teilstichproben unabhängig voneinander durchgeführt und verglichen werden (sog. Kreuzvalidierung).

```
USE ALL.
COMPUTE filter_$=(uniform(1)<=.50).
VARIABLE LABELS filter_$ 'Ungefähr 50 % der Fälle (SAMPLE)'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.
EXECUTE.
```

- **Nach Zeit- oder Fallbereich**

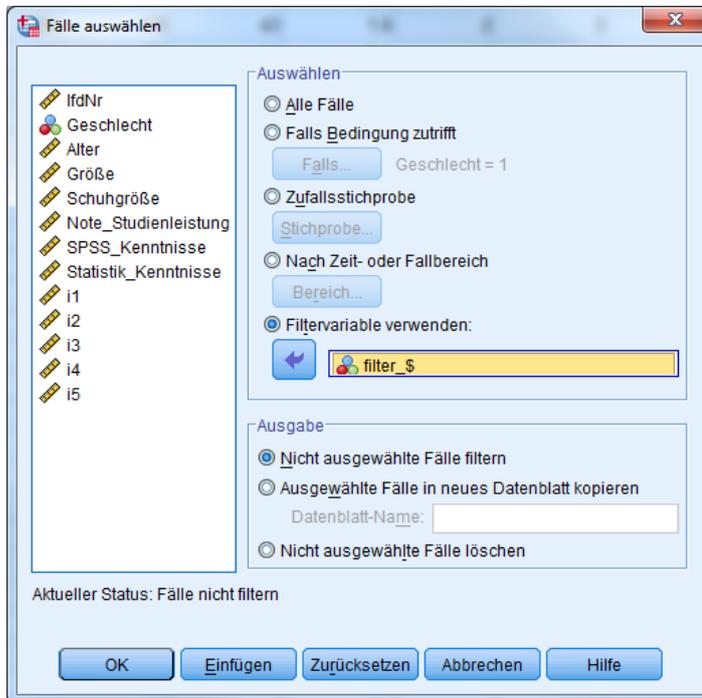
Hierüber können Fälle anhand ihrer Zeilennummer ausgewählt werden. Achtung: Um die Auswahl später replizieren zu können, muss die Reihenfolge der Fälle wieder hergestellt werden.



```
FILTER OFF.
USE 10 thru 21.
EXECUTE.
EXECUTE.
```

- **Filtervariable verwenden**

Diese Option kann gewählt werden, wenn bereits eine Variable im Datenblatt vorliegt, in der alle nicht auszuwählenden Fälle eine 0 oder einen fehlenden Wert aufweisen. Dann kann diese Option aktiviert und die Filtervariable direkt aus der Variablenliste links ausgewählt werden.



FILTER BY filtervariable.
EXECUTE.

Im Anschluss an die Definition der Auswahlregel kann im Haupt-Dialogfenster zum *Fälle auswählen* angegeben werden, wie mit den ausgewählten bzw. nicht ausgewählten Fällen zu verfahren ist:

- **Nicht ausgewählte Fälle filtern**

Bei den nicht ausgewählten Fällen wird in der Datenansicht die Fallnummer zu Beginn der Zeile durchgestrichen. Zudem wird in der Datendatei unten rechts „Filter aktiv“ angezeigt.

Beispieldaten.sav [DatenSet1] - IBM SPSS Statistics Daten-Editor

1: Code 1 Sichtbar: 14 von 14 Variablen

	Code	alter	geschlecht	i1	i2	i3	i4	i5
1	1	18	1	2	3	2	2	
2	2	20	1	1	1	1	3	
3	3	22	2	4	3	3	4	
4	4	19	1	3	2	4	3	
5	5	21	2	2	1	3	3	
6	6	24	2	5	4	4	5	
7	7	17	2	1	2	2	1	
8	8	23	1	1	3	2	3	
9	9	22	1	3	4	3	2	
10	10	21	2	2	3	3	3	
11	11	18	1	4	4	3	4	
12	12	19	2	2	1	1	2	

Datenansicht Variablenansicht

IBM SPSS Statistics Prozessor ist bereit Filter aktiv

Innerhalb der Ausgabedatei wird in den „Anmerkungen“ berichtet, ob und welcher Filter ggf. aktiv war und wie viele Fälle bei den Berechnungen berücksichtigt wurden („Anzahl der Zeilen in der Arbeitsdatei“).

*Ausgabe1 [Dokument1] - IBM SPSS Statistics Viewer

Häufigkeiten

Anmerkungen		
Ausgabe erstellt		26-Apr-2011 15:00:49
Kommentare		
Eingabe	Daten	K:\Lehre\SPSS\Beispieldaten.sav
	Aktiver Datensatz	DatenSet1
	Filter	geschlecht = 1 (FILTER)
	Gewichtung	<keine>
	Aufgeteilte Datei	<keine>
	Anzahl der Zeilen in der Arbeitsdatei	9
Behandlung fehlender Werte	Definition von fehlenden Werten	Benutzerdefinierte fehlende Werte werden als fehlend behandelt.
	Verwendete Fälle	Statistik basiert auf allen Fällen mit gültigen Daten.
Syntax		FREQUENCIES VARIABLES=alter /ORDER=ANALYSIS.
Ressourcen	Prozessorzeit	00 00:00:00,000
	Verstrichene Zeit	00 00:00:00,016

[DatenSet1] K:\Lehre\SPSS\Beispieldaten.sav

IBM SPSS Statistics Prozessor ist bereit

In der Syntax werden Auswahl-Bedingungen zum Filtern mittels „FILTER BY ...“ aktiviert.

```
FILTER BY ...  
EXECUTE.
```

- **Ausgewählte Fälle in neues Datenblatt kopieren**

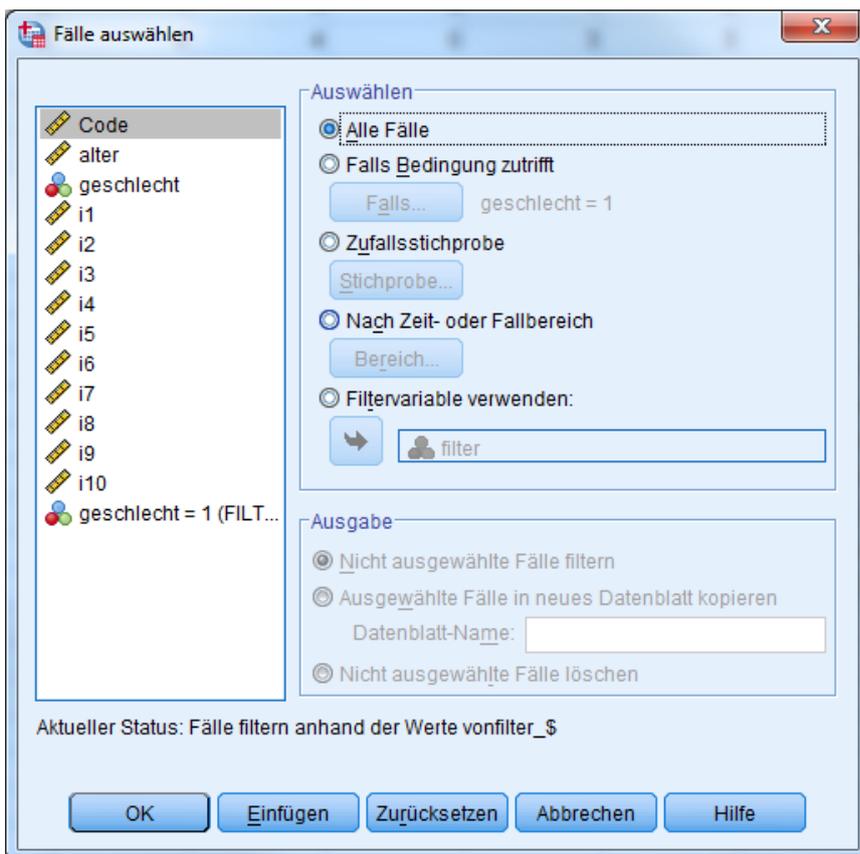
Es wird eine neue Datendatei erstellt, die nur die ausgewählten Fälle enthält.

- **Nicht ausgewählte Fälle löschen**

Die nicht ausgewählten Fälle werden aus der Datendatei gelöscht. Achtung: Es erfolgt keine Sicherheitsfrage! Ggf. zuvor Sicherheitskopie der Datendatei anlegen!

```
SELECT IF ...  
EXECUTE.
```

Soll die Filterfunktion beendet werden, so kann im Hauptdialogfenster ganz oben die Option „Alle Fälle“ gewählt werden.



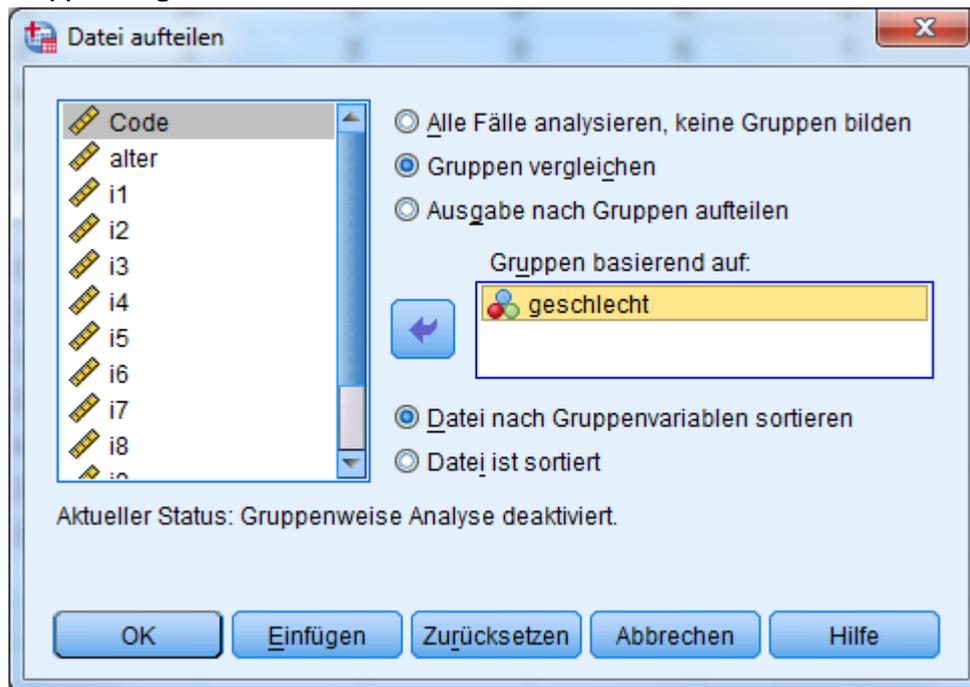
```
FILTER OFF.  
USE ALL.  
EXECUTE.
```

4.2 Datei aufteilen

Auch diese Funktion dient dazu, Berechnungen für eine oder mehrere Teilgruppen der Gesamtstichprobe durchzuführen und die Ergebnisse zu vergleichen (s. [Fälle auswählen](#)). Voraussetzung für den Datei-aufteilen-Befehl ist mindestens eine Variable (z. B. eine Filtervariable, s. [Fälle auswählen](#)), die verschiedene Werte für die Teilgruppen aufweist.

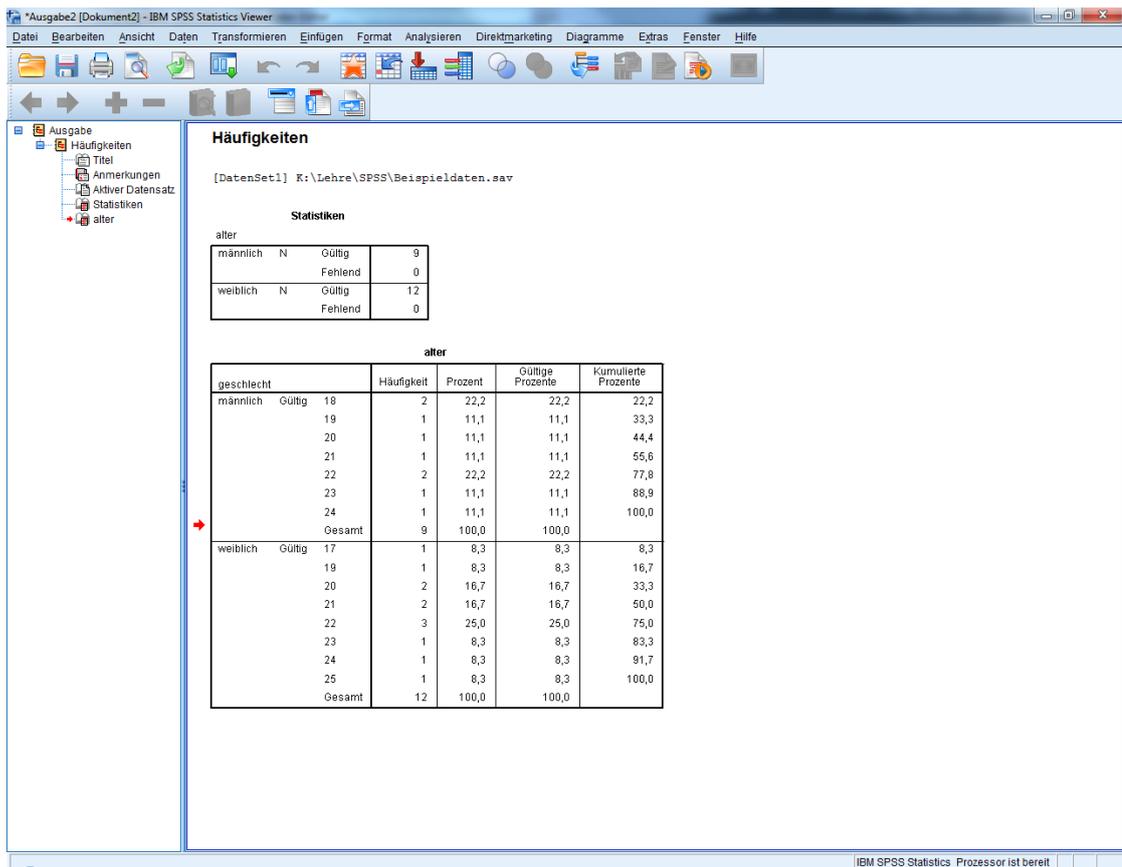
Unter *Daten/Datei aufteilen* stehen zwei Optionen zur Verfügung, bei denen jeweils eine oder mehrere Gruppenvariablen ausgewählt werden können, nach denen die Datei aufgeteilt werden soll.

- **Gruppen vergleichen**

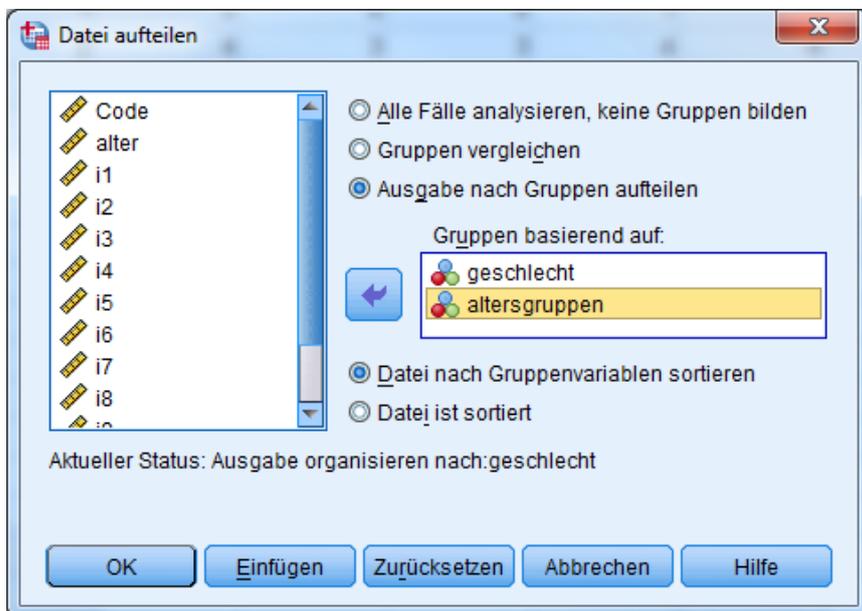


SORT CASES BY geschlecht.
SPLIT FILE LAYERED BY geschlecht.
EXECUTE.

