

Reader

zum Programm

IBM SPSS

Februar 2019 (Vers.1.2_Bachelor)
Gaissmaier, R. unter Mitarbeit von Arnold, L.

Inhalt

Aufbau	1
Dateneditor	2
Ausgabe	3
Syntaxeditor	3
Syntaxeinführung	4
Umgang mit Daten	6
Skalenniveau	6
Kodierung	7
Werte(label)	7
Fehlende Werte.....	8
Umgang mit Filterfragen	8
Umgang mit „Sonstiges“	8
Umgang mit Mehrfachnennungen	8
Umgang mit Freitext.....	9
Dateneingabe	9
Syntax: Pflege der VA	10
Transformieren – Modifikation von Variablen	12
Variable berechnen	12
Umcodieren in dieselbe/andere Variable	13
Visuelles Klassieren	14
z- Standardisierung.....	14
Invertieren von Werten.....	14
Syntax: Transformieren	15
Daten – Selektionsmöglichkeiten	17
Fälle auswählen	17
Aufgeteilte Datei	18
Syntax: Selektion	18
Analysieren – Deskription	20
Häufigkeiten	20
Deskriptive Statistik.....	21
Kreuztabellen (Kontingenztafeln).....	21

Analysieren – Zusammenhänge	22
Lineare Korrelationen.....	22
Regressionsanalysen	24
Lineare Regression (einfach/multipel)	24
Analysieren – Unterschiede	27
1-Stichproben-t-Test	27
t-Test bei unabhängigen Stichproben	28
t-Test bei abhängigen Stichproben	29
Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummentest.....	30
Dokumentation	32
Codebuch	32
SPSS-Webbericht.....	32
Zusatzmaterial.....	33
Syntax: Operatoren	33
Effektgrößen.....	33
Literatur.....	34
Primärliteratur.....	34
Weiterführende Literatur.....	34
Serie zur Bewertung wissenschaftlicher Publikationen	35

Vorwort

SPSS Statistics, von IBM, ist ein weitverbreitetes Programm für statistische Analysen. Es wird vorwiegend in den Sozial- und Gesundheitswissenschaften genutzt, darüber hinaus aber auch im Bereich der Bildungswissenschaften und spezifischen Marketingbereichen.

Das Funktionsspektrum erstreckt sich hierbei von einfachen deskriptiven über verschiedene bivariate bis hin zu multivariaten Methoden und enthält eine Vielzahl von relevanten Testungen sowie tiefergehende Analyseprozeduren.

Neben der statistischen Analyse beinhaltet das Programm verschiedene Möglichkeiten der Datenverwaltung sowie deren Dokumentation sowie in neueren Versionen ein Marketing- und Businessmodul um Auswertungsdaten schneller bzw. leichter publizieren zu können.

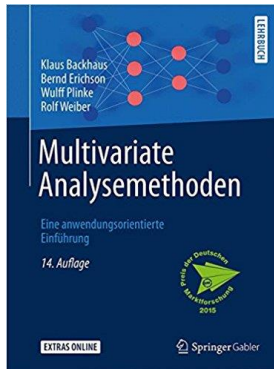
Der vorliegende Reader bezieht sich auf eine Programmversion von „SPSS Statistics“ im vollen Funktionsumfang (vormals „Premium“) und wird den Studierenden seitens der Hochschule zur Verfügung gestellt. Der vorliegende Reader ist darüber hinaus mit der Programmversion 25.0 erstellt worden, welche im Vergleich zu älteren Versionen u.U. aktualisierte Begrifflichkeiten enthält.

Es gibt an der Hochschule Ravensburg-Weingarten verschiedene Möglichkeiten das Programm zu nutzen:

- Das Programm ist auf den meisten Poolrechnern (D004, H104, ...) installiert
- Immatrikulierte Studierende haben die Möglichkeit, eine Pendlerlizenz zu erhalten. Hiermit kann das Programm auf dem eigenen Laptop genutzt werden. Eine Anleitung zur Installation finden Sie im Moodle-Auftritt der Veranstaltung
- IBM stellt Testversionen zur Verfügung, die nach Bezug auf der IBM-Seite (www.ibm.com), 14 Tage kostenlos genutzt werden können. Dies kann pro Nutzer bzw. pro PC nur einmal (!) erfolgen.

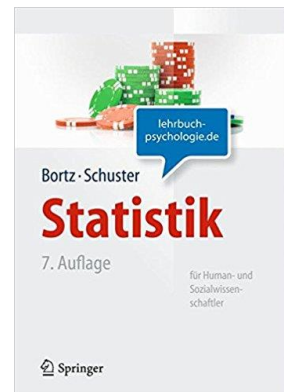
Neben den Inhalten der Vorlesung sowie dem vorliegenden Reader sei an dieser Stelle auch auf Inhalte von www.youtube.com sowie einschlägige Internetseiten verwiesen (bspw. www.spss-tutorials.com), die eine Vielzahl an Tutorial-Videos anbieten, die v.a. spezifische Analysen aber auch die Grundlagen des Programms dezidiert erläutern.

Verwendete Primärliteratur



Backhaus, Klaus; Erichson, Bernd; Plinke, Wulf; Weiber, Rolf (2016): Multivariate Analysemethoden. Eine anwendungsorientierte Einführung. 14., überarbeitete und aktualisierte Auflage. Berlin, Heidelberg: Springer Gabler (Lehrbuch).

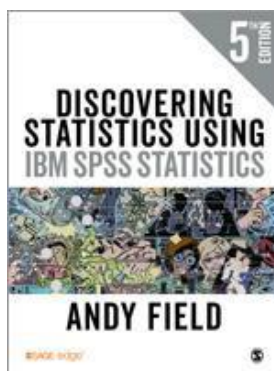
Bortz, Jürgen; Schuster, Christof (2010): Statistik für Human- und Sozialwissenschaftler. 7., vollständig überarbeitete und erweiterte Auflage. Berlin: Springer (Springer-Lehrbuch).



Eid, Michael; Gollwitzer, Mario; Schmitt, Manfred (2011): Statistik und Forschungsmethoden. Lehrbuch. 3. Aufl., deutsche Erstausgabe Weinheim: Beltz.




Eid, Michael; Gollwitzer, Mario; Schmitt, Manfred; Hohmann, Louisa; Lawes, Maria; Wang, Tianqi (2016): Formelsammlung Statistik und Forschungsmethoden. Weinheim, Basel: Beltz.

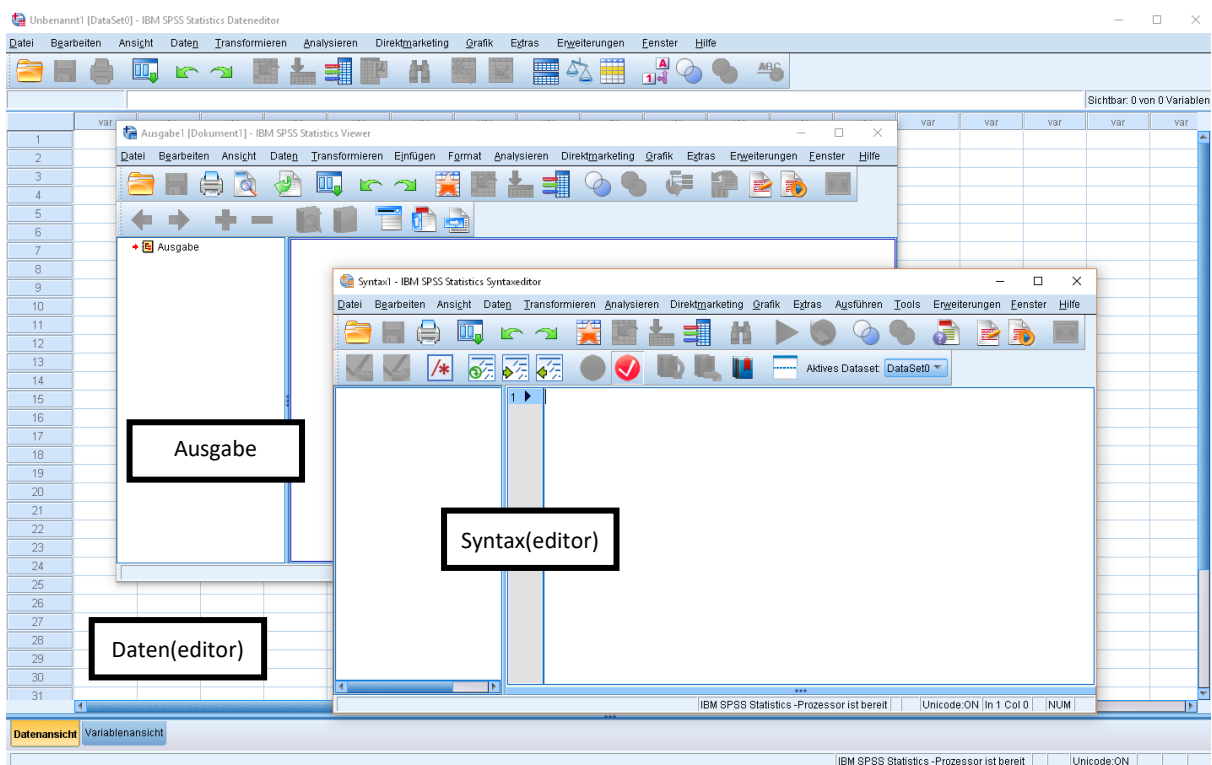


Field, Andy (2018): Discovering statistics using IBM SPSS statistics. 5th edition. Los Angeles, London, New Delhi, Singapore, Washington DC: Sage (Mobile-Study).

