

DATENAUSWERTUNG MIT SPSS

Claudia Krüger, Lars Borgmann, Tobias Antonik, Ann-Kathrin Meyer, Mathias Diebig,
Alexander Kay, Yannick Schwinger

Lehrstuhl für Personalentwicklung und Veränderungsmanagement

Zentrum für Hochschulbildung

TU Dortmund

Version 1.6 (Sommersemester 2020)

Inhaltsverzeichnis

1 Grundlegendes zur Software SPSS	5
1.1 Ziel und Beschaffung	5
1.2 Programmstruktur	5
1.2.1 Exkurs Syntax-Editor	5
1.3 Unterstützung	6
2 Dateneingabe	7
2.1 Variablenansicht	7
2.1.1 Variablenname und Beschriftung	7
2.1.2 Typ	8
2.1.3 Breite und Dezimalstellen	8
2.1.4 Werte	8
2.1.5 Fehlend	9
2.1.6 Messniveau	9
2.1.6.1 Exkurs Messniveau	9
2.2 Datenansicht	10
2.3 Datenimport aus Excel	11
2.4 Daten zusammenfügen: Variablen hinzufügen	13
3 Datensichtung	16
3.1 Häufigkeiten	17
3.2 Deskriptive Statistik	18
3.2.1 Exkurs Maße der zentralen Tendenz (Lagemaße)	20
3.2.2 Exkurs Streuungsmaße	21
3.3 Diagramme	21
3.3.1 Kreisdiagramm	21
3.3.2 Säulendiagramm	22
3.3.3 Histogramm	23
3.3.4 Boxplot	23
3.4 Überprüfung auf Normalverteilung	24
3.4.1 Exkurs Normalverteilung	27

3.5 Kreuztabelle.....	28
4 Datenstrukturierung	30
4.1 Fälle auswählen	30
4.2 Datei aufteilen	34
5 Datenaufbereitung.....	38
5.1 Umcodieren.....	38
5.2 Visuelle Klassierung.....	39
5.3 z-Standardisierung.....	40
6.1 Reliabilitätsanalyse.....	42
6.1.1 Exkurs Reliabilität.....	43
6.2 Variable berechnen	44
7 Prüfung von Zusammenhangshypothesen.....	46
7.1 Bivariate Korrelation	46
7.1.1 Exkurs Pearson-Korrelation.....	49
7.1.2 Streudiagramm.....	50
7.2 Partielle Korrelation	51
7.3 Lineare Regression	52
7.3.1 Exkurs Multiple Regression	54
7.3.2 Hierarchische Regression	57
7.3.3 Kontrollvariablen.....	59
7.3.4 Dummy-Variablen	61
7.4 Mediatoreffekt	64
7.5 Moderatoreffekt.....	65
7.5.2.1 Exkurs Vergleich von Korrelationskoeffizienten.....	72
7.6 Exkurs: SPSS-Makro PROCESS.....	74
7.6.1 Installationsanleitung.....	74
7.6.2 Mediation und Moderation mit PROCESS.....	75
7.6.2.1 Mediation anhand eines Beispiels	76
7.6.2.2 Moderation anhand eines Beispiels	79
8 FAKTORENANALYSE.....	84

8.1 Grundidee der Faktorenanalyse.....	84
8.2 Unterschiedliche Verfahren der Faktorenanalyse.....	84
8.3 Ablauf einer exploratorischen Faktorenanalyse (EFA)	85
8.3.1 Faktorisierbarkeit	85
8.3.2. Extraktion der Faktoren	87
8.3.3 Extraktionsmethoden.....	87
8.3.4 Faktoranzahl.....	88
8.4 Rotation der Faktoren	88
8.4.1 Rotationsverfahren	89
8.5. Interpretation der Faktoren	89
8.6. Faktorinterpretation.....	89
8.7 Praxisbeispiel: Exploratorische Faktorenanalyse	90

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 (<https://www.itmc.tu-dortmund.de/cms/de/dienste/hard-software/index.html>, mit UniAccount einloggen)

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.

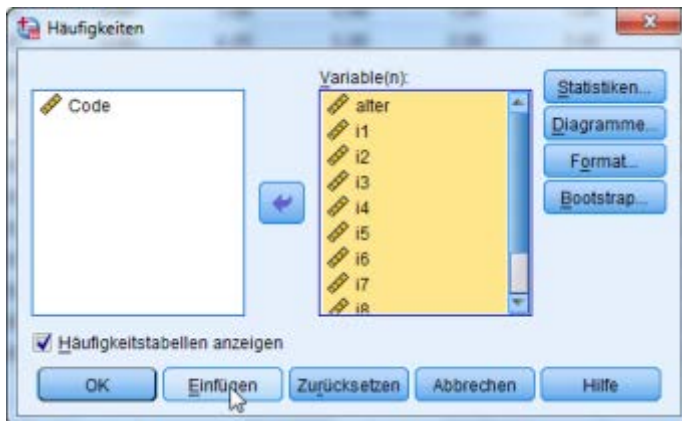
Jede dieser Ebenen zeigt eine typische Menüleiste am oberen Rand. Unter *Bearbeiten/Optionen* können zahlreiche Anpassungen vorgenommen werden, u.a. zur Darstellung und Nutzung der verschiedenen Programmebenen.

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/Befehlssyntaxreferenz*) 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 <https://www.ibm.com/products/spss-statistics/support> 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/beratung.html>).

2 DATENEINGABE

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

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

2.1.1 Variablenname und Beschriftung

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 *Beschriftung* 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 Beschriftung 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. Werte). Für freie Texteinträge bietet sich der Typ „Zeichenfolge“ an.

2.1.3 Breite und Dezimalstellen

Die *Breite* 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 Werte

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 sowie unter „Beschriftung“ die inhaltliche Bedeutung eingetragen und abschließend auf „Hinzufügen“ geklickt.

The screenshot shows the IBM SPSS Statistics Dateneditor interface. The main window displays a list of variables with columns for Name, Typ, Breite, Dezimal..., Beschriftung, Werte, Fehlend, Spalten, Ausrichtung, Messniveau, and Rolle. The 'Variablenansicht' tab is active. A dialog box titled 'Wertbeschriftungen' is open, showing the 'Wert' field set to 5,00 and the 'Beschriftung' field set to 'ich stimme völlig zu'. The dialog also contains a list of existing value labels: 1,00 = "stimme gar nicht zu", 2,00 = "stimme eher nicht zu", 3,00 = "stimme teilweise zu", and 4,00 = "stimme eher zu". Buttons for 'Hinzufügen', 'Ändern', 'Entfernen', 'OK', 'Abbrechen', and 'Hilfe' are visible.

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

2.1.5 Fehlend

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

The screenshot shows the IBM SPSS Statistics Dateneditor interface. The main window displays a list of variables with columns for Name, Typ, Breite, Dezimal..., Beschriftung, Werte, Fehlend, Spalten, Ausrichtung, Messniveau, and Rolle. The 'Fehlende Werte' dialog box is open, showing options for handling missing values. The 'Einzelne fehlende Werte' option is selected, and the value '-77' is entered in the input field.

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

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 „Metrisch“ (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. Variablen. Je nach 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, wie 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 ordinalskalierbares 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 gleichgroß ist wie der zwischen Platz 3 und Platz 4.

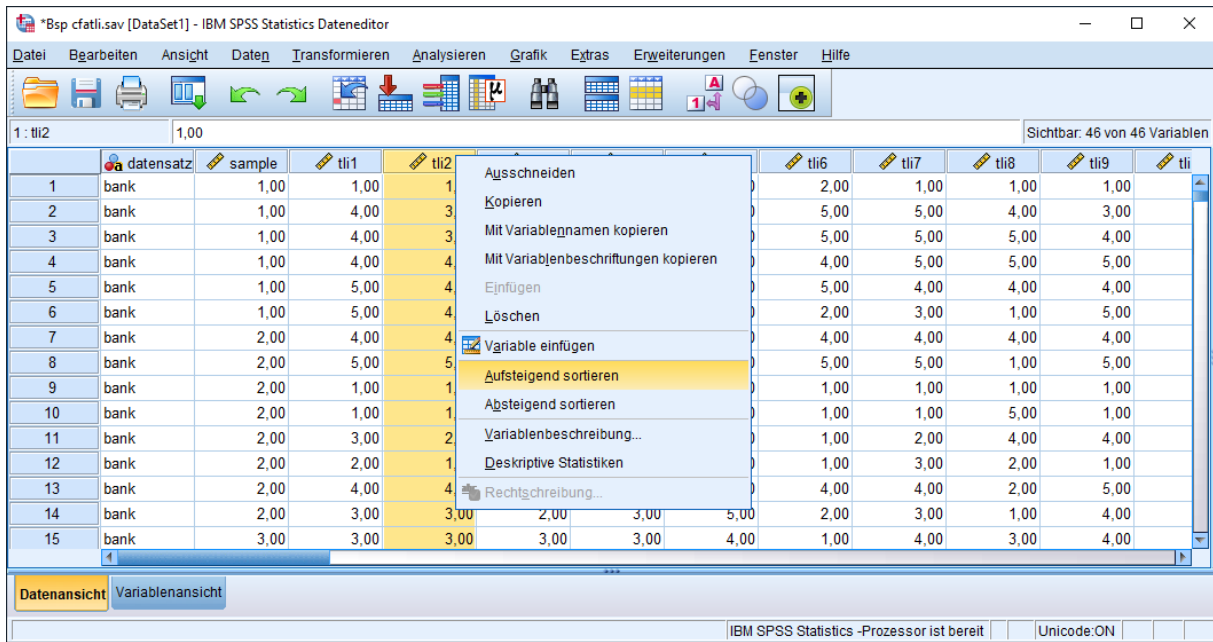
Intervallskala: Die Reihenfolge der Merkmalswerte ist festgelegt, und die Größe des Abstandes zwischen zwei Werten lässt sich sachlich begründen. Die Ausprägungen dieses Skalenniveaus lassen sich quantitativ mittels Zahlen darstellen. Es ist für die metrische Skala charakteristisch, dass die Abstände 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 3 (manchmal) beantwortet, das zweite TLI-Item (tli2) wurde mit einer 4 (oft) beantwortet.

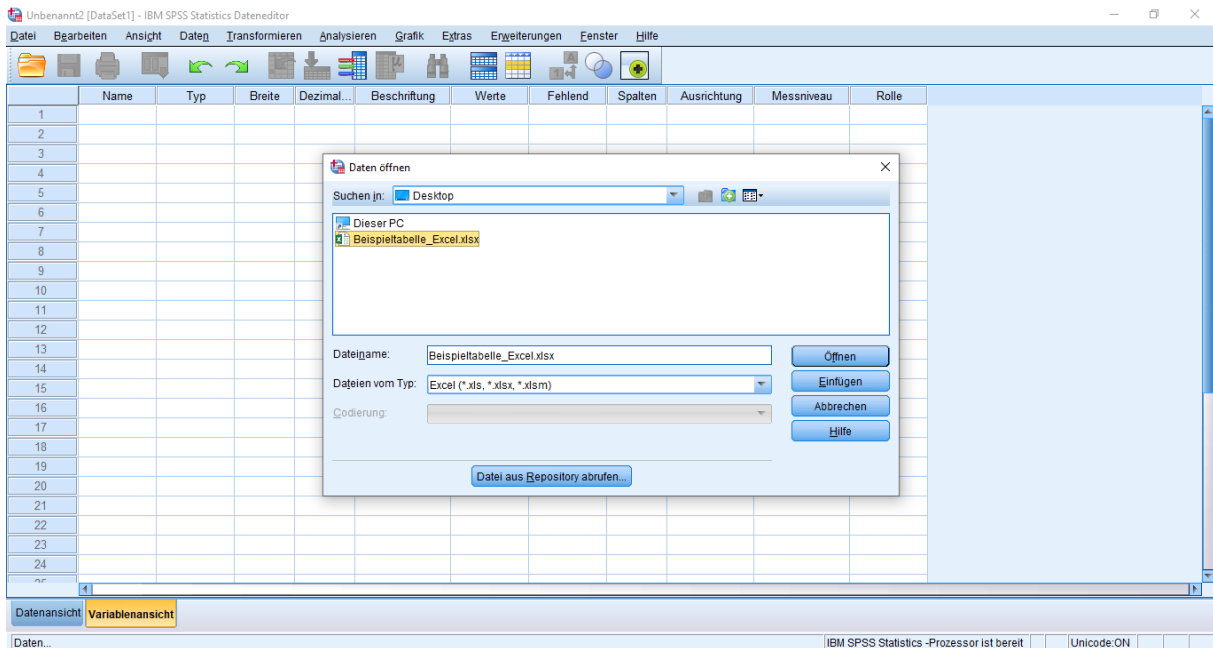
	datensatz	sample	tli1	tli2	tli3	tli4	tli5	tli6	tli7	tli8	tli9	tli10	tli11	tli12
307	consulting	1,00	4,00	5,00	4,00	3,00	3,00	3,00	2,00	1,00	4,00			
308	consulting	1,00	2,00	1,00	2,00	1,00	1,00	2,00	4,00	3,00	4,00			
309	consulting	1,00	5,00	2,00	4,00	4,00	4,00	3,00	3,00	2,00	4,00			
310	consulting	1,00	3,00	4,00	4,00	4,00	5,00	3,00	3,00	4,00	4,00			
311	consulting	1,00	3,00	3,00	2,00	4,00	4,00	2,00	2,00	1,00	5,00			
312	consulting	1,00	3,00	2,00	1,00	1,00	2,00	3,00	3,00	1,00	2,00			
313	consulting	1,00	3,00	2,00	2,00	2,00	3,00	3,00	4,00	4,00	4,00			
314	consulting	1,00	3,00	3,00	3,00	2,00	3,00	3,00	3,00	4,00	3,00			
315	consulting	1,00	3,00	3,00	4,00	5,00	5,00	5,00	4,00	2,00	3,00			
316	consulting	1,00	3,00	4,00	3,00	2,00	2,00	4,00	3,00	2,00	2,00			
317	consulting	1,00	2,00	2,00	3,00	2,00	4,00	3,00	3,00	3,00	4,00			
318	consulting	1,00	3,00	4,00	3,00	2,00	3,00	2,00	2,00	3,00	2,00			
319	consulting	1,00	3,00	3,00	4,00	4,00	4,00	4,00	4,00	2,00	3,00			
320	consulting	1,00	3,00	4,00	4,00	4,00	4,00	3,00	4,00	2,00	4,00			
321	consulting	1,00	4,00	4,00	4,00	3,00	5,00	5,00	3,00	2,00	4,00			

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.



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/Daten* als Dateityp Excel ausgewählt werden.



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

Excel-Datei lesen

C:\Users\IPVM\Desktop\Beispieltable_Excel.xlsx

Arbeitsblatt: Tabelle1 [A1:F2]

Bereich:

Variablennamen aus erster Datenzeile lesen

Prozentsatz der Werte, die den Datentyp festlegen


Ausgeblendete Zeilen und Spalten ignorieren

Führende Leerzeichen aus Zeichenfolgewert entfernen

Nachfolgende Leerzeichen aus Zeichenfolgewert entfernen

Vorschau

	a	b	c	d	f
1	12	13	14	15	16

 Der endgültige Datentyp basiert auf allen Daten und kann vom Datentyp der Vorschau abweichen, die auf den ersten 200 Datenzeilen basiert. Die Vorschau zeigt nur die ersten 500 Spalten an.

2.4 Daten zusammenfügen: Variablen hinzufügen

Ausgangslage

Verschiedene Variablen (v1, v2 versus v3, v4) derselben Personen wurden in verschiedenen Datendateien gesammelt, wie beispielsweise bei der Datenerhebung zu unterschiedlichen Zeitpunkten. Die Zusammenführung der Daten ist erforderlich, um Zusammenhänge zwischen den Variablengruppen zu ermitteln.

	Code	v1	v2	var	var	var	var	var
1	1	1	2					
2	2	1	2					
3	3	1	2					
4	4	1	2					
5	5	1	2					
6	6	1	2					
7	7	1	2					
8	8	1	2					
9	9	1	2					
10	10	1	2					

	Code	v3	v4
1	1	3	4
2	2	3	4
3	3	3	4
4	4	3	4
5	5	3	4
6	6	3	4
7	7	3	4
8	8	3	4
9	9	3	4
10	10	3	4

Datenvorbereitung

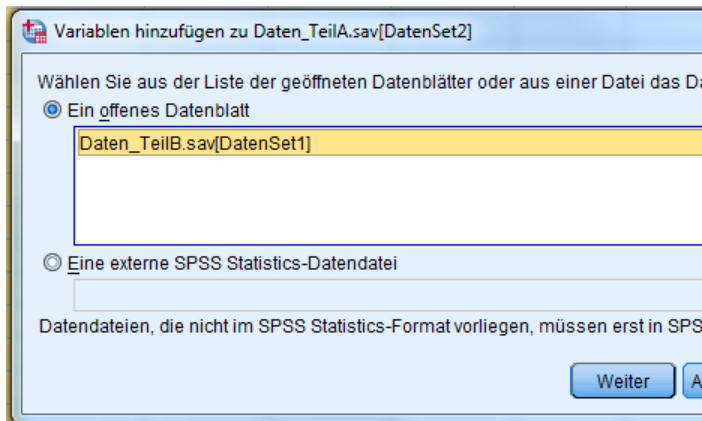
Zunächst müssen alle Teil-Datendateien geöffnet werden, sodass im Anschluss in allen Teil-Datendateien die Fälle aufsteigend nach ihrem jeweiligen Code sortiert werden können. Abschließend müssen alle Teil-Datendateien gespeichert werden.

	Code	v1	v2	var
1				
2				
3				
4				
5				
6				
7				
8				
9				
10	10	1	2	
11				
12				

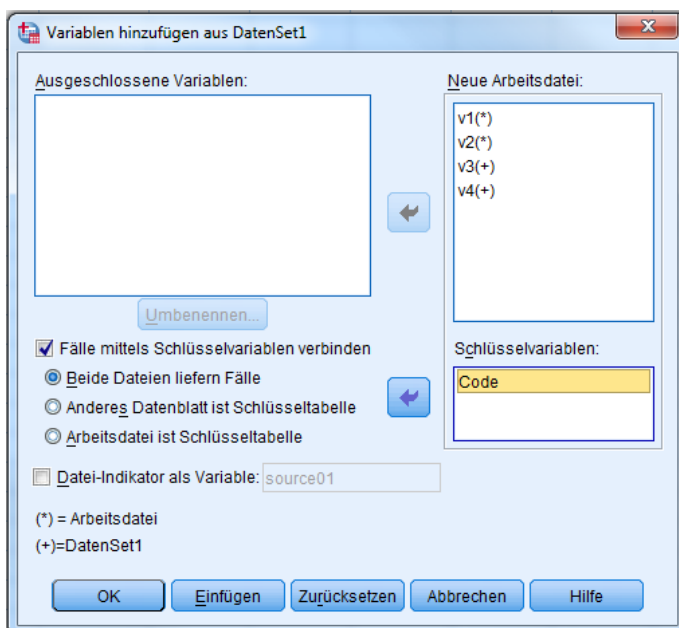
	Code	v3	v4	var
1				
2				
3				
4				
5				
6				
7				
8				
9				
10	10	3	4	
11				
12				

Daten zusammenfügen

Zur Zusammenführung der Daten wird in der Teil-Datendatei A unter *Daten/ Dateien zusammenfügen/ Variablen hinzufügen* das für die Zusammenführung der Daten relevante Dialogfenster geöffnet. Anschließend muss die hinzuzufügende Teil-Datendatei B ausgewählt werden.



Im nächsten Schritt wird im nachfolgenden Dialogfenster die Option, *Fälle mittels Schlüsselvariable verbinden*, ausgewählt. Als Schlüsselvariable wird hierbei die Code-Variable verwendet (nach rechts klicken). Nun kann kontrolliert werden, ob alle gewünschten Variablen in die neue Arbeitsdatei übernommen werden. Abschließend werden die gewünschten Einstellungen mit *OK* gewählt und bestätigt, dass alle Dateien nach der Schlüsselvariable aufsteigend sortiert sind.



Zielzustand und Datensicherung

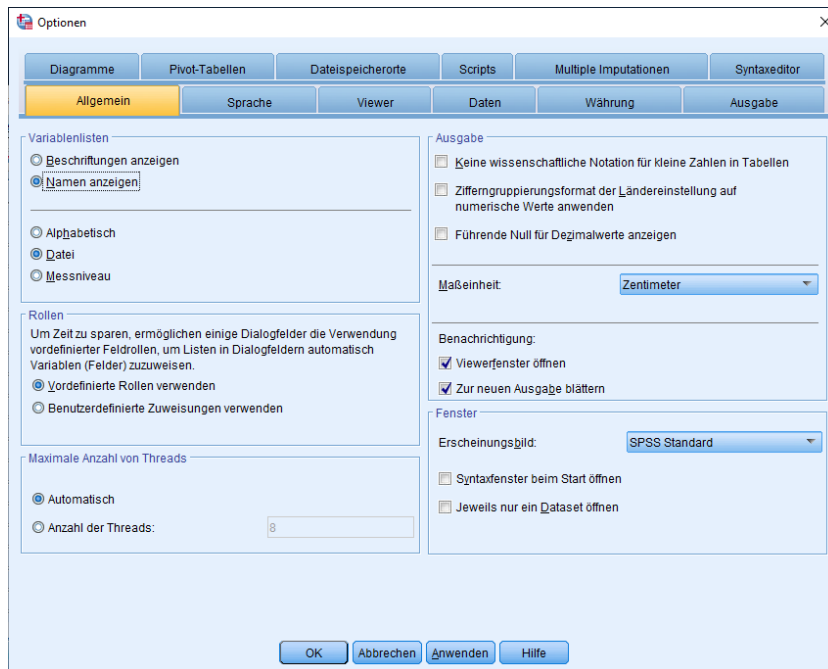
SPSS ordnet nun automatisch die Werte der beiden Datendateien über die Codevariable zu. Die gemeinsame Datendatei kann jetzt unter neuem Namen abgespeichert und verwendet werden.

	Code	v1	v2	v3	v4
1	1	1	2	3	4
2	2	1	2	3	4
3	3	1	2	3	4
4	4	1	2	3	4
5	5	1	2	3	4
6	6	1	2	3	4
7	7	1	2	3	4
8	8	1	2	3	4
9	9	1	2	3	4
10	10	1	2	3	4

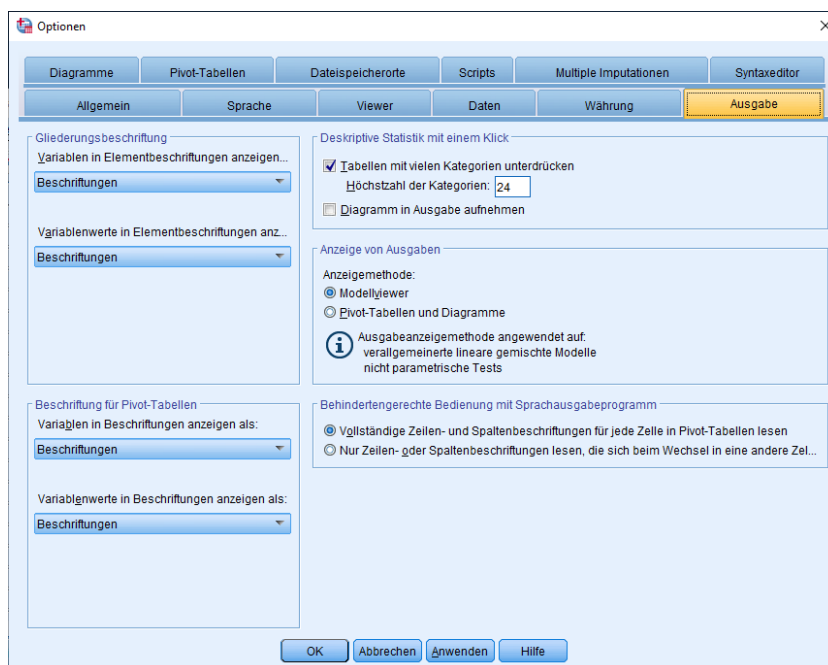
Achtung: SPSS unterscheidet bei der Zuordnung der Codevariable zwischen Groß- und Kleinbuchstaben, d.h. man sollte vorher kontrollieren, ob die Groß- und Kleinschreibweise der Codes in beiden Datendateien identisch ist.

3 DATENSICHTUNG

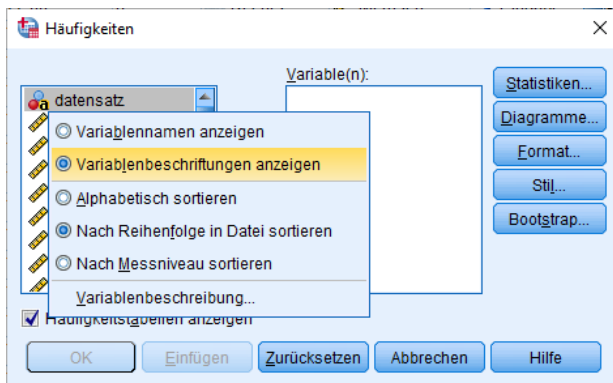
Um die Übersichtlichkeit bei der Ausführung der statistischen Analysen zu verbessern, kann man Einstellungen bezüglich der Variablennamen tätigen. Unter *Bearbeiten/Optionen* gibt es die Möglichkeit nur die Namen (und nicht die Beschriftung) der Variablen in den Dialogfenstern anzeigen zu lassen.



Des Weiteren kann man die Beschriftung der Ausgabe unter *Ausgabe/Variablen in Elementbeschriftungen anzeigen als* anpassen. Hier kann beispielsweise die Auswahl *Namen und Beschriftungen* getroffen werden.



Optional besteht die Möglichkeit in jedem Analyse-Dialogfenster mit einfachem Rechtsklick auf die Variablen zwischen Namen oder Beschriftungen umzuschalten.

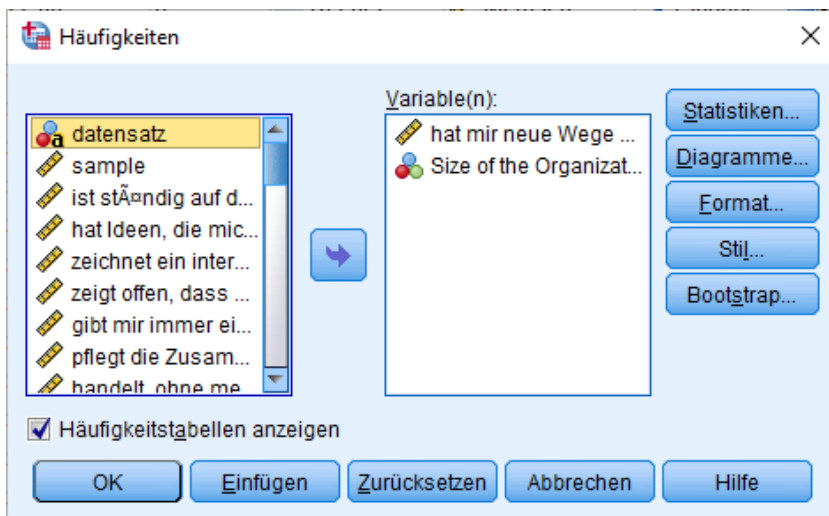


3.1 Häufigkeiten

Einer der ersten Schritte beim Verwenden von neuen Datensätzen ist die Daten zu beschreiben, um einen Überblick über die Daten zu erhalten und mögliche Unregelmäßigkeiten erkennen zu können.

Hierfür kann zum einen der Befehl *Häufigkeiten* genutzt werden, der sich vor allem bei kategorialen Variablen (Nominal-/ Ordinalskalenniveau) 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 absolute und die relative Häufigkeit angezeigt wird.

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



Statistiken

hat mir neue Wege gezeigt, an Dinge heranzugehen, die für mich unverständlich waren.

N	Gültig	4715	1002
	Fehlend	0	3713

Häufigkeitstabelle

hat mir neue Wege gezeigt, an Dinge heranzugehen, die für mich unverständlich waren.