Wird die Option „Gruppen vergleichen“ gewählt und wie im Beispiel (s. Screenshot) Geschlecht als Gruppenvariable ausgewählt, so werden alle folgenden Berechnungen getrennt für die beiden Geschlechtergruppen durchgeführt und die Ergebnisse in der Ausgabe (s. Screenshot, hier Häufigkeitstabelle des Alters) vergleichend untereinander dargestellt.



- Ausgabe nach Gruppen aufteilen



SORT CASES BY geschlecht altersgruppen.
 SPLIT FILE SEPARATE BY geschlecht altersgruppen.
 EXECUTE.

Das Beispiel (s. Screenshots) zeigt wie mehrere Variablen, hier Geschlecht und Altersgruppen, zum Aufteilen der Datei herangezogen werden. Bei der Option „Ausgabe nach Gruppen aufteilen“ werden die Ergebnisse für die einzelnen Teilgruppen (hier 4

Gruppen aufgrund von 2 Geschlechter- und 2 Altersgruppen) getrennt voneinander und nacheinander in der Ausgabe berichtet (s. Screenshot, hier Häufigkeitstabelle für Variable i1).

Häufigkeiten

Anmerkungen

Ausgabe erstellt	26-Apr-2011 16:35:47
Kommentare	
Eingabe	Daten K:\Lehre\SPSS\Beispieldaten.sav
	Aktiver Datensatz DatenSet1
	Filter <keine>
	Gewichtung <keine>
	Aufgeteilte Datei geschlecht, altersgruppen
	Anzahl der Zeilen in der Arbeitsdatei 21
Behandlung fehlender Werte	Definition von fehlenden Werten Benutzerdefinierte fehlende Werte werden als fehlend behandelt.
	Verwendete Fälle Statistik basiert auf allen Fällen mit gültigen Daten.
Syntax	FREQUENCIES VARIABLES=i1 /ORDER=ANALYSIS.
Ressourcen	Prozessorzeit 00 00:00:00,000
	Verstrichene Zeit 00 00:00:00,012

[DatenSet1] K:\Lehre\SPSS\Beispieldaten.sav

geschlecht = männlich, altersgruppen = 1

Statistiken^a

i1		
N	Gültig	4
	Fehlend	0

a. geschlecht = männlich, altersgruppen = 1

i1^a

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig 1	1	25,0	25,0	25,0
2	1	25,0	25,0	50,0
3	1	25,0	25,0	75,0
4	1	25,0	25,0	100,0
Gesamt	4	100,0	100,0	

a. geschlecht = männlich, altersgruppen = 1

Die Datei muss nach den Gruppenvariablen sortiert sein. Daher wird das Sortieren standardmäßig ebenfalls durchgeführt.

Wird der Befehl zum Datei aufteilen aktiviert, erscheint in der Datendatei unten rechts der Hinweis „Aufteilen nach ... [Gruppenvariable]“; in der Ausgabe wird innerhalb der Anmerkungen ebenfalls berichtet, ob und ggf. nach welchen Variablen die Datendatei zum Zeitpunkt der Berechnungen aufgeteilt war.

Die im Dialogfenster zu oberst aufgeführte Option „Alle Fälle analysieren, keine Gruppen bilden“ dient dazu, den Befehl zum Aufteilen der Datei zu deaktivieren.

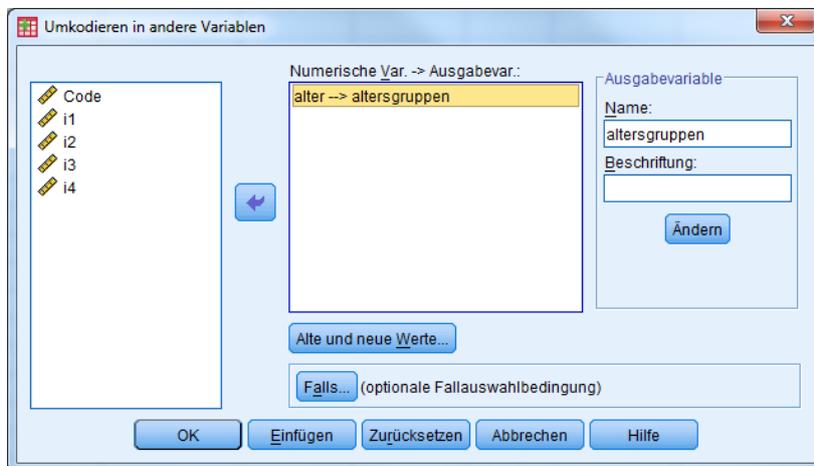
SPLIT FILE OFF.

5 DATENAUFBEREITUNG

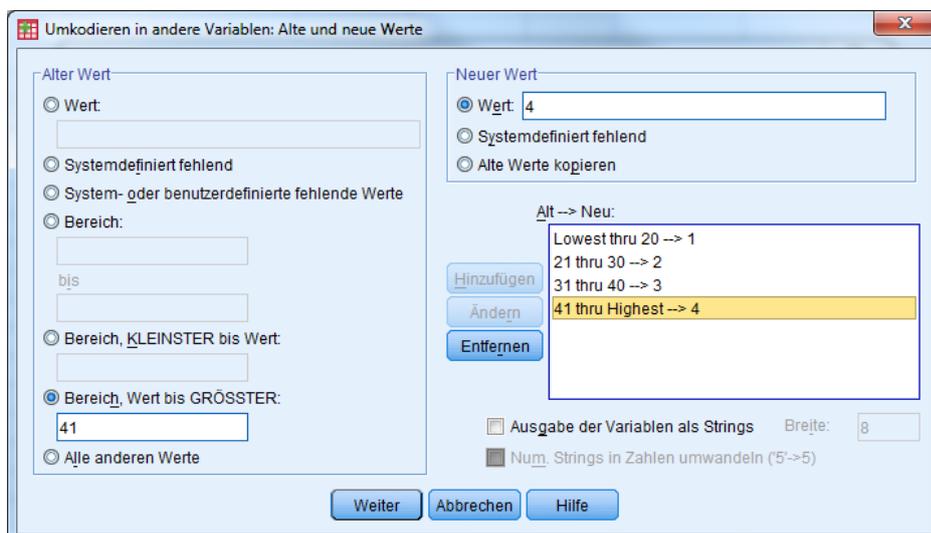
5.1 Umkodieren

Das Umkodieren von Werten wird relativ häufig angewendet, vor allem um Daten so aufzubereiten, dass bestimmte Berechnungen möglich werden. Typische Beispiele sind die Bildung von Altersgruppen anhand des Alters in Jahren und das Umdrehen der Antwortskala bei negativ gepolten Items.

SPSS bietet hierzu unter *Transformieren* zwei Optionen: *Umkodieren in dieselben Variablen* und *Umkodieren in andere Variablen*. In der Regel ist hier die zweite Option zu empfehlen, da den ursprünglichen Variablen später nicht mehr anzusehen ist, ob bzw. wie häufig sie umkodiert wurden.



Im Dialogfenster zum Umkodieren in andere Variablen sind auf der linken Seite die umzukodierenden Variablen auszuwählen und in das mittlere Feld „Numerische Var. -> Ausgabevar.“ zu verschieben. Rechts ist der Name der neuen Variablen anzugeben und mittels „Ändern“ der in der Mitte ausgewählten Variablen zuzuordnen.



Unter „Alte und neue Werte“ können die Umkodierungsregeln festgelegt werden. Hierzu wird links unter „Alter Wert“ ein Wert bzw. ein Wertebereich der ursprünglichen Variablen ausgewählt und rechts unter „Neuer Wert“ der Zielwert in der neuen Variable eingegeben. Jede auf diese Weise erstellte Regel muss mittels „Hinzufügen“ in das rechte untere Feld „Alt --> Neu:“ übernommen werden.

Als Syntax für das in den Screenshots illustrierte Beispiel zur Bildung von Altersgruppen ergibt sich:

```
RECODE alter (Lowest thru 20=1) (21 thru 30=2) (31 thru 40=3) (41 thru Highest=4) INTO
  altersgruppen.
EXECUTE.
```

Soll eine Antwortskala vollständig umgedreht werden, z. B. um negativ formulierte Items entsprechend einer hohen Skalenausprägung zu kodieren, ergibt sich als Umkodierungsregel häufig eine Syntax nach diesem Muster (hier für eine 5-stufige Antwortskala von 1 bis 5):

```
RECODE i1 (1=5) (2=4) (3=Copy) (4=2) (5=1) INTO i1_u.
EXECUTE.
```

Das gleiche Ergebnis kann in solchen Fällen auch mit einem Befehl zur [Berechnung einer neuen Variable](#) erzielt werden:

```
COMPUTE i1_u = 6 - i1.
EXECUTE.
```

5.2 Visuelles Klassieren

Unter *Transformieren/Visuelles Klassieren* bietet SPSS verschiedene komfortable Möglichkeiten eine kontinuierlich verteilte Variable in Gruppen einzuteilen und eine neue Variable entsprechend der Kategorisierung zu erstellen. Im Unterschied zum [Umkodieren](#) bietet sich diese Option vor allem dann an, wenn die empirischen Häufigkeiten der kontinuierlichen Variablen bei der Kategorisierung berücksichtigt werden sollen.

Nach Aufrufen des Dialogfensters sind zunächst die kontinuierlichen Variablen auszuwählen und in das Feld „Variablen für Klassierung“ zu verschieben. Im folgenden Dialogfenster wird die Verteilung der Werte für die jeweils links ausgewählte Variable als [Histogramm](#) dargestellt. Über „Trennwerte erstellen“ können Trennwerte definiert werden, die die vorliegenden kontinuierlichen Daten anhand von „Intervallen mit gleicher Breite“, in „gleiche Perzentile“ oder „bei [Mittelwert](#) und ausgewählten [Standardabweichungen](#)“ klassifizieren. Um zum Beispiel eine Variable „Altersgruppen“ zu erstellen, die die ursprüngliche Variable am [Median](#) teilt und so genau die Hälfte der vorliegenden Fälle als „jung“ und die andere als „alt“ klassifiziert, kann hier festgelegt

werden, dass gleiche Perzentile anhand von einem Trennwert gebildet werden sollen. Beide Perzentile umfassen dann automatisch 50% der Fälle (s. Screenshot).

Trennwerte erstellen

Intervalle mit gleicher Breite

Intervalle - mindestens zwei Felder ausfüllen:

Position des ersten Trennwerts:

Anzahl der Trennwerte:

Breite:

Position des letzten Trennwerts:

Gleiche Perzentile auf der Grundlage der durchsuchten Fälle

Intervalle - eines der beiden Felder ausfüllen:

Anzahl der Trennwerte:

Breite (%):

Trennwerte bei Mittelwert und ausgewählten Standardabweichungen auf der Grundlage der durchsuchten Fälle

+/- 1 Std.-Abw.

+/- 2 Std.-Abw.

+/- 3 Std.-Abw.

i Durch "Zuweisen" werden die Trennwertdefinitionen durch diese Spezifikation ersetzt.
Ein letztes Intervall enthält alle übrigen Werte: N Trennwerte führen zu N+1 Intervallen.

Durch „Zuweisen“ gelangt man zum vorherigen Dialogfenster, in dem jetzt zusätzlich der Trennwert im Histogramm abgetragen ist. Der Name der neuen, klassierten Variable ist anzugeben und es können [Wertelabels](#) für die Werte der klassierten Variable vergeben werden.

Visuelles Klassieren

Liste der durchsuchten Variablen: alter

Name: Label:

Aktuelle Variable:

Klassierte Variable:

Minimum: Nicht fehlende Werte Maximum:

i Geben Sie Intervall-Trennwerte ein oder klicken Sie auf "Trennwerte erstellen", um automatische Trennwerte zu erstellen. Ein Trennwert von 10 beispielsweise definiert ein Intervall, das über dem vorangegangenen Intervall beginnt und bei 10 endet.

Durchsuchte Fälle: Fehlende Werte:

Klassen kopieren:

Gitter	Wert	Beschriftung
1	21,0	jung
2		HOCH alt
3		

Obere Endpunkte:

Eingeschlossen (<=)

Ausgeschlossen (<)

Skala umkehren

5.3 z-Standardisierung

Die Standardisierung von Werten dient der besseren Vergleich- und Interpretierbarkeit von Einzelwerten. Zu diesem Zweck wird hier nicht nur die zentrale Tendenz einer Verteilung, sondern auch die Streuung der Verteilung als Vergleichsmaßstab mit einbezogen. Die z-standardisierte Variable hat stets einen Mittelwert von 0 und eine Standardabweichung von 1.

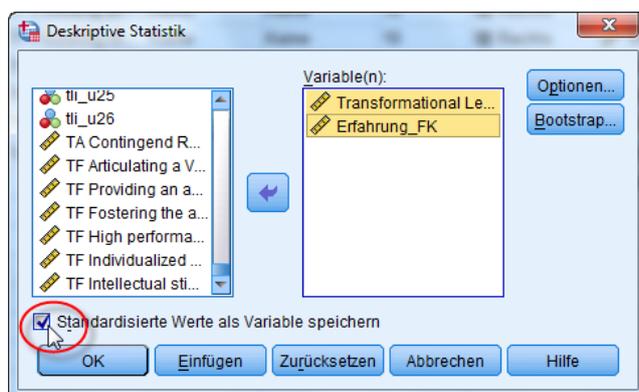
Rechnerisch ergibt sich der Wert z der z-standardisierten Variable aus der ursprünglichen Variable X zu

$$z = \frac{x - m_x}{s_x}$$

mit m_x = Mittelwert der Variable X und
 s_x = Standardabweichung der Variable X.

In SPSS können Variablen z-standardisiert werden, in dem eine neue Variable manuell anhand der o.g. Formel berechnet wird (s. [Variable berechnen](#)) oder automatisiert wie im Folgenden beschrieben:

Unter *Analysieren/Deskriptive Statistiken/Deskriptive Statistik* werden die zu standardisierenden Variablen ausgewählt. Anschließend ist der Haken bei „Standardisierte Werte als Variable speichern“ zu setzen. Die z-standardisierten Werte erscheinen nun mit einem vorstehenden „Z“ im Variablennamen als neue Variablen in der Datendatei.



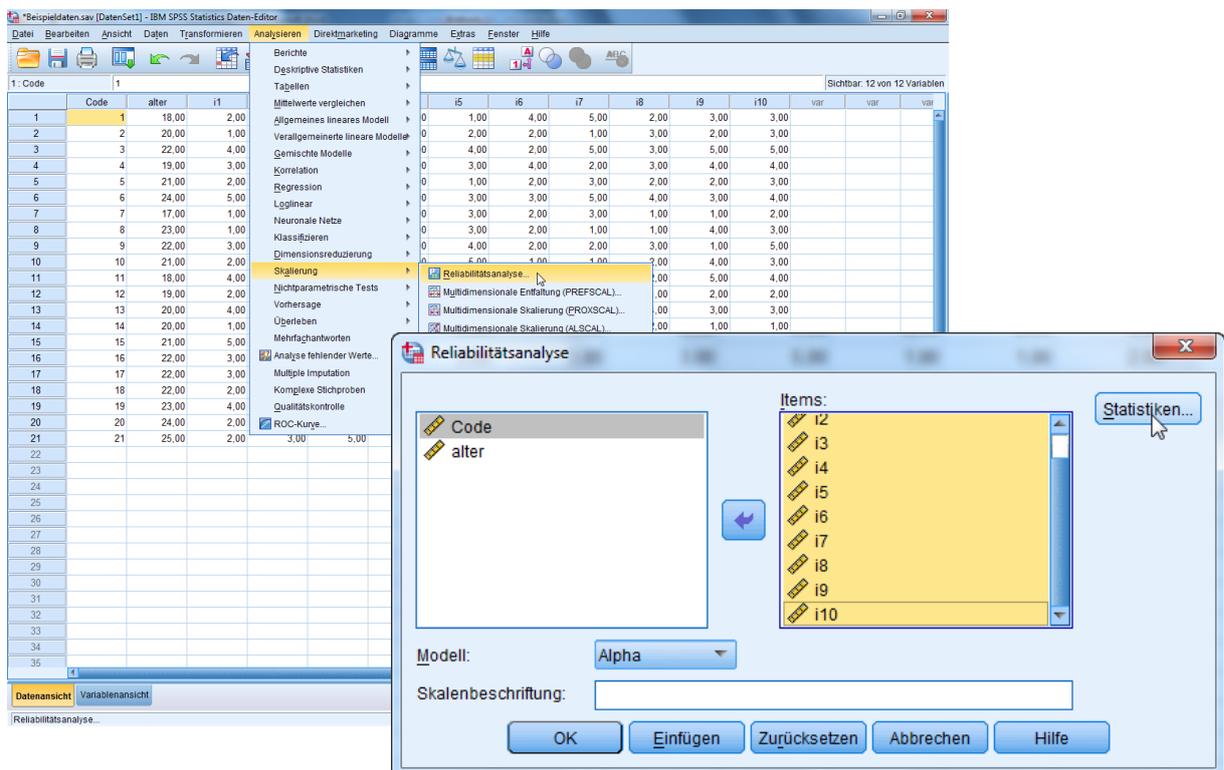
Z-standardisierte Variablen werden beispielsweise zur Durchführung einer [Regressionsanalyse mit Moderator](#) benötigt.