Aufbau

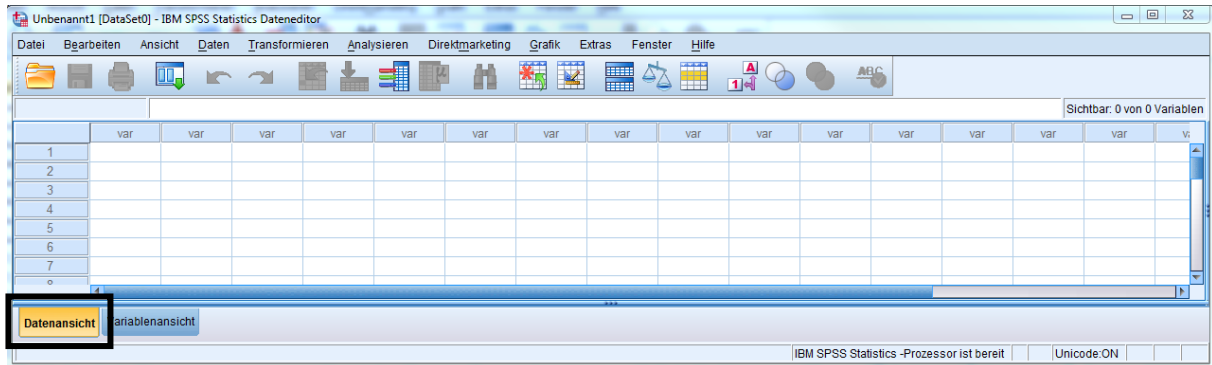
In SPSS wird mit bis zu drei Fenstern gearbeitet: Dateneditor, Ausgabe und Syntax.

- **Dateneditor:** Dieses Fenster besteht aus zwei Ansichten, der Datenansicht und Variablenansicht, welche die Daten des Datensatz beinhalten bzw. die hinterlegten Variablen und deren Attribute. Beim Speichern des Dateneditors wird die Dateiendung *.sav* vergeben.
- **Ausgabe:** In der Ausgabe werden sämtliche Ergebnisse angezeigt, welche aufgrund der Datenlage berechnet wurden. Das Fenster öffnet sich automatisch sobald das Programm eine Prozedur durchlaufen hat, welche Informationen für die Ausgabe bereitstellt. Ausgabe-Dateien haben die Dateiendung *.spv*.
- **Syntaxeditor:** Kann Bei Bedarf über Datei/Neu/Syntax geöffnet werden. Hier können Befehle direkt eingegeben werden (also nicht notwendigerweise über das Menü herausgesucht) bzw. Befehle zur Datenverarbeitung hinterlegt werden. Die Befehle werden schriftlich dargestellt und können nach Markierung mit einem Klick auf  ausgeführt werden. Die Dateiendung des Save-Files lautet hier *.spss*.

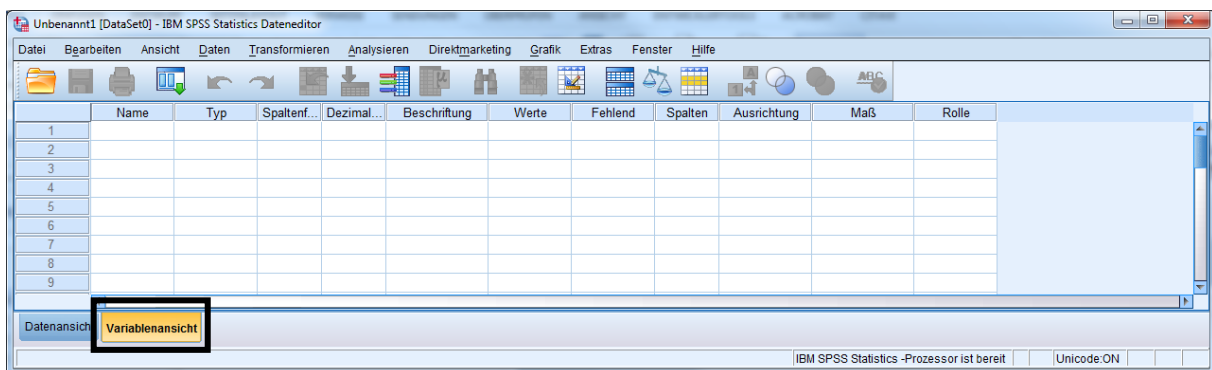


Dateneditor

Der Dateneditor ist zweigeteilt. In der *Datenansicht* (DA) sind alle Daten des Datensatzes enthalten. Jede Zeile entspricht hier einem Teilnehmer bzw. einem Fragebogen oder einem Versuchsobjekt – jede Spalte entspricht einer Variablen.



Durch einen Klick auf den Reiter *Variablenansicht* (VA) kann die zweite Ebene des Dateneditors aufgerufen werden. In der Variablenansicht entspricht jede Zeile einer Variablen. Die Spalten beinhalten die verschiedenen Attribute der jeweiligen Variablen.



Die Spalte **Name** beinhaltet den Kurznamen der Variablen, welcher den „wahren Namen“ einer Variable symbolisiert, was spätestens für das Arbeiten im Syntax relevant ist! Dieser darf bspw. keine Leerzeichen enthalten, nicht mit einer Zahl beginnen und nur wenige Sonderzeichen enthalten. **Typ** definiert ob eine Variable bspw. *numerisch* ist und somit verrechnet werden kann oder ob es sich bspw. um eine *Zeichenfolge* (*string*) also um Text handelt. Unter **Spaltenformat** kann die Anzahl der maximalen Zeichen eingegeben werden und unter **Dezimal** die Anzahl der (angezeigten) Dezimalstellen. Sofern eine **Beschriftung** (*Variablenlabel*) gepflegt ist – welche keine Restriktionen hinsichtlich einzelner Zeichen aufweist – ersetzt diese den Namen der Variable in den Auswertungen im Ausgabe-Fenster. Hinter **Werte** (*-label*) verbirgt sich die Kodierung der Variablenausprägungen – welche dann ebenfalls in der Ausgabe durch den gepflegten Text ersetzt wird. Unter **Fehlend** können fehlende Werte definiert werden, welche vom System aus der Analyse ausgeschlossen werden. Unter **Spalten** kann die Spaltenbreite definiert werden und ferner die **Ausrichtung** der Werte in der DA. **Maß** ermöglicht schlussendlich eine Hinterlegung des Skalenniveaus der Variable und unter **Rolle** kann hinterlegt werden ob spezifischen Variablen (im aufgestellten Modell) eine bspw. unabhängige oder abhängige Rolle einnehmen.

Die Datenansicht bzw. der Dateneditor ist gleichzeitig das Hauptfenster von SPSS.

Ausgabe

Im Ausgabefenster von SPSS werden sämtliche Ergebnisse festgehalten. Von diesen in Tabellen und/oder Grafiken dargestellten Daten wird zusätzlich jedwede Transformation der Daten mitaufgenommen bzw. schriftlich dokumentiert (→ Syntax).

Das Bild zeigt das SPSS-Ausgabefenster mit folgenden Elementen:

- Syntax:** Zeigt die Befehlszeile mit `FREQUENCIES VARIABLES=LQsf369a LQsf369b LQsf369c LQsf369d LQsf369e LQsf369f LQsf369g LQsf369h` und `/ORDER=ANALYSIS.`
- Gliederung:** Die linke Navigationsleiste zeigt die Hierarchie der Ausgabe: Ausgabe > Häufigkeiten > Häufigkeitstabelle > SF36: Voller Leben i.d.l. 4 Wo.
- Analyse:** Die Hauptdarstellung zeigt die Häufigkeitstabelle für 'SF36: Voller Leben i.d.l. 4 Wo.' mit den folgenden Daten:

		SF36: Voller Leben i.d.l. 4 Wo.	SF36: Sehr nervös i.d.l. 4 Wo.	SF36: Niedergeschlagen i.d.l. 4 Wo.	SF36: Ruhig und gelassen i.d.l. 4 Wo.	SF36: Voller Energie i.d.l. 4 Wo.	SF36: Entmutigt und traurig i.d.l. 4 Wo.	SF36: Erschöpft i.d.l. 4 Wo.	SF36: Glückliche i.d.l. 4 Wo.	SF36: Müde i.d.l. 4 Wo.
N	Gültig	19023	19248	19229	19240	19232	19193	19223	19122	19221
	Fehlend	271	46	65	54	62	101	71	172	73

SF36: Voller Leben i.d.l. 4 Wo.				
	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	19023	32.7	33.1	33.1
Immer				

Jeder Analyseschritt wird in die Gliederung der Ausgabe mitaufgenommen. Bevor auf die Ergebnisse eingegangen wird, werden vorab die zugrundeliegenden Anweisungen in Syntax-Schreibweise sowie (hier bei der Häufigkeitsanalyse) eine zusammenfassende Statistik angezeigt.

Die Gliederung kann einerseits zur schnelleren Navigation in der Ausgabe genutzt werden, andererseits aber auch um einzelne Teilschritte aus dieser zu entfernen. Hierfür kann der betreffende Punkt mit der Maus angewählt und mit **entf** (Entfernen) gelöscht werden.

Einzelne Elemente der Ausgabe können ebenfalls angewählt und entfernt werden, allerdings auch für den Übertrag in bspw. MS Excel oder MS Word kopiert, bzw. mit einem Doppelklick in SPSS angepasst bzw. abgeändert werden.

Syntaxeditor

Nahezu sämtliche Applikation in SPSS können entweder über die Menüführung ausgeführt oder durch die in SPSS hinterlegte Programmiersprache ausgeführt werden. Ein großer Vorteil in der Verwendung von Syntax liegt in der Transparenz und der Reproduzierbarkeit, da hiermit automatisch alle Bearbeitungsschritte dokumentiert sind/werden und ggf. auch nochmals per Klick wiederholt bzw. nachträglich angepasst werden können.

Für das Arbeiten mit Syntax sind keine Programmierkenntnisse erforderlich! Nahezu alle SPSS-Befehle können später über die Schaltfläche  in den Editor eingefügt werden.