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	stimme gar nicht zu	581	12,3	12,3	12,3
	stimme eher nicht zu	1005	21,3	21,3	33,6
	stimme teilweise zu	1573	33,4	33,4	67,0
	stimme eher zu	1091	23,1	23,1	90,1
	stimme vollständig zu	465	9,9	9,9	100,0
Gesamt		4715	100,0	100,0	

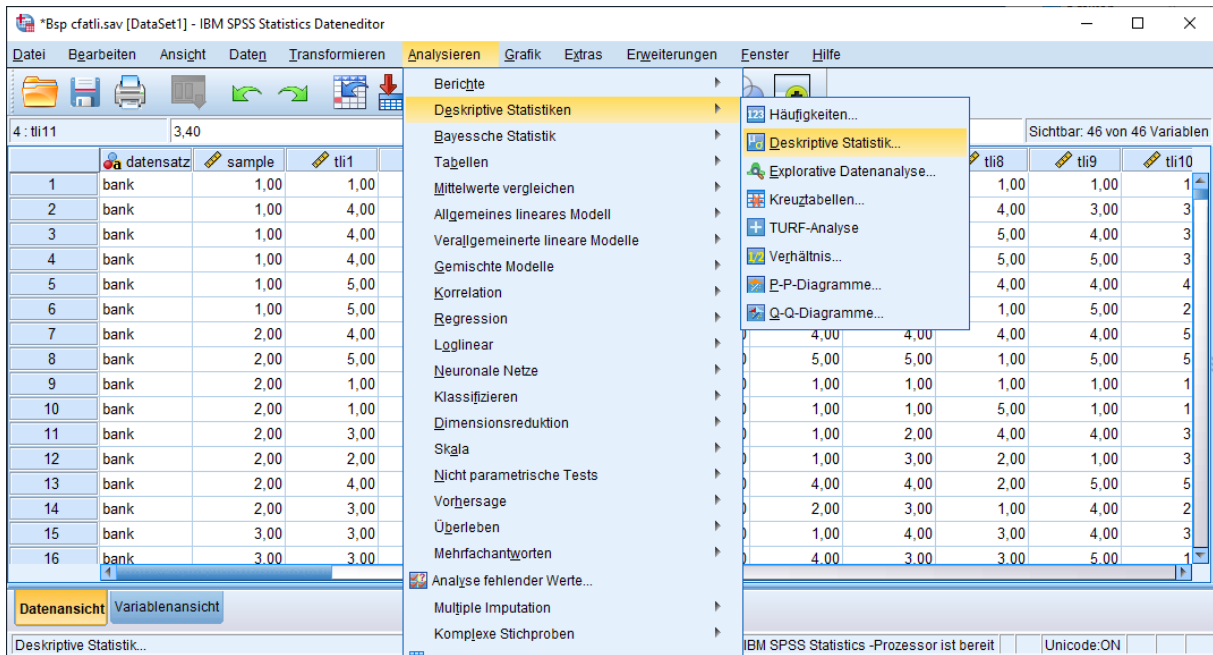
Size of the Organization

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	less than 50	361	7,7	36,0	36,0
	between 50 and 250	280	5,9	27,9	64,0
	more than 250	361	7,7	36,0	100,0
	Gesamt	1002	21,3	100,0	
Fehlend	System	3713	78,7		

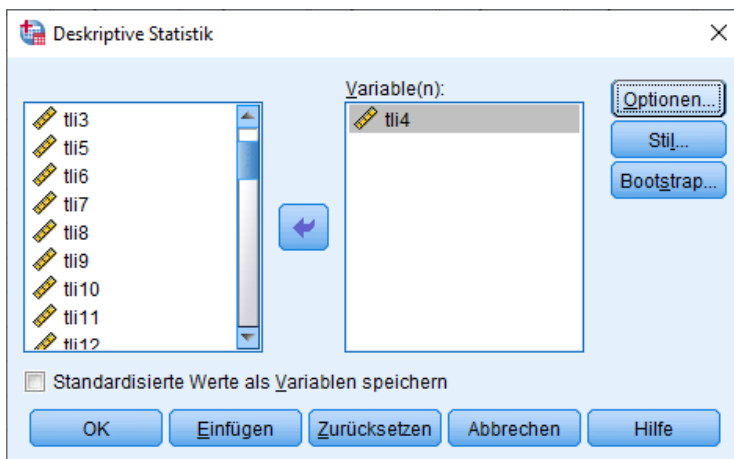
Innerhalb der Ausgabedatei findet sich für jede Variable, die ausgewählt wurde, eine Häufigkeitstabelle. In der oberen Häufigkeitstabelle im Screenshot ist ersichtlich, dass „stimme gar nicht zu“ von 581 Personen angekreuzt wurde bzw. „stimme gar nicht zu“ eine (absolute) Häufigkeit von 581 hat. Bezogen 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 dem betrachteten Wert ist. 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.

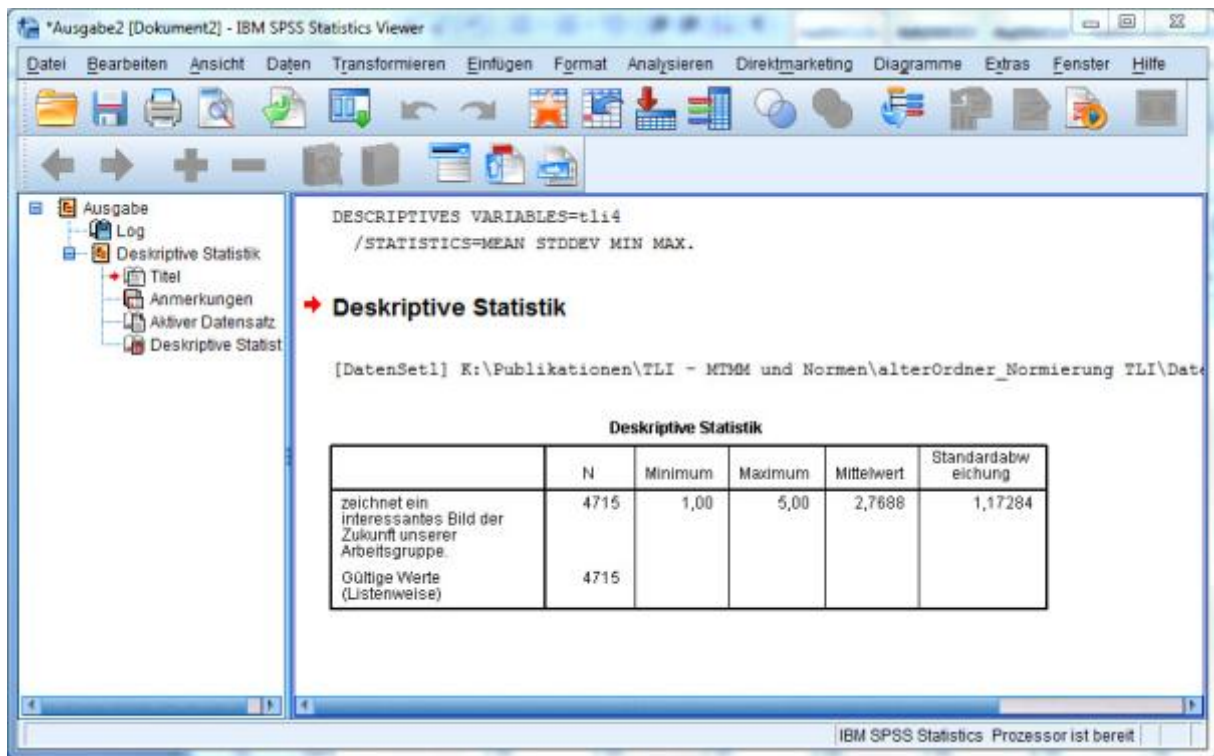


Die gewünschten Variablen sind auszuwählen und in das Feld Variable(n) zu ziehen.



Unter „Optionen“ können die gewünschten Lagemaße (hier Mittelwert) sowie die Maße der Streuung ausgewählt werden (hier Standardabweichung und Varianz).

```
DESCRIPTIVES VARIABLES=tli4
  /STATISTICS=MEAN STDDEV MIN MAX.
```



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 heißt in der Statistik arithmetisches 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 der Größe nach geordnet (136, 146, 156, 167, 177), so ist der Median der Wert, der in der Mitte liegt: In diesem Beispiel ist der Median = 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: 156, 146, 136, 167 und 257 Zentimeter (statt 177). Der Median bliebe von dem Ausreißer unberührt, er wäre immer noch 156. Das arithmetische Mittel läge hier hingegen bei 172.6, obwohl nur zwei Personen über 1.70 Meter groß sind.

Modus: Der Modus oder Modalwert ist der, bei einer empirischen Häufigkeitsverteilung, am häufigsten auftretende Wert bei einer Variablen. Der Modus entspricht somit immer einem tatsächlich beobachteten Wert und ist zudem unabhängig von potenziellen Ausreißern. Im Gegensatz zum Arithmetischen Mittel besteht also keine Gefahr der Verzerrung des Modalwertes durch Ausreißer. Berechnet wird der Modus durch den Vergleich der absoluten Häufigkeiten einer Variablen. Demnach ist der Modalwert derjenige beobachtete Wert mit der größten absoluten Häufigkeit.

3.2.2 Exkurs Streuungsmaße

Standardabweichung: Häufig streuen die beobachteten Werte um den Mittelwert der Verteilung. Das bekannteste Maß, um das Ausmaß dieser Streuung zu beschreiben, ist die Standardabweichung. Sie berechnet sich als Wurzel aus der Varianz bzw. als

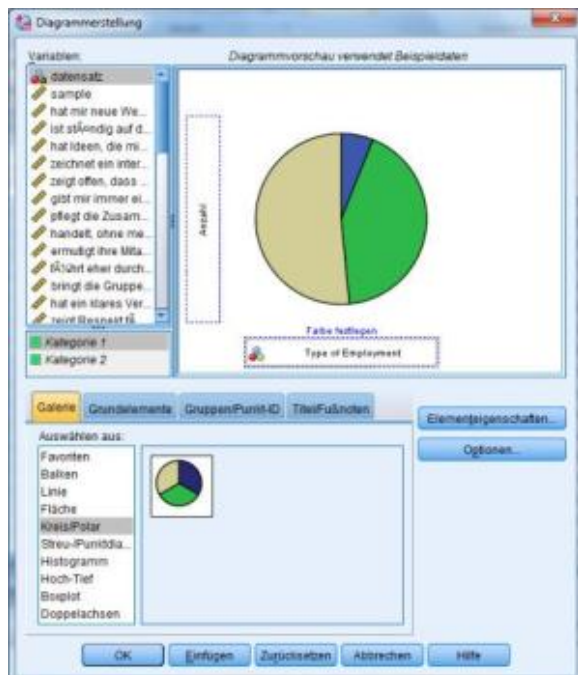
$$s_x = \sqrt{\frac{\sum_{i=1}^n (x_i - m_x)^2}{n}}$$

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 *Grafik/Diagrammerstellung* erzeugt werden.

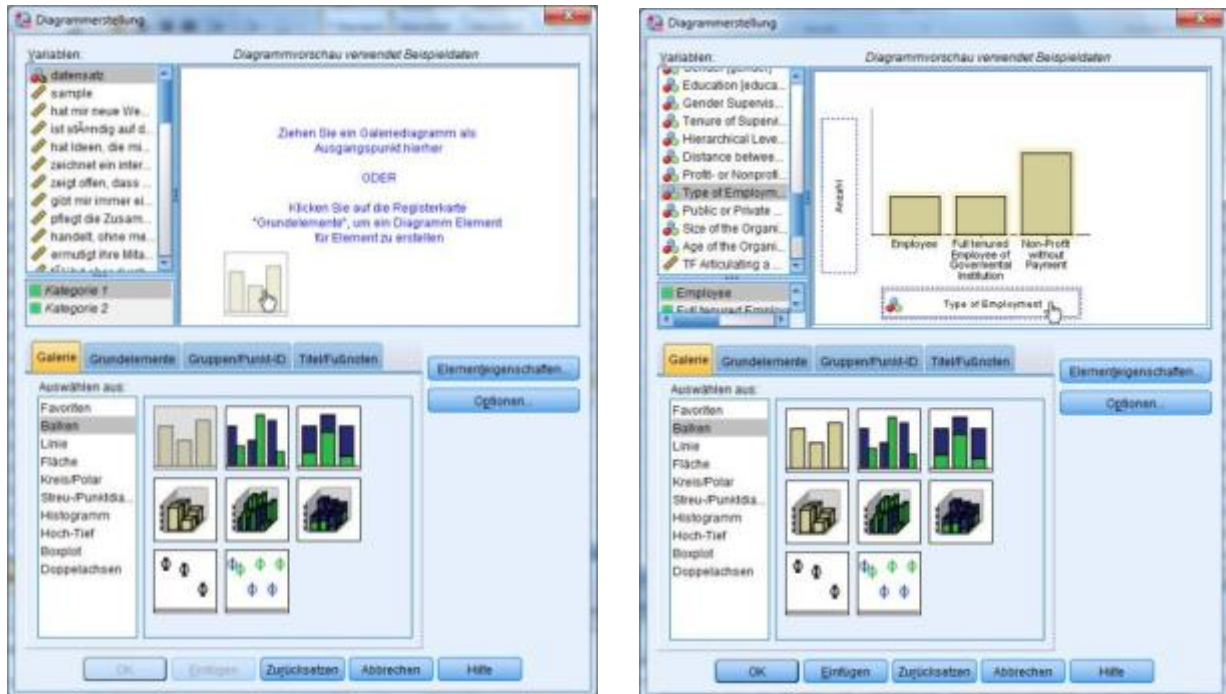
3.3.1 Kreisdiagramm

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

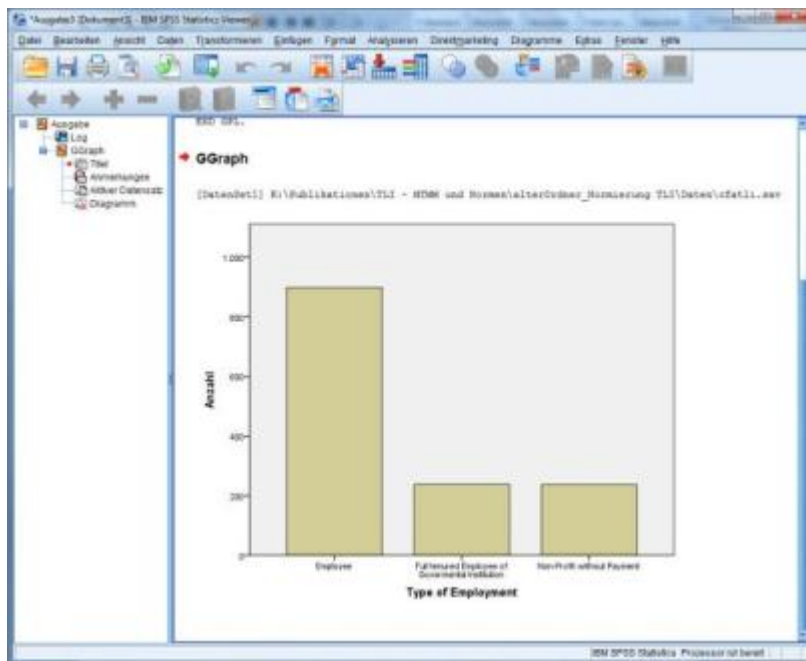


3.3.2 Säulendiagramm

Das Säulendiagramm ist ein Diagramm, das durch die Häufigkeitsverteilung einer diskreten Variablen vertikaler Säulen veranschaulicht. Das Säulendiagramm eignet sich besonders, um die Häufigkeit weniger verschiedener Werte zu veranschaulichen. Im Dialogfenster zur Diagrammerstellung wird mit der Maus der Diagrammtyp, hier Säulendiagramm, in das obige Feld gezogen.



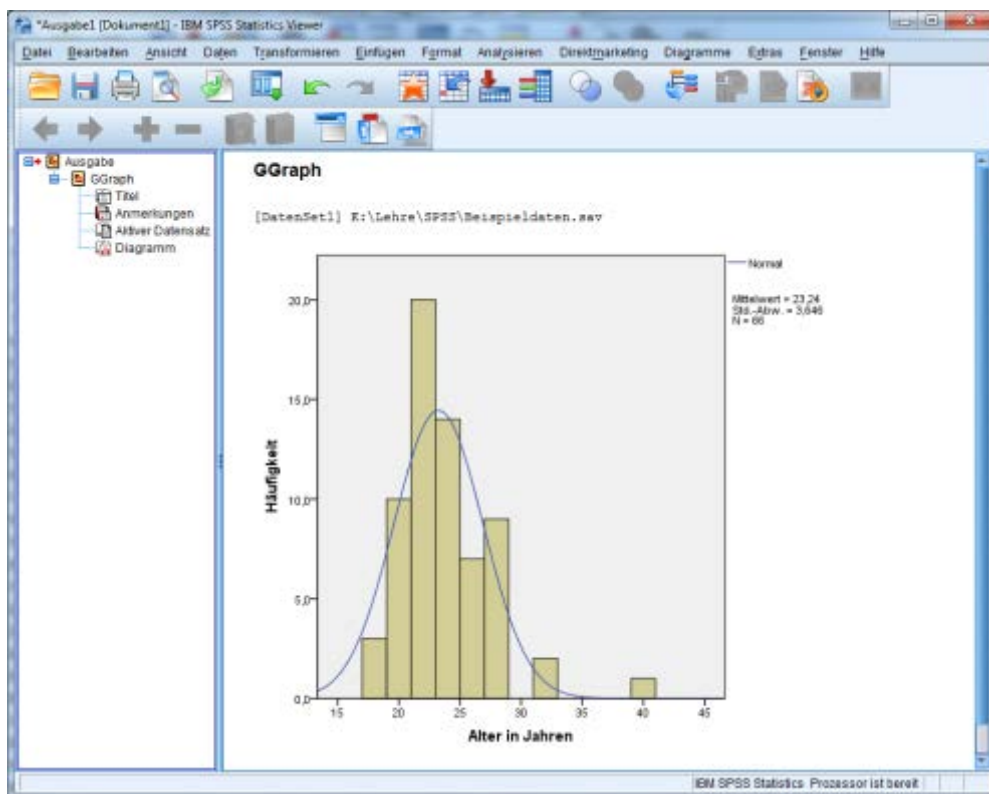
Anschließend wird die Variable, für die ein Diagramm angefordert werden soll, in das Feld „X-Achse?“ gezogen.



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 metrische Skala der Variable. Nur das Histogramm ist daher geeignet die empirische Verteilung mit einer [Normalverteilung](#) grafisch zu vergleichen.

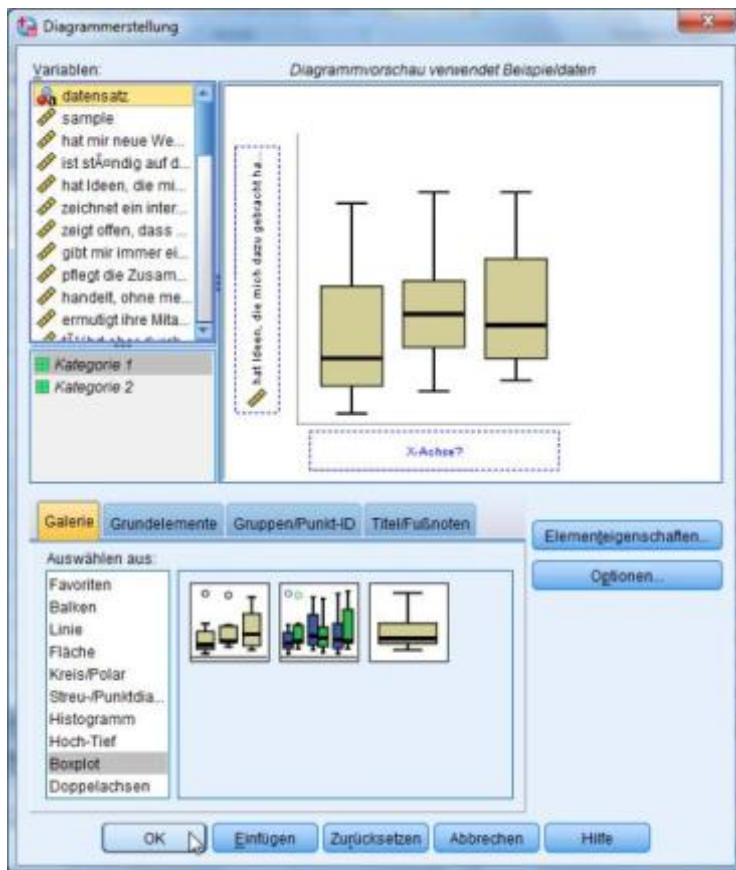
Es kann über das Dialogfenster zur [Diagrammerstellung](#) oder über den Befehl [Häufigkeiten](#) erzeugt werden. In beiden Fällen kann eine Normalverteilungskurve zum Vergleich angefordert werden.



3.3.4 Boxplot

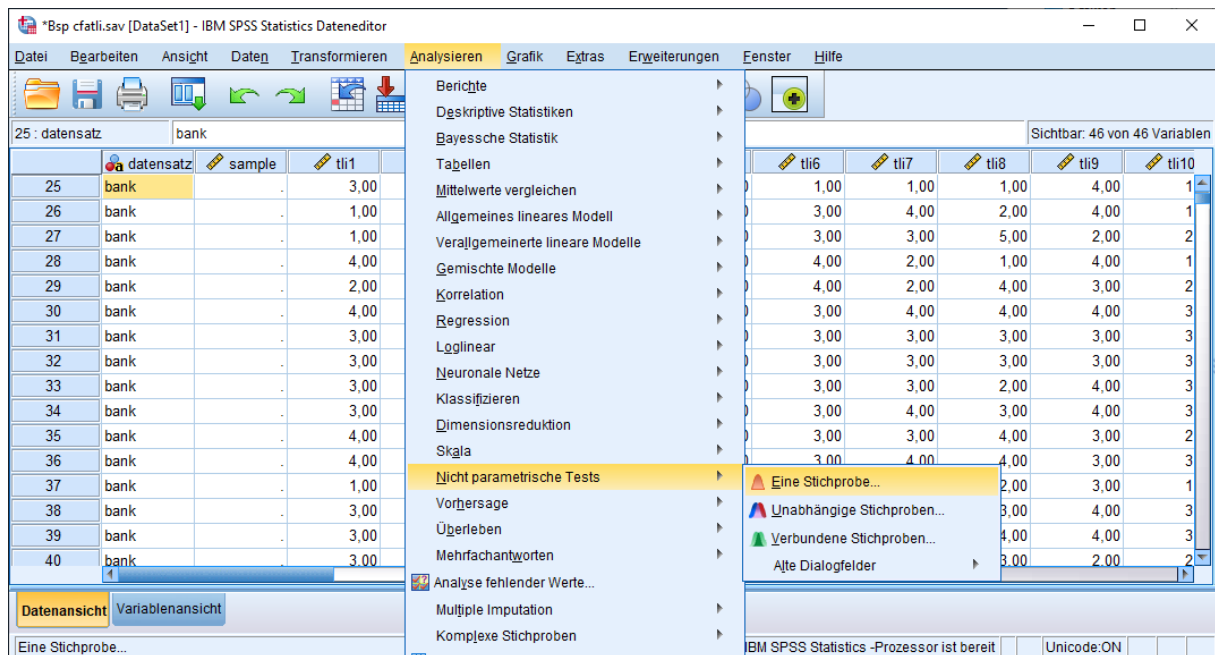
Der Boxplot ist ein Diagramm, das zur grafischen Darstellung der Verteilung statistischer Daten verwendet wird. Es fasst dabei verschiedene robuste Streuungs- und Lagemaße in einer Darstellung zusammen. Ein Boxplot soll schnell einen Eindruck darüber vermitteln, in welchem Bereich die Daten liegen und 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.

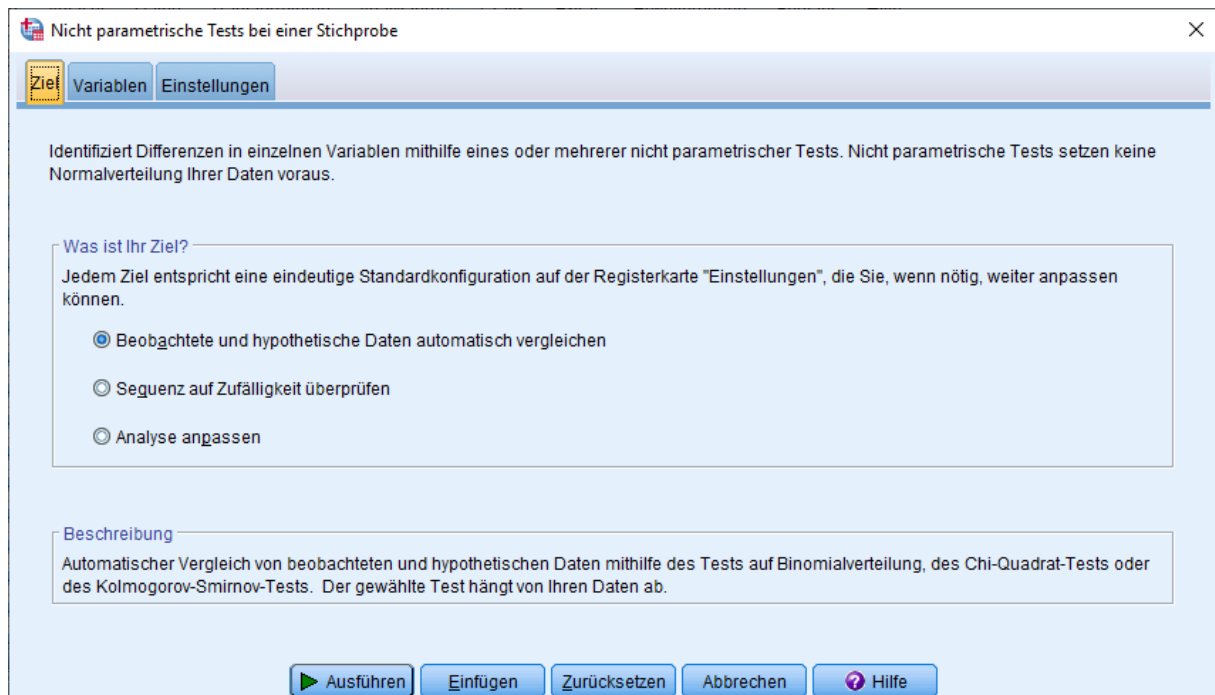


3.4 Überprüfung auf Normalverteilung

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 wird die beobachtete kumulative Verteilungsfunktion für eine Variable mit einer festgelegten theoretischen Verteilung verglichen (meist mit der Normalverteilung). Der K-S-Test prüft, ob eine Abweichung zu der angenommenen Verteilung vorliegt.



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

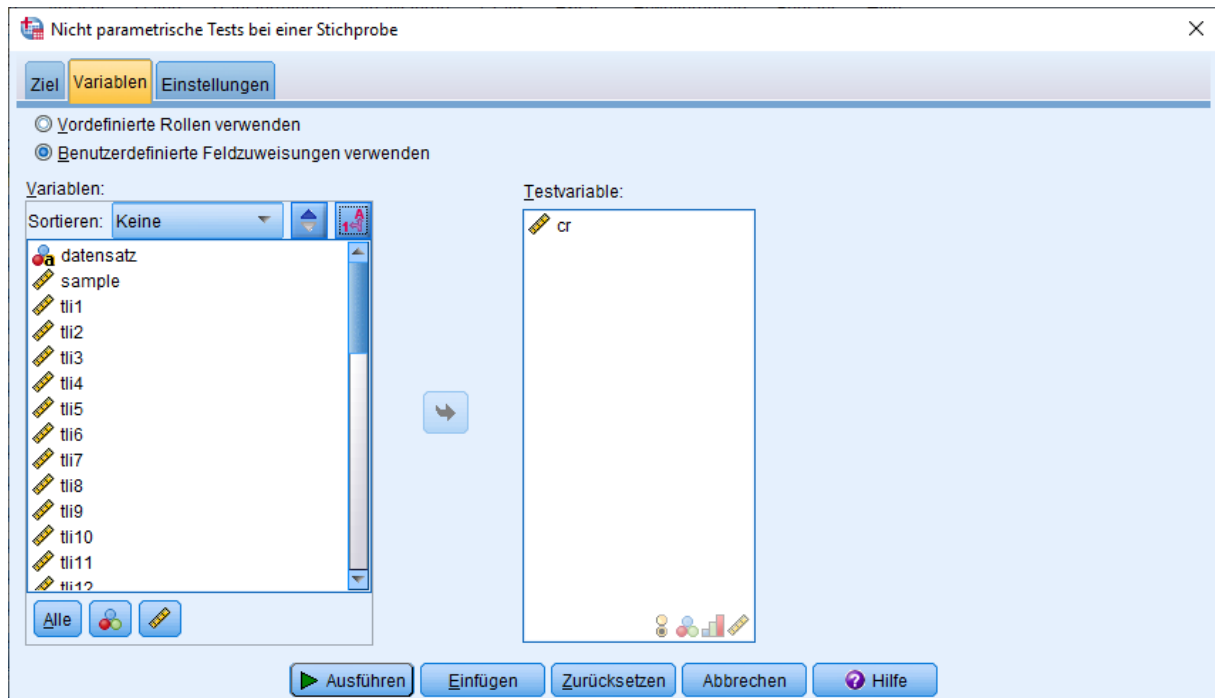


```

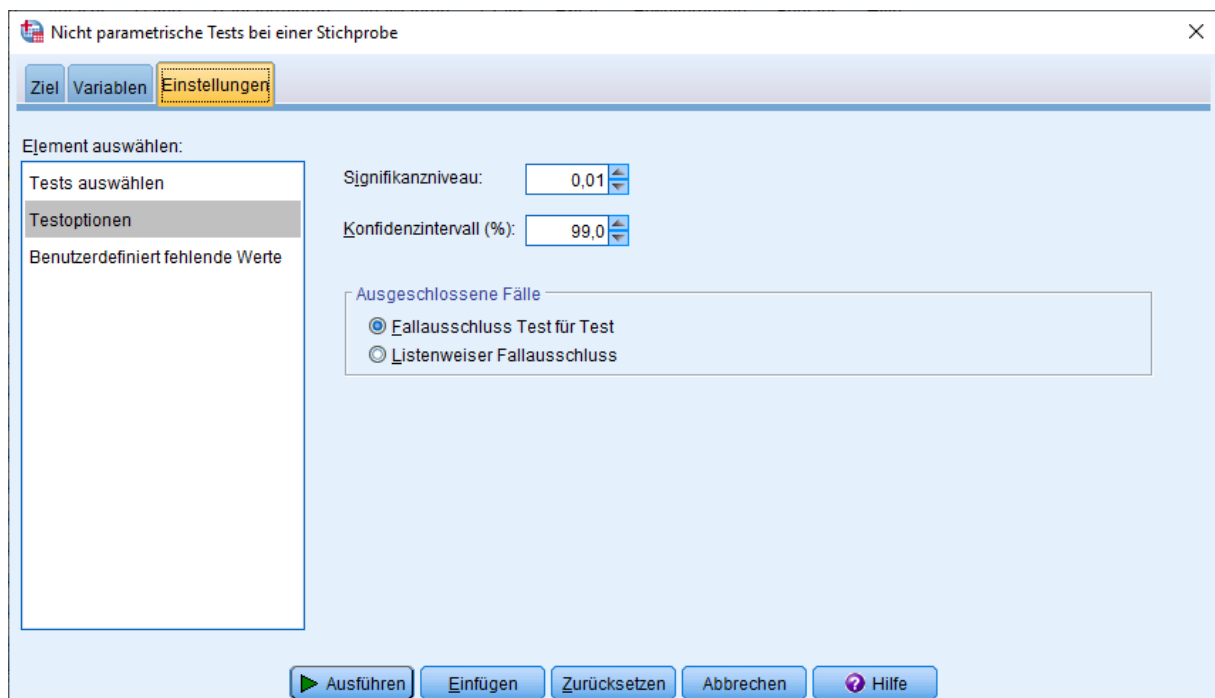
NPTESTS
/ONESAMPLE TEST (datensatz sample tli1 (...) isn cr)
KOLMOGOROV_SMIRNOV(NORMAL=SAMPLE )
/MISSING SCOPE=ANALYSIS USERMISSING=EXCLUDE
/CRITERIA ALPHA=0.05 CILEVEL=95.
EXECUTE.