6 SKALENBILDUNG

Unter Skalenbildung ist die Zusammenfassung mehrerer Einzelfragen (sog. Items) zu einer Skala gemeint, um ein Konstrukt zuverlässiger ([reliabler](#)) zu erfassen als es mit einer einzelnen Frage möglich ist.

6.1 Reliabilitätsanalyse

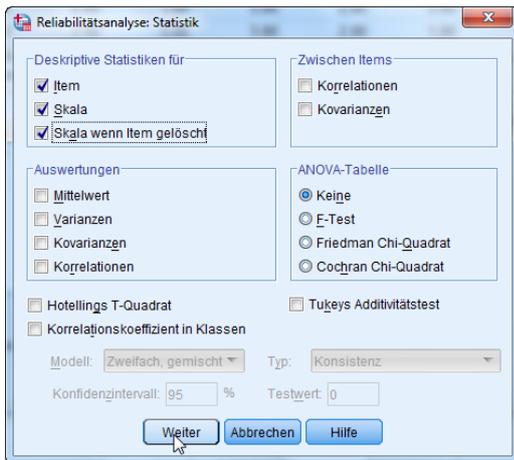
SPSS bietet unter *Analysieren/Skalierung/Reliabilitätsanalyse* Prozeduren an, um Maße der [internen Konsistenz](#) zu bestimmen.



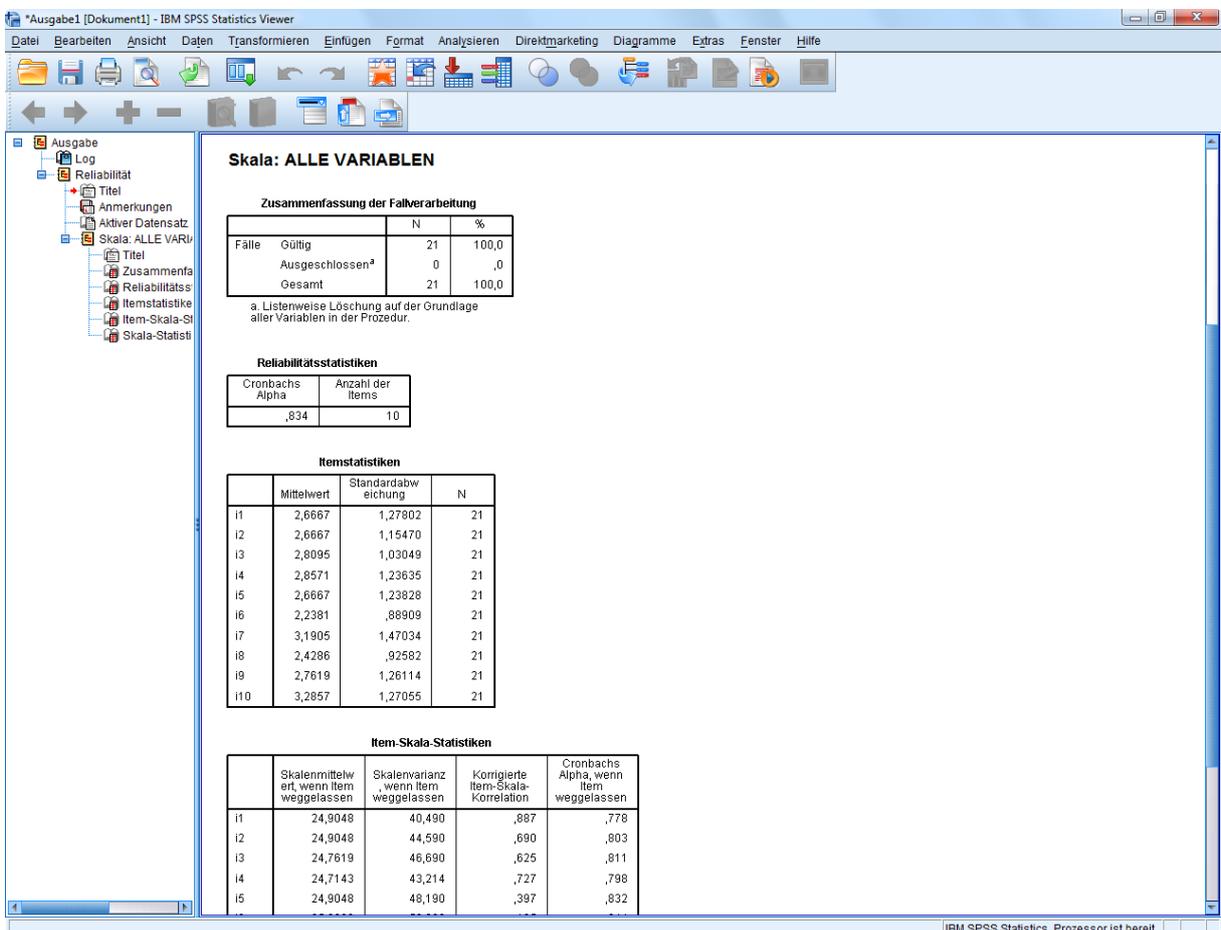
The screenshot shows the SPSS Reliability Analysis dialog box. The 'Items' list on the right contains variables i2, i3, i4, i5, i6, i7, i8, i9, and i10. The 'Modell' dropdown is set to 'Alpha'. The 'Skalenbeschriftung' field is empty. The 'Statistiken...' button is highlighted. In the background, the SPSS main window shows a data editor with variables 'Code', 'alter', and 'i1'.

Unter „Modell“ steht u. a. „Alpha“ zur Bestimmung von [Cronbachs Alpha](#) zur Verfügung. Die Items, die die Skala bilden sollen, sind im linken Teil auszuwählen und in den rechten Teil zu verschieben. Unter „Statistiken“ können verschiedene zusätzliche Maße angefordert werden, zu empfehlen

sind die deskriptiven Statistiken für „Item“, „Skala“ und „Skala wenn Item gelöscht“.



In der Ausgabe wird Cronbachs Alpha berichtet sowie (falls angefordert) deskriptive Statistiken zu den einzelnen Items und der gesamten Skala. In der Tabelle „Item-Skala-Statistiken“ kann unter „Korrigierte Item-Skala-Korrelation“ die sog. Trennschärfe der Items abgelesen werden, sie gibt die Korrelation des Items mit dem Skalenwert wieder, der aus den jeweils anderen Items berechnet wurde. Items mit einer negativen Trennschärfe müssen evtl. noch umkodiert werden, Items mit einer sehr niedrigen Trennschärfe messen offenbar etwas anderes als die anderen Items. Unter „Cronbachs Alpha, wenn Item weggelassen“ kann abgelesen werden, ob sich Cronbachs Alpha verbessert, wenn ein einzelnes Item ausgeschlossen wird.



6.1.1 Exkurs Reliabilität

Die Reliabilität beschreibt die Zuverlässigkeit einer Messung. Ausgehend von der klassischen Testtheorie gibt die Reliabilität den Anteil der Varianz in den Testwerten wieder, der nicht auf die Fehlervarianz sondern auf die Varianz der wahren Werte zurückgeht.

Anhand vorliegender Daten kann die Reliabilität einer Skala auf verschiedene Arten bestimmt werden:

- **Interne Konsistenz**

Die interne Konsistenz beschreibt die Homogenität der Skala, d.h. inwiefern die einzelnen Items dasselbe Konstrukt erfassen. Es wird davon ausgegangen, dass die Messung der Skala umso reliabler ist, je homogener die einzelnen Items sind.

- Bei der Berechnung der **Split-Half-Reliabilität** werden die Items der Skala in zwei Hälften geteilt (z. B. Items mit ungeraden Nummern vs. Items mit geraden Nummern) und zu zwei Kennwerten zusammengefasst (z. B. Mittelwert oder Summe der jeweiligen Items). Die Korrelation dieser beiden Skalenhälften-Kennwerte ergibt die Split-Half-Reliabilität.
- Bei der Berechnung von **Cronbachs Alpha** wird ein übergreifendes Maß für die Homogenität der Skala bestimmt, indem (grob gesagt) alle möglichen Testhalbierungen durchgeführt und die jeweiligen Split-Half-Reliabilitäten gemittelt werden. Cronbachs Alpha hat sich in den letzten Jahren als Standardmaß zur Reliabilität einer Skala etabliert und wird praktisch in jedem psychologischen Fachartikel für die eingesetzten Skalen berichtet. In der Regel wird dabei ein Wert von .70 als Mindestgröße postuliert, wobei bei sehr kurzen Skalen (mit weniger als 4 Items) oder bei sehr breiten Konstrukten ggf. geringere Werte akzeptiert werden.

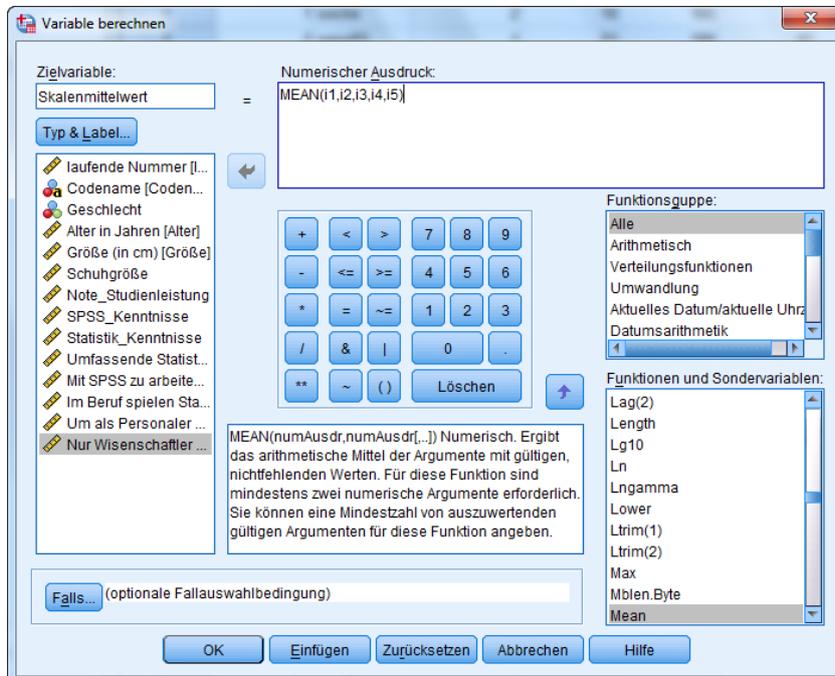
- **Retest-Reliabilität**

Die Retest-Reliabilität beschreibt die Zuverlässigkeit der Messung über verschiedene Messzeitpunkte hinweg. Sie wird bestimmt, indem die Skalenwerte von zwei Messzeitpunkten miteinander korreliert werden.

Beide Arten der Reliabilitätsbestimmung sind sehr geläufig und können unterschiedlich angemessen sein. Bildet die Skala ein eher heterogenes Konstrukt ab (wie z. B. die Big Five, die inhaltlich sehr breite Persönlichkeitseigenschaften beschreiben), bietet es sich an, die Reliabilität mittels einer Retest-Studie zu untersuchen. Handelt es sich beim zu messenden Konstrukt hingegen um ein Merkmal, das starken zeitlichen Schwankungen unterworfen ist (wie z. B. Stimmungen oder Emotionen) sollte die Reliabilität anhand der internen Konsistenz ermittelt werden.

6.2 Variable berechnen

Ist nachgewiesen worden, dass mehrere Items ausreichend [homogen](#) sind und somit dasselbe Konstrukt messen, können sie zu einem Gesamtwert verrechnet werden, typischerweise wird hierzu die Summe oder der Mittelwert der Itemwerte bestimmt.



Unter *Transformieren/Variable berechnen* ist zunächst der Name und ggf. der Typ der Zielvariable anzugeben. Im rechten oberen Fenster „Numerischer Ausdruck“ kann die Formel eingegeben werden, die angewendet werden soll, um die Werte der Zielvariablen für jeden Fall zu bestimmen. Hierbei können sowohl bestehende Variablen (linke Spalte) als auch eine Vielzahl zur Verfügung stehender Berechnungsfunktionen (rechte Spalten „Funktionsgruppe“ und „Funktionen und Sondervariablen“) eingebunden werden. Der Screenshot zeigt die Berechnung eines Skalenwertes als Mittelwert (MEAN) der Items i1 bis i5. Als Syntax ergibt sich:

```
COMPUTE Skalenmittelwert=MEAN(i1,i2,i3,i4,i5).  
EXECUTE.
```

Wird der Befehl ausgeführt, wird eine neue Variable in der Datendatei angelegt, in der für jeden Fall anhand der Formel ein Wert berechnet und eingetragen wird.

Hinweise zu den Berechnungsfunktionen und ihrer Verwendung finden sich in der SPSS-Hilfe unter „Functions“ sowie in der Befehlssyntax-Referenz unter dem Befehl COMPUTE.

7 PRÜFUNG VON ZUSAMMENHANGSHYPOTHESEN

Bei der Prüfung von Zusammenhangshypothesen wird untersucht, inwieweit zwischen zwei metrisch skalierten Variablen ein Zusammenhang besteht. Grundlegende Methode hierfür bildet die [bivariate Korrelation](#). Werden zugleich weitere Variablen als Kontrollvariablen berücksichtigt, handelt es sich um [partielle Korrelationen](#). Soll untersucht werden, inwieweit mehrere (unabhängige) Variablen mit einer (abhängigen) Variablen zusammenhängen, kann das Verfahren der [linearen Regression](#) genutzt werden.

7.1 Bivariate Korrelation

Eine Korrelation beschreibt eine lineare Beziehung zwischen Variablen. Handelt es sich um den Zusammenhang zwischen zwei Variablen, wird von einer *bivariaten* Korrelation gesprochen, bei mehr Variablen von *multipler* Korrelation.

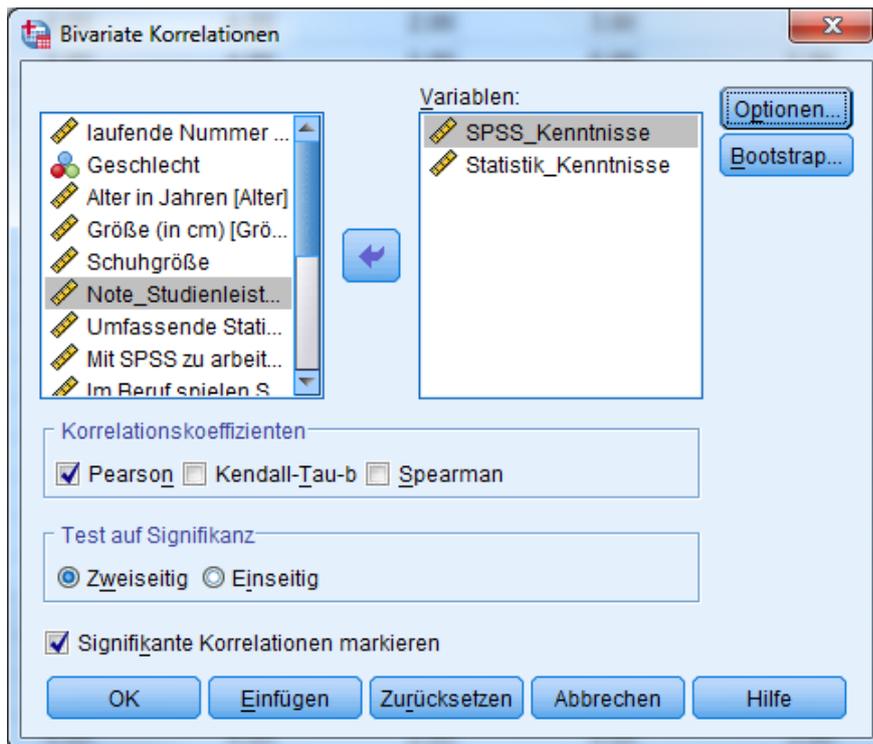
Ein linearer Zusammenhang ist gegeben, wenn die Ausprägungen der Variablen nach einem je-desto-Prinzip variieren. Der *Korrelationskoeffizient* r als Maß für einen linearen Zusammenhang beschreibt sowohl die Richtung als auch die Größe des Zusammenhangs:

- r kann Werte zwischen -1 und 1 annehmen.
- Bei $r = 0$ ist kein linearer Zusammenhang gegeben. (Es kann jedoch durchaus ein nicht-linearer Zusammenhang bestehen.)
- Bei einer positiven Korrelation ($r > 0$) variieren die beiden Variablen gleich gerichtet, d. h. je größer Variable A, desto größer Variable B.
- Bei einer negativen Korrelation ($r < 0$) variieren die beiden Variablen entgegen gerichtet, d. h. je größer Variable A, desto kleiner Variable B.