Der Arbeitsgang über Syntax ist der reinen Arbeit mit der Menüführung vorzuziehen.

Syntaxeinführung

Einer der größten Vorteile mit Syntax zu arbeiten, besteht in der späteren **Nachvollziehbarkeit und Reproduzierbarkeit** aller im Datensatz getätigten Aktionen bzw. Analysen. Einerseits natürlich für den/die Bearbeiter der Daten aber eben auch bspw. für die Mitarbeiter der Methodenberatung oder den/die Betreuer der Arbeit.


Das Gefühl, dass durch den ersten Blick auf eine SPSS-Syntax ausgelöst wird, ist fast immer Panik, denn obwohl SPSS einstmals für Sozial- und Gesundheitswissenschaftler konzipiert wurde, ist diese Berufsgruppe landläufig eher weniger bekannt für ihr Programmierwissen. Allerdings ist die Panik unbegründet, da die Erzeugung der SPSS-Syntax nahezu vollkommen vom Programm selbst bzw. automatisiert geschrieben werden kann und der Nutzer somit kein Vorwissen in Programmierung an den Tag legen muss.

Die meisten Operationen in SPSS können auf zwei verschiedene Arten abgeschlossen werden. Einerseits mit einer sofortigen Durchführung (Abschluss mit **OK**), andererseits mit einer aufgeschobenen Durchführung (Abschluss mit **Einfügen**). Bei der Wahl von **OK** öffnet sich das Ausgabefenster mit der Ausgabe der jeweiligen Ergebnisdarstellung und bei der Wahl von **Einfügen** das Syntaxfenster mit der skriptartigen Anweisung der jeweiligen Auswertung (ggf. wird ein bereits geöffnetes Fenster um den Inhalt erweitert).

Im vorliegenden Skript ist mit „abschließen“ zumeist die Syntax-Variante mit **Einfügen gemeint!**

Jede Operation in Syntax-Schreibweise besteht im Wesentlichen aus einem Befehlswort (was), mehreren Kriterien bzw. Spezifikationen (wie), einem abschließenden Ausführungsbefehl (wann) und einem Punkt (.) als Stoppsymbol für die jeweilige Aktion. Je nach Operation weist das wie unterschiedliche Komplexität auf, allerdings besteht ein weiterer Vorteil darin, dass die formulierte Syntax nicht zu 100% verstanden sein muss, da der Arbeitsschritt bereits vorab über die Menüführung definiert wurde.

Die Struktur der Syntax wird vom Programm farblich gekennzeichnet. Falsche, nicht erkannte oder noch nicht abgeschlossene Befehlswörter werden **hellrot** angezeigt – **dunkelblau** wenn Sie abgeschlossen sind. **Grün** steht für Unterbefehle und **dunkelrot** für Schlüsselwörter innerhalb einer Operation. Alles andere wie bspw. Variablennamen wird in **Schwarz** angezeigt.

Um einen dokumentierten Arbeitsschritt auszuführen bzw. in die Tat umzusetzen, wird der betreffende Teil markiert und mit einem Klick auf  oder mit der Tastenkombination **Strg + R** ausgeführt.

Neben den automatisierten Eintragungen von SPSS, kann natürlich auch direkt im Syntaxeditor gearbeitet werden. Neben Veränderungen der bestehenden Syntax, Kopiervorgängen und Anpassungen sind hier vor allem Kommentierungen der Arbeit sinnvoll und empfehlenswert. Damit das Programm eine Kommentierung nicht als Befehlswort auffasst, wird ihr ein * vorangestellt. Zusätzlich muss dem Programm vermittelt werden wann die Kommentierung endet, was mit einem Punkt (.) am Ende des Kommentars geschieht. Kommentierungen werden in **Grau** angezeigt.

In der Syntax wird kein Wert auf Groß- und Kleinschreibung gelegt, mehrere Leerzeichen werden als ein einzelnes interpretiert und jede Aktion muss mit einem Punkt abgeschlossen werden.

Über einen Rechtsklick auf ein Befehlswort kann die zugehörige und sehr umfangreiche SPSS Hilfe aufgerufen werden. Hier findet man neben einer Auflistung aller notwendigen Operanden auch Beispiele zum Befehl und weiterführende Informationen.

Neben der bereits angesprochenen Tastenkombination Strg + R können im Rechtsklickmenü (je nach SPSS Version) weitere Tastenkombinationen bzw. Shortkeys abgerufen werden. Hervorzuheben ist bspw. die Tastenkombination Strg + Leertaste welche eine Liste aller möglichen Befehle in alphabetischer Reihenfolge liefert.

Eine mögliche Analyse bzw. Syntaxeintragung könnte beispielsweise wie folgt aussehen:

*Häufigkeitsanalyse der Variable Alter.

```
FREQUENCIES VARIABLES=age  
  /STATISTICS=STDDEV MEAN MEDIAN  
  /ORDER=ANALYSIS.
```

Mit **FREQUENCIES** wird eine Häufigkeitsanalyse eingeleitet, welche durch das Schlüsselwort **VARIABLES** genauer definiert wird (hier für die Variable **age**). Die Unterbefehle **/STATISTICS** und **/ORDER** geben (hier) zusätzlich an, welche Statistiken mit ausgegeben und in welcher Reihenfolge die Analysen angezeigt werden sollen. Um den Inhalt der Prozedur später schneller erfassen zu können, wird nachträglich noch eine Kommentierung vorgenommen.

Ohne weit vorgreifen zu wollen, sei an dieser Stelle nochmals darauf hingewiesen, dass die Prozedur nicht selbst verfasst werden muss, sondern (hier) über *Analysieren / Deskriptive Statistik / Häufigkeiten* ausgewählt, spezifiziert und automatisch über **Einfügen** in das Syntaxfenster aufgenommen wird! Einzig die Kommentierung muss nachträglich manuell vorgenommen werden.

Umgang mit Daten

Datensätze resultieren zumeist aus Befragungen bspw. auf Grundlage von Fragebögen. Jede Frage eines Fragebogens misst hierbei ein bestimmtes Merkmal der Teilnehmer bzw. allgemein, eine spezifische Eigenschaft eines Untersuchungsobjekts. Die erhobenen Merkmale werden anschließend einer Skala zugeordnet. Dabei wird zwischen nominalen (klassifikatorischen), ordinalen (komparativen) und metrischen (intervall-, verhältnisskalierten) Merkmalen unterschieden.

Skalenniveau

Skalenart		Mögliche Aussagen	Beispiel
Nominalskala		Gleichheit Verschiedenheit	Geschlecht, Krankheitsklassifikation,
Ordinalskala		Größer-kleiner-Relationen	Windstärken, Ränge in der Siegerehrung
Metrisch	Intervallskala	Gleichheit von Differenzen	Temperatur in C°, IQ-Skala
	Verhältnisskala	Gleichheit von Verhältnissen	Längenmessung, Gewichtsmessung

1 Die vier wichtigsten Skalentypen nach Bortz/Schuster (2010) (eigene Darstellung)

Merkmale können (reduziert betrachtet) bis zu drei Attribute aufweisen: Quantitative Ausprägung, Konstante Abstände zwischen aufeinanderfolgenden Skalenwerten, Absoluter Nullpunkt. Eine **Nominalskala** besitzt keines dieser Attribute – die Ausprägungen besitzen schlichtweg einen Unterschied. Wenn dieser Unterschied zusätzlich noch in eine Rangfolge gebracht werden kann, also quantifizierbar ist und Aussagen wie *größer/kleiner, besser/schlechter* möglich sind, handelt es sich um eine **Ordinalskala**. Wenn diese Rangfolge konstante Abstände aufweist, bspw. der Anstieg von 1 Punkt auf der IQ-Skala immer gleichviel Wertigkeit besitzt, liegt eine **Intervallskala** vor. Das höchstmögliche Skalenniveau wird schließlich durch einen zusätzlichen absoluten Nullpunkt erreicht, wie u.a. bei Längen- oder Gewichtsmessungen. In diesem Fall spricht man von einer **Verhältnis-** oder **Ratioskala**. Vom Niveau der Skala hängt u.a. ab, welche Analysen durchgeführt bzw. welche Maßzahlen berechnet werden können. In mancher Literatur wird zusätzlich noch von einer fünften Skalierung gesprochen, der **Absolutskala**, welche vorliegt, wenn zusätzlich eine natürliche Einheit vorliegt, es sich also bspw. um Stückzahlen handelt. Die letzten zwei (bzw. drei) Skalarten werden in der Gruppe der metrischen Skalen zusammengefasst.

Eine Variable kann nur ein Skalenniveau besitzen (!), dennoch sind vor allem Ordinalskala und Intervallskala nicht immer ohne weiteres zu differenzieren weshalb auch oft von quasi-metrischen Skalen gesprochen wird – Skalen die augenscheinlich ein ordinales Skalenniveau aufweisen, aber bedingt durch theoretisch gleichgroße Abstände der Skalenpunkte auch als Intervallskala verstanden werden könnten. Im Zweifelsfall sollte im Methodenteil einer Arbeit hinterlegt werden welches Skalenniveau für welche Variable angenommen wird, da bspw. manche Verfahren ein intervallskaliertes Merkmal voraussetzen.

Eine Variable die ein solches „Problem“ mit sich ziehen würde, ist im folgenden Gliederungspunkt (→ Kodierung) als Beispiel dargestellt.

Kodierung

Für die Auswertung mithilfe computergestützter Programme ist es günstig (meist unabdingbar), dass die Merkmale bzw. deren Ausprägungen mit einer Abbildungsvorschrift versehen werden – kodiert, also durch Zahlen repräsentiert werden. Hierbei wird jeder Merkmalsausprägung eine eindeutige Zahl zugeordnet. Dabei ist es wichtig, dass vorliegende Informationen, bspw. die Rangfolge, weiterhin erhalten bleiben.