```

Im obigen Dialogfeld können unter *Variablen/Benutzerdefinierte Feldzuweisungen verwenden* die Variablen, welche auf Normalverteilung geprüft werden sollen, einzeln ausgewählt werden. Ohne diese Auswahl werden sämtlich Variablen überprüft.



Des Weiteren kann unter *Einstellungen/Testoptionen* das Signifikanzniveau sowie das entsprechende Konfidenzintervall eingestellt werden. In der Regel wählt man ein Signifikanzniveau von .01, da man beim K-S-Test die Nullhypothese beibehalten möchte.



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 .01 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 .01 sollten non-parametrische Verfahren verwendet werden. Im Folgenden ist die Überprüfung zweier Variablen auf Normalverteilung anhand des Signifikanzniveaus .01 dargestellt.

Übersicht über Hypothesentest

	Nullhypothese	Test	Sig.	Entscheidung
1	Die Verteilung von TA Contingend Reward ist normal mit Mittelwert 3,325 und Standardabweichung 1,11.	Kolmogorov-Smirnov-Test einer Stichprobe	.134	Nullhypothese behalten.
2	Die Verteilung von Z-Wert: Anzahl der Mitarbeiter im Team ist normal mit Mittelwert -0,000 und Standardabweichung 1,00.	Kolmogorov-Smirnov-Test einer Stichprobe	.000	Nullhypothese ablehnen.

Asymptotische Signifikanzniveaus werden angezeigt. Das Signifikanzniveau ist ,01.

Die erste Variable ist normalverteilt, da der K-S-Test nicht signifikant ist, d.h. es liegt keine Abweichung zur Normalverteilung vor. Hier wird die Nullhypothese behalten bzw. nicht verworfen. Die Verteilung der zweiten Variablen unterscheidet sich signifikant von einer Normalverteilung, die Nullhypothese wird daher abgelehnt.

Die Prüfung auf Normalverteilung der gewählten Variablen ist nur auf Skalenebene sinnvoll.

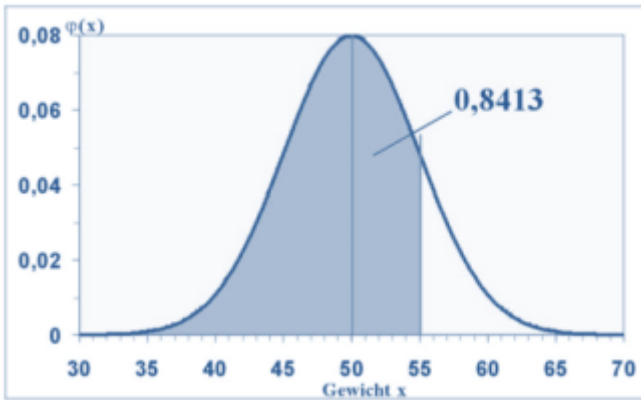
Bei sehr großen Stichproben führt der K-S-Test auch dann zu signifikanten Ergebnissen, wenn die Werte de facto normal verteilt sind. In diesen Fällen sind ggf. deskriptive Verteilungsmaße (Schiefe) und ein Histogramm (oder Boxplot) zu nutzen, um auf Normalverteilung zu prüfen.

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

3.4.1 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 dessen relative Häufigkeit. Es zeigt sich, dass sich das Gewicht der meisten Eier in der Nähe des Mittelwerts 50 befindet 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.

3.5 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.

The screenshot shows the IBM SPSS Statistics interface. The main window displays a list of variables with their names, types, widths, and decimals. The 'gender' and 'education' variables are highlighted. The 'Kreuztabellen' dialog box is open, showing the 'Zeile(n):' field containing 'Gender Supervisor [gendersup]' and the 'Spalten:' field containing 'Education [education]'. The dialog box also includes options for 'Exakt...', 'Statistiken...', 'Zellen...', 'Format...', 'Stil...', and 'Bootstrap...'. The 'OK' button is visible at the bottom.

```

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

```

The screenshot shows the IBM SPSS Statistics Viewer interface. The main window displays the following output:

```

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

```

→ Kreuztabellen

Verarbeitete Fälle

	Gültig		Fälle 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

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

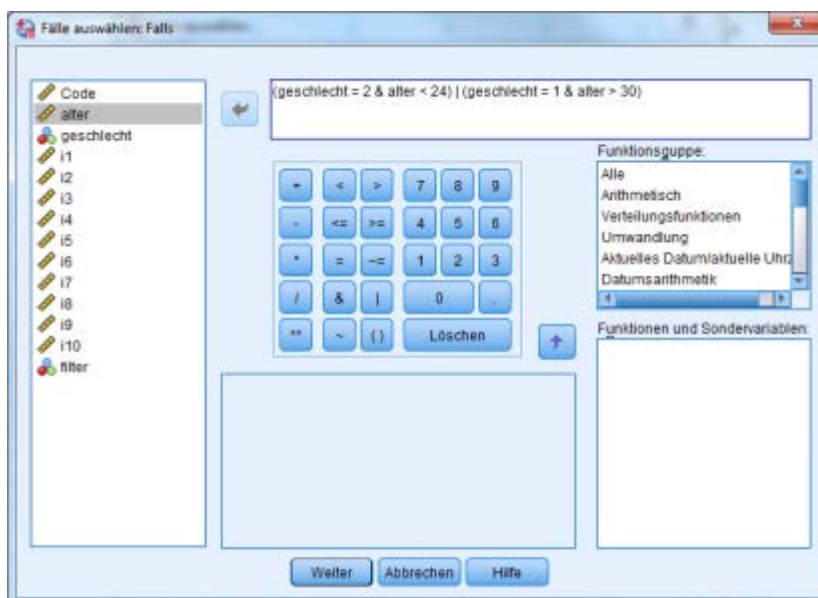
Wichtig: Ausgewählt werden stets jene Fälle, für die anschließend Berechnungen durchgeführt werden sollen. Nicht ausgewählte Fälle werden nicht berücksichtigt.

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.



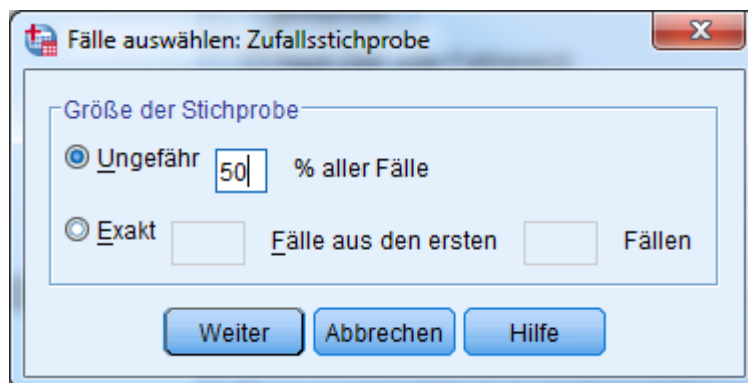
SPSS erstellt hierzu eine neue Variable (mit dem Namen „filter_“), die für die ausgewählten Fälle eine 1 und für die nicht ausgewählten eine 0 enthält; die Bedingung wird zudem als Beschriftung hinterlegt. In der

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.



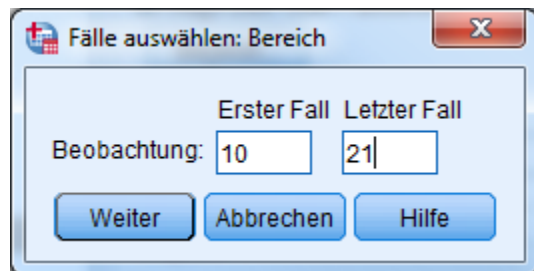
Diese Funktion wird genutzt, um zum Beispiel die Stabilität von Ergebnissen zu prüfen. Sofern die Gesamtstichprobe ausreichend groß ist, können die Berechnungen an zufällig ausgewählten Teilstichproben unabhängig voneinander durchgeführt und verglichen werden (sog. Kreuzvalidierung).

```
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**

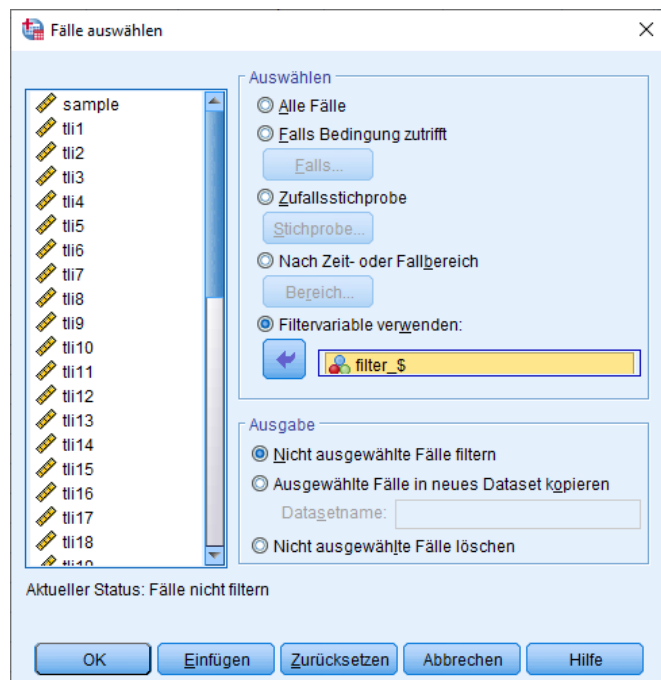
Hierüber können Fälle anhand ihrer Zeilennummer ausgewählt werden. Achtung: Um die Auswahl später replizieren zu können, muss die Reihenfolge der Fälle wiederhergestellt werden.

```
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.

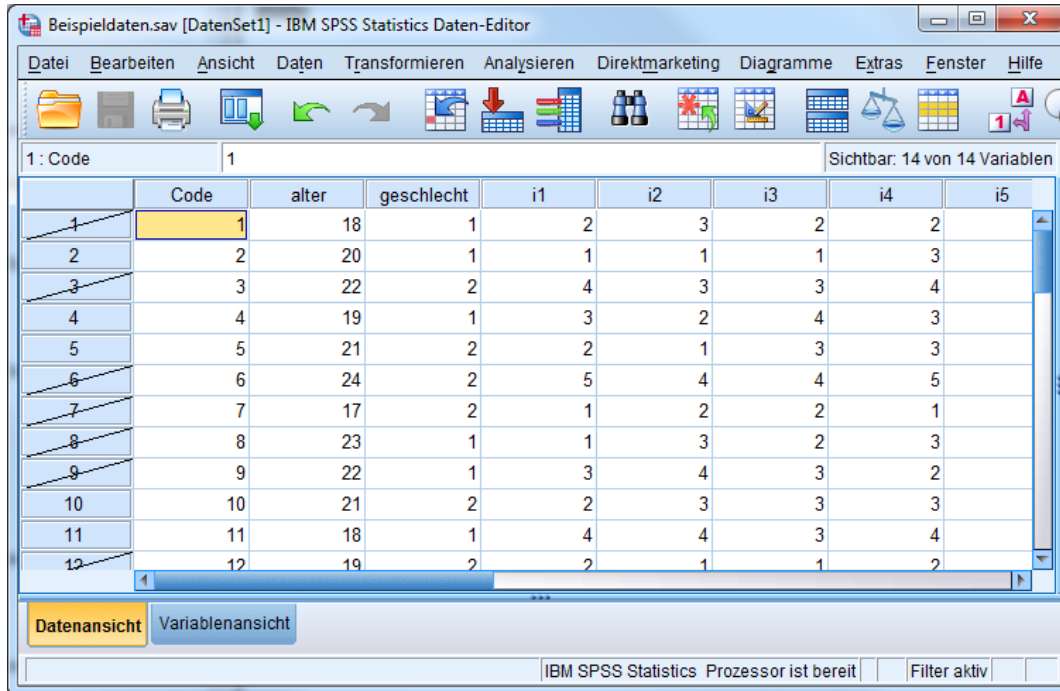


```
FILTER BY filtervariable.
EXECUTE.
```

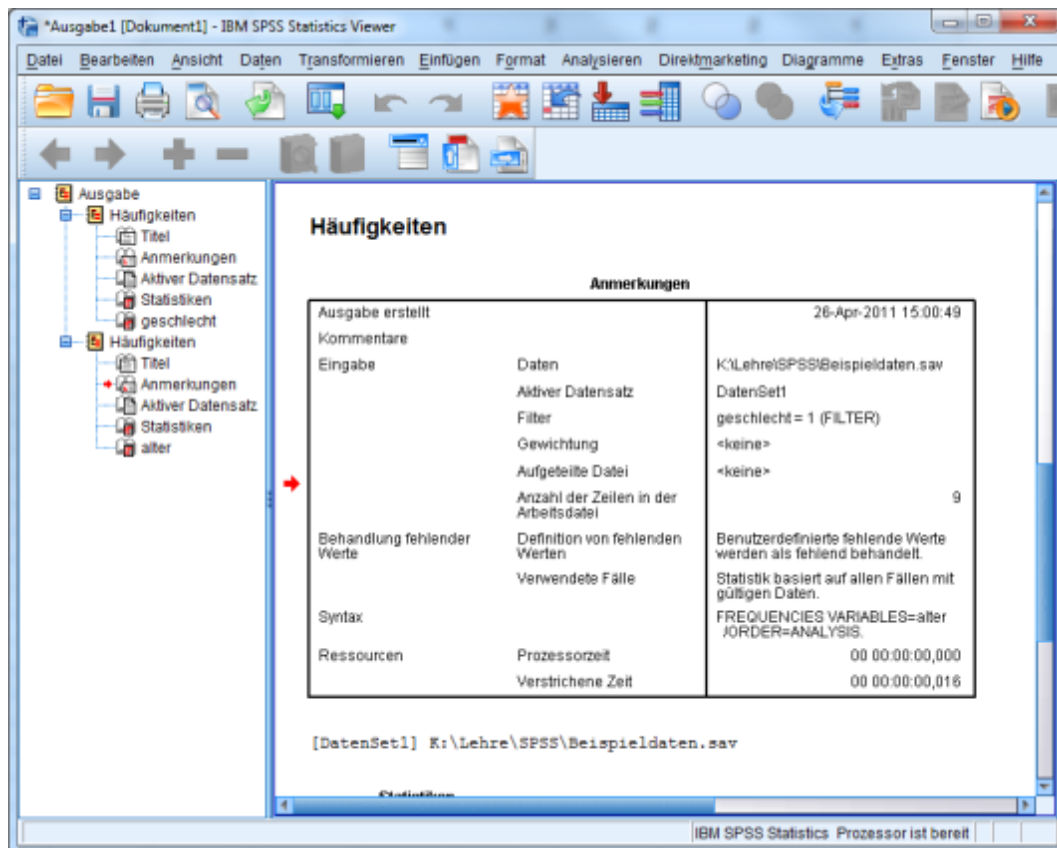
Im Anschluss an die Definition der Auswahlregel kann im Haupt-Dialogfenster zum *Fälle auswählen* angegeben werden, wie mit den ausgewählten bzw. nicht ausgewählten Fällen zu verfahren ist:

- **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.



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“).



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

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

- **Ausgewählte Fälle in neues Dataset kopieren**

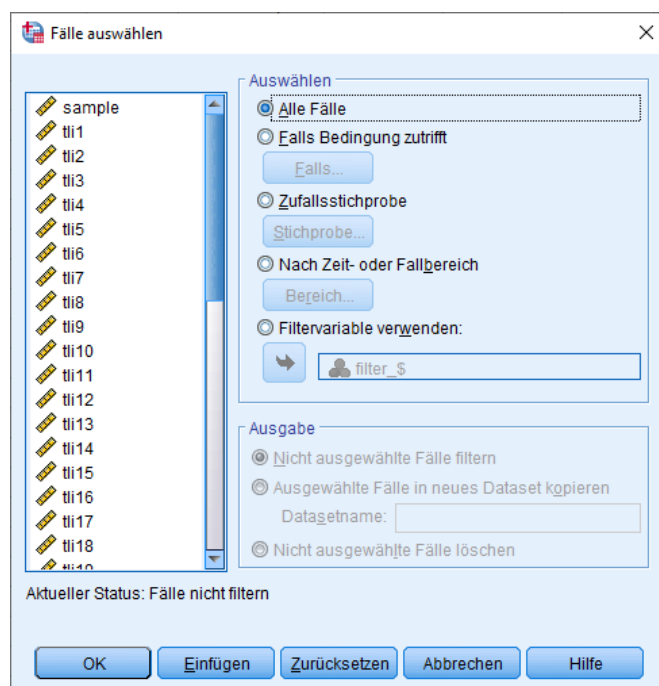
Es wird eine neue Datendatei erstellt, die nur die ausgewählten Fälle enthält.

- **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.



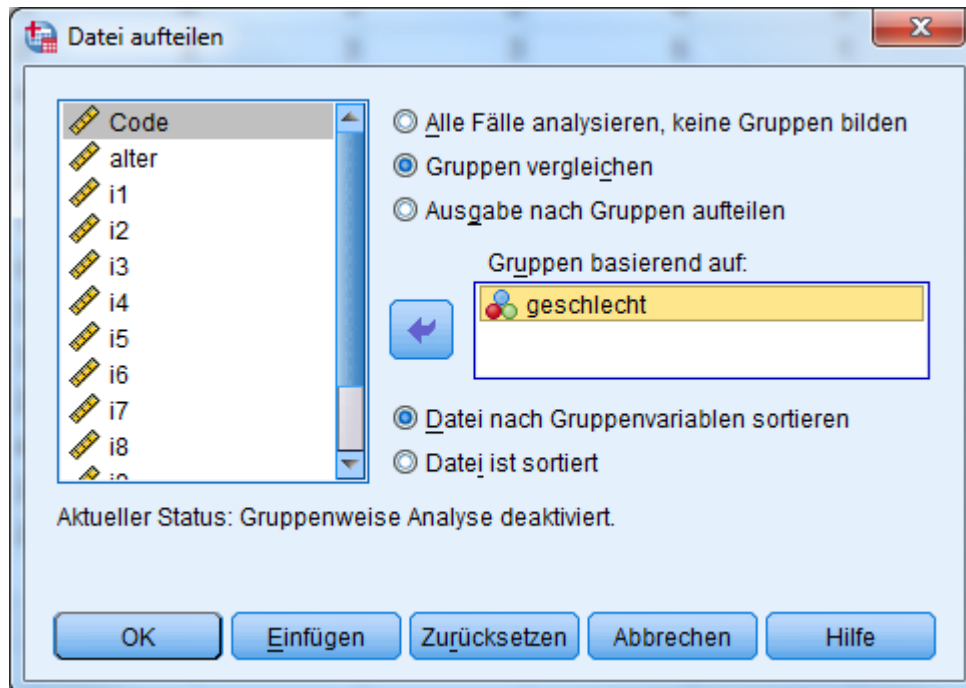
```
FILTER OFF.  
USE ALL.  
EXECUTE.
```

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/Aufgeteilte Datei* 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.

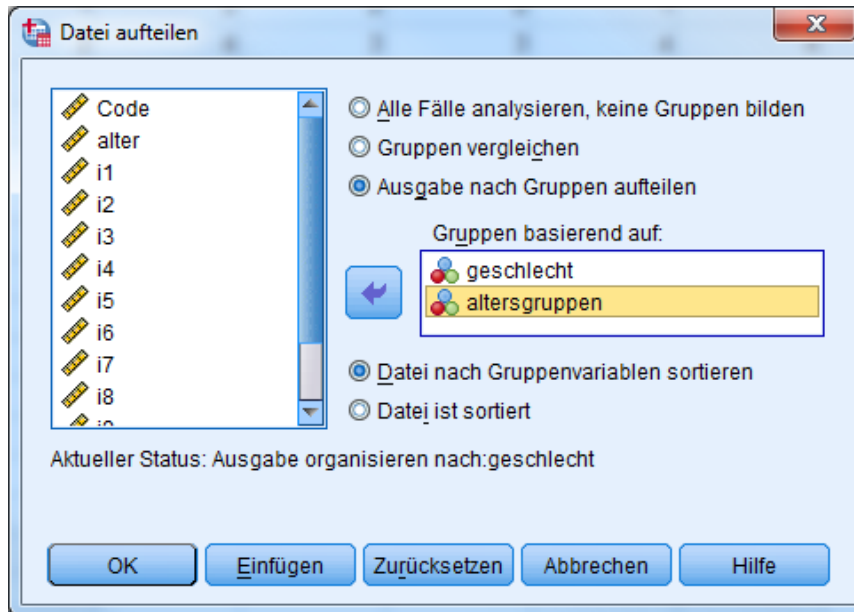
Häufigkeiten
 [Datenrec1] K:\Lehre\SPSS\Beispiel\data1.sav

Statistiken

alter					
männlich	N	ausg	Fehlend		
	9		0		
weiblich	N	ausg	Fehlend		
	12		0		

alter					
geschlecht	ausg	Häufigkeit	Prozent	ausg	kumulierte
				Prozente	Prozente
männlich	18	2	22,2	22,2	22,2
	19	1	11,1	11,1	33,3
	20	1	11,1	11,1	44,4
	21	1	11,1	11,1	55,6
	22	2	22,2	22,2	77,8
	23	1	11,1	11,1	88,9
	24	1	11,1	11,1	100,0
	Gesamt	9	100,0	100,0	
weiblich	17	1	8,3	8,3	8,3
	19	1	8,3	8,3	16,7
	20	2	16,7	16,7	33,3
	21	2	16,7	16,7	50,0
	22	3	25,0	25,0	75,0
	23	1	8,3	8,3	83,3
	24	1	8,3	8,3	91,7
25	1	8,3	8,3	100,0	
	Gesamt	12	100,0	100,0	

- 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:	29-Apr-2011 16:35:47
Kommentare:	
Eingabe:	Daten: K:\Lehre\SPSS\Beispiel\dates.sav
Filter:	Datenfall: <keine>
Gewichtung:	<keine>
Aufgeteilt: Datei:	geschlecht, altersgruppen
Anzahl der Zeilen in der Arbeitstabelle:	21
Behandlung fehlender Werte:	Benutzerdefinierte fehlende Werte werden als fehlend behandelt. Statistik basiert auf allen Fällen mit gültigen Daten.
Syntax:	FREQUENCIES VARIABLES=i1 ROWS=COLS=YES
Ressourcen:	Prozesszeit: 00:00:00,000
Verdichtete Zeit:	00:00:00,012

[DatenSet1] K:\Lehre\SPSS\Beispiel\dates.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 zuoberst aufgeführte Option „Alle Fälle analysieren, keine Gruppen bilden“ dient dazu, den Befehl zum Aufteilen der Datei zu deaktivieren.

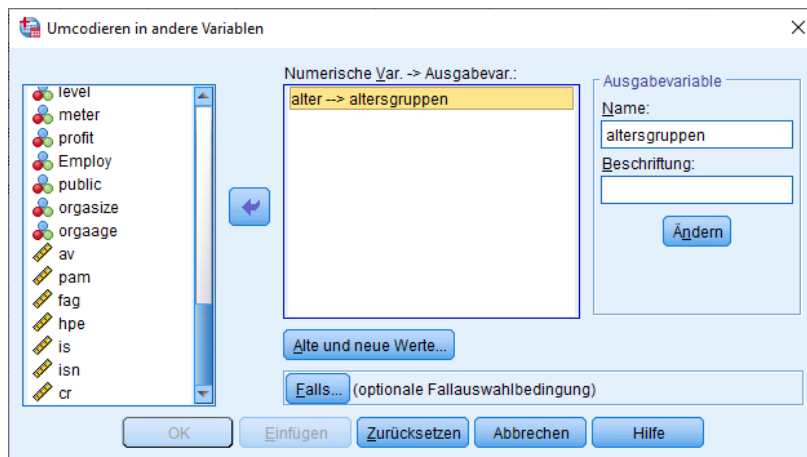
SPLIT FILE OFF.

5 DATENAUFBEREITUNG

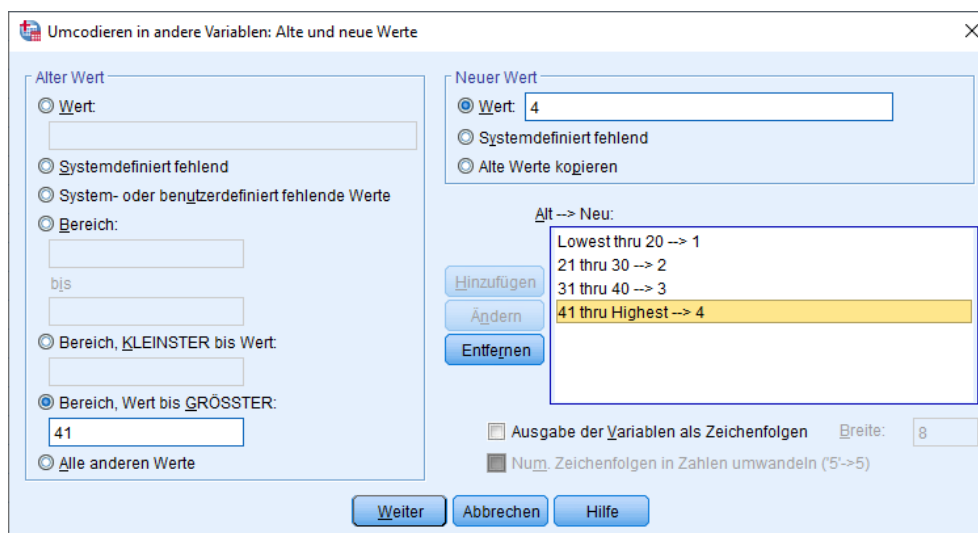
5.1 Umcodieren

Das Umcodieren 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: *Umcodieren in dieselben Variablen* und *Umcodieren 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 umcodiert wurden.