Wichtig: Das Vorliegen einer Korrelation sagt nichts über die Kausalität aus – weder über die Richtung, noch ob überhaupt eine direkte Kausalität zwischen A und B besteht, ob beide von einer dritten Variable beeinflusst werden, oder ob sie nur zufällig kovariieren.

Mindestvoraussetzung für das sinnvolle Berechnen einer Korrelation ist das Ordinalskalenniveau der Variablen (Rangkorrelation), klassischerweise werden jedoch erst ab Intervallskalenniveau Korrelationen berechnet (s. [Pearson-Korrelation](#)).

Mit SPSS können Korrelationskoeffizienten über *Analysieren/Korrelationen/Bivariat* berechnet werden. Hier sind die zu korrelierenden Variablen auszuwählen. Die Voreinstellungen (Pearson-Koeffizient; 2-seitiger Signifikanztest; signifikante Korrelationen kennzeichnen) können beibehalten werden.

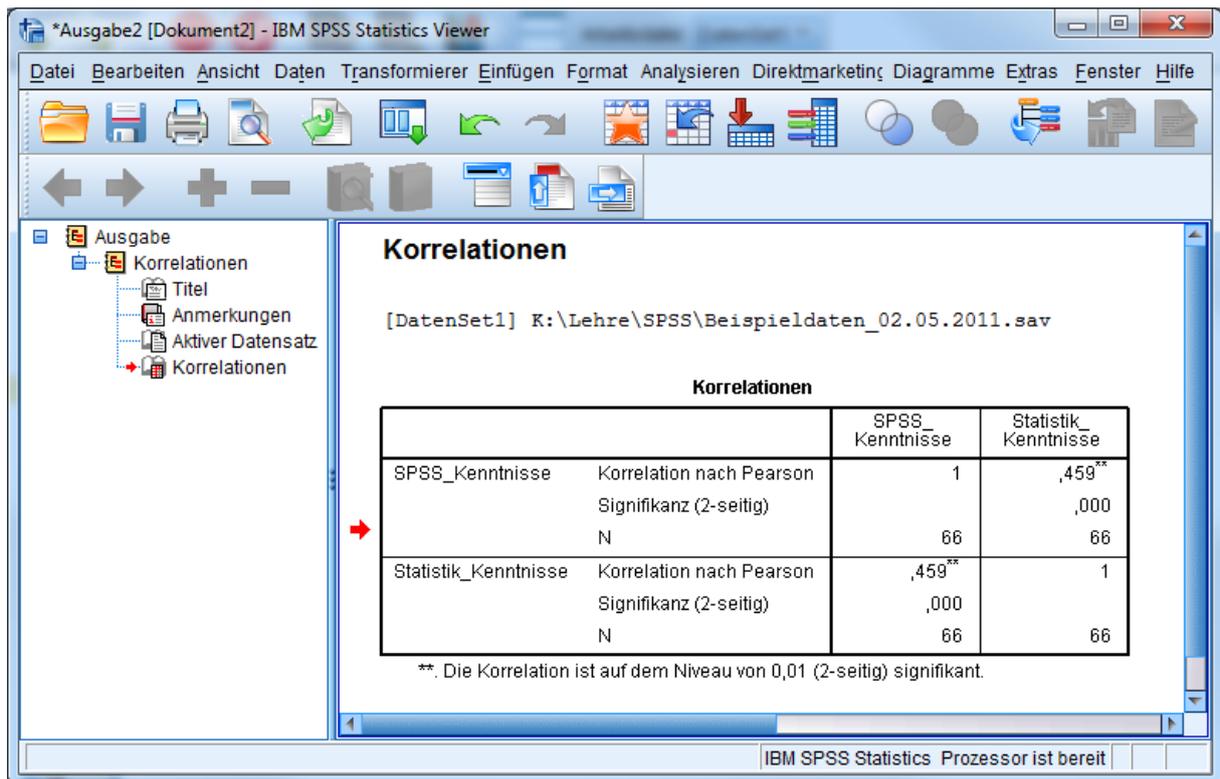


```

CORRELATIONS
/VARIABLES=SPSS_Kenntnisse Statistik_Kenntnisse
/PRINT=TWOTAIL NOSIG
/MISSING=PAIRWISE.
EXECUTE.

```

Die Ausgabe liefert eine symmetrische Tabelle, in deren Zellen über und unter der Diagonalen jeweils die Korrelation des Variablenpaares der jeweiligen Zeile und Spalte zu finden ist.

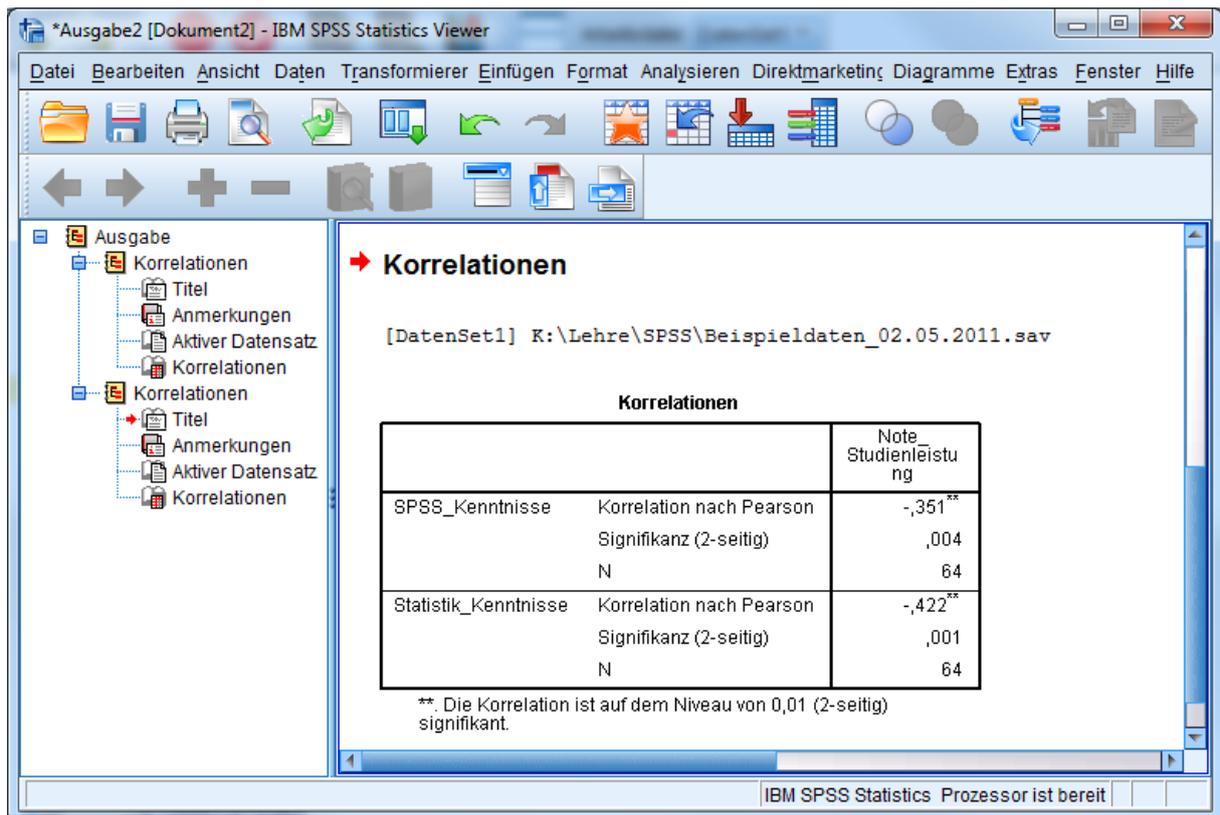


Es können im Dialogfenster auch mehr als zwei Variablen ausgewählt werden, um die Korrelationen aller enthaltenen Variablenpaare zu berechnen. Interessieren nur die Korrelationen zwischen einem Teil der Variablen mit einem anderen Teil der Variablen, kann im Syntaxbefehl das Wort „WITH“ zwischen diesen beiden Variablengruppen eingefügt werden. Es resultiert eine nicht symmetrische Tabelle, mit den vor dem „WITH“ genannten Variablen in Zeilen und den nach dem „WITH“ genannten Variablen als Spalten.

```

CORRELATIONS
/VARIABLES=SPSS_Kenntnisse Statistik_Kenntnisse WITH Note_Studienleistung
/PRINT=TWOTAIL NOSIG
/MISSING=PAIRWISE.
EXECUTE.

```



7.1.1 Exkurs Pearson-Korrelation

Der Pearson-Korrelationskoeffizient r , auch Produkt-Moment-Korrelation, berechnet sich für die n Wertepaare der Variablen x und y mit den [Mittelwerten](#) m_x und m_y und den [Standardabweichungen](#) s_x und s_y zu

$$r = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - m_x)(y_i - m_y)}{s_x s_y}$$

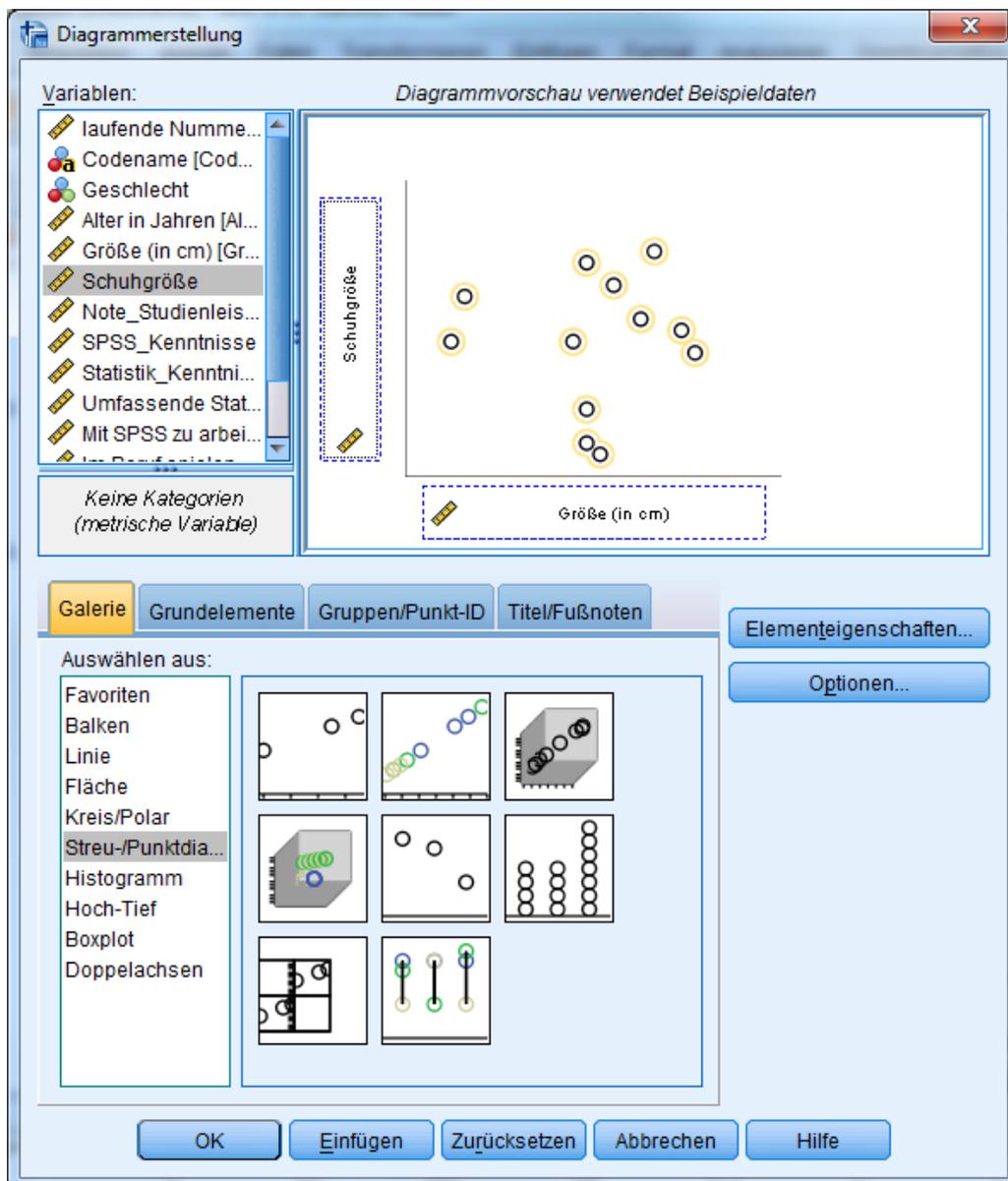
Voraussetzung für die Berechnung des Pearson-Korrelationskoeffizienten ist das [Intervallskalenniveau](#) der beiden Variablen. Für die sinnvolle Interpretation des in SPSS mitgelieferten Signifikanztest zum Korrelationskoeffizienten ist zusätzlich eine annähernde [Normalverteilung](#) der Variablen Voraussetzung.

Dieser Signifikanztest prüft, ob die ermittelte Korrelation signifikant von Null verschieden ist. Nur wenn dies gegeben ist, wenn also die Wahrscheinlichkeit für die H_0 , dass die Korrelation gleich Null ist, geringer als z. B. 5% ist, darf der ermittelte Korrelationskoeffizient interpretiert werden. Ein signifikantes Ergebnis ist umso wahrscheinlicher, je größer die Stichprobe ist (also umso mehr Wertepaare für x und y vorliegen). Bei sehr kleinen Stichproben, bei denen somit auch große Korrelationen nicht signifikant werden, kann zur Interpretation die Heuristik von Bortz und Döring (Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Auflage). Berlin: Springer.) herangezogen werden:

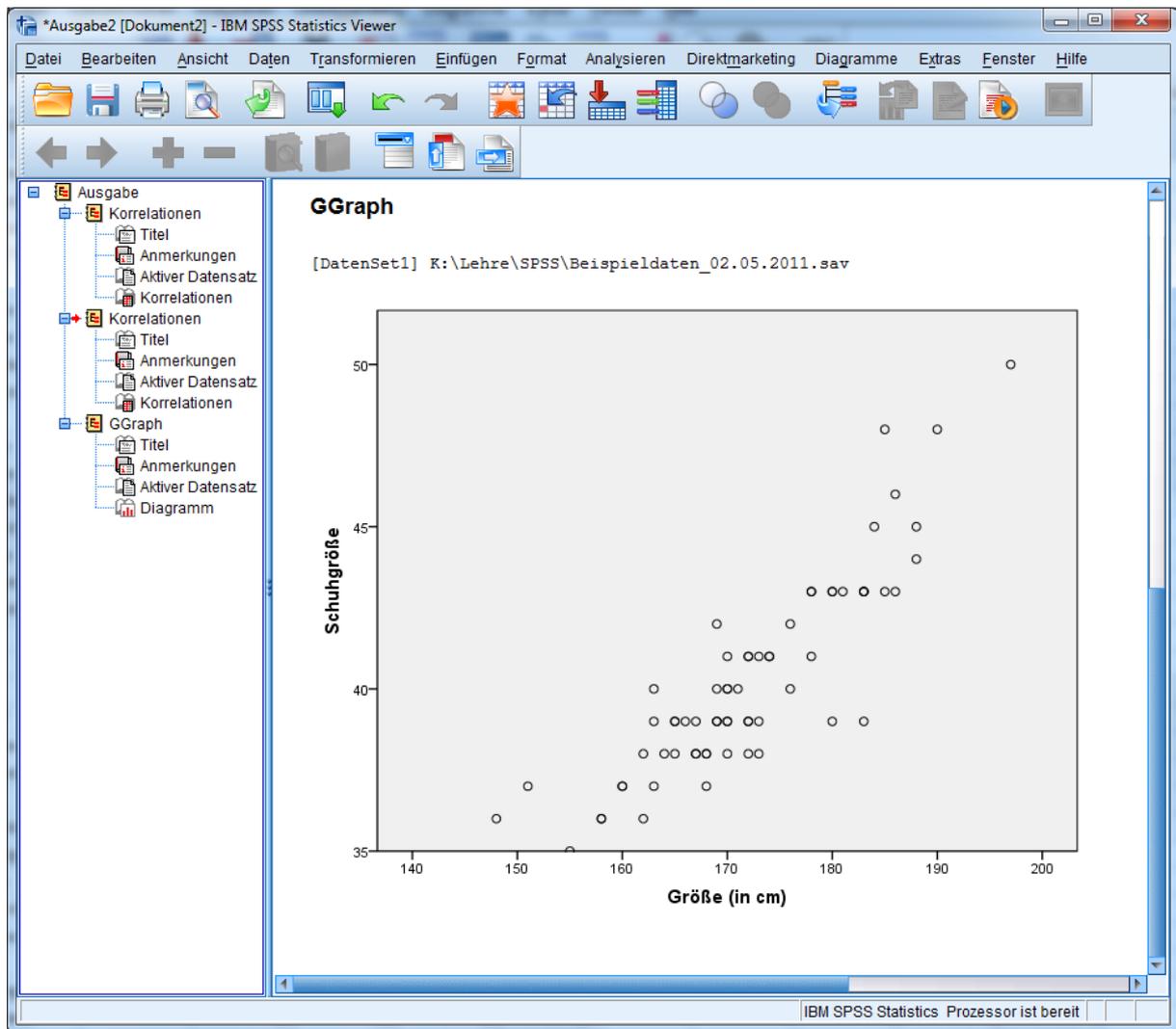
- $r = .10$ kleine Korrelation
- $r = .30$ mittlere Korrelation
- $r = .50$ große Korrelation

7.1.2 Streudiagramm

Lineare Zusammenhänge zwischen zwei Variablen können anhand eines Streudiagramms veranschaulicht werden. Unter *Diagramme/Diagrammerstellung* kann in der Galerie ein einfaches Streu-/Punktediagramm gewählt und nach oben in die Diagrammvorschau gezogen werden. Aus der Variablenliste kann dann jeweils eine Variable für die X- und für die Y-Achse ausgewählt und in die Vorschau gezogen werden.



Das resultierende Streudiagramm zeigt die Punktwolke der Wertepaare der gewählten Variablen. Je näher sich die Punktwolke einer Geraden annähert, desto größer ist der lineare Zusammenhang der Variablen.



7.2 Partielle Korrelation

Eine Korrelation zwischen zwei Variablen X und Y kann unter Umständen auf eine Drittvariable U zurückgeführt werden. Das klassische Beispiel hierzu: Die Korrelation zwischen der Häufigkeit von Störchen und der Geburtenzahl geht auf die Variable Ländlichkeit zurück. Um solche Zusammenhänge aufzudecken, kann die partielle Korrelation genutzt werden. Die partielle Korrelation von X und Y unter U ergibt sich zu

$$r_{(X,Y)/U} = \frac{r_{XY} - r_{XU} * r_{YU}}{\sqrt{(1 - r_{XU}^2)(1 - r_{YU}^2)}}$$

In SPSS können partielle Korrelationen berechnet werden unter *Analysieren/Korrelation/Partiell*, wo die zu korrelierenden Variablen ebenso wie die (eine oder mehrere) Kontrollvariablen auszuwählen sind.



```
PARTIAL CORR
/VARIABLES=Alter Schuhgröße BY Größe
/SIGNIFICANCE=TWOTAIL
/MISSING=LISTWISE.
EXECUTE.
```

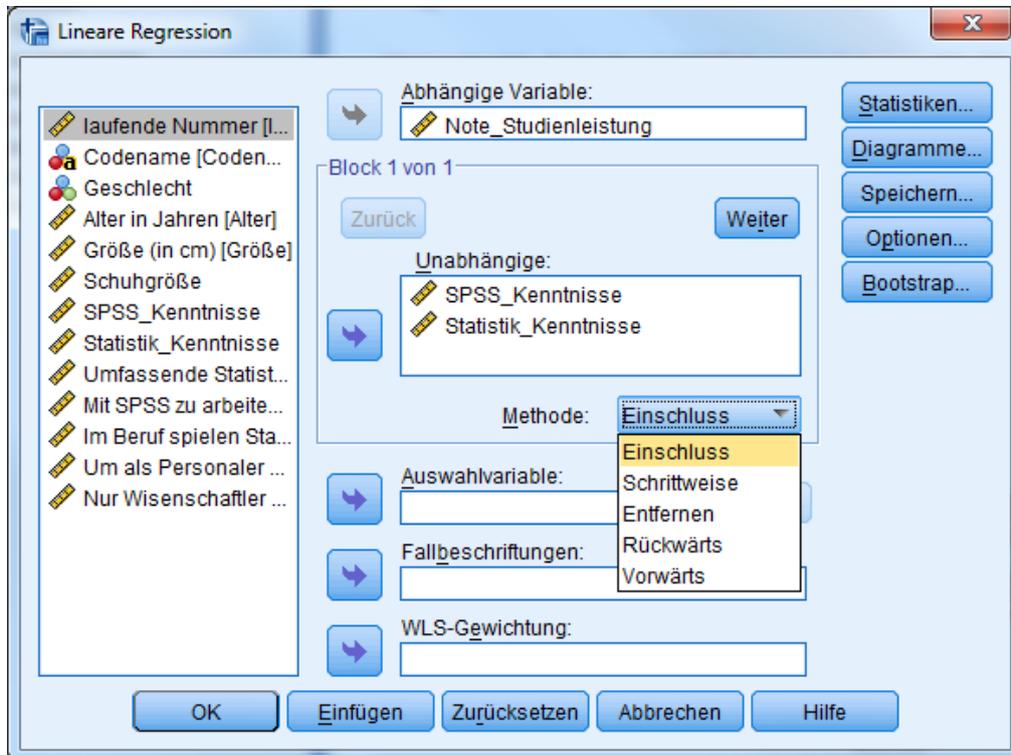
7.3 Lineare Regression

Im Unterschied zur bivariaten Korrelation ermöglicht es die Regressionsanalyse Zusammenhänge zwischen *mehreren* unabhängigen Variablen (uVs, auch Prädiktoren) und einer abhängigen Variable (aV, auch Kriterium) zu ermitteln. Beispielsweise kann untersucht werden, inwieweit die Studienleistung (als aV) durch Prädiktoren wie Abiturnote, Intelligenz und Studierzeit je Woche erklärt werden kann.

Unter *Analysieren/Regression/Linear* können daher eine abhängige und mehrere unabhängige Variablen ausgewählt werden. Zusätzlich ist die Methode anzugeben, nach der die uVs in die Regressionsgleichung (vgl. [Exkurs Regression](#)) aufgenommen werden:

- Einschluss (Standardeinstellung): Alle uVs werden in einem Schritt aufgenommen.
- Schrittweise: Die uVs werden einzeln aufgenommen, wobei bei jedem Schritt die uV mit der kleinsten F-Wahrscheinlichkeit aufgenommen, sofern diese Wahrscheinlichkeit klein genug ist. Bereits in der Regressionsgleichung enthaltene Variablen werden entfernt, wenn ihre F-Wahrscheinlichkeit hinreichend groß wird.
- Entfernen: Alle uVs werden in einem Schritt ausgeschlossen.

- Rückwärts: Alle uVs werden aufgenommen und anschließend sequenziell ausgeschlossen, jeweils mit der kleinsten Teilkorrelation zur aV.
- Vorwärts: Variablen werden schrittweise aufgenommen, jeweils mit der größten Korrelation bzw. (ab der zweiten uV) partiellen Korrelation zur aV.



Bei Bedarf können die uVs zudem in Blöcke gegliedert werden, die jeweils einzeln gemäß der blockweise gewählten Methode aufgenommen werden. Die Blöcke können über die Schaltflächen „Zurück“ und „Weiter“ ausgewählt und entsprechend spezifiziert werden.

Über die Schaltflächen „Statistiken“, „Diagramme“, „Speichern“ und „Optionen“ können weitere Einstellungen vorgenommen und Ausgabe-Elemente angefordert werden. Die Regression einer aV auf zwei uVs nach dem Einschluss-Verfahren ergibt in den Standardeinstellungen eine Ausgabe mit folgenden Elementen:

- Aufgenommene/Entfernte Variablen:
Es wird berichtet, welche uVs aufgenommen wurde.
- Modellzusammenfassung:
 - Die multiple Korrelation R ist ein Maß für den Zusammenhang, der insgesamt zwischen den uVs und der aV besteht.
 - Der Determinationskoeffizient R^2 gibt an, welcher Anteil der Varianz der aV durch die aufgenommenen uVs erklärt werden kann und ist somit ein Maß für die Modellgüte. Mit dem korrigierten R^2 wird versucht abzuschätzen, welche Güte das Modell in der Grundgesamtheit (und nicht nur in der vorliegenden Stichprobe) aufweist. In der Regel wird das korrigierte R^2 als konservative Schätzung berichtet.

- ANOVA:

Es wird anhand eines F-Tests geprüft, ob die uVs insgesamt einen signifikanten Beitrag zur Vorhersage der aV leisten können. Werte $< .05$ in der Spalte „Sig.“ werden in der Regel als signifikant bewertet.

- Koeffizienten:

- Die nicht standardisierten Regressionskoeffizienten der uVs entsprechen den Gewichten, mit denen die [Regressionsgleichung](#) gebildet wird.
- Die standardisierten Regressionskoeffizienten können genutzt werden, um die Bedeutsamkeit der einzelnen uVs zur Vorhersage der aV abzulesen, unabhängig von ihrer ursprünglichen Skalierung.
- T-Wert und Signifikanz geben an, ob die standardisierten Regressionskoeffizienten signifikant von Null verschieden sind.

Regression

Aufgenommene/Entfernte Variablen^a

Modell	Aufgenommene Variablen	Entfernte Variablen	Methode
1	Statistik_Kenntnisse, SPSS_Kenntnisse	.	Einschluß

a. Alle gewünschten Variablen wurden eingegeben.
b. Abhängige Variable: Note_Studienleistung

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,457 ^a	,208	,183	,6391

a. Einflußvariablen : (Konstante), Statistik_Kenntnisse, SPSS_Kenntnisse

ANOVA^b

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	6,563	2	3,282	8,034	,001 ^a
	Nicht standardisierte Residuen	24,916	61	,408		
	Gesamt	31,480	63			

a. Einflußvariablen : (Konstante), Statistik_Kenntnisse, SPSS_Kenntnisse
b. Abhängige Variable: Note_Studienleistung

Koeffizienten^a

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Sig.
		Regressionskoeffizient B	Standardfehler	Beta		
1	(Konstante)	2,908	,206		14,110	,000
	SPSS_Kenntnisse	-,110	,072	-,197	-1,532	,131
	Statistik_Kenntnisse	-,140	,055	-,330	-2,561	,013

a. Abhängige Variable: Note_Studienleistung

7.3.1 Exkurs Regression

Ziel der linearen Regression ist es in der Regel festzustellen, inwiefern die Prädiktoren in der Lage sind, die Kriteriumswerte vorherzusagen. Zur Vorhersage wird eine Regressionsgleichung ermittelt, in die diejenigen Prädiktoren (x) aufgenommen werden, die einen signifikanten Beitrag zur Vorhersage des Kriteriums (y) leisten können. Sie werden (zusätzlich zu einer Konstanten c) addiert und mit spezifischen Gewichten (β) versehen, die ihren relativen Anteil an der Vorhersageleistung widerspiegeln:

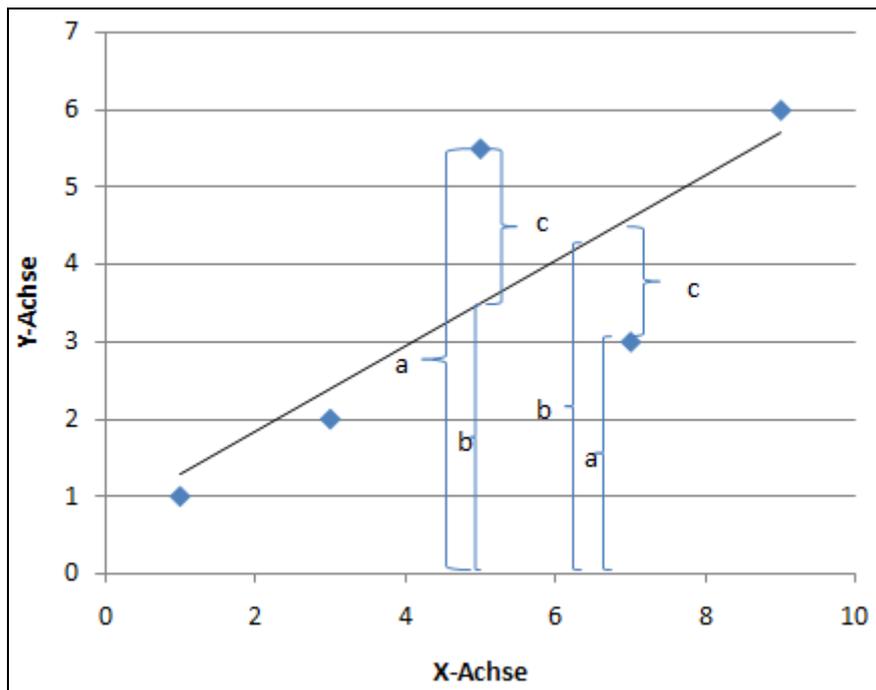
$$y = c + \beta_1 x_1 + \beta_2 x_2 + \dots + e$$

Die nicht erklärten Anteile der Kriteriumswerte (e) werden als Residuen bezeichnet.

Voraussetzungen:

- aV ist mind. intervallskaliert
- uVs sind intervallskaliert oder dichotom und korrelieren nicht perfekt miteinander (Multikollinearität)
- Residuen sind normalverteilt, haben einen Erwartungswert von Null, korrelieren nicht miteinander und nicht mit den uVs und zeigen eine konstante Varianz über verschiedene Ausprägungen der uVs hinweg (Homoskedastizität)

Das Vorgehen der Regressionsanalyse kann für den Fall von nur einer uV graphisch veranschaulicht werden. Hierzu dient ein Streudiagramm mit der uV auf der X-Achse und der aV auf der Y-Achse. Mit Hilfe der linearen Regression wird eine Gerade durch die Punktwolke der X-Y-Wertepaare gelegt, so dass die Abstände (senkrechtes Lot) aller Punkte zur Geraden insgesamt minimiert werden. Diese Abstände, die sog. Residuen, werden berechnet als Summe der Abweichungsquadrate (SAQ).



Der Y-Wert eines Punktes ist gleich dem Abstand des Punktes von der X-Achse. Mithilfe der Regressionsgeraden bzw. der Regressionsgleichung wird für jeden X-Wert ein Y-Wert vorhergesagt, \hat{y} . Dieser Wert liegt auf der Regressionsgeraden senkrecht oberhalb des X-Wertes auf der X-Achse. Durch die Regressionsgerade wird der Gesamtabstand eines Punktes von der X-Achse (in der Abbildung mit „a“ bezeichnet) in zwei Abschnitte geteilt:

- \hat{y} = der Abschnitt zwischen der X-Achse und der Regressionsgeraden (in der Abbildung „b“)
- e = der Abschnitt zwischen der Regressionsgeraden und dem Punkt (in der Abbildung „c“)

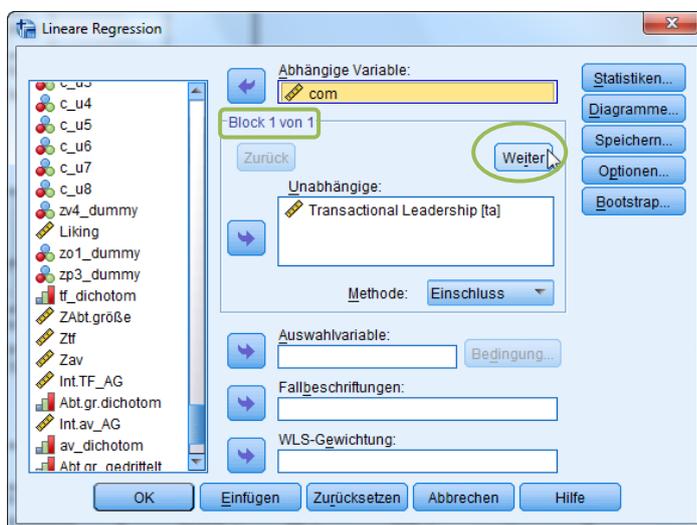
Bei der sog. Varianzzerlegung wird der erst genannte Teil als *erklärter* Varianzanteil bezeichnet, da dieser Varianzanteil der Y-Werte durch die Regressionsgerade vorhergesagt werden kann. Der zweite Abschnitt, also die Abweichungen der Y-Werte von der Geraden, können nicht vorhergesagt/erklärt werden, bleiben sozusagen übrig (*Residuen*).