2. Ich lese Ihnen eine Reihe von Gefühlen vor. Geben Sie bitte jeweils an, wie häufig oder selten Sie dieses Gefühl in den letzten vier Wochen erlebt haben.

Wie oft haben Sie sich ...

	Sehr selten	Selten	Manchmal	Oft	Sehr oft
- ärgerlich gefühlt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
- ängstlich gefühlt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Sehr selten könnte in diesem Beispiel mit 1 kodiert werden, *selten* mit 2, *manchmal* mit 3, etc. . Eine Kodierung mit 1, 3, 2, 5, 4 wäre hingegen unzulässig, da hierdurch die Rangfolge der vorliegenden (augenscheinlichen) Ordinalskala nicht mehr abgebildet ist.

Werte(label)

Die vorgenommene Kodierung wird in der Variablenansicht hinterlegt. Auf das Beispiel bezogen, könnte ein Teil des Eintrags in der VA folgendermaßen aussehen:

	Name	Typ	Breite	Dezimal...	Beschriftung	Werte	Fehlend	Spalten	Ausrichtung
1	F2_Emo_Ärger	Numerisch	8	2	F2 Wie oft haben Sie sich ärgerlich gefühlt?	{1,00, sehr ...	-7,00,-8,00,...	8	Rechts
2	F2_Emo_Angst	Numerisch	8	2		Keine	Keine	8	Rechts

Unter *Werte* wurde bereits die Kodierung hinterlegt. Die Subfenster können jeweils über den im Feld vorhandenen Button aufgerufen werden.



Wertbeschriftungen

Wert:

Beschriftung:

Hinzufügen Ändern Entfernen

1,00 = "sehr selten"
2,00 = "selten"
3,00 = "manchmal"
4,00 = "oft"
5,00 = "sehr oft"

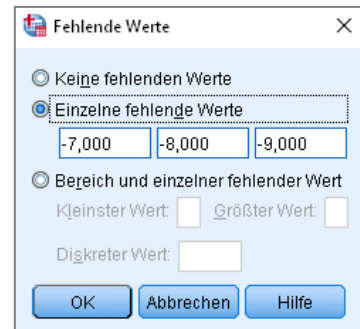
Rechtschreibung...

OK Abbrechen Hilfe

Fehlende Werte

Sofern im Fragebogen fehlende oder falsche Nennungen vorliegen, sollten diese ebenfalls numerisch vercodet werden. U.U. kann es sogar von Interesse sein, zwischen falschen und fehlenden Werten zu differenzieren. Gebräuchlicherweise werden fehlende Werte mit -8 (*auch 88 oder 888*) und falsche Werte mit -9 (*auch 99 oder 999*) gekennzeichnet. Diese Werte, die nicht in die Analyse mitaufgenommen werden sollen, können je Variable unter *Fehlend* definiert werden.

Soll im Text der späteren Analyse bspw. -8 durch *Fehlender Wert* ersetzt werden, kann dies zusätzlich in der Wertbeschriftung hinterlegt werden.



Umgang mit Filterfragen

In manchen Fällen können leere Felder auch bewusst vorliegen. Fragen welche mit „wenn ja, ...“ einleiten sollen bspw. bewusst von einem Teil der Befragten übersprungen werden, Nichtraucher würden die Frage zur „Anzahl der gerauchten Zigaretten pro Tag“ nicht beantworten und so weiter. In Fällen wie diesen könnte eine weitere Kodierung vorgenommen werden – bspw. -7 für *Filterbedingt fehlend*.

Sollte allerdings eine Antwort vorliegen obwohl der Proband filterbedingt ausgeschlossen wurde, darf diese Angabe natürlich nicht gewertet werden und im Sinne der Datenbereinigung bspw. mit einer -9 (für falsche Angabe) kodiert werden!

Umgang mit „Sonstiges“

Eine Variable im Datensatz besteht entweder aus Zahlen oder (seltener) aus Text – eine Mischung beider Datentypen ist hinsichtlich der Eingabe nicht möglich (auch weil sich die Auswertungsmöglichkeiten von Zahlen und Text teils deutlich voneinander unterscheiden).

Sollten Fragen vorliegen, welche als Skalenwert bspw. ein „Sonstiges: _____“ enthalten, müssen diese als zwei unterschiedliche Variablen gehandhabt werden! Die erste Variable symbolisiert die „Ankreuzmöglichkeit“, die zweite Variable den Freitext bei angekreuztem „Sonstiges“.

Dementsprechend haben auch alle Personen die etwas Anderes als „Sonstiges“ gewählt haben in der neuen Variable ein „*sollte nicht beantwortet werden*“ bzw. „*filterbedingt fehlend*“ (-7).

Umgang mit Mehrfachnennungen

Je nach Anzahl der möglichen Angaben vervielfältigen sich die Kombinationsmöglichkeiten bei möglichen Mehrfachnennungen. Jede Kombinationsmöglichkeit als einzelne Aufprägung zu behandeln, ist weder effizient noch zielführend im Sinne der Analyse (bspw. gibt es bei 10 Kästchen 3.628.800(!) verschiedene Kombinationen). Im Falle von Mehrfachnennungen wird dieses Problem damit umgangen, indem jede Antwortmöglichkeit als einzelne Frage (Variable) aufgefasst und mit „angekreuzt“ bzw. „nicht angekreuzt“ kodiert wird. Ein Informationsverlust besteht hierbei nicht, da die Kombinationsmöglichkeiten immer noch durch Filter (→ Selektion), Stratifizierungen (→ Kreuztabellen) oder spezifische Operationen (→ Mehrfachantworten) angezeigt werden können.

Umgang mit Freitext

Zwar arbeiten Statistikprogramme vorwiegend mit numerischen Angaben, allerdings ist die Eingabe von Freitext im Datensatz ebenfalls möglich. Die Variable – standardmäßig numerisch – muss in diesem Fall auf *Zeichenfolge (string)* gesetzt sowie u.U. die Anzahl der möglichen Zeichen (*Spaltenformat*) erhöht werden.

Um Text quantifizierbar zu machen bzw. mithilfe quantitativer Verfahren analysieren zu können, werden aus den vorhandenen Nennungen Kategorien gebildet. Diese sollten in sich so homogen wie möglich sein sowie zueinander so heterogen wie möglich. Anschließend wird jede Textnennung einer Kategorie zugeordnet sowie eine weitere Variable eingefügt, in der jeder Kategorie eine Kodierung zugewiesen wird.

Um Freitextnennungen quantitativ auswerten bzw. prägnant darstellen zu können, wird versucht, die einzelnen Nennungen unter Oberbegriffen zusammenzufassen. Hierbei gilt der Grundsatz, so viele Oberbegriffe wie nötig und so wenig Oberbegriffe wie möglich zu bilden bzw. dass die resultierenden Gruppen in sich so homogen wie möglich und gleichzeitig untereinander so heterogen wie möglich sind. Anschließend können Aussagen getroffen werden wie häufig Begriffe einer Kategorie (Oberbegriff) angegeben wurden. Diese (Neu-)Gruppierung wird als neue nominale Variable in SPSS hinterlegt um die einzelnen Nennungen den jeweiligen Begriffen zuordnen zu können.

Beispielsweise könnten die Begriffe *Apfel*, *Banane*, *Salat*, *Veganismus* und *Gurke* unter den Oberbegriffen *Obst* (Apfel, Banane (Kodierung 1)), *Gemüse* (Gurke, Salat(2)) und *Sonstiges* (Veganismus (3)) zusammengefasst werden.

Dateneingabe

Als erster Schritt der Dateneingabe sollte die VA anhand des Fragebogens gepflegt werden. Jede Variable wird aufgeführt, näher spezifiziert sowie mit Ihrer individuellen Kodierung hinterlegt. Zusätzlich sollte eine Variable *Fragebogennummer (Case)* als laufende Nummer hinterlegt werden, da nur so ein späterer Rückschluss auf den Fragebogen (auf dem die Nummer natürlich auch hinterlegt sein sollte) möglich ist.

Nachdem die VA vollständig hinterlegt wurde, kann in der DA mit der Eingabe der Daten begonnen werden. Wie bereits an anderer Stelle erwähnt, entspricht eine Zeile in der DA einem Teilnehmer bzw. einem Untersuchungsobjekt. Jede Spalte einer Variablen und jede Zelle dementsprechend einer spezifischen Ausprägung eines Teilnehmers in der jeweiligen Variablen.



Durch einen Klick auf **Wertbeschriftung** in der oberen Leiste, kann zwischen dem hinterlegten Text und der Kodierung gewechselt werden.

Datei Bearbeiten Ansicht Daten Transformieren Analysieren Direktmarketing Grafik Extras Erweiterungen Fenster Hilfe											
6 : KHkhkmyo 2											
	PIDNR	age10B	age5B	sex	ktyp4	reg7	KHhyp	KHhyp1 2	KHhyp medA	KHkhk	
1	1201000573	2	4	1	3	3	2	-96	-96	2	
2	1201000676	6	12	2	1	5	1	1	2	2	
3	1201001066	1	1	1	1	2	2	-96	-96	2	
4	1201001072	7	14	2	1	5	2	-96	-96	1	
5	1201001352	5	9	2	2	3	1	1	1	2	
6	1201002609	2	4	2	3	6	2	-96	-96	2	
7	1201004146	2	3	1	1	5	2	-96	-96	2	
8	1201004200	3	5	1	2	3	2	-96	-96	2	
9	1201004899	4	7	1	1	1	2	-96	-96	2	
10	1201006038	7	13	2	2	3	1	1	1	2	
11	1201006681	3	6	2	2	6	2	-96	-96	2	
12	1201006954	6	11	1	2	7	1	1	1	2	

Syntax: Pflege der VA

Die Pflege der VA mittels Syntax kann bestimmte Arbeitsschritte beschleunigen bzw. erleichtern. Bestimmte Variablen (die zumindest einen Namen hinterlegt haben) können bspw. auch mittels Syntax näher spezifiziert werden.