Im Dialogfenster zum *Umcodieren in andere Variablen* sind auf der linken Seite die umzucodierenden 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 Umcodierungsregeln 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 Variablen 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 codieren, ergibt sich als Umcodierungsregel 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.
```

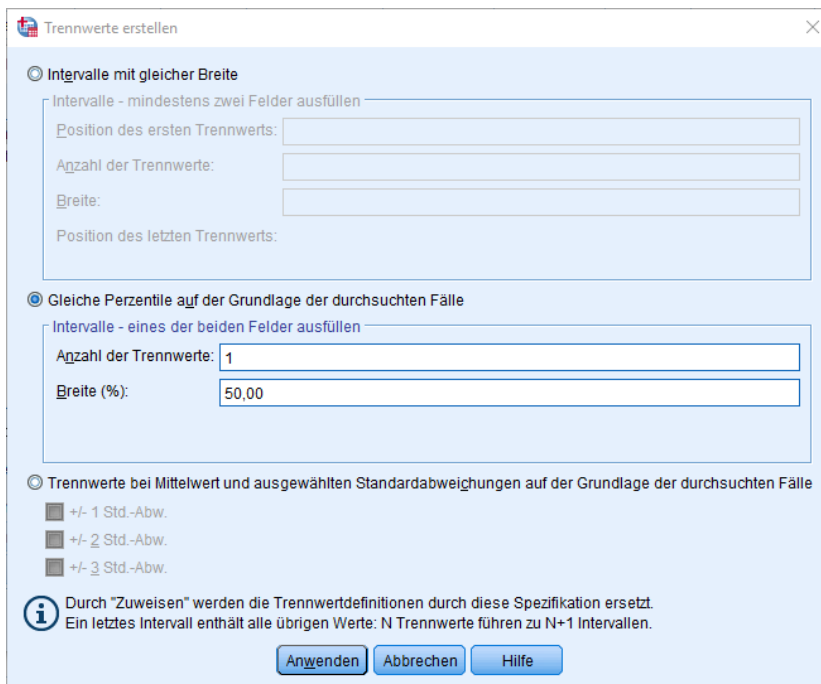
Das gleiche Ergebnis kann in solchen Fällen auch mit einem Befehl zur [Berechnung einer neuen Variablen](#) erzielt werden:

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

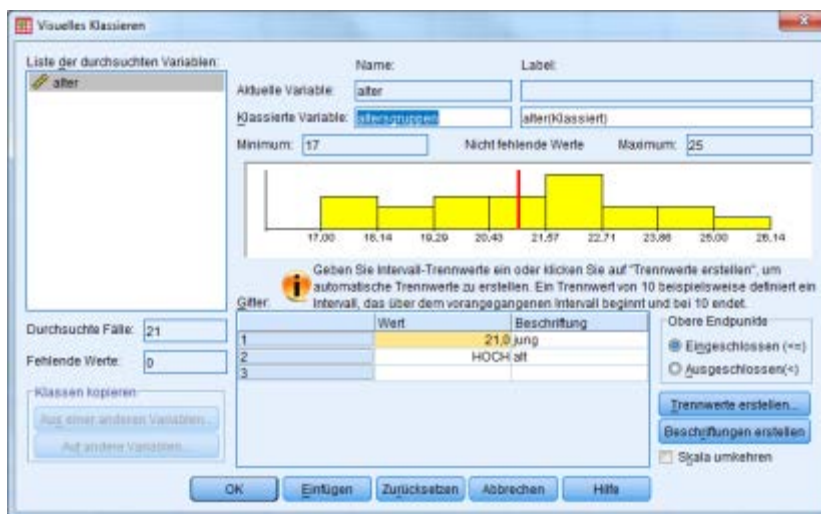
5.2 Visuelle Klassierung

Unter *Transformieren/Visuelle Klassierung* bietet SPSS verschiedene komfortable Möglichkeiten eine kontinuierlich verteilte Variable in Gruppen einzuteilen und eine neue Variable entsprechend der Kategorisierung zu erstellen. Im Unterschied zum [Umcodieren](#) 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).



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



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.

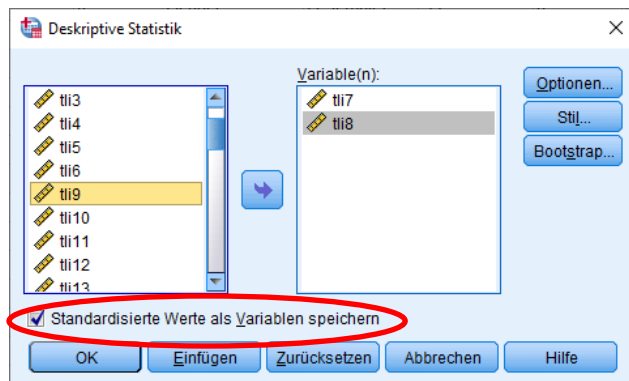
Rechnerisch ergibt sich der Wert z der z-standardisierten Variable aus der ursprünglichen Variable X zu

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

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.



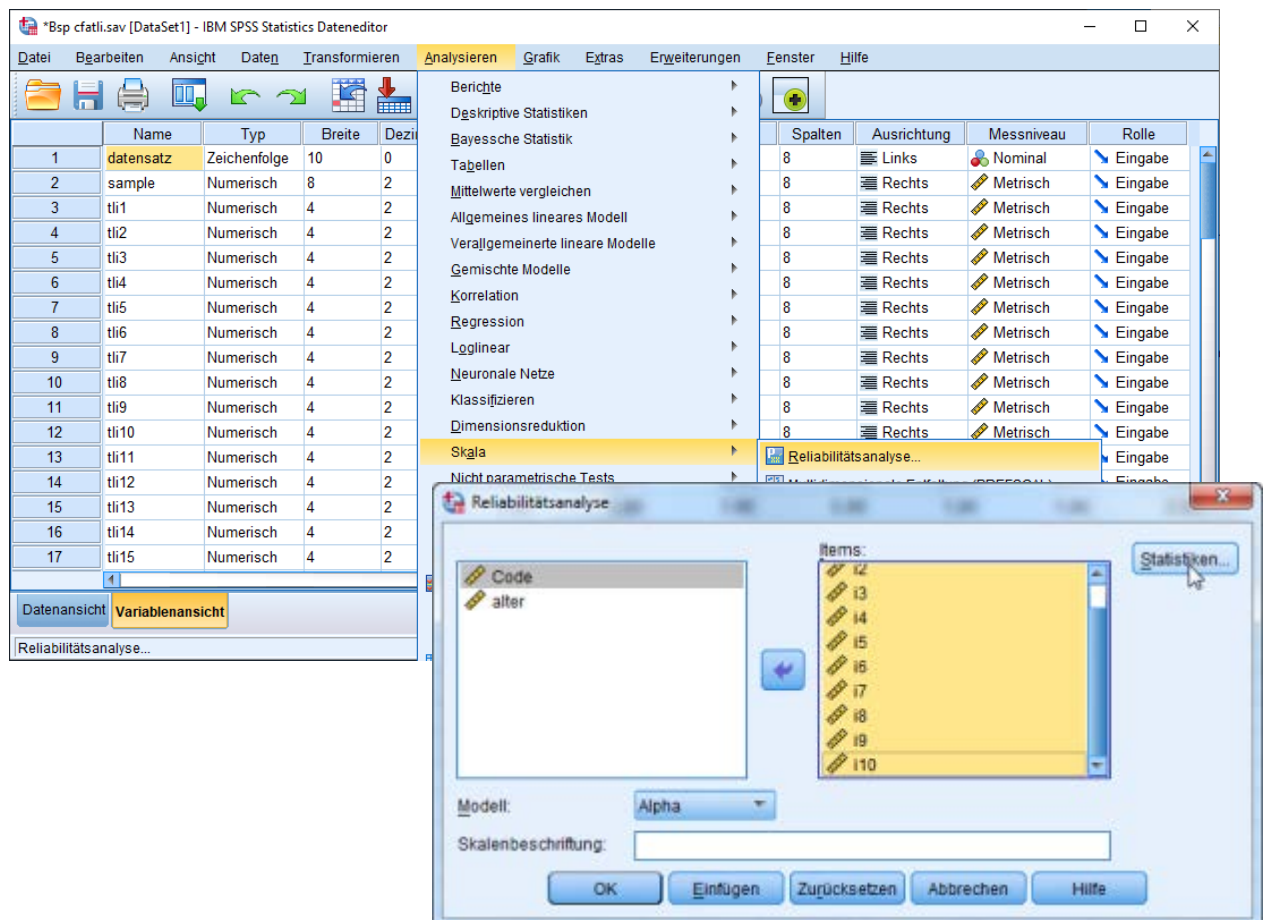
Z-standardisierte Variablen werden beispielsweise zur Durchführung einer [Regressionsanalyse mit Moderator](#) benötigt.

6 SKALENBILDUNG

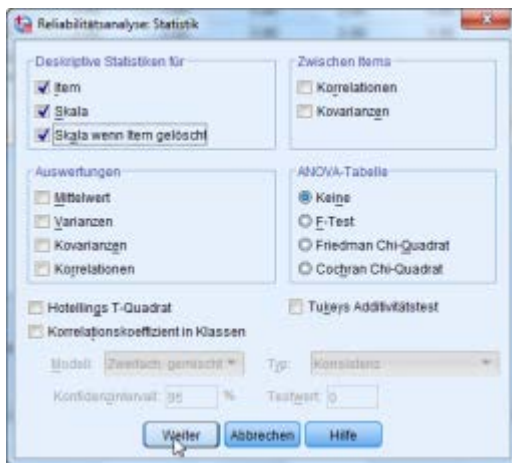
Unter Skalenbildung ist die Zusammenfassung mehrerer Einzelfragen (sog. Items) zu einer Skala gemeint, um ein Konstrukt zuverlässiger ([reliabler](#)) zu erfassen, als es mit einer einzelnen Frage möglich ist.

6.1 Reliabilitätsanalyse

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

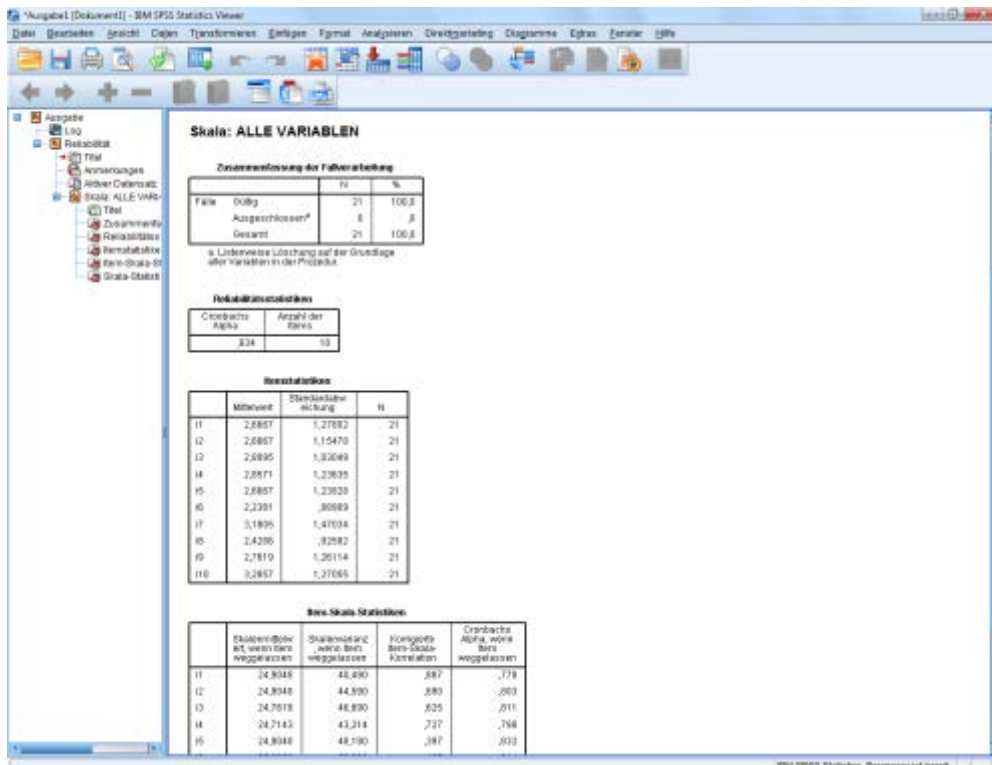


Unter „Modell“ steht u. a. „Alpha“ zur Bestimmung von [Cronbachs Alpha](#) zur Verfügung. Die Items, die die Skala bilden sollen, sind im linken Teil auszuwählen und in den rechten Teil zu verschieben. Unter „Statistiken“ können verschiedene zusätzliche Maße angefordert werden, zu empfehlen sind die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

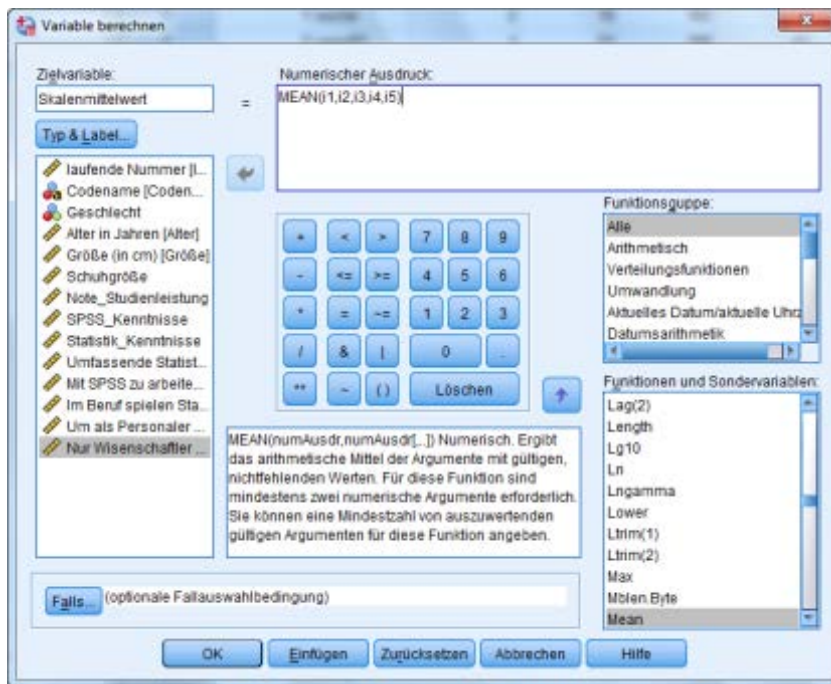
- **Retest-Reliabilität**

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 bei dem 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 Befehlssyntaxreferenz unter dem Befehl COMPUTE.

7 PRÜFUNG VON ZUSAMMENHANGSHYPOTHESEN

Bei der Prüfung von Zusammenhangshypothesen wird untersucht, inwieweit zwischen zwei metrisch skalierten Variablen ein Zusammenhang besteht. Die 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.

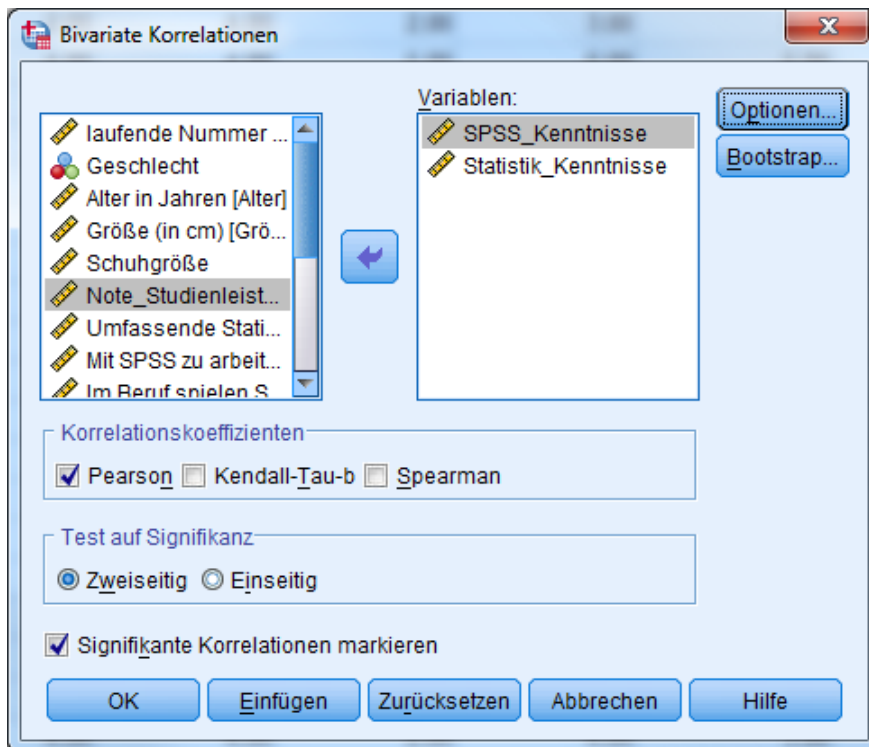
Ein linearer Zusammenhang ist gegeben, wenn die Ausprägungen der Variablen nach einem je-desto-Prinzip variieren. Der *Korrelationskoeffizient* r als Maß für einen linearen Zusammenhang beschreibt sowohl die Richtung als auch die Größe des Zusammenhangs:

- 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, denn es ist möglich, dass beide Variablen von einer dritten Variable beeinflusst werden oder dass 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/Korrelation/Bivariat* berechnet werden. Hier sind die zu korrelierenden Variablen auszuwählen. Die Voreinstellungen (Pearson-Koeffizient; 2-seitiger Signifikanztest; signifikante Korrelationen kennzeichnen) können beibehalten werden.

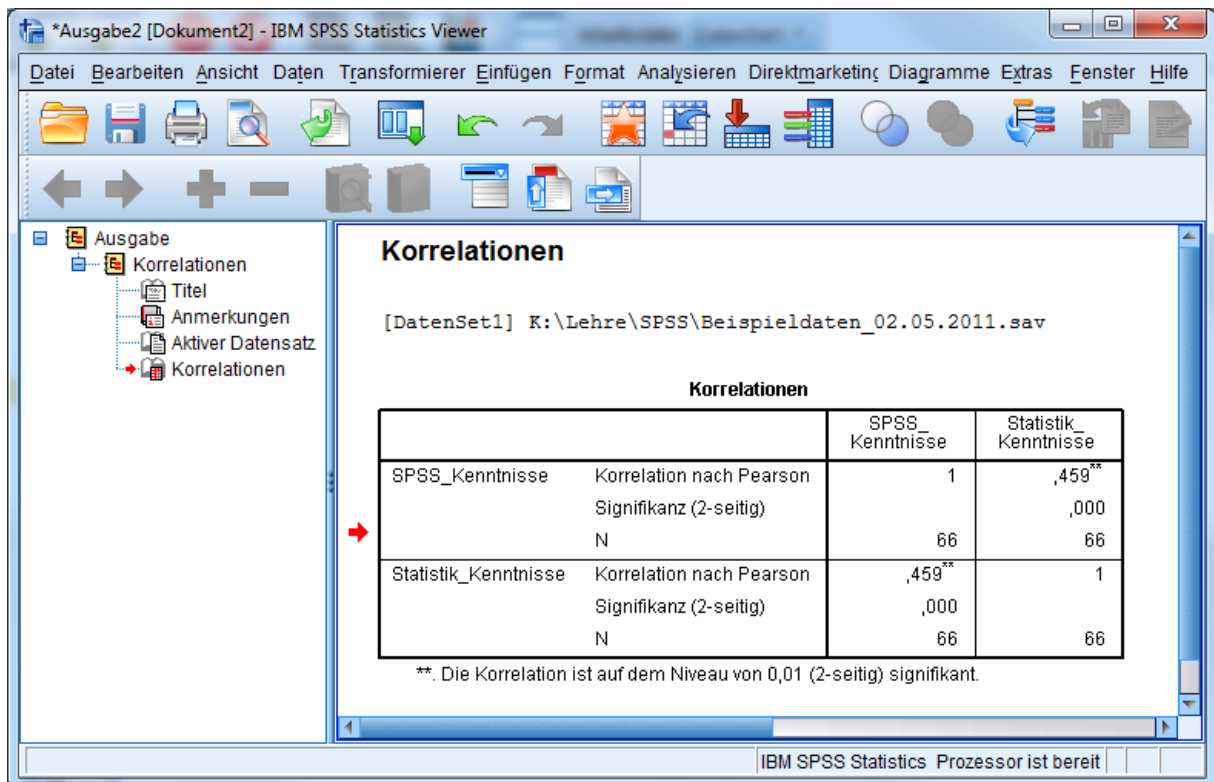


```

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 Variablenpaars der jeweiligen Zeile und Spalte zu finden ist.

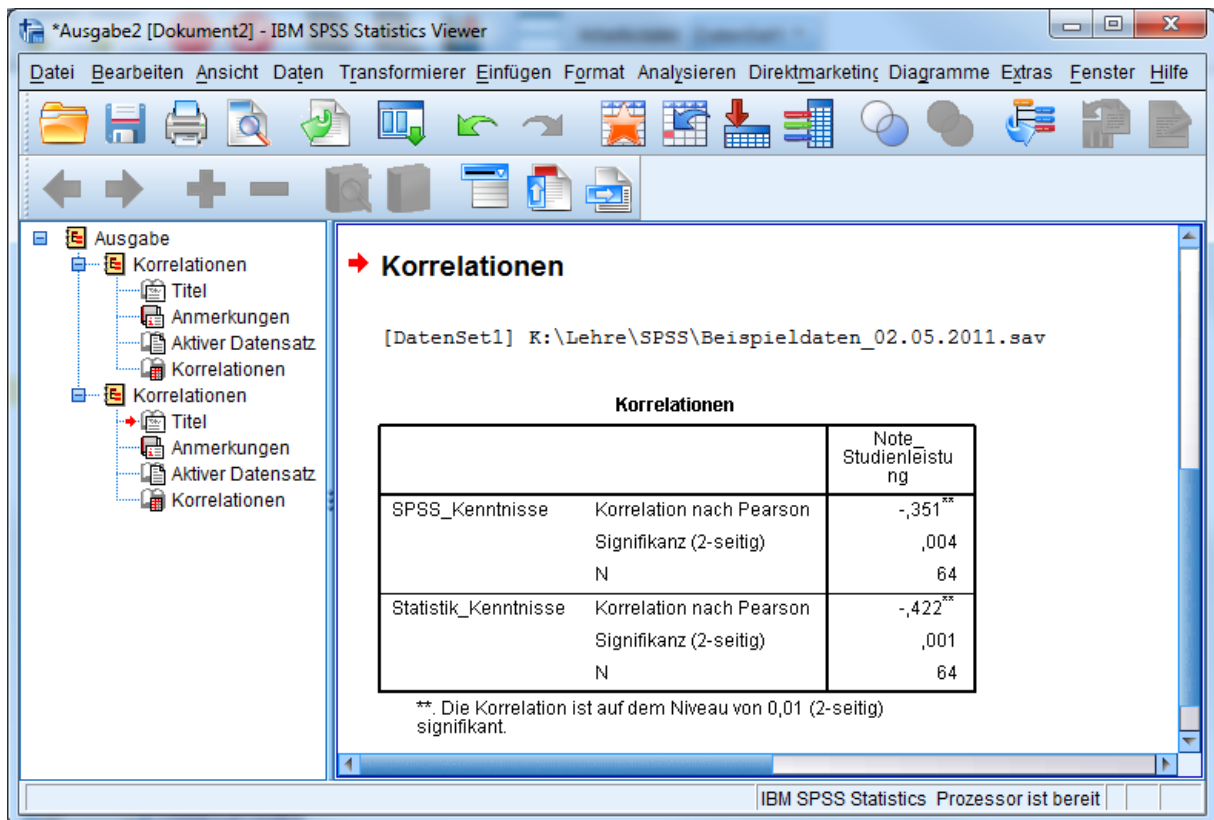


Es können im Dialogfenster auch mehr als zwei Variablen ausgewählt werden, um die Korrelationen aller enthaltenen Variablenpaare zu berechnen. Interessieren nur die Korrelationen zwischen einem Teil der Variablen mit einem anderen Teil der Variablen, kann im Syntaxbefehl das Wort „WITH“ zwischen diesen 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

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

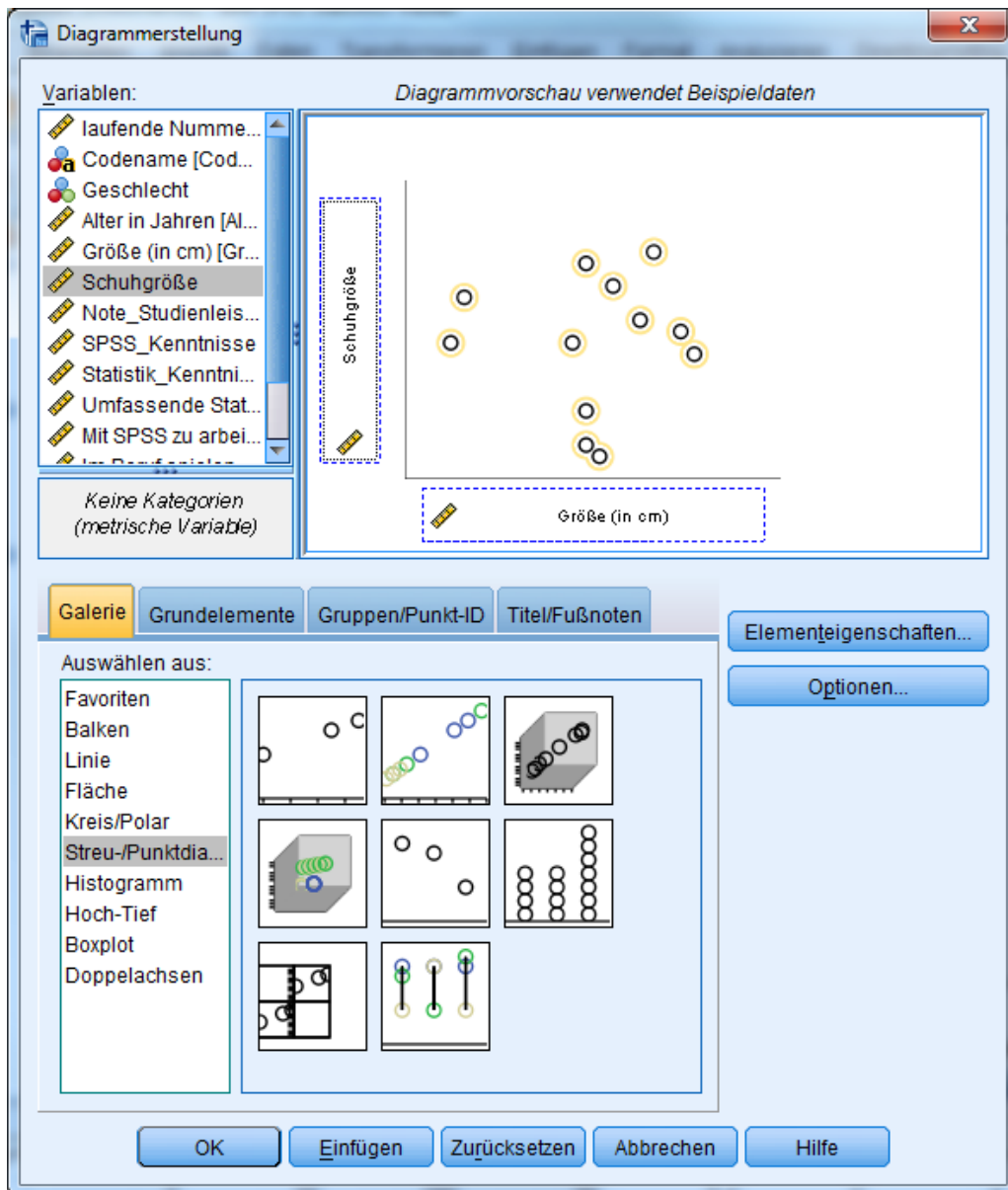
Voraussetzung für die Berechnung des Pearson-Korrelationskoeffizienten ist das [Intervallskalenniveau](#) der beiden Variablen. Für die sinnvolle Interpretation des in SPSS mitgelieferten Signifikanztest zum Korrelationskoeffizienten ist zusätzlich eine annähernde [Normalverteilung](#) der Variablen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

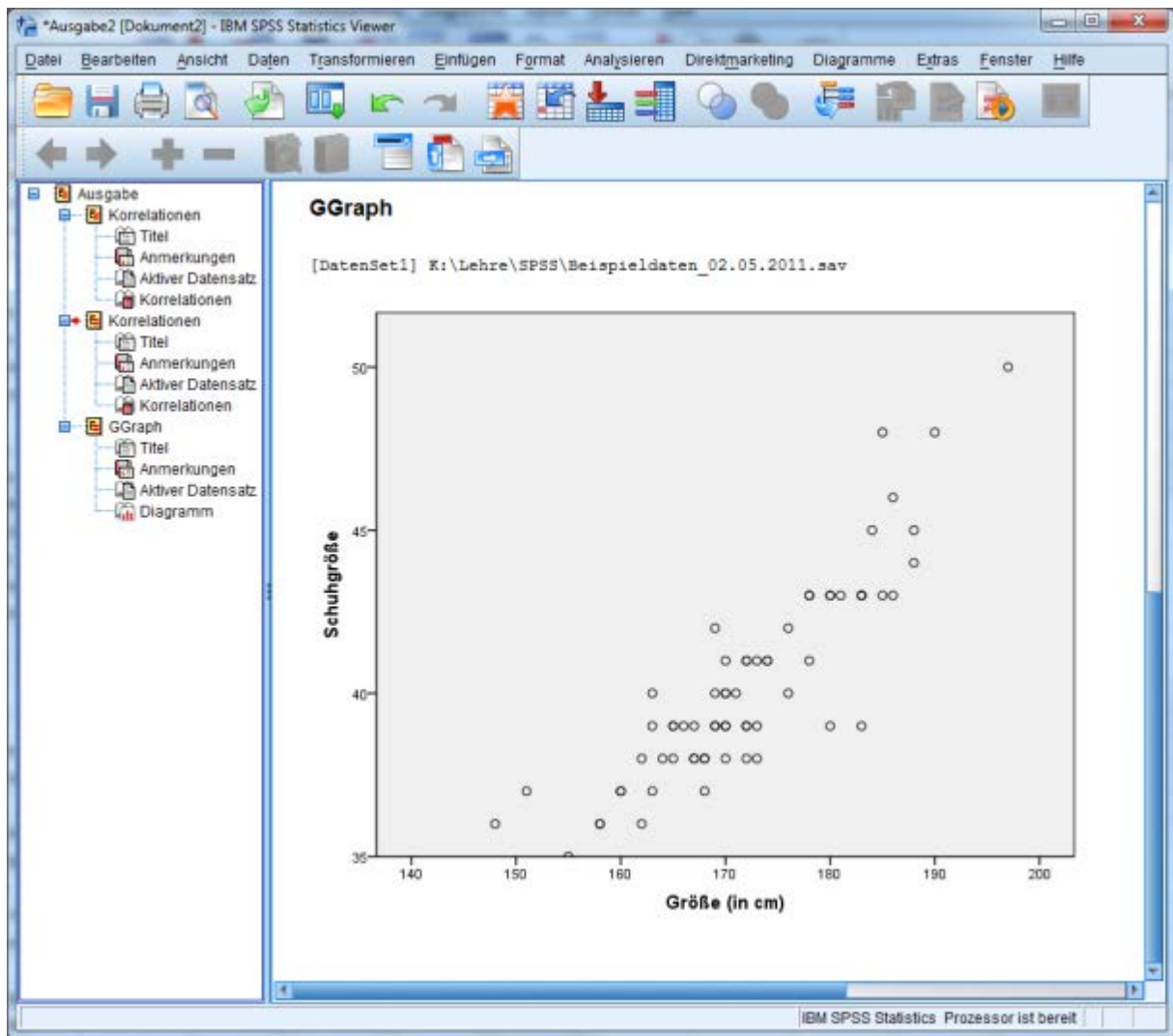
- $r = .10$ kleine Korrelation
- $r = .30$ mittlere Korrelation
- $r = .50$ große Korrelation

7.1.2 Streudiagramm

Lineare Zusammenhänge zwischen zwei Variablen können anhand eines Streudiagramms veranschaulicht werden. Unter *Grafik/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.