Der Determinationskoeffizient R^2 setzt den Anteil erklärter Varianz ins Verhältnis zur Gesamtvarianz und ist somit ein Maß für die Güte des Vorhersagemodells. Die Wurzel aus R^2 wird als multipler Korrelationskoeffizient R bezeichnet und kann (analog zum bivariaten Korrelationskoeffizienten r) die Größe des Zusammenhangs im multivariaten Fall beziffern.

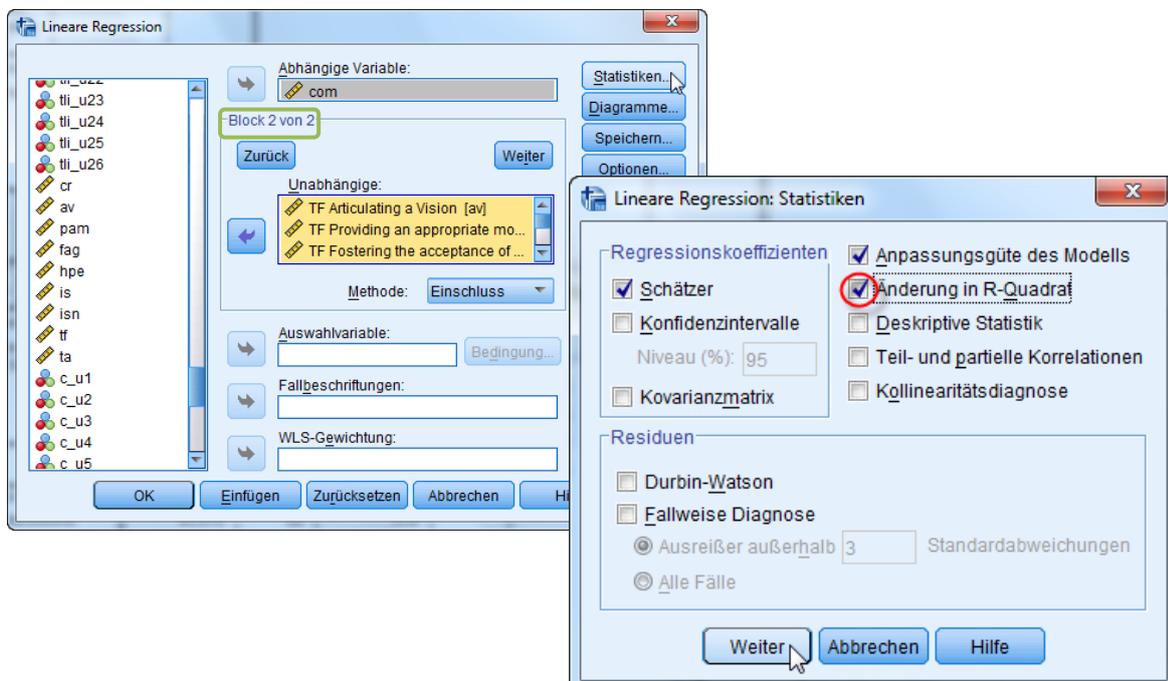
7.3.2 Hierarchische Regression

Bei einer hierarchischen Regressionsanalyse werden die uVs blockweise in die Gleichung aufgenommen. Auf diese Weise kann die Vorhersageleistung einzelner uVs oder Gruppen von uVs ermittelt werden. Sollen beispielsweise der Einfluss transaktionaler Führung („ta“) und der Einfluss der Facetten transformationaler Führung („av“, „pam“, „fag“, „hpe“, „is“, „isn“) auf das Commitment („com“) untersucht werden, bietet es sich an in einem ersten Block nur transaktionale Führung und in einem zweiten Block die Facetten der transformationalen Führung einzufügen.

Unter *Analysieren/Regression/Linear* werden „ta“ als uV und „com“ als aV ausgewählt. Durch klicken auf „Weiter“ wird anschließend der nächste Block definiert.



Hier werden nun die Facetten „av“, „pam“, „fag“, „hpe“, „is“ und „isn“ als unabhängige Variablen eingefügt. Im Anschluss wird unter „Statistiken“ für die Ausgabe „Änderung in R-Quadrat“ angefordert.



Es ergibt sich u. a. folgende Ausgabe:

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Änderungsstatistiken				
					Änderung in R-Quadrat	Änderung in F	df1	df2	Sig. Änderung in F
1	,387 ^a	,150	,138	,72650	,150	12,185	1	69	,001
2	,645 ^b	,416	,351	,63046	,266	4,770	6	63	,000

a. Einflußvariablen : (Konstante), Transactional Leadership

b. Einflußvariablen : (Konstante), Transactional Leadership, TF High performance expectations, TF Providing an appropriate model , TF Individualized Support, TF Intellectual stimulation, TF Fostering the acceptance of group goals , TF Articulating a Vision

ANOVA^a

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	6,431	1	6,431	12,185	,001 ^b
	Nicht standardisierte Residuen	36,418	69	,528		
	Gesamt	42,849	70			
2	Regression	17,808	7	2,544	6,400	,000 ^c
	Nicht standardisierte Residuen	25,041	63	,397		
	Gesamt	42,849	70			

a. Abhängige Variable: com

b. Einflußvariablen : (Konstante), Transactional Leadership

c. Einflußvariablen : (Konstante), Transactional Leadership, TF High performance expectations, TF Providing an appropriate model , TF Individualized Support, TF Intellectual stimulation, TF Fostering the acceptance of group goals , TF Articulating a Vision

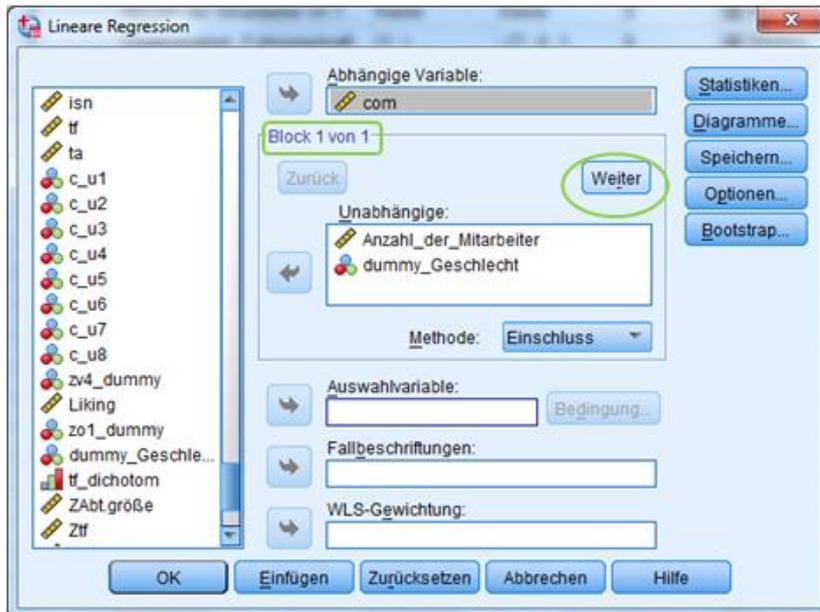
In der Ausgabe werden 2 Modell unterschieden: Modell 1, das nur die uVs des Block 1 enthält, und Modell 2, das beide Blöcke von uVs umfasst. Der Vergleich der beiden Modelle hinsichtlich der erklärten Varianz der aV (R^2) zeigt, dass durch „ta“ alleine nur 15.0 % der Varianz der aV erklärt werden können. Werden zusätzlich die sechs Facetten der transformationalen Führung als uVs miteinbezogen, erhöht sich die Varianzerklärung auf 41.6 %, was einer Änderung um 26.6 % entspricht (s. Spalte „Änderung in R-Quadrat“). Bei der hierarchischen Regression wird sowohl die Signifikanz der beiden einzelnen Modelle betrachtet (s. Ausgabe ANOVA) als auch die Signifikanz der Änderung von Modell 1 zu Modell 2 (s. „Sig. Änderung in F“ in der Tabelle „Modellzusammenfassung“). In diesem Fall sind beide Modelle ebenso wie die Modellverbesserung auf dem 1%-Niveau signifikant.

7.3.3 Kontrollvariablen

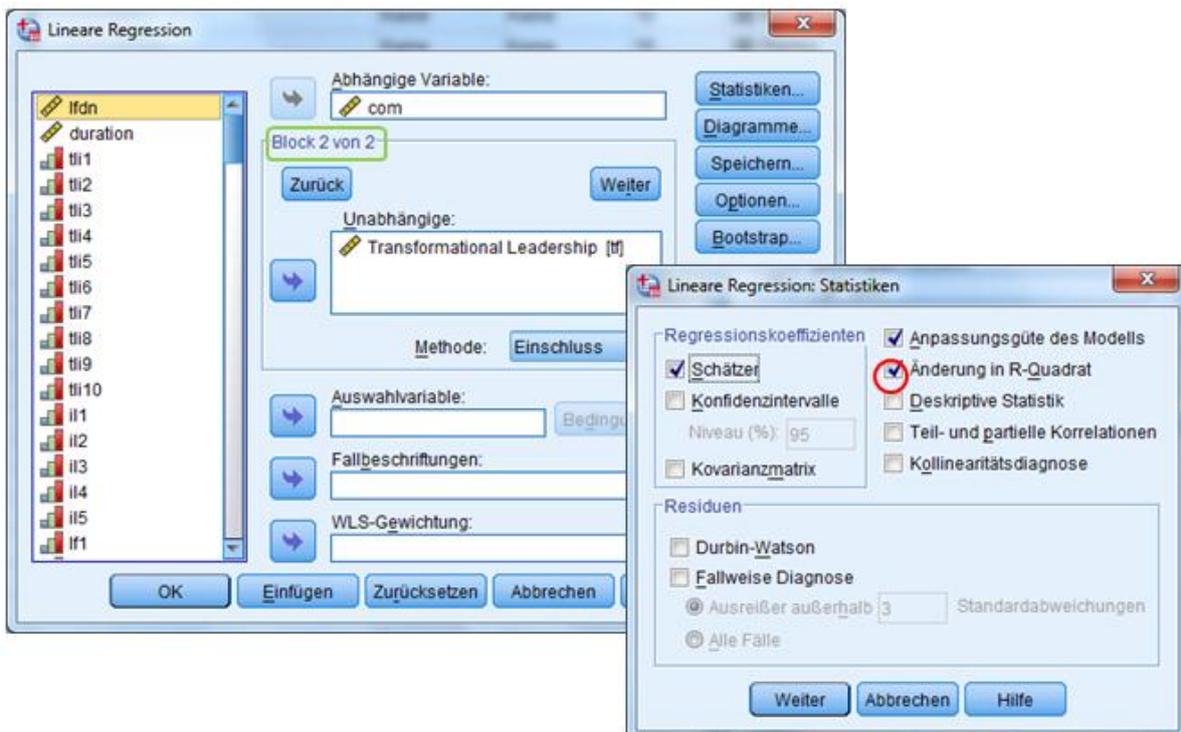
Das Verfahren der hierarchischen Regression kann genutzt werden, um den Einfluss sog. Kontrollvariablen auf die untersuchten Zusammenhänge zu prüfen. Als Kontrollvariable wird eine Variable bezeichnet, die sich in der Forschungsfrage nicht wiederfindet, aber dennoch den Zusammenhang der uV(s) auf die aV beeinflussen könnte. Hierzu ein Beispiel: Es soll der Einfluss von transformationaler Führung auf Commitment untersucht werden. Als Kontrollvariablen

könnten die Anzahl der Mitarbeiter im Team und das Geschlecht der Mitarbeiter eingesetzt werden, da sie den oben genannten Zusammenhang beeinflussen könnten.

Um den Einfluss der Kontrollvariablen zu prüfen, werden in einer hierarchischen Regressionsanalyse zunächst die Kontrollvariablen als erster Block aufgenommen.



Danach werden die eigentlichen uVs in einem zweiten Block hinzugefügt. So lässt sich überprüfen, ob die Kontrollvariablen zusätzliche Varianzerklärung liefern und somit zur Vorhersage der aV geeignet sind.



Folgende Ausgabe entsteht:

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Änderungsstatistiken				
					Änderung in R-Quadrat	Änderung in F	df1	df2	Sig. Änderung in F
1	,214 ^a	,046	,018	,77542	,046	1,632	2	68	,203
2	,637 ^b	,405	,379	,61674	,359	40,495	1	67	,000

a. Einflußvariablen : (Konstante), dummy_Geschlecht, Anzahl der Mitarbeiter im Team

b. Einflußvariablen : (Konstante), dummy_Geschlecht, Anzahl der Mitarbeiter im Team, Transformational Leadership

ANOVA^a

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	1,962	2	,981	1,632	,203 ^b
	Nicht standardisierte Residuen	40,887	68	,601		
	Gesamt	42,849	70			
2	Regression	17,365	3	5,788	15,218	,000 ^c
	Nicht standardisierte Residuen	25,485	67	,380		
	Gesamt	42,849	70			

a. Abhängige Variable: com

b. Einflußvariablen : (Konstante), dummy_Geschlecht, Anzahl der Mitarbeiter im Team

c. Einflußvariablen : (Konstante), dummy_Geschlecht, Anzahl der Mitarbeiter im Team, Transformational Leadership

Es zeigt sich, dass die Kontrollvariablen allein nur 4.6 % der Varianz der aV erklären. Wird „transformationale Führung“ mitaufgenommen, erhöht sich die Varianzaufklärung um 35.9 % auf insgesamt 40.5 %. Die zweite Tabelle zeigt, dass die Kontrollvariablen in Modell 1 nicht signifikant zur Varianzaufklärung beitragen können. Modell 2 hingegen weist eine auf dem 1%-Niveau signifikante Varianzaufklärung auf. Auch die Modellverbesserung ist auf dem 1%-Niveau signifikant. Inhaltlich bedeutet dies, dass die Kontrollvariablen „Anzahl der Mitarbeiter“ und „Geschlecht“ die Forschungsfrage, wie sich transformationale Führung auf Commitment auswirkt, nicht signifikant beeinflussen.

Es ist wichtig die Kontrollvariablen zu berücksichtigen, da andernfalls die Ergebnisse verzerrt werden können. Angenommen die Kontrollvariablen wären signifikant geworden, wären aber nicht berücksichtigt worden, dann könnte es der Fall sein, dass der Einfluss von transformationaler Führung auf Commitment verfälscht worden wäre. Beispielsweise könnte ein größerer Einfluss als tatsächlich vorhanden nachgewiesen worden sein.

7.3.4 Dummy-Variablen

Mit Hilfe der Regressionsanalyse lassen sich nicht nur intervallskalierte uVs untersuchen, sondern auch der Einfluss nominaler oder ordinaler uVs auf eine intervallskalierte aV.

Nominale bzw. ordinale uVs können jedoch nur in Form von sog. Dummy-Variablen berücksichtigt werden und müssen daher ggf. zunächst umkodiert werden. Dummy-Variablen sind binäre Variablen, nehmen also lediglich zwei Ausprägungen (0 und 1) an. Um eine Variable mit k Kategorien abzubilden, werden k-1 Dummy-Variablen benötigt, da eine Kategorie als Referenzkategorie gewählt wird. Ziel ist es, die übrigen Kategorien mit der Referenzkategorie sowie auch untereinander zu vergleichen, um mögliche Unterschiede festzustellen.