Die gängigsten Befehlswörter sind in diesem Zusammenhang...

VARIABLE LABELS *VALUE LABELS* *VARIABLE LEVEL* *MISSING VALUES*

Und das wichtigste Schlüsselwort... *TO*

Mithilfe des Syntax-Befehls *VARIABLE LABELS* kann die Beschriftung einer Variablen hinterlegt werden...

*Variablenbeschriftung age – Alter in Jahren.

VARIABLE LABELS age 'Alter in Jahren'.

...mit *VALUE LABELS* die Wertbeschriftungen und mit *VARIABLE LEVEL* das Skalenniveau.

*Wertbeschriftung smoke - Rauchstatus.

VALUE LABELS smoke 1 'täglich' 2 'gelegentlich' 3 'selten' 4 'nie'.

*Skalenniveau age und smoke.

VARIABLE LEVEL age (*scale*).

VARIABLE LEVEL smoke (*ordinal*).

Mit *MISSING VALUES* können fehlende Werte definiert werden.

*Fehlende Werte age.

MISSING VALUES age (-9).

Mithilfe des Schlüsselworts *to*, können Variablen-Bereiche abgesteckt werden („von“ „bis“). Entscheidend ist hierfür die Reihung die in der Variablenliste vorliegt (verständlicherweise funktioniert der Befehl bspw. nicht bei *VARIABLE LABELS*).

*Fehlende Werte von *smoke* bis *nutrition*.

MISSING VALUES *smoke to nutrition* (-7, -8, -9).

to ist bspw. dann sinnvoll, wenn eine vorhandene Kodierung über mehrere Items vergessen wurde, schlichtweg falsch geschrieben wurde oder bspw. nachträglich ein weiterer fehlender Wert definiert werden muss.

Um nachträglich ein Wertelabel hinzuzufügen kann der Befehl *ADD VALUE LABELS* genutzt werden, welcher mit *EXECUTE* abschließend bestätigt werden muss.

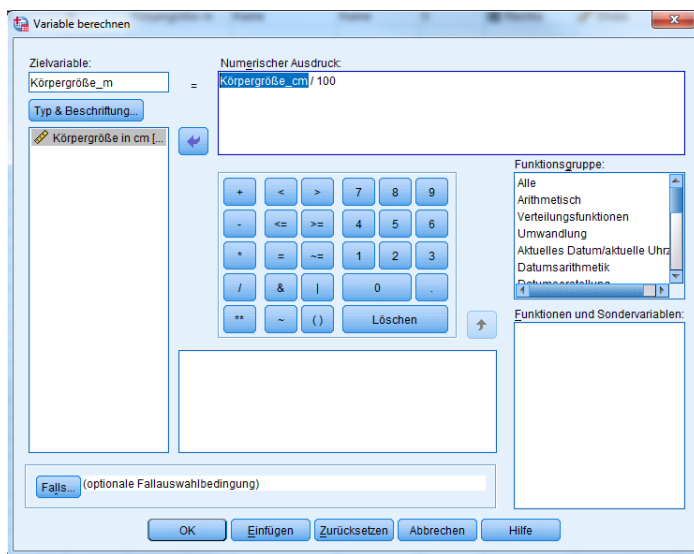
*Zusätzliches Wertelabel *smoke* – ehemalige Raucher.

ADD VALUE LABEL *smoke* 5 'nicht mehr'.

Transformieren – Modifikation von Variablen

Sobald Variablen über mehrere Untersuchungsobjekte hinweg verändert werden sollen, neue Variablen mithilfe bestehender Variablen berechnet werden sollen, wird das Menü *Transformieren* herangezogen.

Variable berechnen



Unter Umständen müssen Daten für einzelne Analysen zusammengefasst werden bzw. bereits erhobene Merkmale miteinander verrechnet werden. Unter *Transformieren / Variable berechnen* könnte bspw. aus einer erhobenen Körpergröße in cm die Körpergröße in m berechnet werden.

Auf der linken Seite wird die Zielvariable – die neue Variable – benannt. In der linken Spalte finden sich alle Variablen des Datensatzes (hier nur eine).

Unter Numerischer Ausdruck wird die zu verwendende Berechnung definiert – dies erfolgt bspw. unter Zuhilfenahme der vorhandenen Variablen, welche mit Doppelklick, Pfeiltaste oder Drag & Drop eingefügt werden können (hier mit Hilfe der Variablen Körpergröße_cm). Nach Fertigstellung der Rechenvorschrift wird die Erstellung der neuen Variable abgeschlossen.

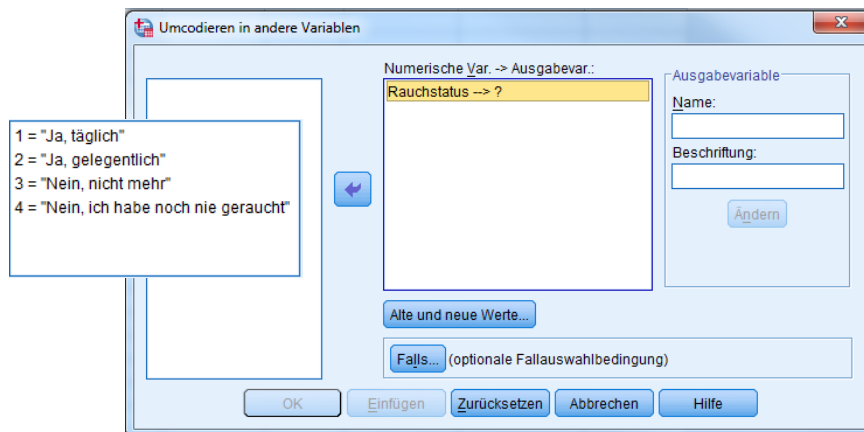
Unter **Typ und Beschriftung** könnte die neue Variable gleich mit einer Variablenbeschriftung versehen werden oder wahlweise der numerische Ausdruck als Beschriftung übernommen werden. Die Variable wird letztlich in der VA an unterster Stelle eingefügt.

Neben der „Neuberechnung“ einzelner Variablen können auch Variablen miteinander verrechnet werden um eine Variable zu erstellen. Beispielsweise könnte der BMI berechnet werden (sofern die Körpergröße in Metern sowie das Gewicht in Kilogramm vorliegt).

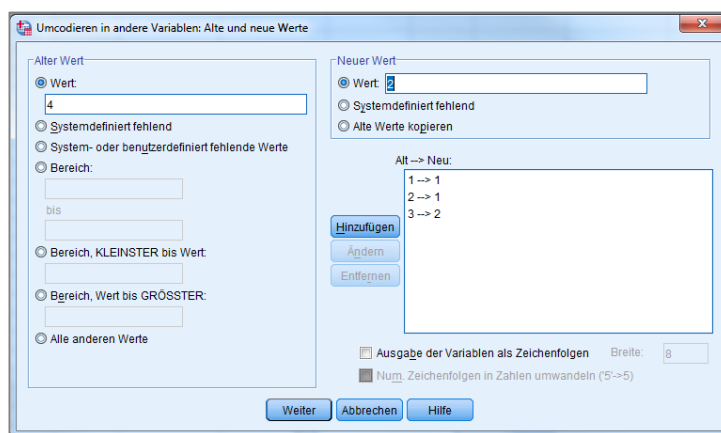
Umcodieren in dieselbe/andere Variable

Für manche Analysen ist es wiederum sinnvoll oder gar notwendig, einzelne Variablen größer zu klassieren oder deren Kodierung im Nachhinein anzupassen. Hierbei handelt es sich um eine Veränderung von vorhandenen Variablen, also ebenfalls um eine Transformation.

Unter *Transformieren / Umcodieren in andere Variable* könnte bspw. ein erhobener Rauchstatus (täglich, gelegentlich, nicht mehr, niemals) in Raucher und Nichtraucher dichotomisiert werden.



Die Variable (hier Rauchstatus) wird in das mittlere Feld gezogen und somit in die Verarbeitung mit aufgenommen. Unter Ausgabevariable wird die neue Variable mit Name und Beschriftung definiert (z.B. Rauch_dich) was anschließend mit einem Klick auf **Ändern** bestätigt wird.



Unter **Alte und neue Werte** werden die Vorschriften der Umcodierung definiert. Auf der linken Seite des Submenüs werden die Kodierungen der alten Kodierung eingegeben, auf der rechten Seite, die der neuen. Über **Hinzufügen** werden die definierten Regeln in das Kodierungsschema übernommen.

Im Beispiel werden die vormaligen 4 Gruppen auf 2 reduziert.

Nach der Bestätigung mit **Weiter** und dem Abschließen wird eine neue Variable anhand der definierten Kodierungsvorschriften erstellt. Schlussendlich findet sich die neue Variable in der VA an letzter Stelle. Bestenfalls sollten die Wertbeschriftungen der neuen Variable gleich im Anschluss gepflegt werden (hier: 1=Raucher; 2= Nichtraucher).

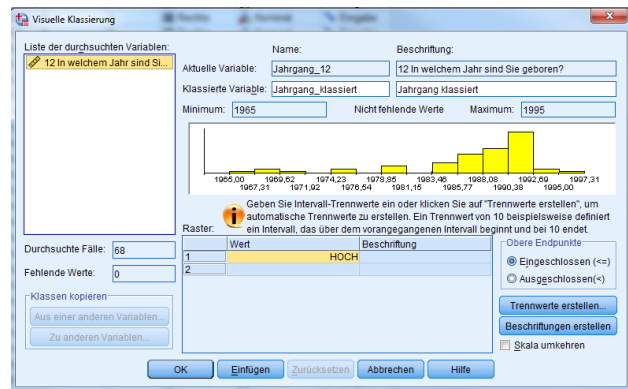
Ebenfalls möglich wäre es, Bereiche einer Variablen abzustecken. Beispielsweise der Bereich vom kleinsten vorliegenden bis zum definierten Wert (KLEINSTER bis Wert) oder ein Bereich von $(\geq) X$ bis $(<) Y$.