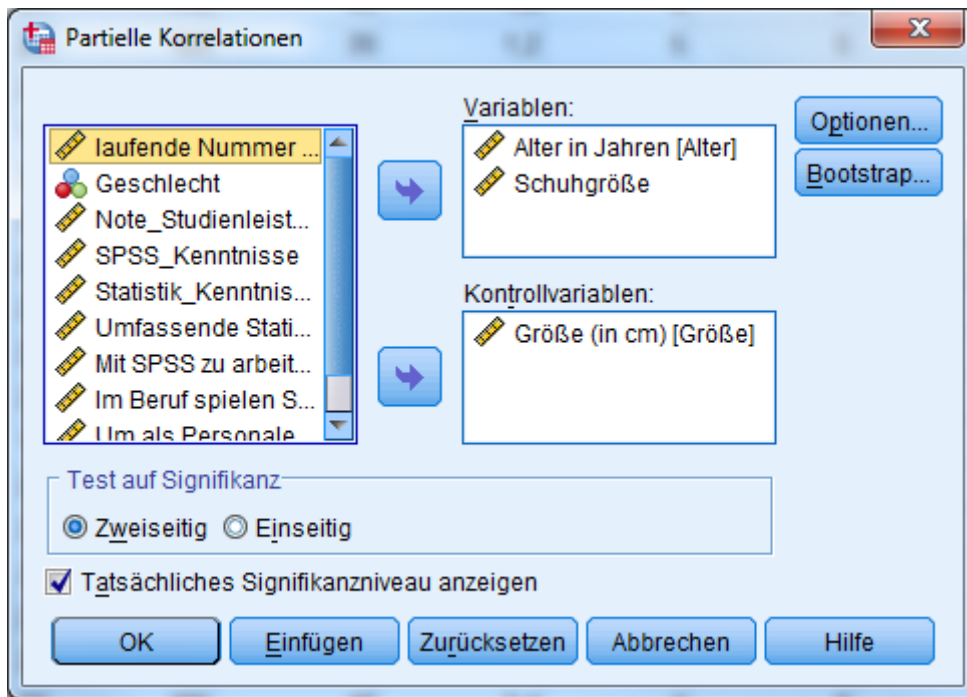


7.2 Partielle Korrelation

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.