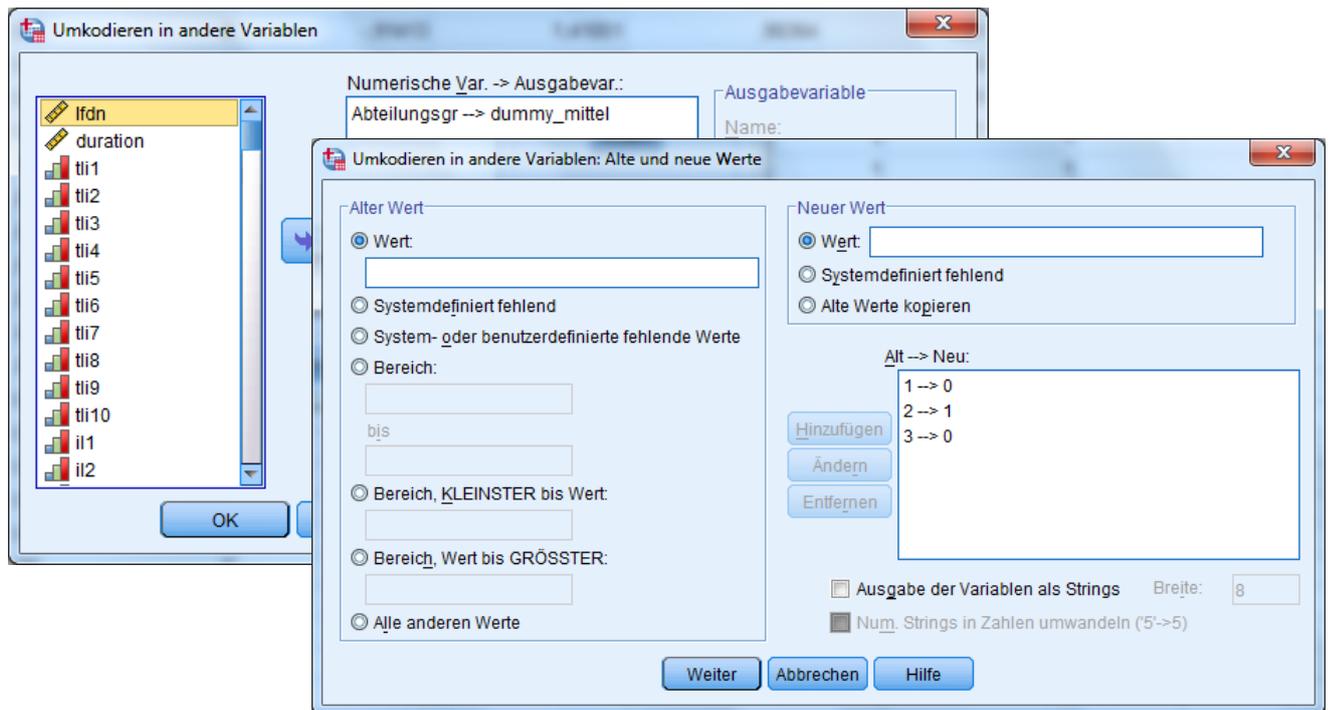
Als veranschaulichendes Beispiel wird der Einfluss der Abteilungsgröße (ordinale uV mit 1=klein, 2=mittel, 3=groß) auf das Commitment (intervallskalierte aV) betrachtet. Wählt man beispielsweise „klein“ als Referenzkategorie, muss man dieser in jeder Dummy-Variablen den Wert 0 zuordnen. Nun müssen also zwei Dummy-Variablen für die Kategorien „mittel“ und „groß“ erstellt werden. Entspricht die Abteilungsgröße der Kategorie „mittel“, so wird für die Variable `dummy_mittel` der Wert 1 vergeben, bei allen anderen Ausprägungen der Wert 0. Bei der Variable `dummy_groß` wird analog verfahren, sprich für „groß“ wird der Wert 1, für alle anderen der Wert 0 vergeben.

<u>y (com)</u>	<u>dummy_mittel</u>	<u>dummy_groß</u>	<u>Abteilungsgr</u>
4.38	0	0	klein
3.50	0	0	klein
4.00	1	0	mittel
3.38	1	0	mittel
2.50	0	1	groß
3.75	0	1	groß

Um also den Einfluss der Abteilungsgröße auf das Commitment zu untersuchen, müssen folgende Schritte vorgenommen werden:

- 1) Auswählen einer Referenzkategorie aus den Kategorien („klein“=1; „mittel“=2; „groß“=3): z.B. „klein“

- 2) Kodieren der Dummy-Variablen „dummy_mittel“ und „dummy_groß“ über *Transformieren/Umkodieren in andere Variablen/Alte und neue Werte*
- Der Referenzkategorie wird auf allen Dummy-Variablen der Wert 0 zugewiesen
 - Jede Dummy-Variable darf nur für eine Kategorie den Wert 1 aufweisen, für alle anderen den Wert 0. Daraus folgt: dummy_mittel (1=0, 2=1, 3=0) und dummy_groß (1=0, 2=0, 3=1).



```

RECODE Abteilungsgr (1=0) (2=1) (3=0) INTO
dummy_mittel.
EXECUTE.
RECODE Abteilungsgr (1=0) (2=0) (3=1) INTO dummy_groß.
EXECUTE.

```

- 3) Durchführen der [Regressionsanalyse](#) über *Analysieren/Regression/Linear* mit „dummy_mittel“ und „dummy_groß“ als uVs und „com“ als aV.

Koeffizienten^a

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Sig.
		Regressionskoeffizient	Standardfehler	Beta		
1	(Konstante)	3,440	,153		22,476	,000
	dummy_mittel	,414	,219	,251	1,894	,062
	dummy_groß	-,008	,219	-,005	-,035	,972

a. Abhängige Variable: com

Es ergibt sich also folgende Regressionsgleichung:

$$com = 3.440 + 0.414 * dummy_mittel$$

Die Regressionskoeffizienten sind hier von besonderer Bedeutung. Die Regressionskonstante ist genau der Mittelwert der aV („com“) in der Referenzkategorie, also in kleinen Abteilungen. Die beiden Regressionskoeffizienten entsprechen jeweils der Differenz aus den Mittelwerten der jeweiligen Kategorie und dem der Referenzkategorie. Also beträgt der Mittelwert für Commitment in mittelgroßen Abteilungen 3.854 und ist um 0.414 Punkte höher als in kleinen. In großen Abteilungen beträgt der Mittelwert 3.432 und ist um 0.008 Punkte geringer als in kleinen Abteilungen. Der geringe Unterschied der Mittelwerte deutet schon darauf hin, dass dieser nicht signifikant von Null verschieden sein kann (Sig.=.972). Daher wird die Variable „dummy_groß“ auch nicht in der Regressionsgleichung berichtet. Neben den Unterschieden zwischen kleinen-mittleren und kleinen-großen Abteilungen, die man direkt an den Regressionskoeffizienten ablesen kann, lässt sich auch der Unterschied mittlere-große Abteilungen ablesen. Dieser ergibt sich aus der Differenz aus den Koeffizienten b_1 und b_2 . Hier: $0.414 - (-0.008) = 0.422$.

Als Ergebnis lässt sich festhalten, dass das Commitment in mittelgroßen Abteilungen am größten ist.

7.4 Mediatoreffekt

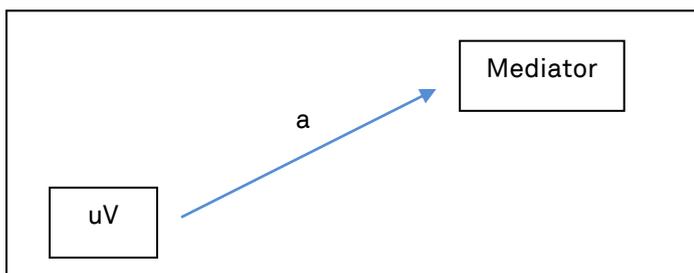
Ein Mediatoreffekt besteht, wenn ein sogenannter Mediator zwischen der unabhängigen und der abhängigen Variable vermittelt, also den Zusammenhang zwischen uV und aV mediiert.

Damit tatsächlich ein Mediator vorliegt, müssen nach Baron und Kenny (Baron, R. M. & Kenny, D. A. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research:

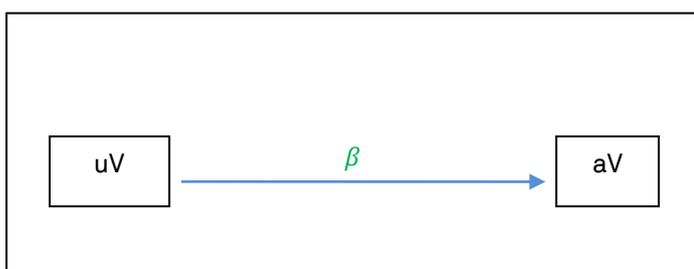
Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.) die folgenden Bedingungen 1.) – 4.) erfüllt sein. Diese lassen sich mit Hilfe mehrerer [Regressionen](#) überprüfen, die hier anhand eines Beispiels veranschaulicht werden. Nachdem die Büroausstattung rundum erneuert wurde, steigt die Leistung der Mitarbeiter einer Abteilung deutlich. Das Management vermutet einen direkten Zusammenhang, nämlich dass die Mitarbeiter mit neuer Ausstattung (uV dichotom -> neue vs. alte Ausstattung) bessere Leistungen (aV) erbringen. Es wäre allerdings auch möglich, dass die neue Ausstattung die Motivation der Mitarbeiter (Mediator) gesteigert hat, da die Mitarbeiter die Aufmerksamkeit der Führungskraft und die neuen Geräte schätzen und dadurch die Leistung steigt.

Folgende vier Bedingungen müssen erfüllt sein, um einen Mediationseffekt nachzuweisen:

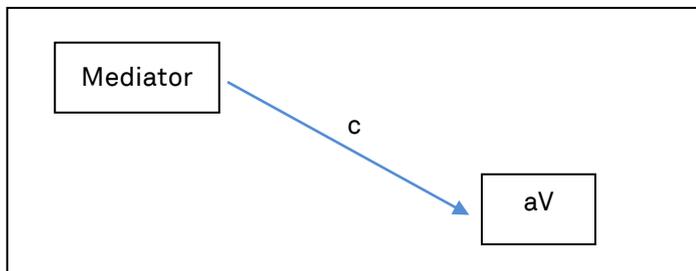
- 1.) Die unabhängige Variable hat einen signifikanten Einfluss (a) auf den mutmaßlichen Mediator. Im Beispiel würde dies bedeuten, dass die Mitarbeiter mit neuer Büroausstattung eine höhere Motivation haben als die, die noch die alte verwenden.



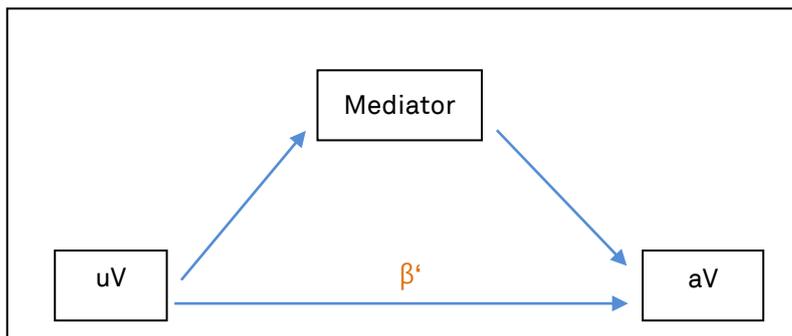
- 2.) Die unabhängige Variable hat einen signifikanten Einfluss (β) auf die abhängige Variable. Wird auch diese Bedingung auf das Beispiel übertragen, müssten die Mitarbeiter mit neuer Büroausstattung bessere Leistungen erbringen.



- 3.) Der Mediator hat einen signifikanten Einfluss (c) auf die abhängige Variable.
Demnach müsste auch die Motivation signifikanten Einfluss auf die Leistung der Mitarbeiter haben.



- 4.) Sind die Bedingungen 1.) - 3.) in den vorausgesagten Richtungen gegeben, so gilt es den Einfluss (β') der unabhängigen Variablen auf die abhängige Variable unter Einbindung des Mediators zu überprüfen. Ist in der gemeinsamen Regression der aV auf die uV *und* den Mediator der Zusammenhang zwischen uV und aV (β') kleiner als der alleinige Einfluss (β) der uV auf die aV, so liegt ein Mediatoreffekt vor ($\beta' < \beta$).



Wenn β' zwar kleiner als β ist, die uV aber auch in der gemeinsamen Regression einen direkten Einfluss auf die aV hat, handelt es sich um einen *partiellen* Mediatoreffekt ($0 < \beta' < \beta$).

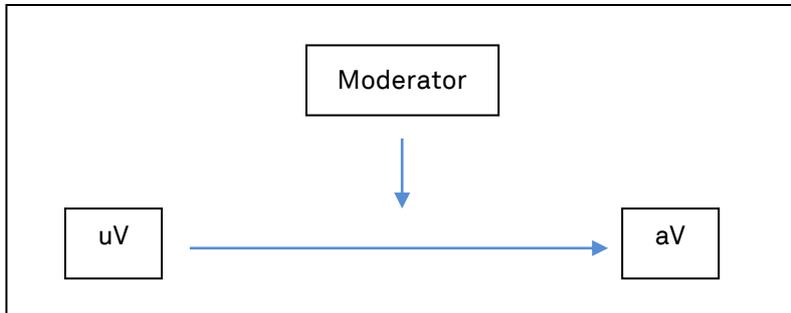
Dies wäre auf das Beispiel bezogen denkbar, wenn sowohl die Motivation für die Leistungssteigerung verantwortlich ist, als auch die neue Ausstattung z.B. durch Effizienzvorteile (z.B. schnellere Rechner, neue Programme) zur Leistungssteigerung beiträgt.

Ein *vollständiger* Mediatoreffekt liegt vor, wenn die uV in der gemeinsamen Regression keinen direkten Effekt mehr auf die aV hat ($\beta' = 0$, Ergebnis nicht signifikant).

In diesem Fall würden die neuen Geräte zwar die Motivation erhöhen, sie würden aber nicht selbst direkt (z.B. durch Effizienzverbesserungen) zur Leistungssteigerung beitragen, sondern ausschließlich mittels der gesteigerten Motivation.

7.5 Moderatoreffekt

Ein Moderatoreffekt besteht, wenn der Einfluss der unabhängigen Variablen auf die abhängige Variable durch die Ausprägung einer dritten Variablen, den Moderator, beeinflusst wird.



Als Beispiel wird der Einfluss transformationaler Führung (TF als uV) auf die Arbeitszufriedenheit (AZ als aV) der Mitarbeiter betrachtet, wobei die Erfahrung der Führungskraft in Jahren (Erfahrung_FK) den Zusammenhang beeinflusst und somit eine Moderatorvariable darstellt. Es soll also folgende Hypothese überprüft werden: Je größer die Erfahrung der Führungskraft, desto größer ist der Einfluss von transformationaler Führung auf die Arbeitszufriedenheit der Mitarbeiter.

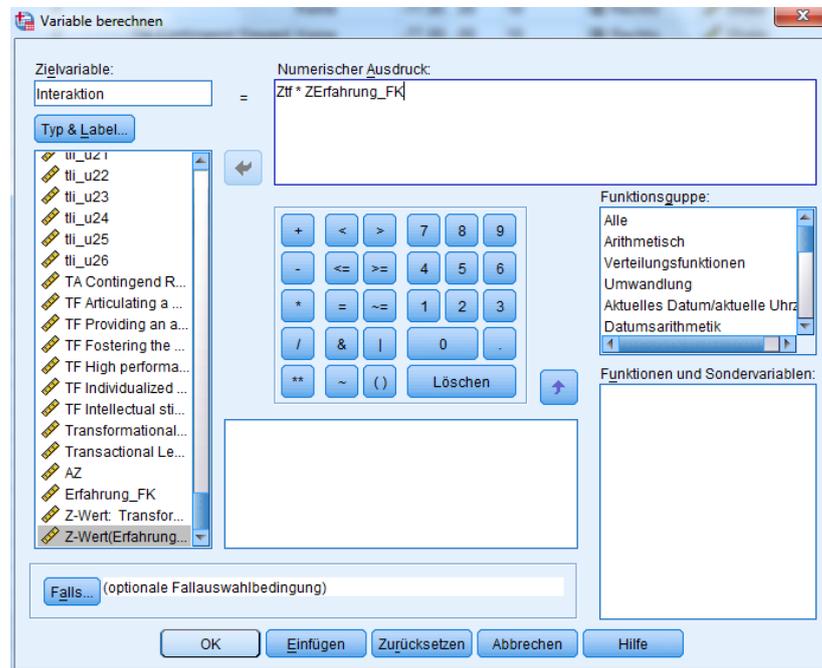
7.5.1 Kontinuierlicher Moderator

Eine Moderationshypothese kann mittels [Regressionsanalyse](#) geprüft werden, wenn der zu überprüfende Moderator kontinuierlich, also mind. intervallskaliert ist. Hierzu muss zunächst in zwei Schritten eine Interaktionsvariable erstellt werden, und zwar als Produkt aus dem z-standardisierten Prädiktor und dem z-standardisierten Moderator.