Neben dem *Umcodieren in andere Variable* besteht auch die Möglichkeit von *Umcodieren in dieselbe Variable*. Hierbei wird keine neue Variable erstellt, sondern die Kodierung der vorhandenen Variable überschrieben! Im Zweifelsfall sollten lieber neue Variablen erstellt werden, damit der Rohdatensatz weiterhin bestehen bleibt und eine Überschreibung der Daten nicht mehr rückgängig gemacht werden kann!

Visuelles Klassieren

Sofern im Vorfeld unklar oder aus der Theorie nicht eindeutig eruierbar wie bestimmte Skalen heruntergebrochen bzw. Variablen gruppiert werden sollen, kann die Aktion *Visuelles Klassieren* benutzt werden. Der Menüpunkt findet sich ebenfalls unter *Transformieren*.

Nachdem die zu gruppierende *Aktuelle Variable* definiert wurde, kann deren Verteilung eingesehen werden sowie die zu erstellende *Klassierte Variable* benannt und beschriftet werden.



Unter **Trennwerte erstellen** kann ausgewählt werden, wieviele Gruppen die neue Variable aufweisen soll bzw. wie viele gleichgroße Gruppen erstellt werden sollen. Unter **Beschriftung erstellen** kann die jeweilige Gruppierung automatisch als Wertbeschriftung übernommen werden. Nach Abschluss wird die neue klassierte Variable am Ende der VA angezeigt.

z- Standardisierung

Eine weitere gängige Transformation liegt in der Standardisierung von Variablen – meist in Form der z-Standardisierung. Rechnerisch wird hierbei von allen Werten der Mittelwert abgezogen und das Ergebnis durch die Standardabweichung der Variablen geteilt.

$$z_i = \frac{x_i - \bar{x}}{s}$$

Die einfachste Möglichkeit z-standardisierte Werte zu erhalten, befindet sich im Menü *Analysieren/Deskriptive Statistik*. Die zu standardisierenden Variablen werden von der linken Liste in den rechten Bereich gezogen der Haken bei **Standardisierte Werte als Variable speichern** wird gesetzt und die Prozedur abgeschlossen. Die neue Variable am Ende der Variablenliste und kann entsprechend betitelt bzw. angepasst werden.

Invertieren von Werten

Oftmals ist es notwendig, einzelne Skalen zu invertieren – beispielsweise bei Kontrollfragen innerhalb einer psychologischen Testbatterie. Invertieren bedeutet hierbei, dass die Skala umgepolt, bzw. die Kodierung umgedreht werden muss.

Theoretisch kann eine Invertierung über *Umcodieren in dieselbe/andere Variable* vorgenommen werden – bspw. 1 wird zu 5; 2 wird zu 4; ... 5 wird zu 1 – allerdings ist dieses Vorgehen bei sehr vielen Skalenausprägungen (bspw. 1-100) nicht sehr effizient.

Eine weitere Möglichkeit liegt im Erstellen einer neuen Variable mit der Berechnung

$$x_{Skala.Min} + x_{Skala.Max} - x_i$$

Übersetzt also: „Der Invertierte Wert jeder Person i berechnet sich aus dem theoretischen Minimum + dem theoretischen Maximum – dem alten Wert der Person i“.

Syntax: Transformieren

Mittels Syntax können ebenfalls Recodierungen (RECODE) und Neuberechnungen (COMPUTE) von Variablen vorgenommen werden.

*Umkodierung der Variable Rauchen zu 2 Ausprägungen (in dieselbe Variable).

```
RECODE smoke (1=1) (2=1) (3=2) (4=2).  
EXECUTE.
```

*Umkodierung der Variable Rauchen zu 2 Ausprägungen (in andere Variable).

```
RECODE smoke (1=1) (2=1) (3=2) (4=2) INTO smoke_dichotom.  
EXECUTE.
```

Hinter RECODE verbergen sich 2 verschiedene Operationen. Im ersten Fall wird *Umcodieren in dieselbe Variable* ausgeführt – im zweiten Fall *in andere Variable*. Der Unterschied der beiden Operationen besteht schlicht im Zusatz von INTO und dem neuen Variablennamen. Wie bereits an anderer Stelle erwähnt, ist im Normalfall das *Umcodieren in andere Variable* zu präferieren.

Für das Kodierungsschema wird je Kodierung eine Klammer verwendet z.B. (1=10) um Kodierungen spezifisch zuzuweisen. Bereiche können ebenfalls definiert werden, bspw. (5 thru 10=2), (LOWEST thru 5=1) oder (10 thru HIGHEST=3). Sollen darüber hinaus alle nicht abgedeckten Werte als (system)fehlend definiert werden kann zusätzlich die Vorschrift (ELSE=SYSMIS) verwendet werden.

Die Berechnung neuer Variablen erfolgt über den Befehl COMPUTE (*dt.: etwas berechnen/kalkulieren*).

*Neue Variable BMI.

```
COMPUTE BMI = weight_kg / (height_m ** 2).  
EXECUTE.
```

Die Rechenvorschrift kann mehrere Variablen enthalten und allgemeine Rechenprioritäten (Potenz > Klammer > Punkt > Strich) finden hier ebenfalls Anwendung.

*Anm.: Wie in vielen EDV-Programmen üblich, wird die Multiplikation durch ein * und die Division durch einen / dargestellt. Zwei * stehen für Potenz bzw. gesprochen für „hoch“ – ergänzt durch die 2 ergibt sich hier die Quadratsumme („hoch 2“).*

Natürlich können die neu zu erstellenden Variablen gleich mit Informationen versehen werden (→ Pflege der VA)...

*Neue Variable BMI.

```
COMPUTE BMI = weight_kg / (height_m ** 2).  
VARIABLE LABELS BMI 'Body-Mass-Index'.  
VARIABLE LEVEL BMI (scale).  
MISSING VALUES BMI (-9).  
EXECUTE.
```

Mehrere Operationen können auch zusammen verfasst bzw. miteinander ausgeführt werden.

*Neue Variable BMI + Klassierung nach WHO.

```
COMPUTE BMI = weight_kg / (height_m ** 2).  
VARIABLE LABELS BMI 'Body-Mass-Index'.  
VARIABLE LEVEL BMI (SCALE).  
MISSING VALUES BMI (-9).  
EXECUTE.
```

```
RECODE BMI (LOWEST thru 18,5=1) (18,5 thru 25=2) (25 thru 30=3)  
          (30 thru HIGHEST=4) INTO BMI_klassiert.  
VARIABLE LABELS BMI_klassiert 'Body Mass Index, klassiert nach WHO  
          (4 Gruppen)'.  
VALUE LABELS BMI_klassiert 1 'untergewichtig' 2 'normalgewichtig'  
          3 'übergewichtig' 4 'adipös'.  
VARIABLE LEVEL BMI (ORDINAL).  
MISSING VALUES BMI (-9).  
EXECUTE.
```

Natürlich ist die Reihenfolge der einzelnen Operationen wichtig. Im Beispiel kann bspw. der BMI nicht klassiert werden, wenn er noch nicht berechnet wurde.

Ist eine Zeile noch nicht abgeschlossen, überschreitet aber eine gewisse Zeichenzahl, muss sie umgebrochen werden. Damit Benutzer und Programm nicht von einer neuen Operation ausgehen, wird die umgebrochene Zeile anschließend eingerückt (je nach Version u.U. notwendigerweise).

Im Normalfall muss jede Zeile mit einem Punkt abgeschlossen werden.

Daten – Selektionsmöglichkeiten

In vielen Fällen kommt es vor, dass Analysen für selektierte Gruppen vorzunehmen sind, bspw. um exponierte Gruppen zu erforschen oder um geschlechtsspezifische Effekte aufzudecken. SPSS bietet hierfür verschiedene Möglichkeiten an, um bspw. eine Gruppe „herauszuselektieren“ (→ *Fälle auswählen*) oder Gruppen miteinander zu vergleichen (→ *Datei aufteilen*).

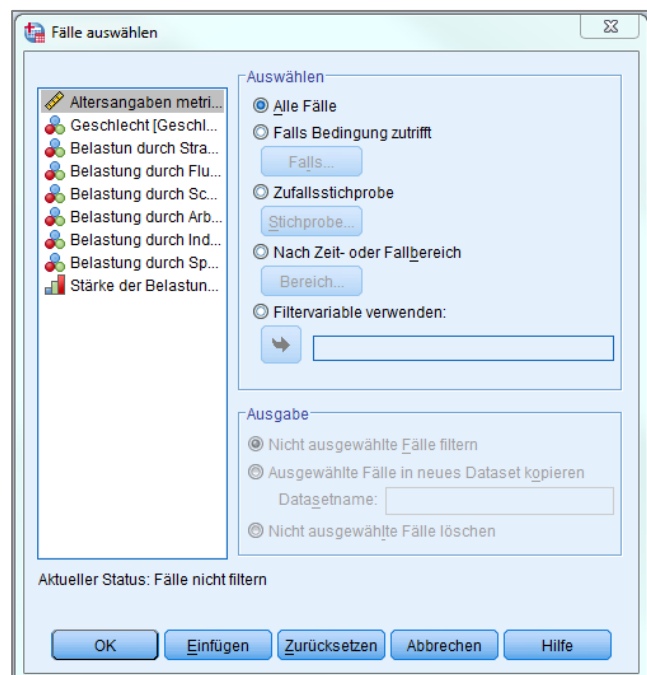
Fälle auswählen

Die Aktion kann über das *Daten*-Menü aufgerufen werden. Wie der Name schon vermutet lässt, können hierdurch bestimmte Fälle selektiert werden, je nachdem ob Sie die definierten Bedingungen erfüllen oder eben nicht.