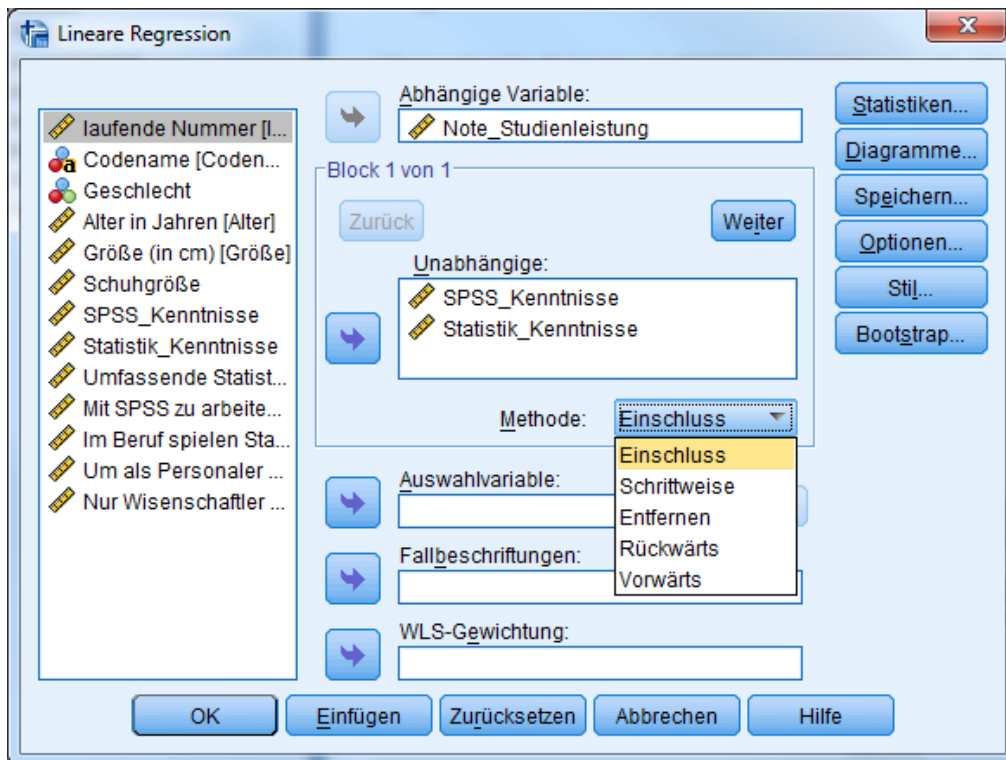
```

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 Studienzzeit 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 dem größten, signifikanten Regressionsgewicht aufgenommen 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“, „Optionen“, „Stil“ und „Bootstrap“ 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.

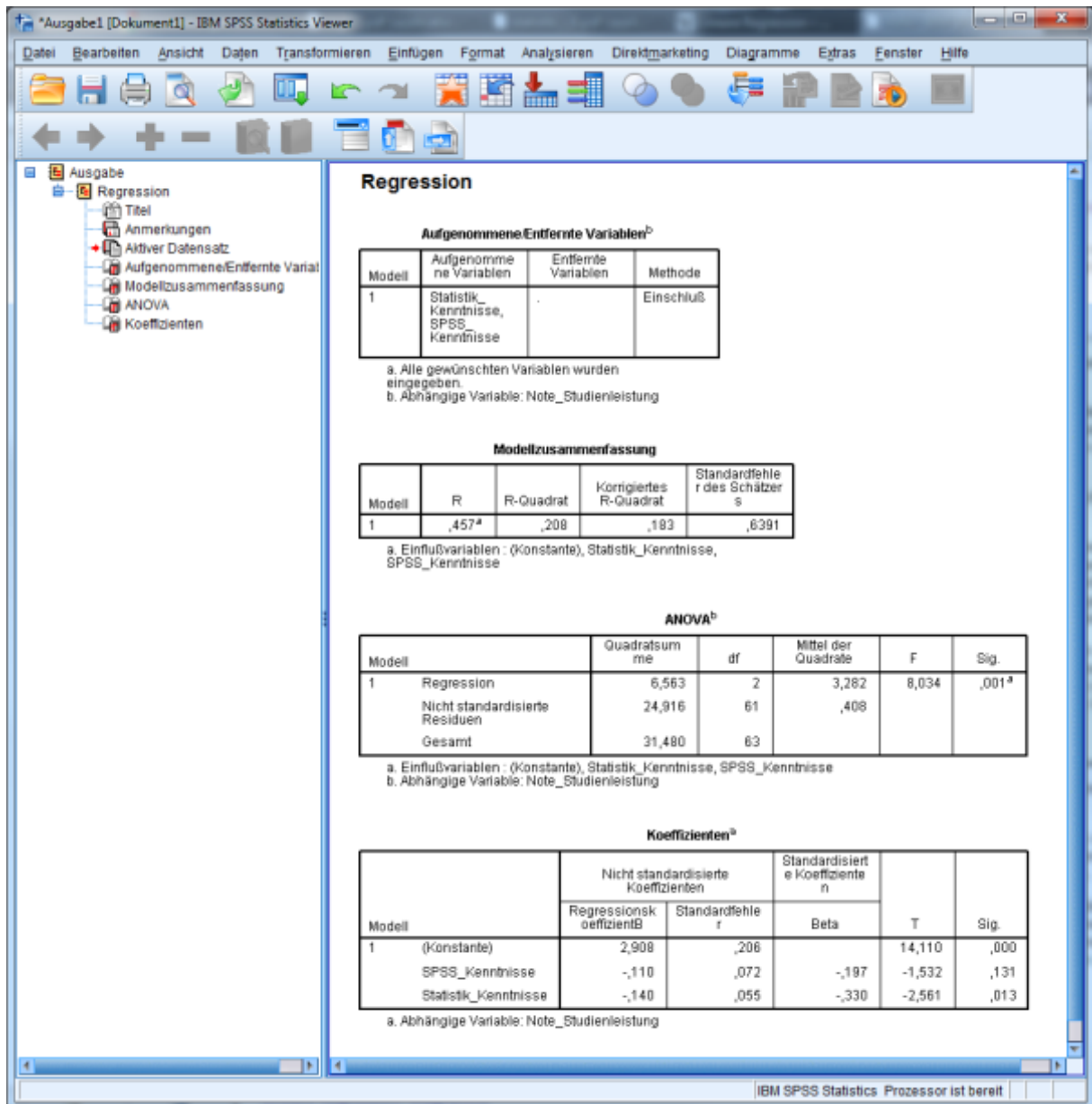
- Modellzusammenfassung:

- 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.



7.3.1 Exkurs Multiple 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

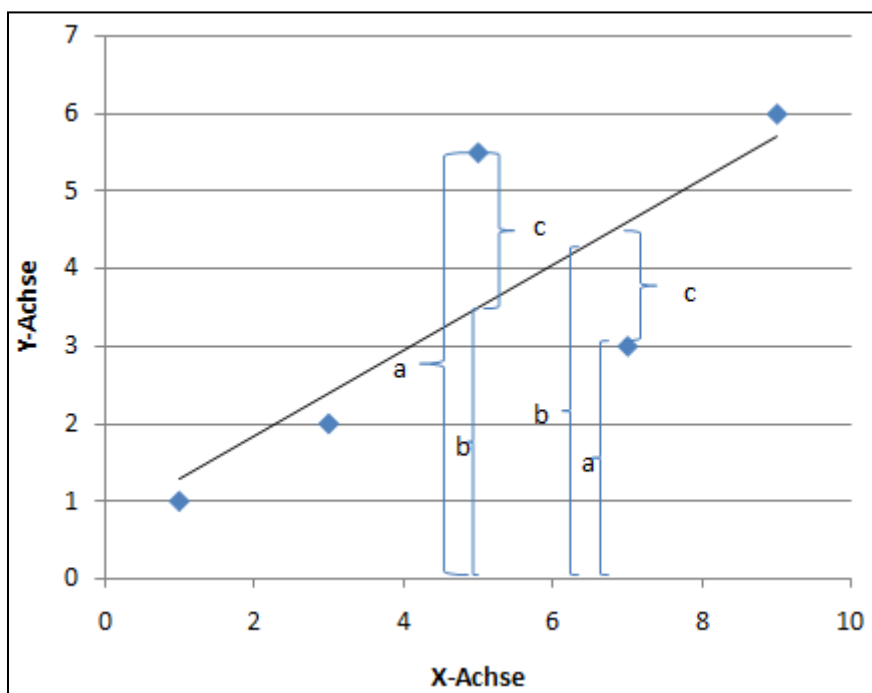
können. Sie werden (zusätzlich zu einer Konstanten c) addiert und mit spezifischen Gewichten (β) versehen, die ihren relativen Anteil an der Vorhersageleistung widerspiegeln:

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

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

- Wichtige Voraussetzungen:
 1. Lineare Zusammenhänge zwischen den Variablen
 2. Homoskedastizität
 3. Normalverteilung der Variablen
 4. Keine Kollinearität
 5. Hohe Reliabilität der Prädiktoren und des Kriteriums

Das Vorgehen der Regressionsanalyse kann für den Fall von nur einer UV grafisch 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 (senkrecht 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.

Beispiel zur Interpretation

Untersucht wird der Einfluss der Facetten der transformationalen Führung (av, pam, fag, hpe, is, isn) und Contingent Reward (cr) auf das Liking.

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,795 ^a	,632	,591	,63679

a. Einflussvariablen: (Konstante), cr, hpe, pam, is, isn, fag, av

➤ 63,2% der Varianz werden durch die uVs erklärt

ANOVA^b

Modell	Quadratsumme	df	Mittel der Quadrate	F	Sig.
1 Regression	43,792	7	6,256	15,428	,000 ^a
Nicht standardisierte Residuen	25,547	63	,406		
Gesamt	69,338	70			

a. Einflussvariablen: (Konstante), cr, hpe, pam, is, isn, fag, av
b. Abhängige Variable: Liking

➤ Das Modell ist signifikant, da .00 < .05

Koeffizienten^a

Modell	Nicht standardisierte Koeffizienten	Standardfehler	Standardisierte Koeffizienten	T	Sig.
1 (Konstante)	,814	,567		1,437	,156
av	,004	,151	,003	,024	,981
pam	,211	,128	,218	1,651	,104
fag	-,005	,109	-,006	-,044	,965
hpe	-,078	,113	-,060	-,687	,495
is	,285	,105	,307	2,715	,009
isn	,348	,137	,318	2,538	,014
cr	,097	,106	,108	,911	,366

a. Abhängige Variable: Liking

➤ nur is und isn signifikant, alle anderen haben keinen Einfluss
➤ isn hat einen größeren Einfluss als is, da $\beta = .318 > \beta = .307$

Sowohl der multiple Korrelationskoeffizient ($R = .795$) als auch der Determinationskoeffizient ($R^2 = .632$) weisen hohe Werte auf. Demnach herrscht ein großer Zusammenhang zwischen den UVs und der AV und es kann ein Großteil der Varianz (63 %) der AV durch die aufgenommenen UVs erklärt werden. Auch die Signifikanz des Gesamtmodells ist gegeben, da der Spalte „Sig.“ in der ANOVA-Tabelle entnommen werden kann, dass $p < .05$ ist. Somit ist R^2 signifikant von Null verschieden.

Bei der Betrachtung der standardisierten Regressionskoeffizienten fällt auf, dass lediglich „is“ (.01 < .05) und „isn“ (.01 < .05) signifikant sind, also nur diese beiden statistisch von Null verschieden sind. Der Vergleich der signifikanten Koeffizienten gibt Aufschluss darüber, welche UV den größeren Einfluss und damit die größere Bedeutsamkeit für die Vorhersage der AV hat. Hier beeinflusst „isn“ ($\beta = .318$) das „Liking“ stärker als „is“ ($\beta = .307$). An dem jeweiligen Vorzeichen lässt sich außerdem die „Richtung“ der Wirkung feststellen, sprich ob die UV positiv oder negativ auf die AV wirkt. In dem vorliegenden Beispiel sind die standardisierten Regressionskoeffizienten von „is“ und „isn“ beide positiv. Als Konsequenz steigt das „Liking“ umso mehr, je höher „is“ und „isn“ ausgeprägt sind. Für alle anderen UVs konnte kein signifikanter Einfluss auf das „Liking“ nachgewiesen werden.

Die Regressionsgleichung kann mithilfe der nicht standardisierten Regressionskoeffizienten aufgestellt werden. Sie ergibt sich als Summe aus der Konstante und den Produkten der jeweiligen Ausprägungen der UVs und dem Regressionskoeffizienten B:

$$liking = 0,814 + 0,004 * av + 0,211 * pam + (-0,005) * fag + (-0,078) * hpe + 0,285 * is + 0,348 * isn + 0,097 * cr$$

Mit dieser Gleichung kann anschließend die Ausprägung des „Liking“ bei vorgegebenen Werten der UVs prognostiziert werden.

Korrelationen

		Liking	TA Contingend Reward	TF Articulating a Vision	TF Providing an appropriate model	TF Fostering the acceptance of group ...	TF High performance expectations	TF Individualized Support	TF Intellectual stimulation
Korrelation nach Pearson	Liking	1,000	,611	,611	,666	,593	-,113	,659	,668
	TA Contingend Reward	,611	1,000	,616	,539	,676	-,121	,605	,611
	TF Articulating a Vision	,611	,616	1,000	,757	,708	,092	,510	,719
	TF Providing an appropriate model	,666	,539	,757	1,000	,672	,032	,559	,694
	TF Fostering the acceptance of group goals	,593	,676	,708	,672	1,000	-,076	,621	,572
	TF High performance expectations	-,113	-,121	,092	,032	-,076	1,000	-,312	,153
	TF Individualized Support	,659	,605	,510	,559	,621	-,312	1,000	,464
	TF Intellectual stimulation	,668	,611	,719	,694	,572	,153	,464	1,000

7.3.2 Hierarchische Regression

Koeffizienten^a

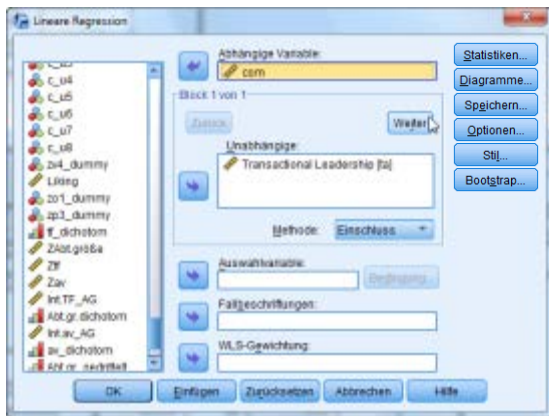
Modell	Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Sig.	Korrelationen		
	Regressionskoeffizient B	Standardfehler	Beta			Nullter Ordnung	Partiell	Teil
1 (Konstante)	,814	,567		1,437	,156			
TA Contingend Reward	,097	,106	,108	,911	,366	,611	,114	,070
TF Articulating a Vision	,004	,151	,003	,024	,981	,611	,003	,002
TF Providing an appropriate model	,211	,128	,218	1,651	,104	,666	,204	,126
TF Fostering the acceptance of group goals	-,005	,109	-,006	-,044	,965	,593	-,006	-,003
TF High performance expectations	-,078	,113	-,060	-,687	,495	-,113	-,086	-,053
TF Individualized Support	,285	,105	,307	2,715	,009	,659	,324	,208
TF Intellectual stimulation	,348	,137	,318	2,538	,014	,668	,305	,194

a. Abhängige Variable: Liking

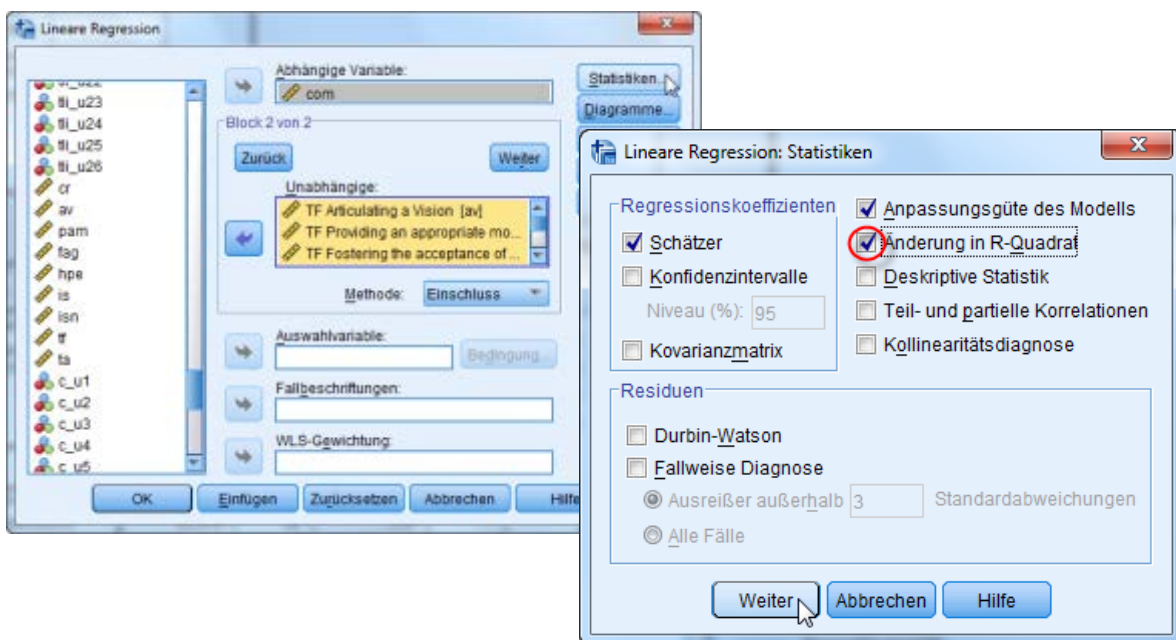
Bei einer hierarchischen Regressionsanalyse werden die UVs blockweise in die Gleichung aufgenommen. Auf diese Weise kann die Vorhersageleistung einzelner UVs oder Gruppen von UVs ermittelt werden. Sollen beispielsweise der Einfluss transaktionaler Führung („ta“) und der Einfluss der Facetten transformationaler 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

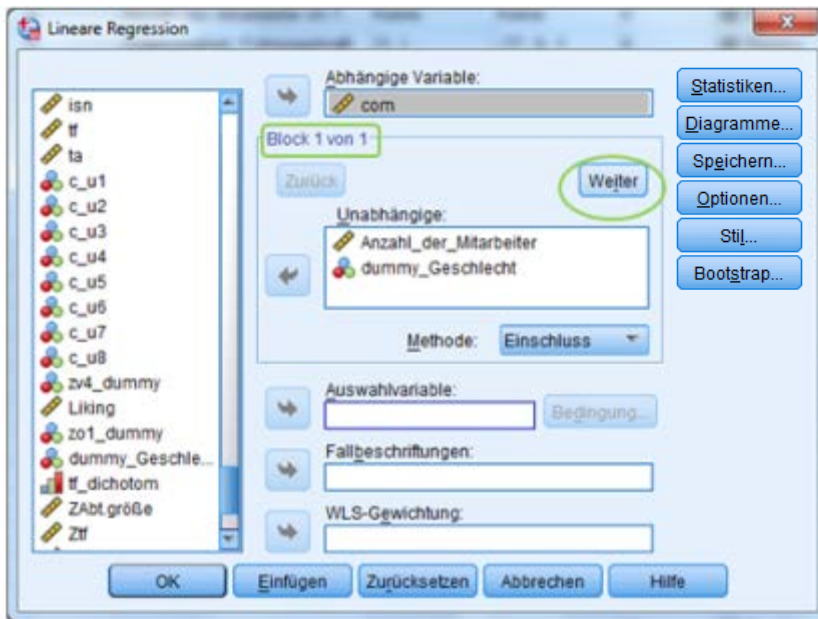
c. Einflußvariablen : (Konstante), Transactional Leadership, TF High performance expectations, TF Providing an appropriate model , TF Individualized Support, TF Intellectual stimulation, TF Fostering the acceptance of group goals , TF Articulating a Vision

In der Ausgabe werden 2 Modelle 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.

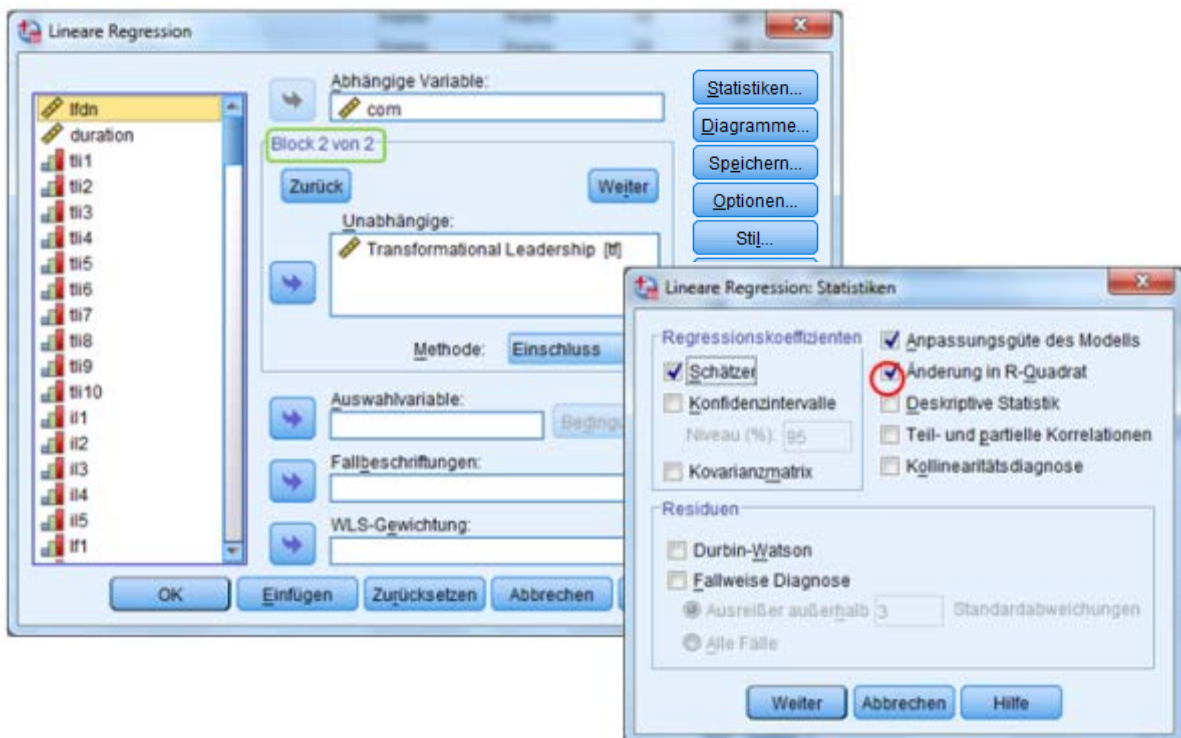
7.3.3 Kontrollvariablen

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 der Einfluss der Kontrollvariablen ist signifikant, würde aber nicht berücksichtigt, könnte es der Fall sein, dass der Einfluss von transformationaler Führung auf Commitment verfälscht wäre.

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.

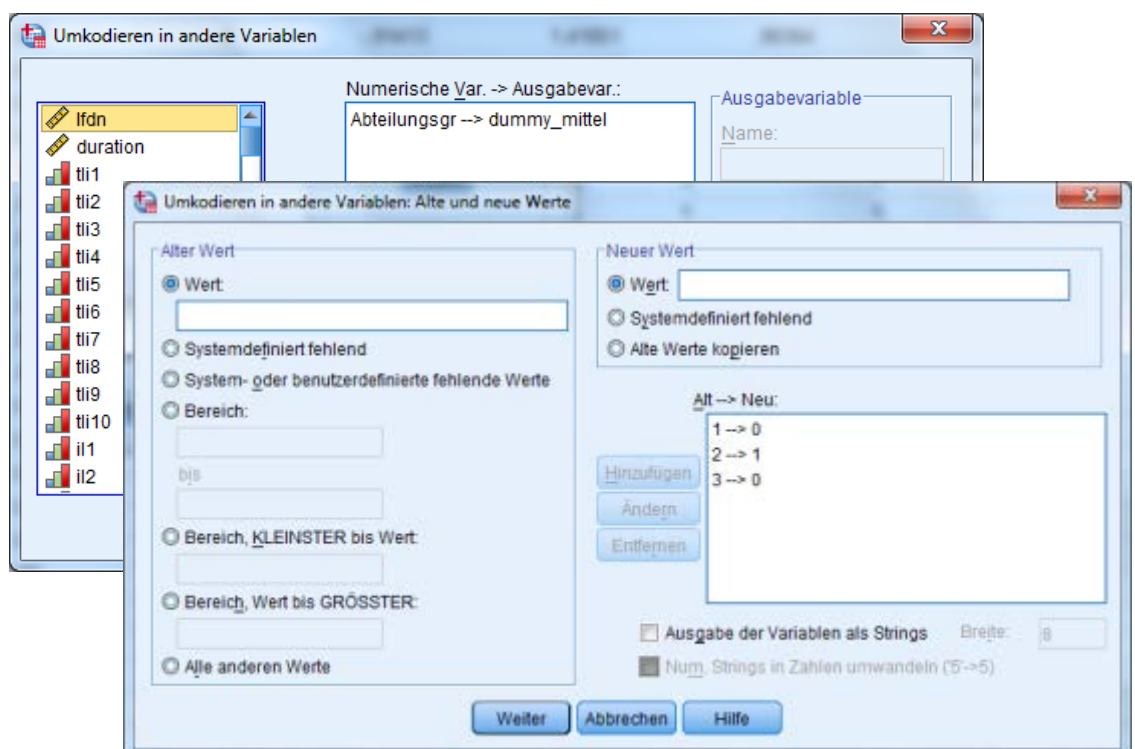
Als veranschaulichendes Beispiel wird der Einfluss der Abteilungsgröße (ordinale UV mit 1=klein, 2=mittel, 3=groß) auf das Commitment (intervallskalierte AV) betrachtet. Wählt man beispielsweise „klein“ als Referenzkategorie, muss man dieser in jeder Dummy-Variablen den Wert 0 zuordnen. Nun müssen also 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.

y (com)	dummy_mittel	dumm_groß	Abteilungsgr
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) [Codieren](#) der Dummy-Variablen „dummy_mittel“ und „dummy_groß“ über *Transformieren/Uncodieren in andere Variablen/Alte und neue Werte*
 - a. Der Referenzkategorie wird auf allen Dummy-Variablen der Wert 0 zugewiesen
 - b. 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, $M = 3.440$. 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 zwischen mittleren-großen Abteilungen ablesen. Dieser ergibt sich aus der Differenz aus den Koeffizienten b_1 und b_2 . Hier: $0.414 - (-0.008) = 0.422$.

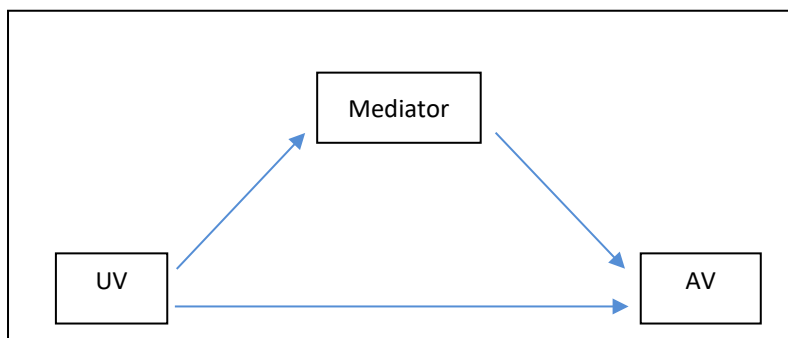
Als Ergebnis lässt sich festhalten, dass das Commitment in mittelgroßen Abteilungen am größten ist.

7.4 Mediatoreffekt

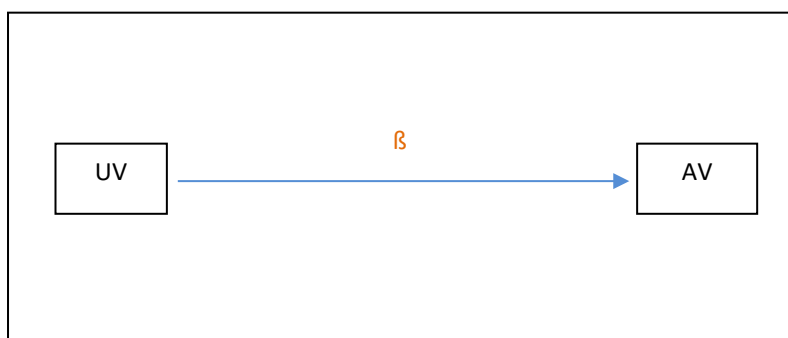
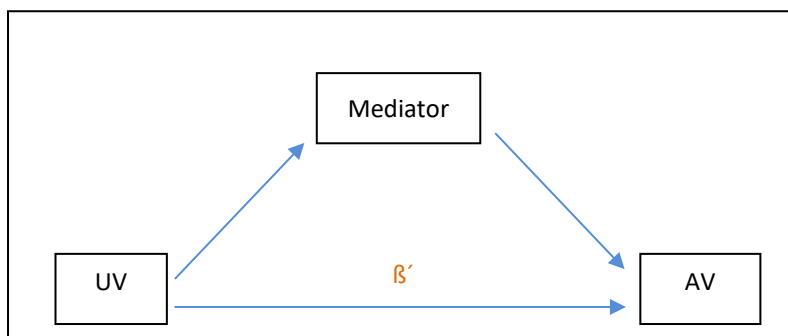
Ein Mediatoreffekt liegt vor, wenn ein sogenannter Mediator zwischen einer unabhängigen und einer abhängigen Variablen vermittelt, den Zusammenhang zwischen UV und AV also mediiert. Zur Überprüfung einer solchen Mediation in SPSS werden drei zentrale Schritte unterschieden.

- 1) Überprüfung, ob zwischen allen Variablen signifikante Korrelationen bestehen.

Das heißt, dass zunächst geprüft wird, ob die unabhängige Variable die Abhängige signifikant beeinflusst. Im Anschluss daran wird dann der Einfluss der unabhängigen Variable auf die Mediatorvariable hinsichtlich der Signifikanz geprüft usw. Ist die Bedingung der Signifikanz erfüllt, sind also Prädiktor, Kriterium und Mediator signifikant miteinander korreliert, kann die Überprüfung der Mediation fortgesetzt werden.



- 2) Im zweiten Schritt wird nun eine lineare multiple Regression berechnet. Hierbei soll mit dem Mediator und dem Prädiktor gemeinsam das Kriterium vorhergesagt werden. Das heißt, dass an dieser Stelle sowohl UV als auch der Mediator als unabhängige Variablen auf die abhängige Variable AV wirken. Ist der Einfluss (β') der unabhängigen Variable auf die abhängige Variable unter Einbindung des Mediators kleiner als der alleinige Einfluss (β), so liegt ein Mediatoreffekt vor.



Falls $\beta' < \beta \rightarrow$ Mediatoreffekt

Falls $\beta' = 0 \rightarrow$ Vollständige Mediation

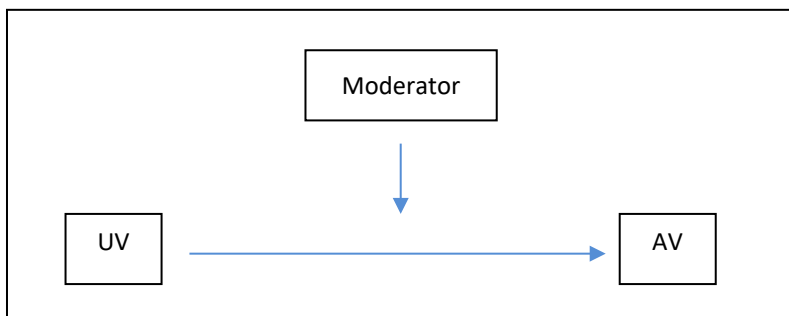
Partieller Mediatoreffekt: Ist der direkte Einfluss der UV auf die AV bei Hinzunahme des Mediators in die Regression (β') zwar kleiner als ohne den Mediator (β), aber immer noch vorhanden, spricht man von einer partiellen Mediation.

Vollständiger Mediatoreffekt: Ist ein direkter Einfluss der UV auf die AV bei Hinzunahme des Mediators nicht mehr vorhanden ($\beta' = 0$, Ergebnis nicht signifikant), liegt eine vollständige Mediation vor. Hierbei wird der Zusammenhang zwischen UV und AV in vollem Umfang über den Mediator vermittelt.

- 3) Liegt keine vollständige Mediation vor, kann die Wirkung des indirekten Pfades (über den Mediator) noch durch den Sobel-Test geprüft werden. Dieser Test überprüft, ob nicht zumindest eine partielle Mediation vorliegt und somit der direkte Pfad (zwischen UV und AV) durch die Kontrolle des Mediators signifikant reduziert wird. Der Sobel-Test lässt sich online unter <http://quantpsy.org/sobel/sobel.htm> durchführen.

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.

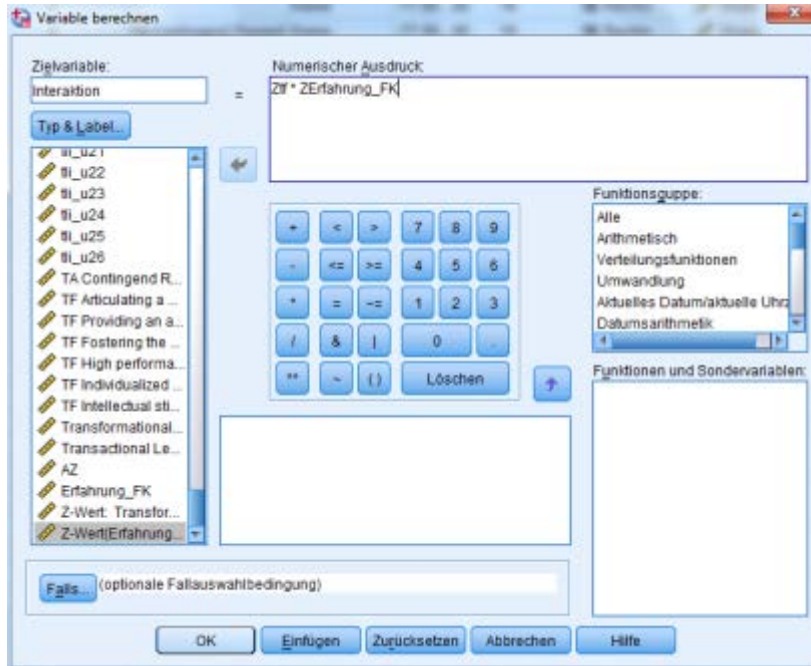
Eine Moderationshypothese wird mithilfe der Regressionsanalyse geprüft, falls der zu prüfende Moderator mindestens intervallskaliert ist. Hierzu sind drei zentrale Schritte notwendig:

Schritt 1: Standardisierung

Bei der Standardisierung handelt es sich um eine sogenannte z-Standardisierung. Hierbei werden die unabhängige Variable (UV) und der Moderator z-standardisiert, das heißt, dass die ursprüngliche Variable in eine neue Variable transformiert wird. An dieser Stelle wird die Differenz zwischen dem beobachteten Wert und dem Mittelwert einer Variablen errechnet und im Anschluss durch die Standardabweichung dividiert. Diese z-standardisierten Variablen besitzen eine Normalverteilung (Mittelwert= 0, Standardabweichung= 1). Diese Standardisierung bildet die Grundlage für den zweiten Schritt.

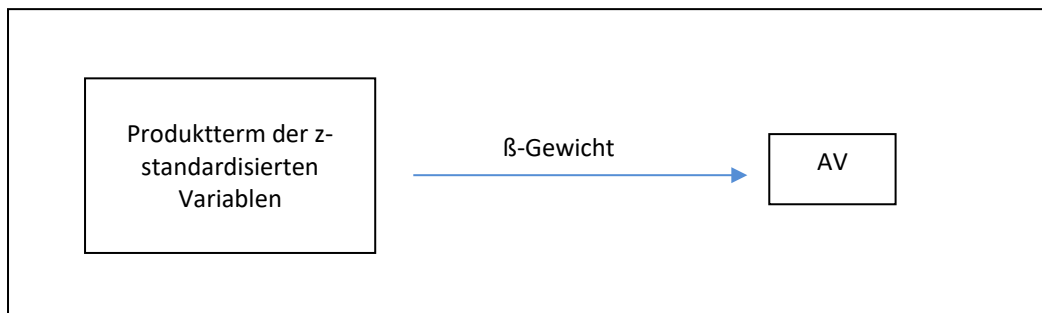
Schritt 2: Berechnung der Interaktionsvariablen

Die Interaktionsvariable wird berechnet als Produkt aus der z-standardisierten UV und dem z-standardisierten potenziellen Moderator (s. [Variable berechnen](#)).

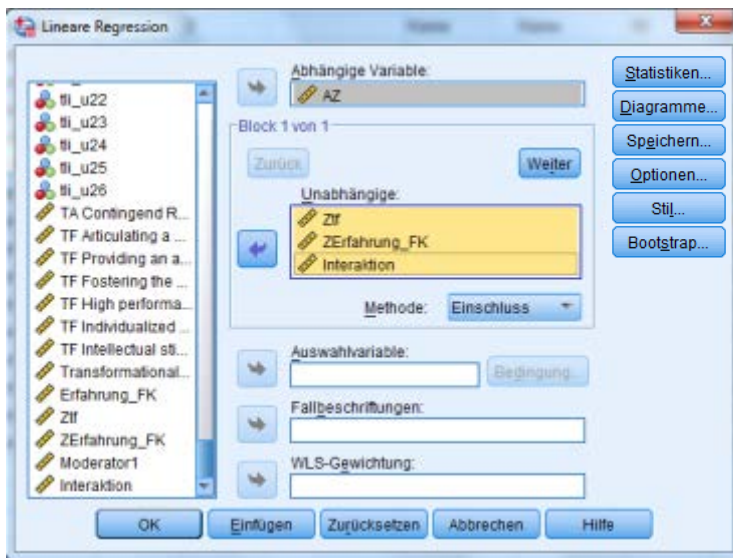


Schritt 3: Regression

Zur Überprüfung der potenziellen Moderation wird eine Regression durchgeführt. Hierbei wird die zuvor berechnete Interaktionsvariable (der Produktterm) sowie die z-standardisierten Variablen als unabhängige Variable verwendet. Ist der Einfluss (β -Gewicht) der Interaktionsvariablen auf die AV signifikant, liegt eine Moderation vor. Das heißt, der Zusammenhang zwischen UV und AV hängt von der Ausprägung des Moderators ab.



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 haben die beiden UVs (Ztf und ZErfahrung_FK) einen signifikanten Einfluss auf die Arbeitszufriedenheit, die Interaktionsvariable „Interaktion“ jedoch nicht. Daher konnte in diesem Beispiel kein Moderationseffekt nachgewiesen werden.

ANOVA^b

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.
		Regressionskoeffizient	Standardfehler			
1	(Konstante)	3,495	,098		35,591	,000
	Ztf	,249	,099	,272	2,517	,014
	ZErfahrung_FK	-,297	,104	-,309	-2,864	,005
	Interaktion	-,063	,092	-,074	-,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, indem 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.

Dichotome (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 (nur bei *nicht dichotomen* Variablen)

Ist der mutmaßliche Moderator kontinuierlich, so muss zuerst eine neue, dichotome Variable erstellt werden (z.B. mittels Mediansplit, s. [visuelle Klassierung](#)). 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

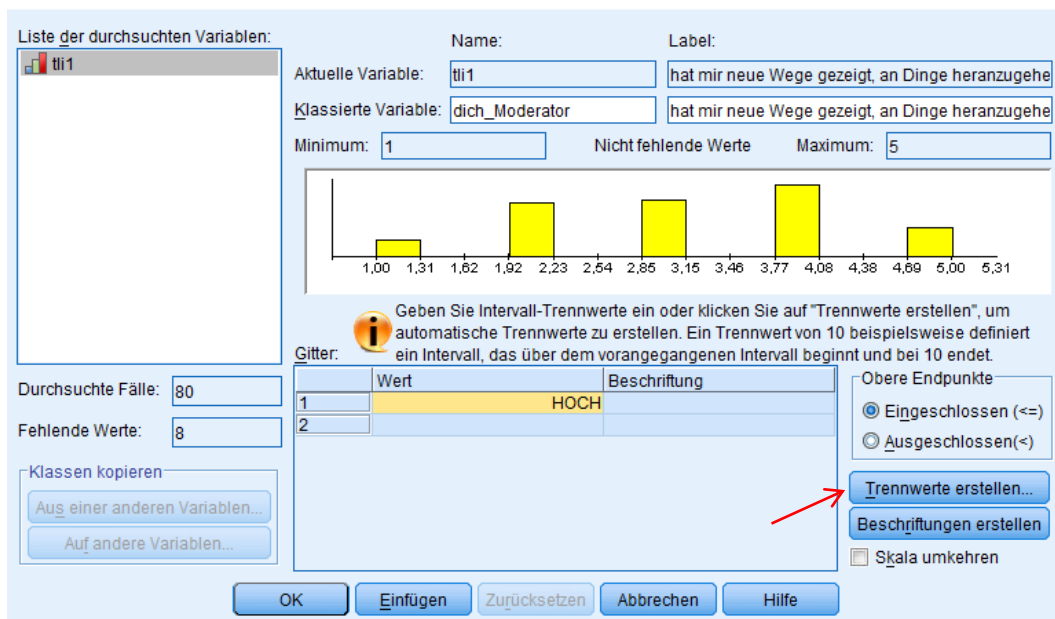
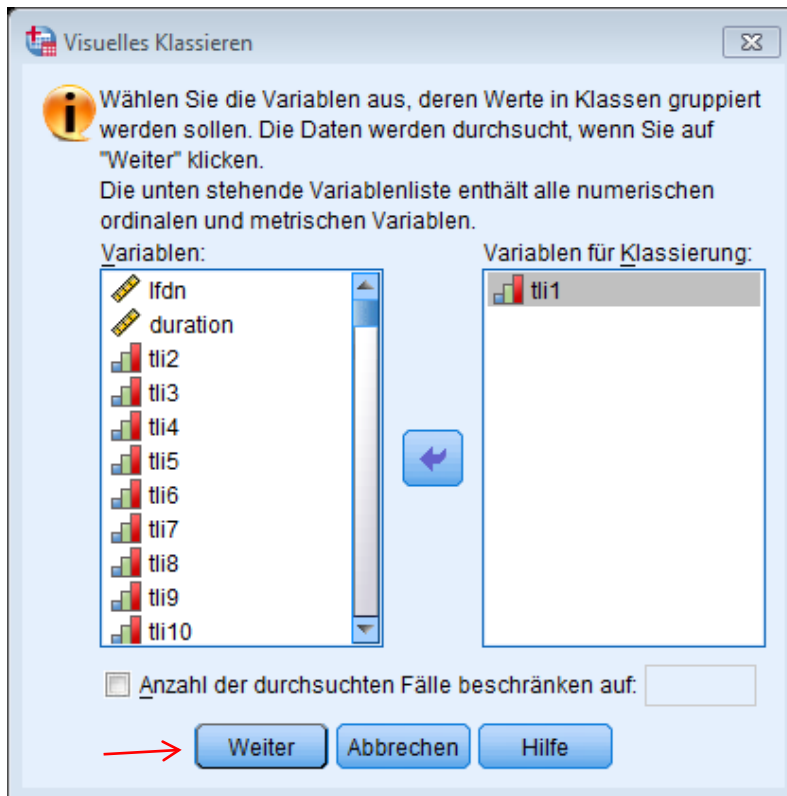
Im dritten Schritt wird geprüft, ob sich die beiden Korrelationskoeffizienten signifikant voneinander unterscheiden (s. [Vergleich von Korrelationskoeffizienten](#)).

Moderation: Grafiken erstellen:

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).

1. Die Moderatorvariable und die UV dichotomisieren (s. [visuelle Klassierung](#)): In SPSS unter Transformieren – Visuelles Klassierung – entsprechende Variable auswählen – oben einen Namen für die „klassierte Variable“ zuweisen (bspw. dich_Moderator). Unter Trennwerte erstellen – Option gleiche Perzentile... wählen – Anzahl der Trennwerte auf 1 setzen – SPSS ergänzt automatisch die

Breite auf 50 %. Abschließend zuweisen und es wird eine neue Variable erstellt. Diesen Schritt sowohl für den Moderator als auch für die UV durchführen.



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.

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

→

2. Anschließend Liniendiagramm zeichnen. Grafik – Diagrammerstellung – Galerie Linie auswählen – „mehrere Linien“ (das Diagramm mit den Farben) auswählen. Oben die Variablen in das Diagramm schieben: Y-Achse = AV, X-Achse = dichotomisierte UV und Farbe festlegen = dichotomisierter Moderator.

Variablen: Diagrammvorschau verwendet Beispieldaten

lfdn
 duration
 tti1
 tti2
 tti3
 tti4
 tti5
 tti6
 tti7
 tti8
 tti9
 tti10
 ti1
 ti2
 ti3

stimme gar nicht zu

Y-Achse P

stimme gar nicht zu [Mehr...]

Farbe festlegen

tt2

Galerie Grundelemente Gruppen/Punkt-ID Titel/Fußnoten

Auswählen aus:

Favoriten

Balken

Linie

Fläche

Kreis/Polar

Streu-/Punktdia...

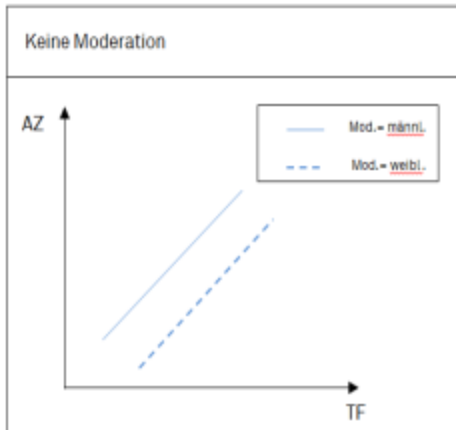
Histogramm

Hoch-Tief

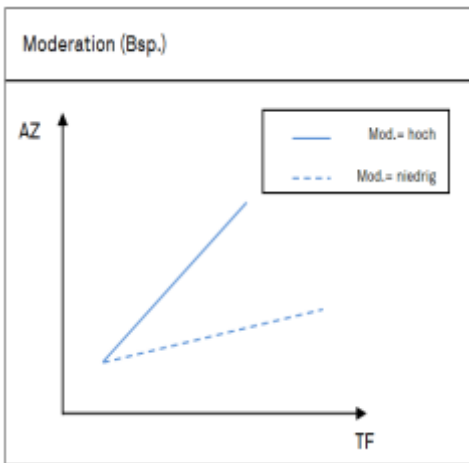
Boxplot

Doppelachsen

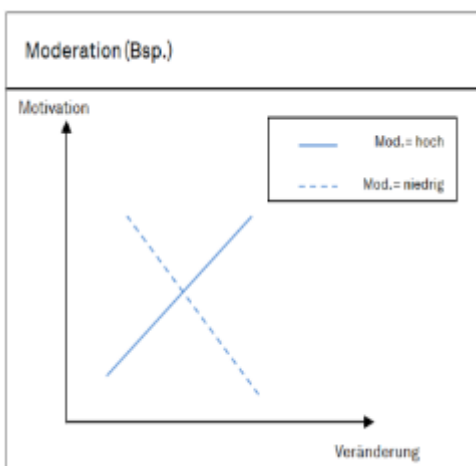
Grafische Darstellung möglicher Moderatoreffekte:



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 nützlichen [Online-Rechner](#)¹ in den lediglich die Pearson-Korrelationskoeffizienten, der Stichprobenumfang und das gewünschte Signifikanzniveau einzutragen sind.

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 errechneten Korrelationen, der Stichprobenumfang sowie das Signifikanzniveau in den Rechner eingetragen werden.

¹ Hemmerich, W. (2017). StatistikGuru: Korrelationen statistisch vergleichen. Retrieved from <https://statistikguru.de/rechner/korrelationen-vergleichen.html>

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.

Rechner

unabh. Gruppen

abh. Gruppen (Drittvariable)

abh. Gruppen (keine Drittvariable)

1. Korrelation 0,517

1. Korrelation

2. Korrelation 0,154

2. Korrelation

n_1 41

Größe der ersten Gruppe

n_2 35

Größe der zweiten Gruppe

Alphaniveau .05

Grenze für ein signifikantes Ergebnis, meist .05 oder .01 .

zweiseitig

Alternativhypothese.

Signifikanzen berechnen

Mit Klick auf die Schaltfläche „Signifikanzen berechnen“ wird die Berechnung ausgeführt. Unterhalb dieser Schaltfläche wird nun automatisch die Nullhypothese formuliert, die besagt, dass sich die Korrelationen für die zwei unterschiedlichen Stichproben nicht signifikant voneinander unterscheiden.

Signifikanzen berechnen

Ergebnis

Nullhypothese H_0 : r_1 ist gleich r_2

Alternativhypothese H_1 : r_1 ist ungleich r_2

Autoren

Statistik, p-Wert

Fisher's z (1925)

$z = 1.7380, p = .082$

Anmerkung: * $p < .05$, ** $p < .01$, *** $p < .001$

Die Nullhypothese muss in diesem Beispiel angenommen werden, da die errechnete Statistik und somit der Unterschied zwischen den Korrelationskoeffizienten nicht signifikant ist. 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.

7.6 Exkurs: SPSS-Makro PROCESS

PROCESS ist ein Modellierungswerkzeug für die Analyse von Mediations- und Moderationsmodellen in SPSS. Geschrieben wurde das Makro von Andrew F. Hayes und wird Nutzern auf seiner Website zum kostenlosen Download bereitgestellt. Mit PROCESS lassen sich unterschiedliche Analysen durchführen, welche in der zugehörigen Literatur genauer beschrieben werden (Andrew F. Hayes (2018): Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach).

Im Folgenden wird nach einer kurzen Installationsanleitung die Umsetzung sowohl einer Mediations- als auch einer Moderationsanalyse in SPSS dargestellt.

7.6.1 Installationsanleitung

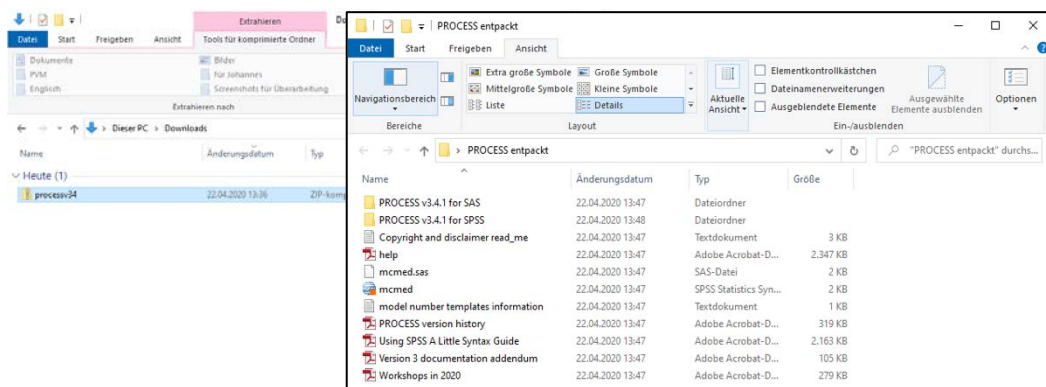
Schritt 1: Installationsdatei herunterladen

Auf dieser Website steht das PROCESS-Makro kostenlos zum Download zur Verfügung.

<http://processmacro.org/download.html>

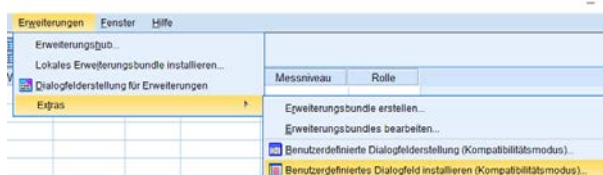
Schritt 2: zip-Datei in Zielordner extrahieren

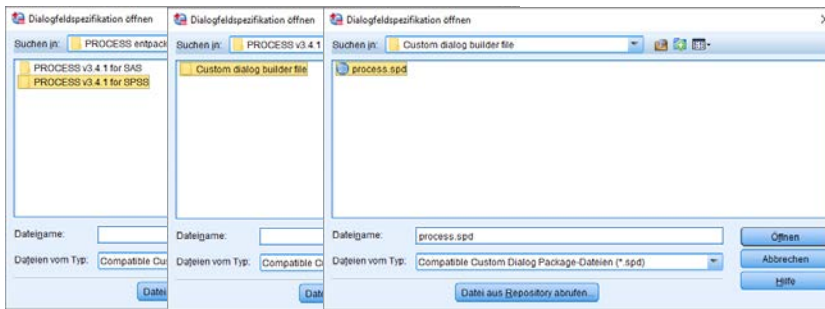
Die heruntergeladene zip-Datei muss nun entpackt werden und in einem gewünschten Ordner gespeichert werden.



Schritt 3: Makro in SPSS installieren

Das Makro wird innerhalb von SPSS installiert. Unter *Erweiterungen/Extras/Benutzerdefiniertes Dialogfeld installieren* wählt man den in Schritt 2 ausgewählten Ordner aus. In dem Ordner *PROCESS v3.4.1 for SPSS* (Bezeichnung kann je nach Version variieren)/*Custom dialog builder file* befindet sich die Installationsdatei *process.spd*. Diese öffnet bzw. führt man aus.

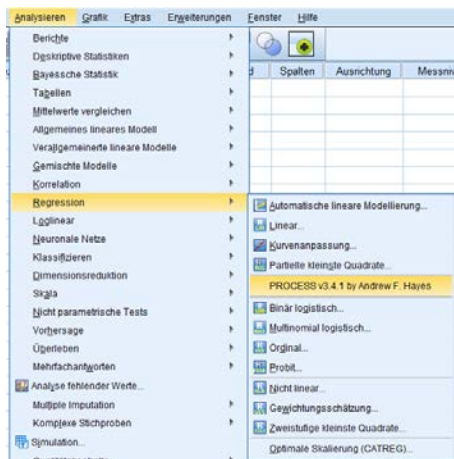




Die erfolgreiche Installation bestätigt SPSS in folgendem Dialogfenster.

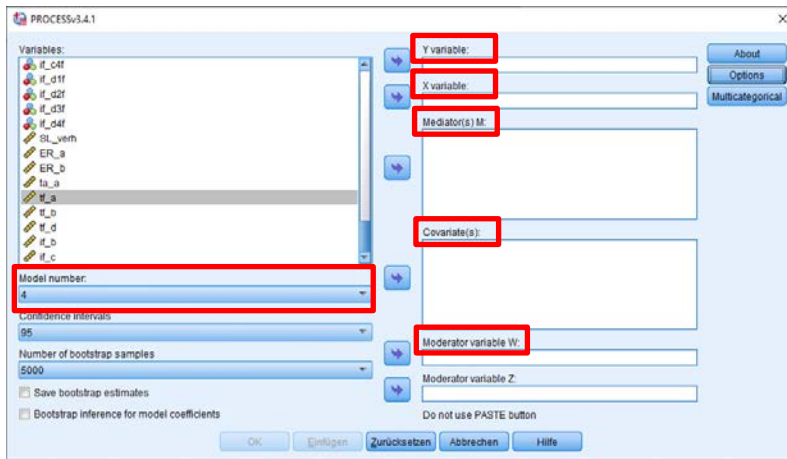


Unter *Analysieren/Regression/PROCESS v3.4.1 by Andrew F.Hayes* kann nun das Makro ausgeführt werden.



7.6.2 Mediation und Moderation mit PROCESS

Um eine Mediation oder Moderation mit PROCESS zu berechnen, ruft man zunächst das Makro, wie bereits beschrieben, unter *Analysieren/Regression/PROCESS v3.4.1 by Andrew F.Hayes* auf. Es erscheint die PROCESS-Oberfläche, in der nun die entsprechenden Eingaben und Einstellungen vorgenommen werden. Anhand der *Model number* wählt man aus, welches Modell berechnet werden soll. Eine Übersicht zu den verschiedenen Modellen befindet sich im Appendix der zugehörigen Lektüre. Alternativ stellen diverse Websites Informationen zu den Modellen zur Verfügung.

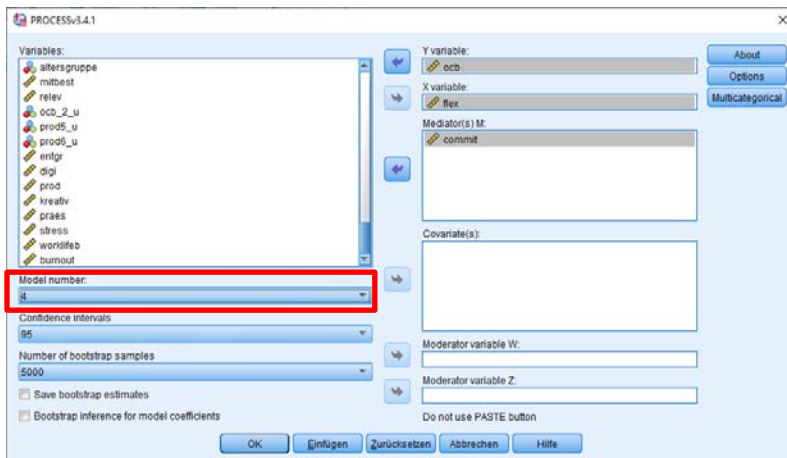


X variable: UV
 Y variable: AV
 Mediator(s) M: Mediatorvariable
 Moderator variable W: Moderatorvariable
 Covariate(s): Kontrollvariablen

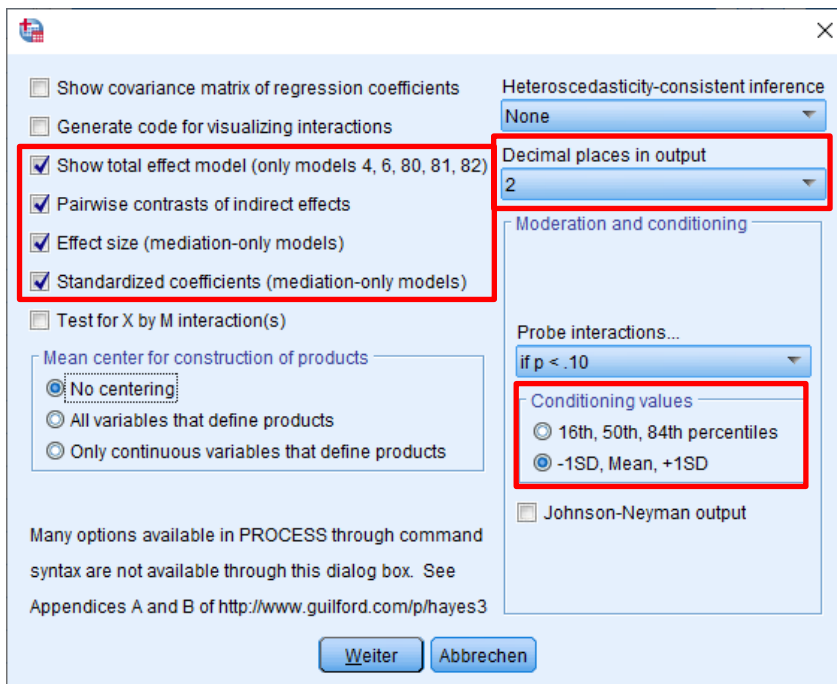
7.6.2.1 Mediation anhand eines Beispiels



In diesem Beispiel wird geprüft, ob der Zusammenhang zwischen der UV (*flex*) mit der AV (*ocb*) über den Mediator (*commit*) vermittelt wird. Für eine einfache Mediation wählt man Modell 4 aus und bestimmt die zu untersuchenden Variablen, in dem genannten Beispiel also *flex* als „X variable“, *ocb* als „Y variable“ und *commit* als „Mediator(s) M“.



Unter *Options* werden nun folgende Einstellungen vorgenommen und mit „Weiter“ bestätigt.



Mit einem Klick auf „Ok“ wird die Berechnung ausgeführt.

Die Ausgabe gliedert sich in Teilmodelle, die im Folgenden dargestellt und erklärt werden.