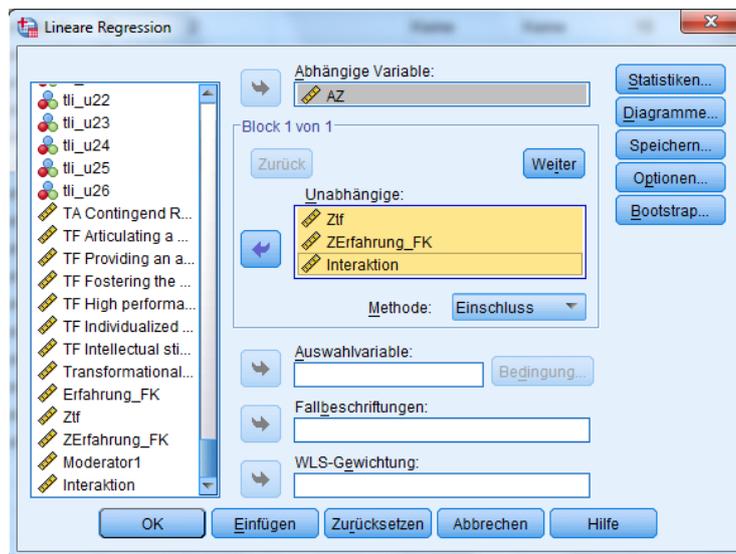
Schritt 1: z-Standardisierung der uV und des mutmaßlichen Moderators (s. [z-Standardisierung](#)).

Schritt 2: Berechnung der Interaktionsvariable

Anschließend kann die Interaktionsvariable als Produkt aus der z-standardisierten uV und dem z-standardisierten mutmaßlichen Moderator berechnet werden (s. [Variable berechnen](#)).



In der anschließenden Regressionsanalyse sind die z-standardisierte uV (im Beispiel: „Ztf“), der z-standardisierte Moderator („ZErfahrung_FK“) und die berechnete Interaktionsvariable („Interaktion“) als uVs auszuwählen.



In der Ausgabe der Regression kann anhand der Signifikanz der einzelnen Koeffizienten geprüft werden, welche Variablen die aV bedeutsam beeinflussen. Ein Moderatoreffekt liegt vor, wenn zusätzlich zum Einfluss der uVs auch das standardisierte Regressionsgewicht der Interaktionsvariablen signifikant von Null verschieden ist.

Im Beispiel weisen die beiden uVs (Ztf und ZErfahrung_FK) einen signifikanten Einfluss auf, die Interaktionsvariable „Interaktion“ jedoch nicht. Daher konnte hier kein Moderationseffekt nachgewiesen werden.

ANOVA^a

Modell	Quadratsumme	df	Mittel der Quadrate	F	Sig.
1 Regression	10,626	3	3,542	4,892	,004 ^a
Nicht standardisierte Residuen	52,135	72	,724		
Gesamt	62,761	75			

a. Einflussvariablen: (Konstante), Interaktion, ZErfahrung_FK, Ztf
b. Abhängige Variable: AZ

Koeffizienten^a

Modell	Nicht standardisierte Koeffizienten	Standardisierte Koeffizienten	T	Sig.
1 (Konstante)	3,495	,098	35,591	,000
Ztf	,249	,099	2,517	,014
ZErfahrung_FK	-,297	,104	-2,864	,005
Interaktion	-,063	,092	-,684	,496

a. Abhängige Variable: AZ

In den empirischen Sozialwissenschaften kommt es sehr häufig vor, dass ein Moderationseffekt in der Regression nicht signifikant wird, vor allem wenn die beiden uVs miteinander korrelieren. Gelegentlich wird dann versucht, den Moderationseffekt dennoch zu beschreiben und/oder nachzuweisen, in dem der Moderator dichotomisiert wird (s. [Dichotomer Moderator](#)). Streng genommen ist hiermit ein unzulässiger Informationsverlust verbunden, der sich in der Scientific Community jedoch als tragbar erwiesen hat.

7.5.2 Dichotomer Moderator

Dichotome (auch binäre) Variablen sind solche, die nur zwei Ausprägungen besitzen. Sie können entweder natürlich gegeben sein oder erzeugt werden. Ein Beispiel für eine von vornherein dichotome Variable wäre das Geschlecht (männlich/weiblich). Es kann aber auch jede beliebige Variable, bspw. die Erfahrung (gering/viel) dichotomisiert werden.

Liegt ein dichotomer mutmaßlicher Moderator vor, wird der Moderatoreffekt geprüft, indem getestet wird, ob sich die Korrelation von uV und aV für die beiden Ausprägungen des Moderators unterscheidet.

Hierzu sind in SPSS die folgenden Schritte notwendig:

Schritt 1: Visuelles Klassieren

Ist der mutmaßliche Moderator kontinuierlich, so muss zuerst eine neue, dichotome Variable erstellt werden (z.B. mittels Mediansplit, s. [visuelles Klassieren](#)). Liegt eine dichotome Variable vor, kann dieser Schritt übersprungen werden.

Schritt 2: Datei aufteilen und Werte berechnen

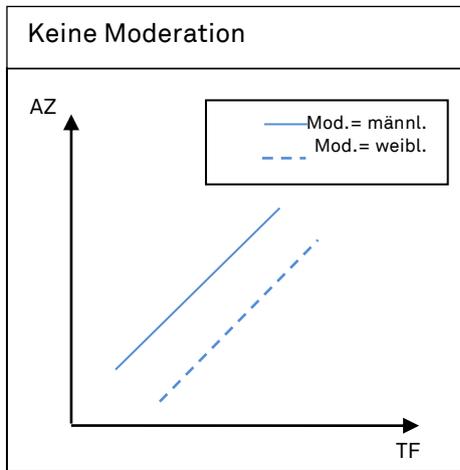
Im zweiten Schritt werden die Daten auf Basis des mutmaßlichen Moderators [aufgeteilt](#), um die [Korrelation](#) zwischen uV und aV getrennt für die beiden Gruppen zu berechnen.

Schritt 3: Vergleich der Korrelationskoeffizienten und Signifikanzprüfung

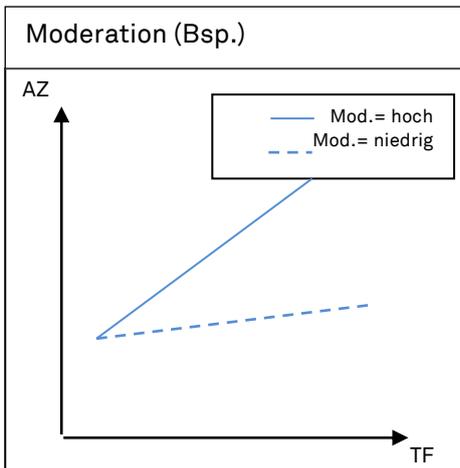
Im dritten Schritt wird geprüft, ob sich die beiden Korrelationskoeffizienten signifikant voneinander unterscheiden (s. [Vergleich von Korrelationskoeffizienten](#)).

Um die Wirkung des Moderatoreffekts auf den Zusammenhang von uV und aV zu beschreiben, werden üblicherweise Liniendiagramme genutzt, bei denen die uV auf der X-Achse, die aV auf der Y-Achse dargestellt wird und für die beiden Ausprägungen des Moderators (z. B. niedrig vs. hoch) einzelne Linien abgetragen werden (s. Beispiele).

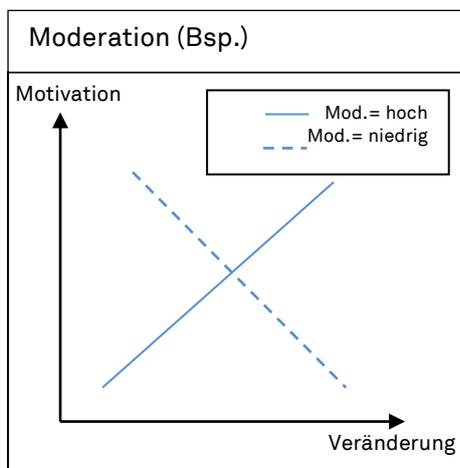
Graphische Darstellung möglicher Moderatoreffekte:



Untersucht wird der Einfluss von transformationaler Führung (uV: TF) auf die Arbeitszufriedenheit (aV: AZ) in Abhängigkeit davon, ob die Führungskraft (FK) männlich oder weiblich ist (Mod: Geschl. d. FK). Es zeigt sich, dass TF die AZ in gleichem Maße fördert, egal ob die FK männlich oder weiblich ist (parallele Geraden). Daher gibt es keinen Moderationseffekt.



Untersucht wird der Einfluss von transformationaler Führung (uV: TF) auf die Arbeitszufriedenheit (aV: AZ) in Abhängigkeit davon, ob Mitarbeiter viel/wenig Kontakt (Mod.) zur Führungskraft (FK) haben. Die Grafik zeigt, dass TF die AZ stark fördert, wenn die Mitarbeiter viel Kontakt mit der FK haben. Haben die Mitarbeiter jedoch wenig Kontakt zur FK, hat TF kaum einen Einfluss auf AZ. In beiden Fällen ist TF für AZ förderlich (gleichgerichtete Wirkung der uV), jedoch unterschiedlich stark.



Die unabhängige Variable sei der Grad der Veränderung, die abhängige die Motivation der Mitarbeiter. Moderator ist die Flexibilität der Mitarbeiter. Bei sehr flexiblen Mitarbeitern steigt die Motivation, je höher der Grad der Veränderung ist. Bei unflexiblen Mitarbeitern führt ein hoher Grad an Veränderung zu einer stark sinkenden Motivation. Das heißt, die Wirkung der uV ist gegenläufig (mal positiv, mal negativ) je nach Ausprägung des Moderators.

7.5.2.1 Exkurs Vergleich von Korrelationskoeffizienten

Um zu überprüfen, ob sich zwei Korrelationskoeffizienten signifikant voneinander unterscheiden, müssen sie zunächst der sog. Fisher Z-Transformation unterzogen werden. Diese Transformation dient dazu, [Pearsons r](#) in eine normalverteilte Variable Z umzuwandeln. Fisher's Z-Wert berechnet sich für den Korrelationskoeffizienten r zu

$$Z = 0.5 * \ln \frac{1+r}{1-r}$$

Wurden beide Korrelationskoeffizienten transformiert, kann anhand der Testgröße z ermittelt werden, ob sie sich signifikant unterscheiden.

Zur Durchführung von Fishers z-Transformation sowie zur Signifikanzprüfung gibt es eine nützliche [Exceldatei](#)¹, in die lediglich die Pearson-Korrelationskoeffizienten, der Stichprobenumfang und das gewünschte Signifikanzniveau einzutragen sind. Die Fisher Z-Transformation, der Vergleich der Korrelationen und die Ergebnisausgabe erfolgen automatisch.

Auf das zu Beginn erwähnte Beispiel angewendet, würde das eine Aufteilung der Daten nach geringer und großer Erfahrung bedeuten. Die Moderationshypothese sei nun „Bei Führungskräften mit großer Erfahrung ist der Einfluss transformationaler Führung auf die Arbeitszufriedenheit größer, als bei Führungskräften mit geringer Erfahrung.“

Nachdem die oben genannten Schritte durchgeführt wurden, können die Variablennamen, der Stichprobenumfang sowie die errechneten Korrelationen in die Tabelle eingetragen werden. Auf der rechten Seite wird nun automatisch die Nullhypothese formuliert, die besagt, dass sich die Korrelationen der Variablen „TF“ und „AZ“ für die zwei unterschiedlichen Stichproben nicht signifikant voneinander unterscheiden.

¹ Hinweis: Der Lehrstuhl für Personalentwicklung und Veränderungsmanagement übernimmt keine Verantwortung für den Inhalt und die Richtigkeit des Links und macht sich dessen Inhalt auch nicht zu Eigen.

Microsoft Excel - corrcomparer1_0_Bsp - Microsoft Excel

Start Einfügen Seitenlayout Formeln Daten Überprüfen Ansicht Acrobat

Normal Seitenlayout Benutzerdef. Ansichten Arbeitsmappenansichten

Lineal Bearbeitungseiste Gitternetzlinien Überschriften Statusleiste Einblenden/Ausblenden

Zoom 100% Zoommodus: Auswahl Fenster

Neues Fenster Teilen Alle anordnen Ausblenden Aufgabenbereich speichern Fenster wechseln Makros

E32

6	Allgemeines				
7	Signifikanzniveau				5%
9	Stichprobe 1				
10	Variable A			geringe Erfahrung_FK	
11	Variable B			TF	
12	Korrelationskoeffizient r_1			AZ	0,517
13	Stichprobenumfang n_1				41
15	Stichprobe 2				
16	Variable A			große Erfahrung_FK	
17	Variable B			TF	
18	Korrelationskoeffizient r_2			AZ	0,154
19	Stichprobenumfang n_2				35
21	Rechnung:				
23	z-Transformation (Fisher)				
25	geringe Erfahrung_FK:	$Z_e(r_1) =$	0,5722		
26	große Erfahrung_FK:	$Z_e(r_2) =$	0,1552		
27	Differenz		0,4170		
28	Sigma	$\sigma =$	0,24		
29	Standardisierte Testgröße	$z =$	1,74		
30	Wahrscheinlichkeit (aus Normalverteilung):	$P =$	8,2%		
32	Test:				
33	vorgegebenes				
34	Signifikanzniveau:	$\alpha =$	5%	(Wahrscheinlichkeit für Fehler 1. Art)	
35	daraus berechneter kritischer Wert der Testgröße:			$Z_{\text{krit}} =$	1,96
36		Korrelationskoeffizienten	Stichprobenumfänge	Fisher's Z-Werte	Testgröße z
37	geringe Erfahrung_FK	0,52	41	0,57	1,74
38	große Erfahrung_FK	0,15	35	0,16	8,2%
39	Interpretation:				
40	Testgröße ist kleiner als der kritische Wert				
41	Nullhypothese annehmen. Der Unterschied zwischen der Stichprobe "geringe Erfahrung_FK" und der Stichprobe "große Erfahrung_FK" ist NICHT signifikant (Signifikanzniveau $\alpha = 5,0\%$!)				
44	CorrComparer 1.0 - ein Tool von markenkunde.de				

Nullhypothese H_0
Ihre Nullhypothese H_0 lautet damit:
Der Korrelationskoeffizient zwischen den beiden Variablen "TF" und "AZ" der Stichprobe "geringe Erfahrung_FK" unterscheidet sich nicht vom Korrelationskoeffizienten der beiden Variablen "TF" u. "AZ" der Stichprobe "große Erfahrung_FK".

Das Ergebnis $P = 8,2\%$ bedeutet:
Bei Gültigkeit der Nullhypothese können (rein zufällig) mit einer Wahrscheinlichkeit von 8,2 % Unterschiede der Korrelationskoeffizienten auftreten, die größer sind als die Unterschiede der beobachteten Werte (0,517 bei "geringe Erfahrung_FK" und 0,154 bei "große Erfahrung_FK"). Anders ausgedrückt: Wenn die Nullhypothese richtig ist, aber aufgrund der beobachteten Unterschiede doch abgelehnt wird, dann macht man mit der Wahrscheinlichkeit $P = 8,2\%$ einen "Fehler 1. Art" ("α - Fehler": richtige Hypothese wird abgelehnt).

Da die errechnete Testgröße $z = 1.74$ kleiner als der kritische Wert $z_{\text{krit}} = 1.96$ ist, muss die Nullhypothese angenommen werden. Mit diesem Ergebnis geht die Ablehnung der Moderationshypothese einher, da sich die Korrelationen nicht signifikant voneinander unterscheiden. Demzufolge macht es für die Wirkung transformationaler Führung auf die Arbeitszufriedenheit keinen Unterschied, ob die Erfahrung der Führungskraft groß oder klein ist.