Unter **Falls** im Punkt **Falls Bedingung zutrifft**, können einzelne Bedingungen definiert werden unter Einbezug der Variablenamen.

Eine Bedingung könnte bspw. lauten „Rauchen = 2 & Geschlecht = 1“; also z.B. alle weiblichen Nichtraucher. Operanden wie <, >, &, OR sind ebenfalls möglich.

Unter **Zufallsstichprobe** kann aus den vorhandenen Daten ein Prozentsatz an zufälligen Fällen selektiert werden. Nach **Zeit- und Fallbereich** selektiert Fälle bspw. von Nummer 1 bis 15.

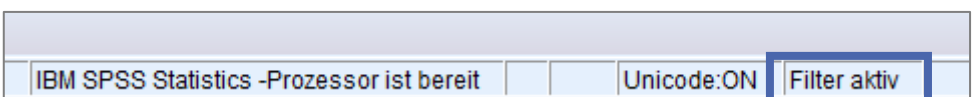


	Nr	G
1	1	g
2	2	g
3	3	g
4	4	sch
5	5	g
6	6	g
7	7	wenig
8	8	sehr
9	9	wenig
10	10	g
11	11	wenig
12	12	g
13	13	g

Filter werden nicht einfach nur angewandt, sondern über eine Filtervariable in der VA angelegt bzw. erstellt. Diese Variablen (*\$filter*) haben 2 Ausprägungen (stimmt/nicht mit Bedingungen überein) und können, falls bereits erstellt, auch hier über *Filtervariable verwenden* angewandt werden oder für spätere Analysen umbenannt werden.

Das ein Filter aktiv ist, kann entweder am rechten unteren Rand abgelesen werden, bzw. ist dadurch ersichtlich, dass bestimmte Fälle als nicht selektiert, durchgestrichen angezeigt werden.

Über *Fälle auswählen/Alle Fälle*, wird ein Filter wieder entfernt.



Solange ein Filter aktiv ist, werden sämtliche Berechnungen nur mit den selektierten Fällen durchgeführt!

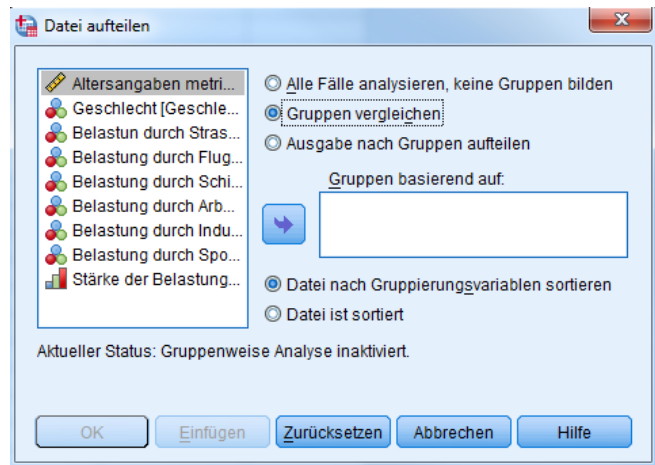
Aufgeteilte Datei

Um Gruppen miteinander vergleichen zu können, kann neben der Filterfunktion alternativ auch die Aktion *Aufgeteilte Datei* genutzt werden.

Unter *Daten / Gruppen vergleichen* kann eine Gruppierungsvariable eingefügt werden, bspw. Geschlecht. Anschließend werden alle folgenden Analysen für alle Ausprägungen der Variablen separat berechnet (hier: je einmal für männlich und weiblich).

Gruppen vergleichen und *Ausgabe nach Gruppen aufteilen* unterscheidet sich darin, dass in der Ausgabe jeweils eine gemeinsame Tabelle pro Analyse erstellt wird oder zuerst die Analyse der ersten Ausprägung durchgeführt werden soll, dann der zweiten, etc. .

Da die Datendatei im Normalfall nicht nach etwaigen Gruppierungsvariablen sortiert ist, sollte *Datei nach Gruppierungsvariablen sortieren* angewählt bleiben.



Syntax: Selektion

Fälle können ebenfalls über Syntax ausgewählt werden, wobei aus Gründen der Komplexität definitiv die „Einfügen-Option“ vorzuziehen ist. Das Entfernen eines Filters erfolgt allerdings allgemein über:

*Aktiven Filter löschen.

FILTER OFF.
USE ALL.
EXECUTE.

Demgegenüber ist die Aktion *Aufgeteilte Datei* über Syntax wesentlich leichter zu bewerkstelligen.

*Gruppenvergleich nach Geschlecht – vergleichend.

SORT CASES BY sex.
SPLIT FILE LAYERED BY sex.

*Gruppenvergleich nach Geschlecht – in Ausgabe geteilt.

SORT CASES BY sex.
SPLIT FILE SEPARATE BY sex.

Für *Aufgeteilte Datei* müssen die Daten nach der Gruppierungsvariablen sortiert werden. Dies geschieht mittels **SORT CASES BY**. Danach wird die Datei mit **SPLIT FILE** aufgeteilt – wahlweise im Gruppenvergleich (**LAYERED**) oder aufgeteilt (**SEPARATE**). Mit **SPLIT FILE OFF** werden die Daten wieder „zusammengeführt“.

Neben der Selektion über die Menüführung, gibt es noch weitere Möglichkeiten, welche nur über Syntax nutzbar sind – bspw. sogenannte Wenn-Dann-Bedingungen.

Vor jedwede Analyse oder Operation kann ein **DO IF (...)** eingefügt werden. Besagte Aktion wird dann nur mit den Fällen durchgeführt, welche die definierten Bedingungen erfüllen.

Sollen Auswertungen für unterschiedliche Gruppen durchgeführt werden, kann das **DO IF** noch um weitere Bedingungen erweitert werden – mithilfe von **ELSE IF**.

*Rekodierung subjektiver Armut aufgrund Haushaltsnetto.

DO IF (Haushaltsnetto = 0 **thru** 1000).

RECODE Subj_Arm (0 **thru** 33=1) (33 **thru** 66=2) (66 **thru** 100=3) **INTO** Subj_Armut_klassiert.

ELSE IF (Haushaltsnetto = 1000 **thru** 2000).

RECODE Subj_Arm (0 **thru** 33=1) (33 **thru** 66=2) (66 **thru** 100=3) **INTO** Subj_Armut_klassiert.

ELSE IF (Haushaltsnetto = 2000 **thru** **HIGHEST**).


RECODE Subj_Arm (0 **thru** 33=1) (33 **thru** 66=2) (66 **thru** 100=3) **INTO** Subj_Armut_klassiert.

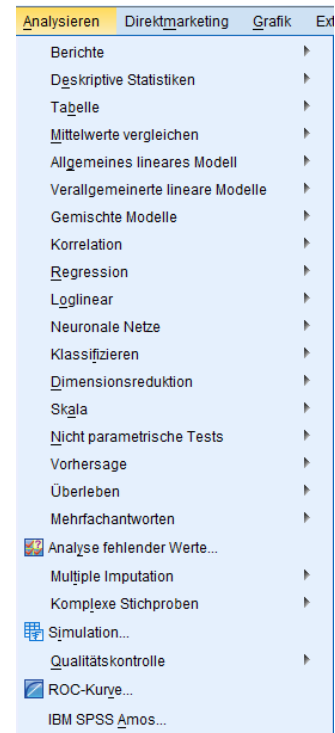
END IF.

EXECUTE.

Analysieren – Deskription

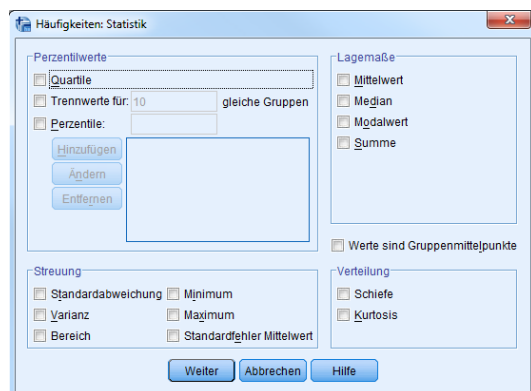
Nahezu alle Analysen sind in SPSS über das gleichnamige Menü aufrufbar. Nach Auswahl der jeweiligen Auswertungsmethode öffnet sich eine Dialogbox, welche folgende, mindestens ähnliche, Bausteine aufweist:

- **Liste der Variablen:** In der linken Spalte der Dialogbox findet sich eine Auflistung aller Variablen des Datensatzes
- **Variable(n):** In die rechte Spalte können diejenigen Variablen eingefügt werden, welche in die Analyse einbezogen werden sollen – entweder über  oder Drag & Drop oder per Doppelklick.
- **Weitere Schaltflächen:** Durch das Anklicken der unterschiedlichen Buttons im rechten und unteren Bereich der Dialogbox werden weitere Submenüs geöffnet oder Aktionen ausgeführt. **OK** startet die Analyse, **Einfügen** fügt die Analyse im Syntaxfenster ein, **Zurücksetzen** setzt alle Felder und Optionen der Analyse zurück, etc. . Drei Punkte ... deuten daraufhin, dass sich hinter dem Button ein weiteres Untermenü befindet.



Häufigkeiten

Für die deskriptive, also die beschreibende Analyse stehen verschiedene Methoden zur Verfügung. Die einfachste Art, Daten deskriptiv auszuwerten, liegt in der Analyse der Häufigkeiten (*Analysieren / Deskriptive Statistik / Häufigkeiten*). Neben der absoluten und relativen Anzahl der Ausprägung einer oder mehrerer Variablen können über **Statistiken** verschiedene Maßzahlen mitberechnet werden.



Das Ergebnis einer Häufigkeitsanalyse wird nach Bestätigung mit **OK** in der Ausgabe angezeigt.