```

*****
Model : 4
Y : ocb
X : flex
M : commit

Sample
Size: 741

*****
OUTCOME VARIABLE:
commit

Model Summary
R      R-sq      MSE      F      df1      df2      p
,22    ,05      ,63     35,96    1,00    739,00    ,00

Model
      coeff      se      t      p      LLCI      ULCI
constant  2,70    ,07    36,11    ,00    2,55    2,84
flex      ,18    ,03     6,00    ,00    ,12    ,23

Standardized coefficients
flex      coeff
          ,22

*****

```

= R², beschreibt hier wie viel Prozent der Varianz in der Mediatorvariable durch UV erklärt wird

Einfluss von UV auf Mediator

flex

↗

commit

signifikant, wenn p < .05 bzw. Wert 0 nicht im Konfidenzintervall [LLCI; ULCI] enthalten

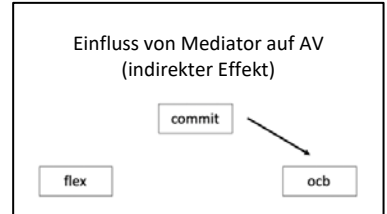
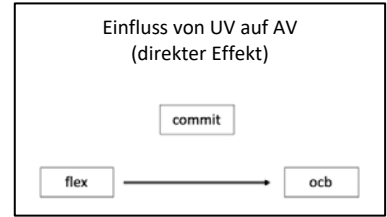
nicht standardisierte Regressionskoeffizienten (B)

standardisierter Regressionskoeffizient (β)

UV (flex) hängt signifikant mit Mediator (commit) zusammen

OUTCOME VARIABLE:
ocb

Model Summary							
	R	R-sq	MSE	F	df1	df2	p
	,48	,23	,77	110,92	2,00	738,00	,00
Model							
	coeff	se	t	p	LLCI	ULCI	
constant	3,36	,14	24,55	,00	3,09	3,63	
flex	,02	,03	,55	,58	-,05	,08	
commit	,59	,04	14,42	,00	,51	,66	
Standardized coefficients							
	coeff						
flex	,02						
commit	,48						

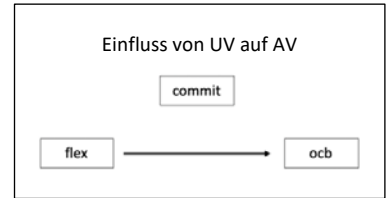


UV (flex) hängt nicht signifikant mit AV (ocb) zusammen
Mediator (commit) hängt signifikant mit AV (ocb) zusammen

***** TOTAL EFFECT MODEL *****

OUTCOME VARIABLE:
ocb

Model Summary							
	R	R-sq	MSE	F	df1	df2	p
	,12	,01	,99	10,95	1,00	739,00	,00
Model							
	coeff	se	t	p	LLCI	ULCI	
constant	4,94	,09	53,00	,00	4,76	5,12	
flex	,12	,04	3,31	,00	,05	,19	
Standardized coefficients							
	coeff						
flex	,12						



UV (flex) hängt signifikant mit AV (ocb) zusammen

***** TOTAL, DIRECT, AND INDIRECT EFFECTS OF X ON Y *****

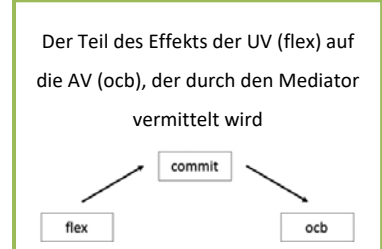
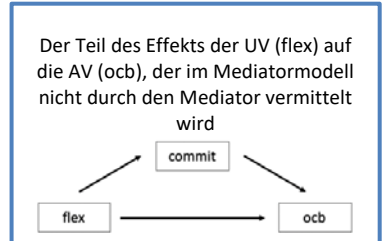
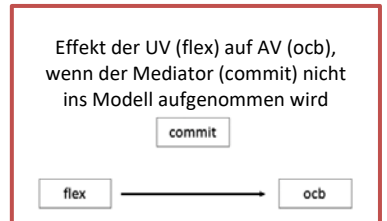
Total effect of X on Y								
Effect	se	t	p	LLCI	ULCI	c_ps	c_cs	
,12	,04	3,31	,00	,05	,19	,12	,12	

Direct effect of X on Y								
Effect	se	t	p	LLCI	ULCI	c'_ps	c'_cs	
,02	,03	,55	,58	-,05	,08	,02	,02	

Indirect effect(s) of X on Y:				
Effect	BootSE	BootLLCI	BootULCI	
commit	,10	,02	,07	,14

Partially standardized indirect effect(s) of X on Y:				
Effect	BootSE	BootLLCI	BootULCI	
commit	,10	,02	,07	,14

Completely standardized indirect effect(s) of X on Y:				
Effect	BootSE	BootLLCI	BootULCI	
commit	,10	,02	,07	,14



Anmerkungen zu TOTAL, DIRECT, AND INDIRECT EFFECTS OF X ON Y:

Sofern nicht anders angegeben, stellen die Werte zu „Effect“ die nicht standardisierten Regressionskoeffizienten dar (siehe. „Completely standardized indirect effect(s) of X on Y).

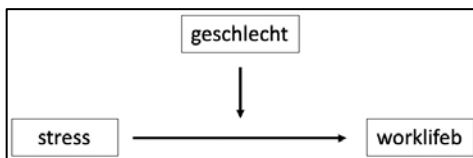
Der „Total effect“ sollte i. d. R. signifikant sein (hier erfüllt).

Der „Direct effect of X on Y“ sollte kleiner ausfallen als der totale Effekt (hier erfüllt). Ist dieser Effekt statistisch signifikant, ist dies ein Indiz für eine partielle Mediation. Ist der Effekt jedoch nicht signifikant (hier der Fall), ist dies ein Indiz für eine vollständige Mediation.

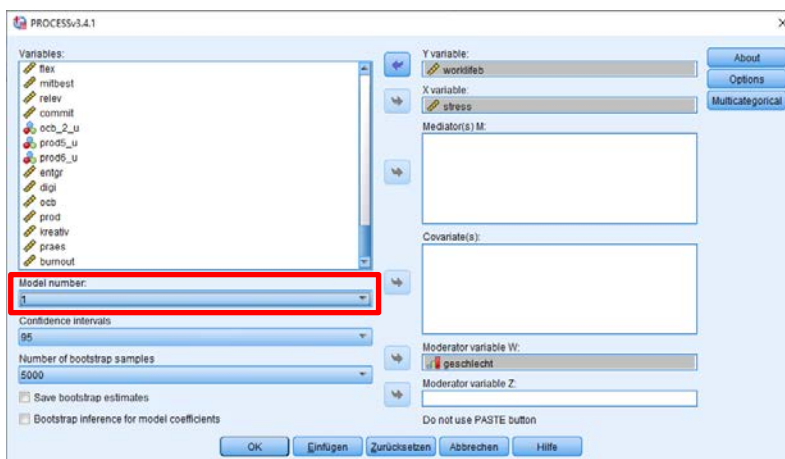
Der „Indirect effect(s) of X on Y“ muss statistisch signifikant sein, damit eine Mediation vorliegt (hier erfüllt, da der Wert 0 nicht im Konfidenzintervall [BootLLCI; BootULCI] enthalten ist).

Für dieses Beispiel lässt sich also bestätigen, dass der Zusammenhang zwischen der UV (flex) mit der AV (ocb) über den Mediator (commit) vermittelt wird.

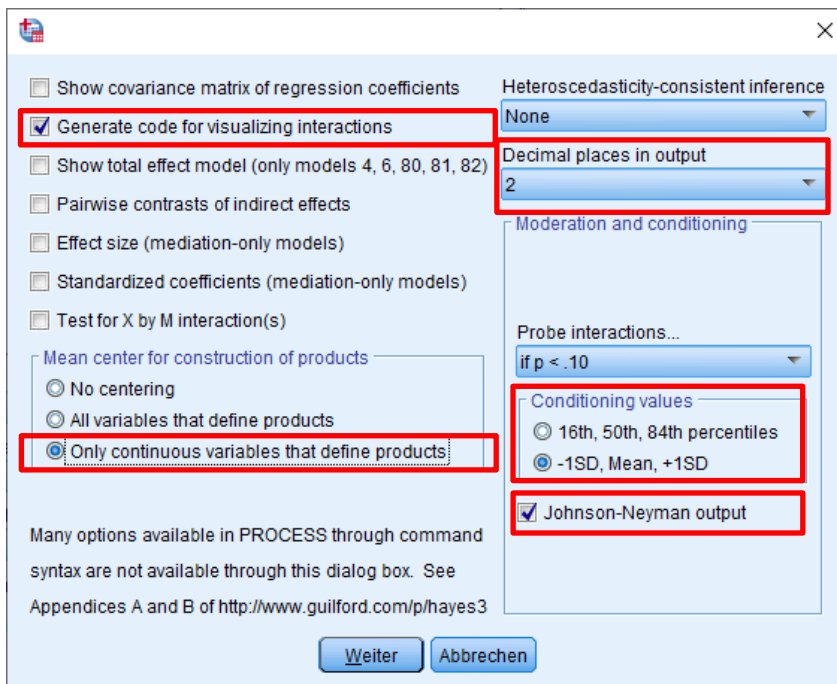
7.6.2.2 Moderation anhand eines Beispiels



In diesem Beispiel wird geprüft, ob der Zusammenhang zwischen der UV (stress) mit der AV (worklifeb) durch die Moderatorvariable (geschlecht) moderiert wird. Für eine Moderation mit einer Moderatorvariable wählt man Modell 1 aus.



Unter *Options* werden nun weitere Einstellungen vorgenommen.



Die Ausgabe wird im Folgenden dargestellt und erklärt.

```
*****
Model : 1
Y : worklife
X : stress
W : geschlec
```

```
Sample
Size: 698
```

```
*****
OUTCOME VARIABLE:
worklife
```

Model Summary	R	R-sq	MSE	F	df1	df2	p
	,28	,08	1,04	19,71	3,00	694,00	,00

= R², beschreibt hier wie viel Prozent der Varianz im Interaktionsmodell erklärt werden

Model	coeff	se	t	p	LLCI	ULCI
constant	2,32	,12	19,15	,00	2,08	2,55
stress	,71	,16	4,58	,00	,41	1,02
geschlec	,11	,08	1,36	,18	-,05	,26
Int_1	-,24	,10	-2,32	,02	-,44	-,04

signifikant, wenn p < .05 bzw. Wert 0 nicht im Konfidenzintervall [LLCI; ULCI] enthalten
Hier: Interaktionsterm signifikant

Product terms key:
Int_1 : stress x geschlec

Zur Überprüfung einer Moderation wird der Interaktionsterm Int_1 zwischen UV (stress) und Moderator (geschlecht) betrachtet

```
Test(s) of highest order unconditional interaction(s):
R2-chng F df1 df2 p
X*W ,01 5,39 1,00 694,00 ,02
```

```
-----
Focal predict: stress (X)
Mod var: geschlec (W)
```

```
Conditional effects of the focal predictor at values of the moderator(s):
```

geschlec	Effect	se	t	p	LLCI	ULCI
1,00	,48	,07	7,00	,00	,34	,61
2,00	,24	,08	3,14	,00	,09	,39

Anmerkung: Bei Moderatoranalysen gibt PROCESS ausschließlich nicht standardisierte Regressionskoeffizienten aus.

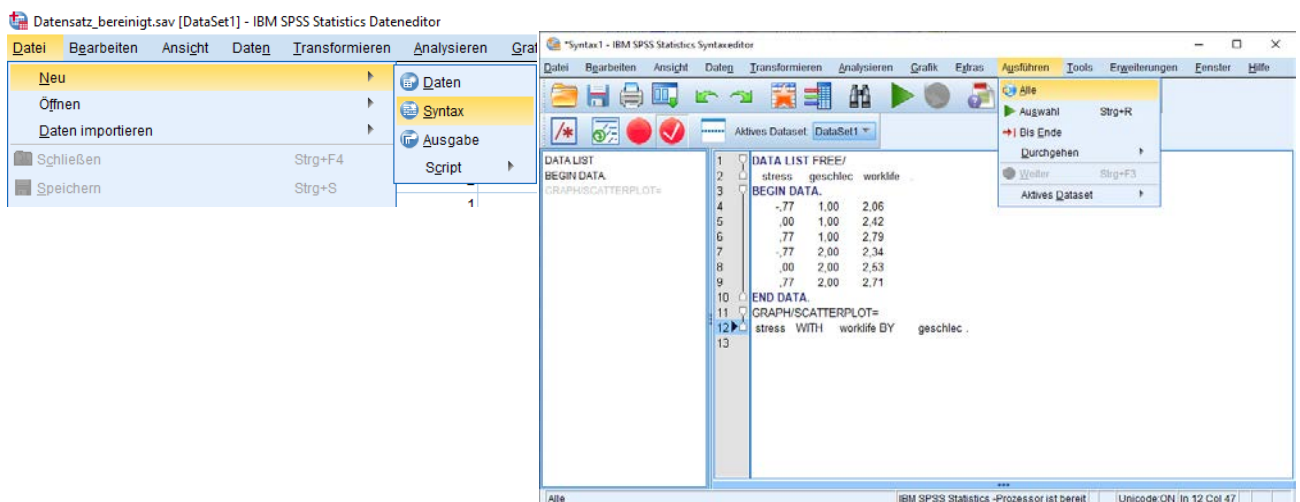
Data for visualizing the conditional effect of the focal predictor:
 Paste text below into a SPSS syntax window and execute to produce plot.

```
DATA LIST FREE/
  stress geschlec worklife .
BEGIN DATA.
  -,77      1,00      2,06
  ,00      1,00      2,42
  ,77      1,00      2,79
  -,77      2,00      2,34
  ,00      2,00      2,53
  ,77      2,00      2,71
END DATA.
GRAPH/SCATTERPLOT=
  stress WITH worklife BY geschlec .
```

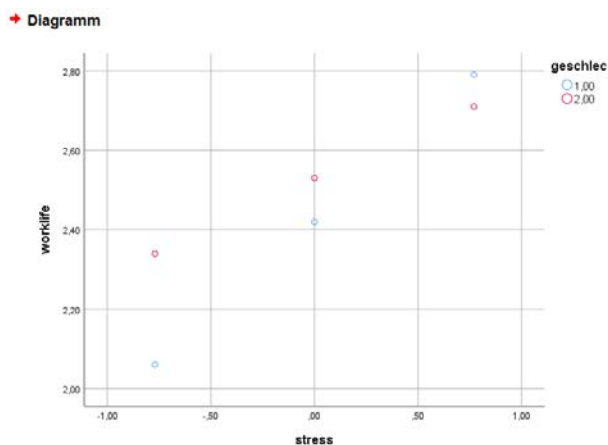
Syntax für die grafische Darstellung
 der Ergebnisse

Grafische Darstellung des Moderatoreffekts mit Hilfe der SPSS-Syntax

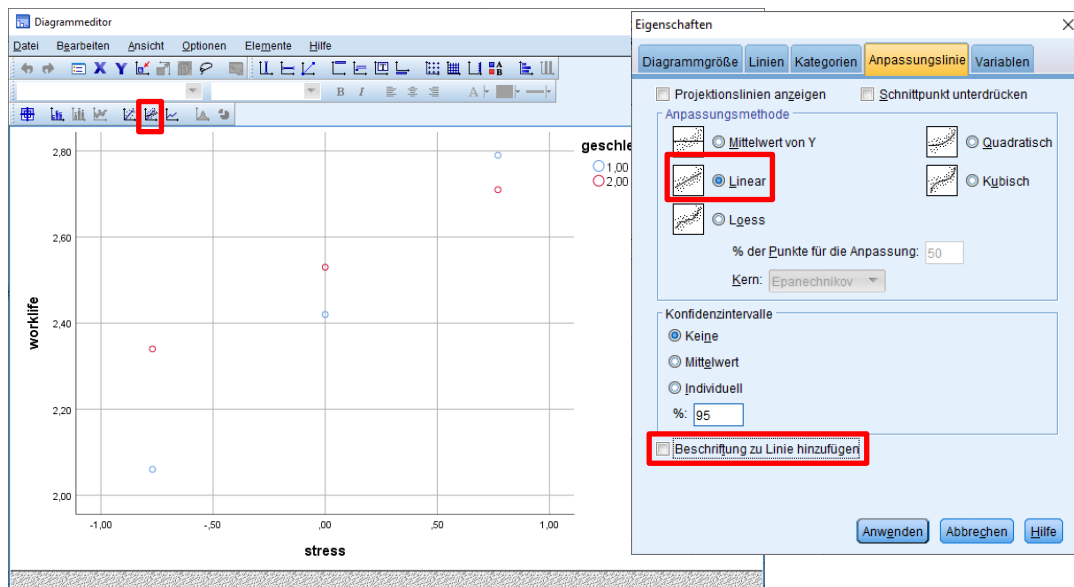
Um den Moderatoreffekt nun grafisch darzustellen, öffnet man zunächst eine neue Syntax unter *Datei/Neu/Syntax*. Als nächstes kopiert man die oben dargestellte Syntax aus der Ausgabe und fügt diese in die neue Syntax ein. Zuletzt führt man die Syntax dann aus unter *Ausführen/Alle*.



Es öffnet sich in der Ausgabe ein Diagramm. Mit einem Doppelklick auf dieses Diagramm öffnet sich der Diagrammeditor.

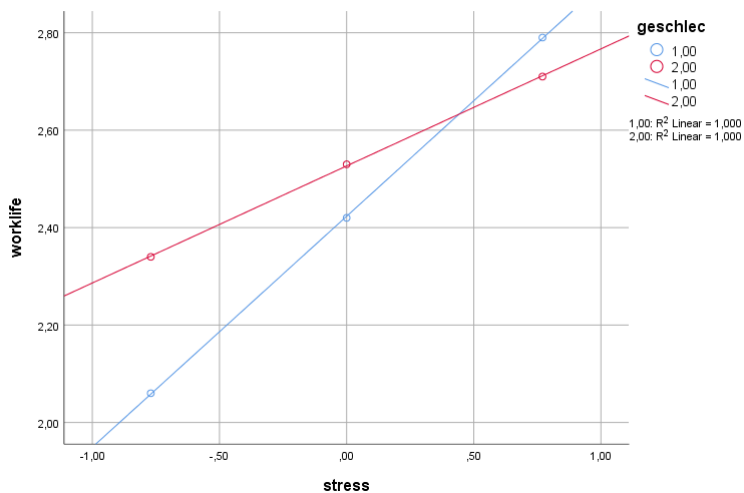


Hier lassen sich weitere Einstellungen vornehmen. Mit einem Klick auf die Schaltfläche „Anpassungslinie bei Untergruppen hinzufügen“ öffnet sich ein weiteres Fenster. Für eine übersichtliche Darstellung wählt man hier die „Anpassungsmethode“ „Linear“ aus und entfernt das Häkchen bei „Beschriftung zu Linie hinzufügen“.



Mit Klick auf „Anwenden“ werden die Änderungen übernommen und nach Schließen des Diagrammeditors erscheint in der Ausgabe das angepasste Diagramm.

Diagramm



Für dieses Beispiel lässt sich feststellen, dass der Einfluss der UV (stress) auf die AV (worklife) für beide Geschlechter zwar positiv ist, jedoch je nach Ausprägung des Moderators (1 oder 2) ein unterschiedlich starker positiver Zusammenhang vorliegt.

Ergänzung: Johnson-Neyman-Signifikanz

Auch wenn die Johnson-Neyman-Signifikanz im vorangegangenen Beispiel im Output nicht erscheint (weil keine nicht signifikanten Effekte vorliegen, s. „conditional effect of focal predictor at values of the moderator“), folgt eine kurze Darstellung bzw. Interpretation dazu. Mit Hilfe der Johnson-Neyman-Signifikanz werden sogenannte Transition Points, an denen die Werte von statistisch signifikant auf statistisch nicht signifikant wechseln, dargestellt.

Im untenstehenden Beispiel liegt der Transition Point bei ,78. D. h. ab diesem Wert für die Moderatorvariable wechselt die Signifikanz des Effekts von bis dahin signifikant auf nicht signifikant. 99,29 Prozent

Product terms key:

Int_1 : stress x geschlec

Test(s) of highest order unconditional interaction(s):

	R2-chng	F	df1	df2	p
X*W	,01	4,96	1,00	699,00	,03

Focal predict: stress (X)
Mod var: geschlec (W)

Conditional effects of the focal predictor at values of the moderator(s):

geschlec	Effect	se	t	p	LLCI	ULCI
-,48	,47	,07	6,95	,00	,34	,60
,00	,36	,05	7,17	,00	,26	,46
,51	,25	,07	3,33	,00	,10	,39

Moderator value(s) defining Johnson-Neyman significance region(s):

Value	% below	% above
,78	99,29	,71

Conditional effect of focal predictor at values of the moderator:

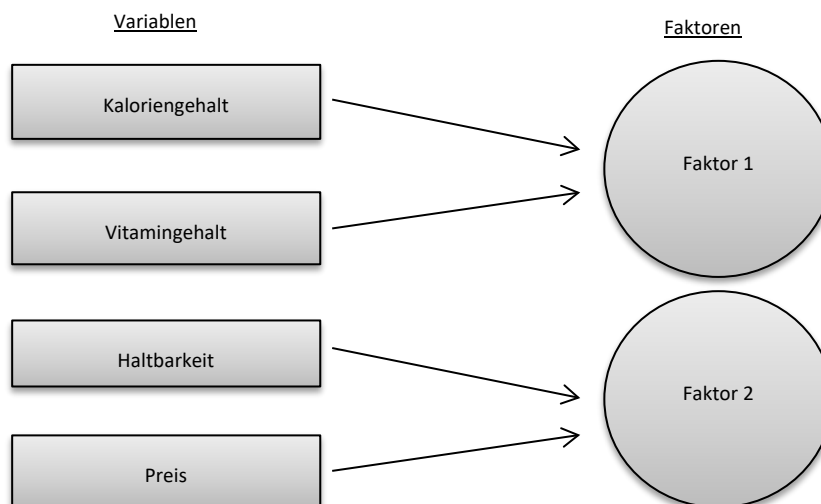
geschlec	Effect	se	t	p	LLCI	ULCI
-,48	,47	,07	6,95	,00	,34	,60
-,38	,45	,06	7,29	,00	,33	,57
-,28	,43	,06	7,55	,00	,32	,54
-,18	,40	,05	7,65	,00	,30	,51
-,08	,38	,05	7,50	,00	,28	,48
,02	,36	,05	7,06	,00	,26	,46
,12	,34	,05	6,38	,00	,23	,44
,22	,31	,06	5,56	,00	,20	,43
,32	,29	,06	4,74	,00	,17	,41
,42	,27	,07	3,97	,00	,14	,40
,52	,25	,07	3,30	,00	,10	,39
,62	,23	,08	2,72	,01	,06	,39
,72	,20	,09	2,23	,03	,02	,38
,78	,19	,10	1,96	,05	,00	,38
,82	,18	,10	1,82	,07	-,01	,38
,92	,16	,11	1,46	,14	-,05	,37
1,02	,14	,12	1,16	,25	-,09	,37
1,12	,11	,13	,90	,37	-,13	,36
1,22	,09	,14	,67	,50	-,17	,36
1,32	,07	,14	,47	,64	-,22	,35
1,42	,05	,15	,30	,76	-,26	,35
1,52	,02	,16	,15	,88	-,30	,35

8 FAKTORENANALYSE

Das Verfahren der Faktorenanalyse ist ein statistisches Instrument, welches Daten strukturiert, reduziert und zusammenfasst. Im Folgenden werden unterschiedliche Vorgehensweisen der Faktorenanalyse unterschieden und in einem Praxisbeispiel mit Hilfe der Statistiksoftware SPSS veranschaulicht.

8.1 Grundidee der Faktorenanalyse

Die Faktorenanalyse gehört zu den multivariaten Analyseverfahren und wird häufig zur Datenreduktion bzw. Datenstrukturierung verwendet. Dies geschieht, indem die Variation einer Vielzahl von Variablen (z.B. Fragebogenitems) auf eine deutlich geringere Zahl von gemeinsamen Dimensionen zurückgeführt wird, welche für weitere Analysen verwendet werden können. Des Weiteren wird die Faktorenanalyse zur Überprüfung der Konstruktvalidität von Fragebögen oder Tests eingesetzt. Hierbei wird häufig entweder die exploratorische Faktorenanalyse (hypothesengenerierendes Verfahren) oder die konfirmatorische Faktorenanalyse (hypothesenprüfendes Verfahren) verwendet.



Das Beispiel verdeutlicht den Grundgedanken der Faktorenanalyse, mehrere Variablen werden durch zwei übergeordnete Faktoren zusammengefasst bzw. strukturiert. Die vorliegende Datenmenge wird auf diese Weise reduziert und somit übersichtlicher.

8.2 Unterschiedliche Verfahren der Faktorenanalyse

Grundsätzlich unterscheidet man zwei faktorenanalytische Methoden: die exploratorische (EFA: Exploratory Factor Analysis) und die konfirmatorische (CFA: Confirmatory Factor Analysis) Faktorenanalyse. Bei der exploratorischen Faktorenanalyse versucht man Faktoren zu finden, wenn vorab noch nicht feststeht, wie viele Fähigkeiten mittels des zu untersuchenden Fragebogens gemessen werden bzw. welche Fragen welche Fähigkeiten messen. Bei der konfirmatorischen Faktorenanalyse überprüft man eine vorab theoretisch angenommene Faktorenstruktur.

Im Mittelpunkt dieses Abschnitts steht die Durchführung einer exploratorischen Faktorenanalyse. Auf die zentralen Merkmale einer konfirmatorischen Faktorenanalyse ist an dieser Stelle nur der Vollständigkeit halber hingewiesen.

Konfirmatorische Faktorenanalyse	Exploratorische Faktorenanalyse
Strukturprüfendes Verfahren	Strukturentdeckendes Verfahren
Festgelegte Anzahl an Faktoren; vorgegebenes Modell	Suche nach latenten Faktoren, die die Korrelation zwischen den Variablen erklären
Konkrete Vorstellungen über den Zusammenhang zwischen den Variablen; Ziel: Modell testen	Keine konkreten Vorstellungen über den Zusammenhang zwischen den Faktoren
Prüfung der Eignung bestehender Faktoren	Extraktion passender Faktoren
Faktoren geben Rückschlüsse über die Zusammenhänge zwischen den Variablen	Faktoren sind ursächlich für die beobachteten Variablen → Linearkombinationen der Faktoren
Prüfung der Übereinstimmung des Modells mit den empirischen Daten	Betrachtung der gemeinsamen Varianz der Variablen, d.h., deren Korrelationen werden durch Faktoren zu erklären versucht
Ziel: Bestätigung der Eignung des zu prüfenden Modells	Ziel: Aufdecken der wesentlichen, voneinander unabhängigen Variablen (Faktoren), die von den beobachteten Variablen gemessen werden und für deren Korrelationen verantwortlich sind

8.3 Ablauf einer exploratorischen Faktorenanalyse (EFA)

Die Durchführung einer Faktorenanalyse gliedert sich in vier zentrale Schritte:

- 1) Prüfung der Voraussetzungen (Faktorisierbarkeit), 2) Extraktion einer angemessenen Faktorenzahl, 3) Rotation der Faktoren und 4) Interpretation

8.3.1 Faktorisierbarkeit

Zunächst wird die Korrelationsmatrix der erhobenen Daten auf ihre Eignung zur Faktorbildung überprüft. Um überhaupt Faktoren extrahieren zu können, müssen unterschiedliche Voraussetzungen erfüllt sein und anschließend verschiedene Eignungsverfahren angewendet werden.

Die Korrelationsmatrix gibt die einzelnen Korrelationen an, welche zwischen den Variablen existieren. Für die Bildung der Faktoren ist eine ausreichend hohe Korrelation zwischen den Variablen notwendig. Diese wird in unterschiedlichen Eignungsverfahren überprüft.

Bartlett-Test

Der Bartlett-Test prüft, ob die Stichprobe aus einer Grundgesamtheit stammt, in der die Variablen unkorreliert sind. D.h. der Test prüft die Nullhypothese, dass alle Korrelationen der Korrelationsmatrix gleich Null sind. Wird die Nullhypothese abgelehnt, bedeutet das, dass ausreichend Korrelation zur Bildung der Faktoren vorhanden sind.

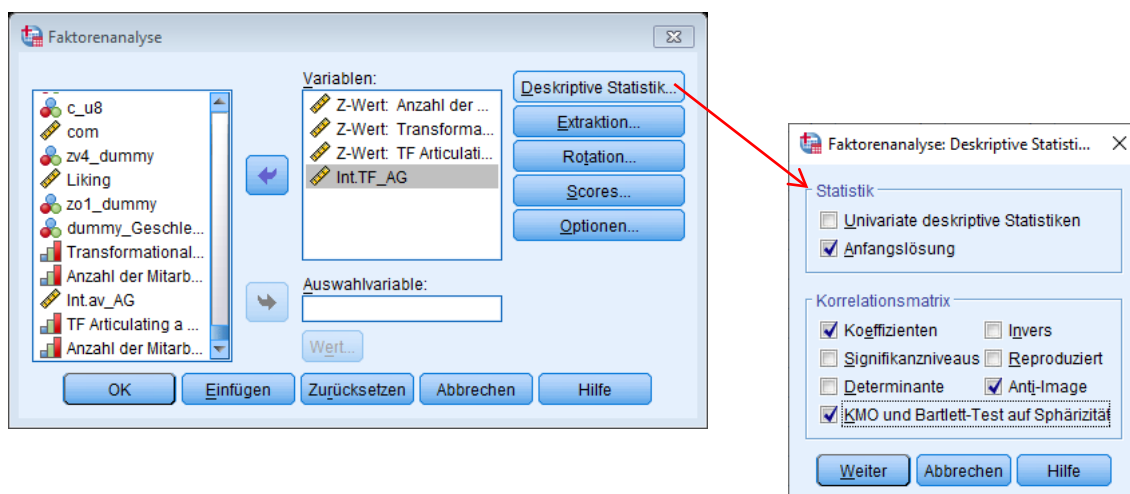
Kaiser-Meyer-Olkin-Kriterium (KMO)

Das KMO-Kriterium zeigt an, in welchem Umfang die Ausgangsvariablen zusammengehören und dient damit als Indikator dafür, ob eine Faktorenanalyse sinnvoll erscheint oder nicht. Das KMO-Kriterium erlaubt hierbei sowohl eine Beurteilung der Korrelationsmatrix insgesamt als auch einzelner Variablen. Der Wertebereich liegt dabei zwischen 0 und 1. Liegt der Wert unter 0,5, ist die Korrelationsmatrix für eine Faktorenanalyse nicht geeignet. Das KMO-Verfahren gilt als das beste Verfahren zur Prüfung der Korrelationsmatrix.

Anti-Image-Kovarianz-Matrix

„Image“ beschreibt den Anteil der Varianz einer Variablen, der durch die verbleibenden Variablen mittels einer multiplen Regressionsanalyse erklärt werden kann. Demnach ist „Anti-Image“ der Teil, der von den übrigen Variablen unabhängig ist. Für die Bildung der Faktoren sollte das Anti-Image möglichst gering ausfallen, d.h. die Werte sollten möglichst nahe Null sein. SPSS gibt in der Diagonalen der Anti-Image-Korrelationsmatrix die Werte für einzelne Variablen an. Liegen diese unter dem Wert 0,5, sollten diese Variablen sukzessiv von der Analyse ausgeschlossen werden. Allerdings darf dies nur einzeln geschehen (beginnend mit der schlechtesten), da sich die Datenstruktur mit jedem Eingriff verändert und die Werte somit möglicherweise bereits über 0,5 liegen.

In SPSS können unter *Analysieren/Dimensionsreduktion/Faktorenanalyse* die entsprechenden Eignungskriterien ausgewählt werden.



8.3.2. Extraktion der Faktoren

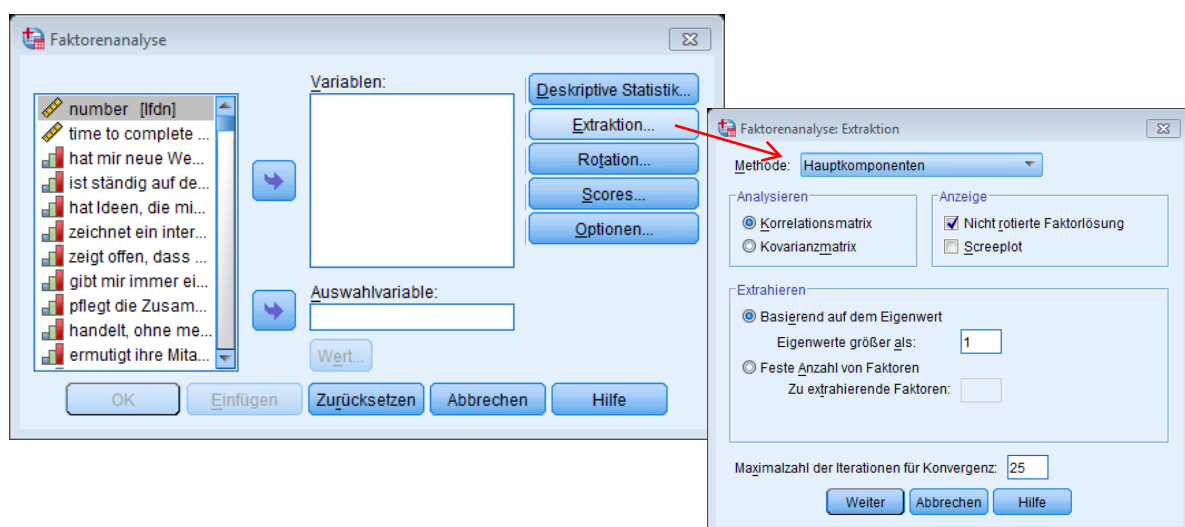
Es gibt unterschiedliche Methoden zur Extraktion der Faktoren: Die Hauptkomponentenanalyse (PCA: Principle Component Analysis), die Hauptachsenanalyse (PAF: Principle Axis Factor Analysis) und die Maximum-Likelihood-Faktorenanalyse (ML).

Bei der Hauptkomponentenanalyse wird die implizite Annahme getroffen, dass die beobachteten Variablen messfehlerfrei erhoben wurden, sodass die gesamte Varianz der Variablen wahre Varianz ist und durch die Faktoren erklärt werden kann. Dieser Fall ist in der Praxis jedoch sehr unwahrscheinlich, sodass die Hauptkomponentenanalyse nur zur Datenreduktion eingesetzt werden sollte, wenn der Fragebogen hoch reliabel ist.

Die Hauptachsenanalyse geht davon aus, dass die einzelnen beobachteten Variablen nicht nur wahre Varianz, sondern auch Messfehlervarianz aufweisen. Ziel hierbei ist es, Faktoren zu identifizieren, welche die Zusammenhänge der Variablen untereinander erklären. Bei dieser Analyseform kann nicht die gesamte Varianz der beobachteten Variablen, sondern nur die wahre Varianz, durch zugrunde liegende gemeinsame Faktoren erklärt werden. Aufgrund dessen werden die wahren Varianzanteile der Variablen zunächst geschätzt und in die Hauptdiagonale der Korrelationsmatrix eingetragen. Hierbei wird als Maß für die geschätzte Reliabilität zumeist die quadrierte multiple Korrelation jeder einzelnen Variablen mit allen anderen Variablen verwendet. Bei der Bildung der Faktoren bleibt auch hier ein unerklärter Anteil (Fehlerterm) der Varianz übrig.

8.3.3 Extraktionsmethoden

Zunächst muss eine Methode zur Extraktion der Faktoren ausgewählt werden, um mit dieser Methode die geeigneten Faktoren zu extrahieren. Hierbei werden die **Faktorladungen**, die **Eigenwerte** und die **Kommunalitäten** bestimmt. Mit Hilfe eines zuvor gewählten Abbruchkriteriums wird dann die Anzahl der Faktoren festgelegt. Im Anschluss daran können die jeweiligen Faktoren interpretiert werden. Zu Interpretationszwecken dient die Rotation der Faktoren.



Faktorladung

Faktorladungen ergeben sich rechnerisch aus der Korrelation der standardisierten Werte der Ausgangsvariablen mit den Faktorwerten. Sie stellen ein Maß der Korrelation der Ausgangsvariablen mit dem Faktor dar und zeigen so gewissermaßen die Gemeinsamkeit von Ausgangsvariable und Faktor an. Eine positive (negative) Faktorladung sagt aus, dass der Faktor positiv (negativ) mit der Variablen korreliert. Die Faktorladung gibt also an, inwieweit ein Faktor durch eine bestimmte Variable erklärt werden kann. An der Höhe der Faktorladung lässt sich ablesen, ob eine Variable viel oder wenig mit einem Faktor gemeinsam hat. Je höher die Ladung, desto größer ist die Bedeutung dieser Variablen für den jeweiligen Faktor. Bei mehreren Faktoren ist allerdings zu prüfen, ob eine Variable auf mehreren Faktoren hoch lädt (unerwünscht). Die Faktorladungen werden in der Mustermatrix als standardisierte semipartielle Regressionsgewichte der Items mit den Faktoren angegeben.

Eigenwert eines Faktors

Der Eigenwert eines Faktors ist der Wert der Gesamtvarianz *aller* Variablen, den dieser Faktor aufklärt. Der Eigenwert ergibt sich unabhängig von der Extraktionsmethode aus der Summe aller quadrierten Faktorladungen des Faktors. Der Eigenwert ist ein Qualitätskriterium für den Faktor.

Kommunalität

Unter dem Begriff Kommunalität versteht man die Summe der quadrierten Faktorladungen *einer* Variablen. Die Kommunalität gibt an, in welchem Maße die Varianz einer Variablen durch *alle* Faktoren aufgeklärt wird. Eine Variable mit einer geringen Kommunalität wird durch das Modell insgesamt schlecht vertreten.

8.3.4 Faktoranzahl

Es existieren mehrere Möglichkeiten, die angemessene Anzahl extrahierter Faktoren zu bestimmen:

- Extrahiere solange, bis X % (z.B. 95 %) der Varianz erklärt sind.
- Extrahiere nur Faktoren mit Eigenwerten größer 1.
- Extrahiere n (z.B. 3) Faktoren.
- Scree-Test: Die Faktoren werden nach Eigenwerten in abfallender Reihenfolge geordnet. An die Faktoren mit den niedrigsten Eigenwerten wird eine Gerade angepasst. Der letzte Punkt auf der Geraden bestimmt die Faktorenzahl.

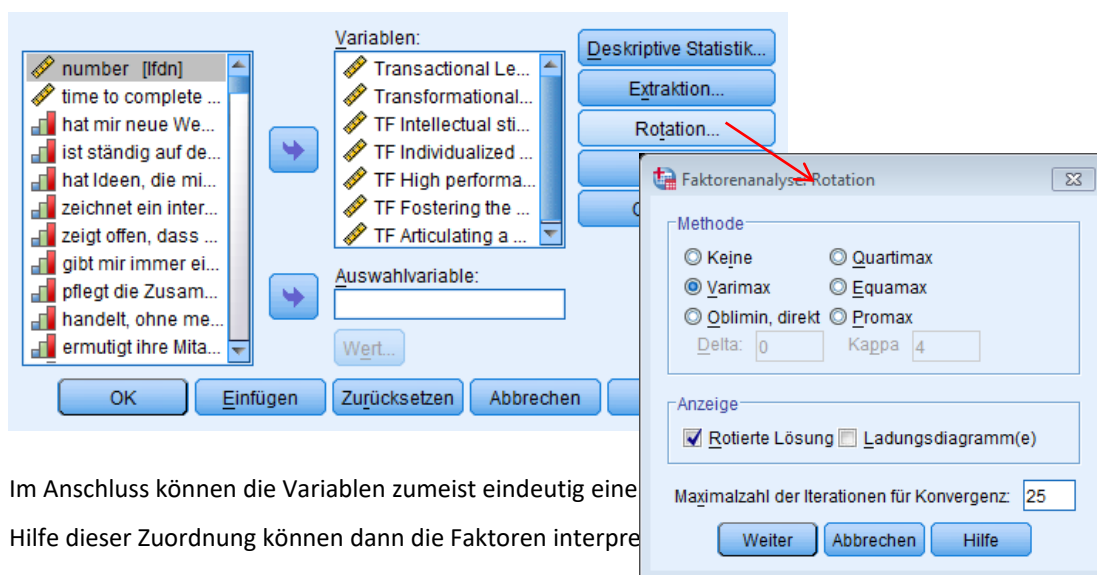
Je nachdem, welche Variante zur Bestimmung der Anzahl der Faktoren gewählt wird, können sich unterschiedlich viele Faktoren ergeben. Die häufigsten Verfahren, um eine angemessene Anzahl extrahierter Faktoren zu bestimmen, sind das Eigenwertkriterium und der Scree-Test. Zudem sollten immer theoretische Überlegungen für die Anzahl der gewählten Faktoren im Vordergrund stehen.

8.4 Rotation der Faktoren

Die Rotation der Faktoren dient der Erleichterung der Interpretation der Faktoren und kann in zwei unterschiedliche Vorgehensweisen unterteilt werden.

8.4.1 Rotationsverfahren

Beim Rotationsverfahren werden zwei grundlegende Vorgehensweisen unterschieden, das *orthogonale* und das *oblique* Rotationsverfahren. Durch die Rotation wird das Koordinatenkreuz, mit den Faktoren als Koordinaten, gedreht. Hierdurch wird die Zuordnung einzelner Variablen zu einem der Faktoren eindeutiger. Die Faktorladungen lassen sich hierdurch optimieren, sodass auf jedem der Faktoren einige Variablen möglichst hoch und die anderen möglichst niedrig laden. Dies erleichtert die Interpretation der Faktoren, da eine eindeutige Zuordnung der Variablen ermöglicht wird. Beim *orthogonalen Verfahren* (Varimax, Quartimax) wird die Unabhängigkeit der Faktoren, also der Abstand zwischen den Achsen, beim Rotieren beibehalten. Im Gegensatz dazu erlaubt das *oblique Verfahren* (Promax) Korrelationen zwischen den Faktoren.



Im Anschluss können die Variablen zumeist eindeutig einer der Faktoren zugeordnet werden. Mit Hilfe dieser Zuordnung können dann die Faktoren interpretiert werden.

8.5. Interpretation der Faktoren

Die Interpretation der Faktoren stellt den abschließenden Schritt der Faktorenanalyse dar. Hier wird die Datenstrukturierung bzw. Datenreduktion dadurch abgeschlossen, dass jede Variable einem bestimmten Faktor zugeordnet wird.

8.6. Faktorinterpretation

Im Anschluss an die Bestimmung der Anzahl der Faktoren, wird geprüft, welche Variable zu welchem Faktor gehört. Hierfür betrachtet man die Faktorladungen. Diese gibt an, wie stark der Zusammenhang zwischen einer Variablen und einem Faktor ist, inwieweit die Variable also mit dem Faktor „etwas zu tun hat“.

In manchen Fällen ist eine eindeutige Zuordnung der Variablen zu einem jeweiligen Faktor jedoch nicht möglich. In diesem Fall lassen die Faktorladungen keine eindeutige Aussage zu, sodass zur Faktorinterpretation das Rotationsverfahren angewendet wird. Die Faktorladungen können Werte zwischen -1 und +1 an. Im Rahmen der Faktorenanalyse wird angestrebt, eine eindeutig interpretierbare Lösung zu erhalten.

Um eine Variable eindeutig einem bestimmten Faktor zuordnen zu können, müssen spätestens im Anschluss an das Rotationsverfahren die Bedingungen der Konvergenzvalidität und der Diskriminanzvalidität erfüllt sein.

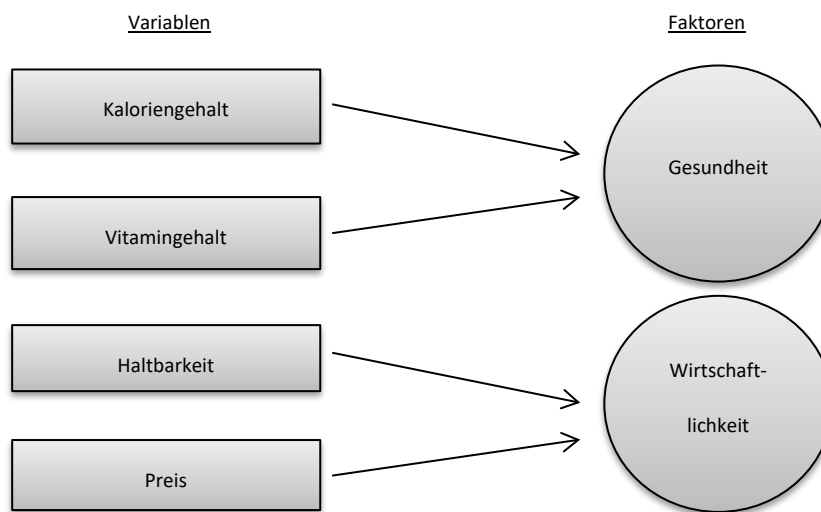
Konvergenzvalidität

Das Kriterium der Konvergenzvalidität ist erfüllt, wenn ausreichend hohe Ladungen der meisten Items auf den jeweiligen Faktor vorliegen (quadrierte Faktorladung > 0.5).

Diskriminanzvalidität

Das Kriterium der Diskriminanzvalidität ist erfüllt, wenn zwischen der höchsten Faktorladung einer Variablen auf einen beliebigen Faktor und der zweithöchsten Faktorladung der gleichen Variablen auf einen der übrigen Faktoren ein Mindestabstand von > 0.2 vorliegt.

Sind beide Kriterien erfüllt, können die Faktoren durch die Variablen interpretiert werden. Als Beispiel der Interpretation eines Faktors, dient hier die zu Beginn der Faktorenanalyse dargestellte Grafik, welche den Grundgedanken der Analyse veranschaulichen sollte. In der Annahme, dass die Variablen Kaloriengehalt und Vitamingehalt beide auf den ersten Faktor laden und zudem die beschriebenen Kriterien erfüllen, lässt sich der erste Faktor allgemein unter dem Oberbegriff „Gesundheit“ zusammenfassen bzw. interpretieren. In Anbetracht der Variablen Haltbarkeit und Preis, welche beide auf den zweiten Faktor laden, lässt sich dieser Faktor durch den Begriff „Wirtschaftlichkeit“ interpretieren.



8.7 Praxisbeispiel: Exploratorische Faktorenanalyse

In diesem Beispiel werden für die einzelnen Items transformationaler Führung Faktoren gebildet, sodass die Items in übergeordnete Gruppen unterteilt werden können, welche die Facetten transformationaler Führung beschreiben. Im ersten Schritt der Analyse werden die zu analysierenden Variablen ausgewählt, welche durch die späteren Faktoren zusammengefasst bzw. strukturiert werden. Im Anschluss daran wird die Eignung der Korrelationsmatrix mithilfe der oben beschriebenen Kriterien überprüft. In diesem Beispiel lässt sich die Eignung der Korrelationsmatrix anhand des KMO- und Bartlett-Tests überprüfen.

KMO- und Bartlett-Test

Maß der Stichprobeneignung nach Kaiser-Meyer-Olkin.		,883
Bartlett-Test auf Sphärizität	Ungefähres Chi-Quadrat	1015,120
	df	231
	Signifikanz nach Bartlett	,000

KMO-Kriterium > 0,5 → Korrelationen ausreichend hoch → Eignung der Korrelationsmatrix

Nullhypothese wird abgelehnt → ausreichende Korrelation vorhanden

Im Anschluss an die Überprüfung der Eignung der Korrelationsmatrix wird die Extraktionsmethode der Faktorenanalyse im SPSS-Programm ausgewählt. In diesem Fall handelt es sich um eine Hauptachsenanalyse.

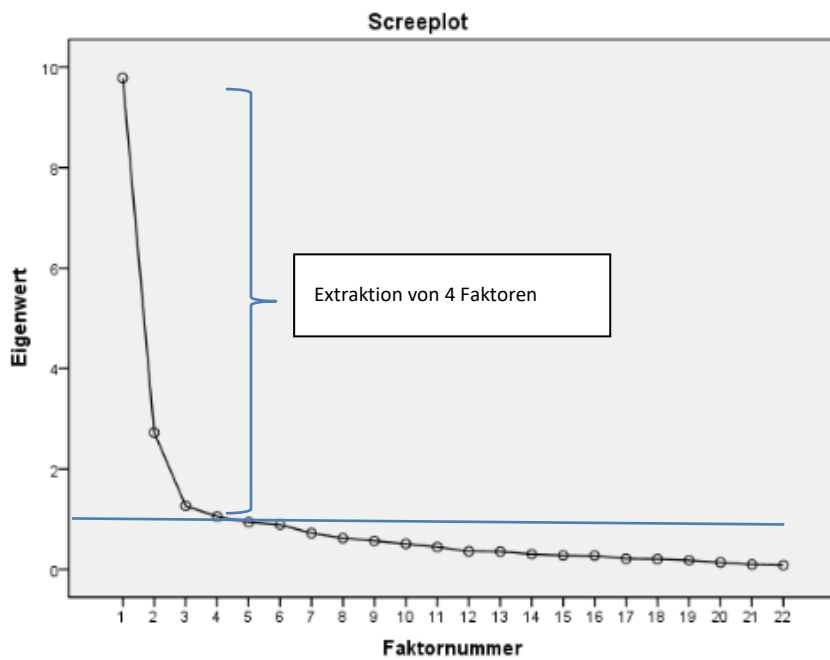
Erklärte Gesamtvarianz

Faktor	Anfängliche Eigenwerte			Summen von quadrierten Faktorladungen für Extraktion			Rotierte Summe der quadrierten Ladungen		
	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
1	9,784	44,475	44,475	9,424	42,837	42,837	6,261	28,459	28,459
2	2,723	12,377	56,852	2,293	10,422	53,259	3,187	14,485	42,944
3	1,266	5,753	62,604	,869	3,950	57,210	1,986	9,027	51,971
4	1,052	4,783	67,388	,707	3,214	60,423	1,859	8,452	60,423
5	,941	4,275	71,663						
6	,890	4,045	75,708						
7	,722	3,281	78,989						
8	,621	2,821	81,810						
9	,567	2,578	84,388						
10	,508	2,308	86,697						
11	,445	2,021	88,717						
12	,359	1,634	90,351						
13	,355	1,614	91,966						
14	,303	1,379	93,345						
15	,276	1,256	94,600						
16	,273	1,239	95,840						
17	,213	,968	96,808						
18	,206	,936	97,744						
19	,180	,818	98,562						
20	,138	,626	99,188						
21	,096	,438	99,626						
22	,082	,374	100,000						

Extraktionsmethode: Hauptachsen-Faktorenanalyse.

Eigenwert > 1

Mit Hilfe der durch die Hauptachsenanalyse ausgegebenen Tabelle kann nun die Anzahl der Faktoren bestimmt werden. Betrachtet man beispielsweise das Eigenwert-Kriterium, werden vier Faktoren extrahiert, da diese einen Eigenwert > 1 besitzen.



Nachdem die Anzahl der zu bildenden Faktoren feststeht, können die Faktoren mit Hilfe der vorhandenen Variablen und deren Faktorladungen interpretiert werden. Die Faktorenmatrix gibt die unrotierten Faktorladungen wieder.

Faktorenmatrix^a

	Faktor			
	1	2	3	4
tli_u1	,701	,095	,233	-,324
tli_u2	,651	,244	-,147	-,170
tli_u3	,656	,258	,058	-,195
tli_u4	,713	,093	-,352	-,072
tli_u5	-,019	,574	,445	,108
tli_u7	,705	-,042	,092	,101
tli_u8	,847	-,356	,187	-,091
tli_u9	,752	,040	,143	,120
tli_u10	,509	-,130	-,085	-,115
tli_u11	,802	-,064	,171	,265
tli_u12	,732	,122	-,114	,173
tli_u13	,704	-,374	,145	,055
tli_u14	,653	,128	-,091	-,060
tli_u15	,628	-,242	-,116	,139
tli_u17	,797	-,236	-,023	,441
tli_u18	,771	,245	-,358	,063
tli_u20	-,173	,686	,143	,298
tli_u21	,742	,197	,124	-,103
tli_u22	,847	,139	,056	-,175
tli_u23	,360	-,517	,362	-,112
tli_u24	,015	,696	,071	-,061
tli_u25	,820	,206	-,087	-,042

Eine Variable wird einem Faktor zugeordnet, wenn der quadrierte Wert > 0.5 ist.

Extraktionsmethode: Hauptachsen-Faktorenanalyse.

a. 4 Faktoren extrahiert. Es werden 11 Iterationen benötigt.

Die Betrachtung der Kommunalitäten, also der Summe der quadrierten Faktorladungen einer Variablen, ist notwendig, da hier geprüft wird, in welchem Maße die Varianz einer Variablen durch alle Faktoren aufgeklärt wird.

Kommunalitäten

	Anfänglich	Extraktion
tli_u1	,689	,660
tli_u2	,577	,534
tli_u3	,604	,538
tli_u4	,674	,647
tli_u5	,499	,540
tli_u7	,682	,517
tli_u8	,672	,588
tli_u9	,706	,602
tli_u10	,538	,297
tli_u11	,802	,747
tli_u12	,626	,593
tli_u13	,724	,661
tli_u14	,537	,455
tli_u15	,661	,487
tli_u17	,822	,887
tli_u18	,822	,786
tli_u20	,544	,609
tli_u21	,675	,616
tli_u22	,840	,770
tli_u23	,523	,541
tli_u24	,554	,493
tli_u25	,832	,725

> 0.1 ansonsten Variablen rausnehmen, da sonst ein zu geringer Anteil der Varianz der Variablen durch die Faktoren erklärt wird.

Extraktionsmethode:
Hauptachsen-
Faktorenanalyse.

Um eine eindeutige Interpretation vornehmen zu können, sollte die gebildete Faktorenmatrix noch rotiert werden. Die Rotation ermöglicht die eindeutige Zuordnung einer Variablen zu einem der vier Faktoren, sodass die jeweiligen Variablen, welche auf einem Faktor laden, dabei helfen, diesen Faktor unmissverständlich zu interpretieren. Im Beispiel wurden die verschiedenen Items transformationaler Führung unterschiedlichen Faktoren zugewiesen. Durch die Faktorladungen einzelner Variablen können die Faktoren anschließend interpretiert und benannt werden. So lässt sich beispielsweise der dritte Faktor unter der Überschrift „individuelle Unterstützung“ zusammenfassen. Demnach handelt es sich bei den gebildeten Faktoren um die Facetten der transformationalen Führung:

Rotierte Faktorenmatrix^a

	Faktor			
	1	2	3	4
tli_u1	,635	,094	,490	,086
tli_u2	,712	,146	,063	,044
tli_u3	,665	,135	,229	,158
tli_u4	,732	,281	-,054	-,171
tli_u5	,012	-,015	,085	,729
tli_u7	,466	,467	,285	-,011
tli_u8	,351	,346	,529	-,256
tli_u9	,510	,494	,302	,083
tli_u10	,422	,196	,191	-,210
tli_u11	,445	,662	,331	,040
tli_u12	,586	,498	,043	,029
tli_u13	,345	,498	,472	-,266
tli_u14	,610	,264	,114	,001
tli_u15	,385	,489	,160	-,272
tli_u17	,371	,830	,181	-,165
tli_u18	,782	,391	-,145	-,025
tli_u20	-,045	,023	-,314	,713
tli_u21	,653	,272	,302	,156
tli_u22	,764	,276	,329	,051
tli_u23	,019	,219	,643	-,282
tli_u24	,269	-,179	-,187	,595
tli_u25	,761	,352	,136	,060

Extraktionsmethode: Hauptachsen-Faktorenanalyse.
 Rotationsmethode: Varimax mit Kaiser-Normalisierung.

1. Faktor → Visionen aufzeigen

Beispielitem: ... inspiriert durch ihre Pläne für die Zukunft.

2. Faktor → Gruppenziele fördern

Beispielitem: ... entwickelt ein Wir-Gefühl und Teamgeist bei den Mitarbeitern ihrer Abteilung.

3. Faktor → Individuelle Unterstützung

Beispielitem: ... behandelt mich auf eine Art und Weise, ohne auf meine persönlichen Gefühle Rücksicht zu nehmen.

4. Faktor → Hohe Leistungserwartung

Beispielitem: ... wird sich nicht mit dem Zweitbesten zufrieden geben.