5 Nun ein paar allgemeine Fragen zu Ihrer Gesundheit. Wie ist Ihr Gesundheitszustand im Allgemeinen? Ist er sehr gut, gut, mittelmäßig, schlecht oder sehr schlecht?					
		Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozente
Gültig	sehr gut	16	23,5	23,5	23,5
	gut	38	55,9	55,9	79,4
	weniger gut	8	11,8	11,8	91,2
	schlecht	2	2,9	2,9	94,1
	sehr schlecht	4	5,9	5,9	100,0
	Gesamtsumme	68	100,0	100,0	

*Häufigkeitsanalyse Gesundheitszustand.

FREQUENCIES VARIABLES=GesZ
/ORDER=ANALYSIS.

Deskriptive Statistik

Sind die Häufigkeiten nicht von Belang, sondern sollen nur Statistische Maßzahlen berechnet werden, kann auch die Aktion *Deskriptive Statistik* genutzt werden. Die Auswahl an Maßzahlen ist hier zwar begrenzter, allerdings ist die Ausgabe kompakter gehalten als bei der *Häufigkeiten*-Analyse.

*Deskription Gesundheitszustand – Mittelwert, Standardabweichung, Min, Max.

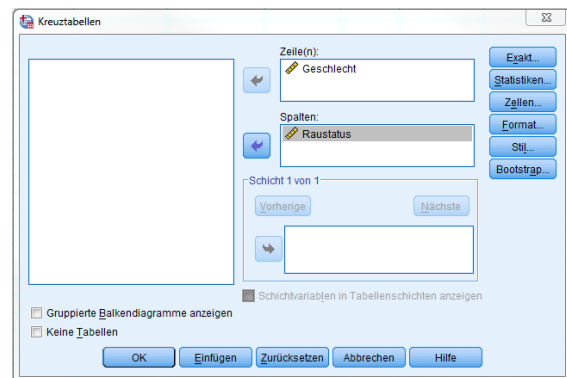
DESCRIPTIVES VARIABLES=GesZ
/STATISTICS=MEAN STDDEV MIN MAX.

Kreuztabellen (Kontingenztafeln)

Ein weiteres Mittel der deskriptiven Auswertung sind Kreuztabellen. Diese Tabellen, stellen Kombinationen von 2 oder mehr Variablen in absoluten und relativen Häufigkeiten dar. Über *Analysieren / Deskriptive Statistik / Kreuztabellen* können vorhandene Variablen – unterteilt in Zeilen und Spalten einer Tabelle – in die Analyse aufgenommen werden.

Unter **Zellen** kann eingestellt werden, ob absolute Häufigkeiten (bspw. die im Datensatz vorhandenen beobachteten Werte) und/oder relative Häufigkeiten (zeilen und/oder spaltenweise) angezeigt werden sollen.

Eine Ausgabe mit beobachteten Werten und Zeilenprozent, könnte bspw. so aussehen:



Geschlecht * Rauchstatus

			Rauchstatus				Gesamtsumme
			Ja, täglich	Ja, gelegentlich	Nein, nicht mehr	Habe noch nie geraucht	
Welches Geschlecht haben Sie?	Männlich	Anzahl	7	0	2	5	14
		% in Welches Geschlecht haben Sie?	50,0%	0,0%	14,3%	35,7%	100,0%
	Weiblich	Anzahl	14	10	7	23	54
		% in Welches Geschlecht haben Sie?	25,9%	18,5%	13,0%	42,6%	100,0%
Gesamtsumme		Anzahl	21	10	9	28	68
		% in Welches Geschlecht haben Sie?	30,9%	14,7%	13,2%	41,2%	100,0%

In das Feld *Schicht* können weitere Variablen eingefügt werden, welche die Analyse verfeinern. Beispielsweise könnte man den Zusammenhang zwischen Geschlecht und Rauchstatus nach Altersgruppen schichten. Also obige Tabelle für mehrere verschiedene Altersgruppen ausgeben lassen.

Zusätzliche spezifische Maße können unter **Statistiken** in die Analyse aufgenommen werden. Bspw. die relative Rate und das Odds Ratio (welche allerdings beide nur bei 4-Felder-Tafeln berechnet werden!).

*Kreuztabelle Geschlecht*Rauchstatus.

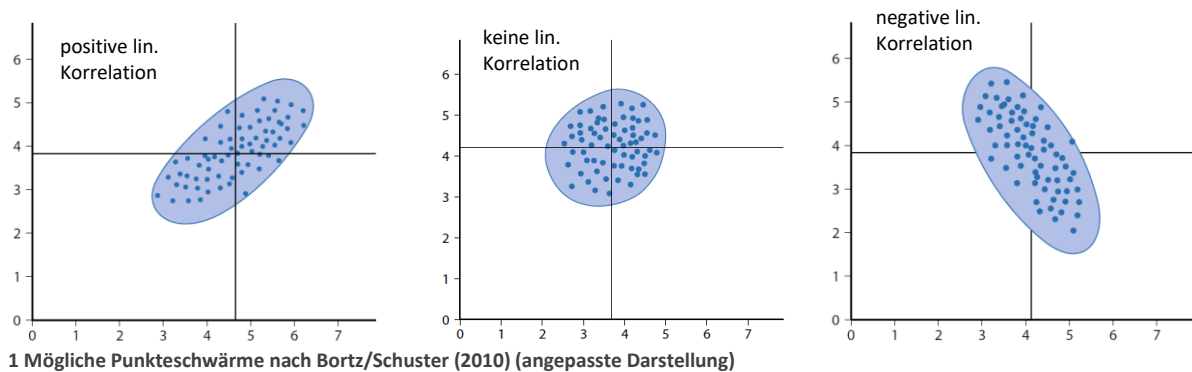
CROSSTABS
/TABLES =sex BY smoke
/format=AVALUE TABLES
/CELLS=COUNT ROW
/COUNT=ROUND CELL.

Analysieren – Zusammenhänge

Werden zwei Variablen in die Analyse miteinbezogen, spricht man von einer bivariaten darüber hinaus von einer multivariaten Analyse. (Streng genommen handelt es sich bei dem Verfahren der Kreuztabellierung auch um ein mindestens bivariates Verfahren.)

Lineare Korrelationen

Lineare Korrelationen stellen ein einfaches Mittel dar, einen Überblick über lineare Zusammenhänge jeweils zweier Variablen zu gewinnen. Vereinfacht gesprochen, zeigt eine lineare Korrelation an, ob der Punkteschwarm zweier Variablen näherungsweise durch eine Gerade (Linie) abbildbar ist.



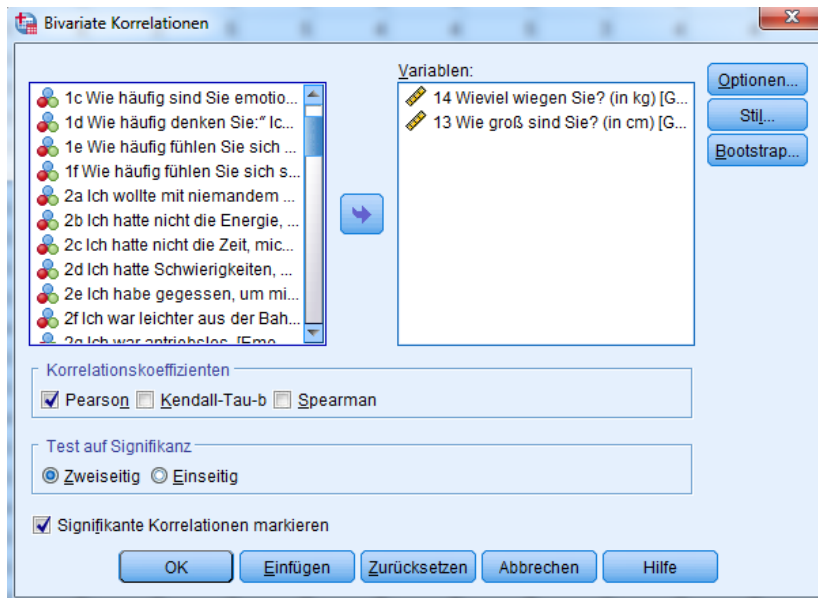
Eine Aussage wie „je mehr..., desto mehr...“ stellt hierbei einen positiven linearen Zusammenhang dar (Abb. 1 links). Ein gegenläufiger Zusammenhang („je mehr..., desto weniger...“) stellt hingegen einen negativen linearen Zusammenhang dar (Abb. 1 rechts). Liegt eine Verteilung wie bei Abbildung 1 Mitte vor, besteht kein linearer Zusammenhang.

Je nach Skalenniveau können verschiedene lineare Korrelationstechniken berechnet werden, welche teilweise in SPSS implementiert sind (siehe Tabelle 2).

	Metrisch	Ordinal	Nominal
Metrisch	Produkt-Moment-Korrel. [PEARSON]	Rangkorrel. [SPEARMAN]	Punktbiseriale Korrel. [PEARSON]
Ordinal		Rangkorrel. [SPEARMAN]	Punktbiseriale Rangkorrel. [NICHT ENTHALTEN]
Nominal			Phi-Koeffizient [KREUZTABELLEN]

2 Korrelationstechniken [in SPSS] (eigene Darstellung)

Das Ergebnis einer linearen Korrelationsanalyse ist der Korrelationskoeffizient r . Der zwischen -1 und +1 variieren kann, wobei das Vorzeichen steht für einen positiven bzw. negativen Zusammenhang steht, die Zahl für dessen Stärke. Diese kann nach Faustregel interpretiert werden. Ab $\pm 0,1$ spricht man von einem kleinen, ab $\pm 0,3$ von einem mittleren und ab $\pm 0,5$ von einem großen Effekt (Cohen 1992). $r = 0$ bedeutet somit, dass kein linearer Zusammenhang besteht, was allerdings nicht heißt, dass kein andersgearteter Zusammenhang besteht!



Lineare Korrelationen können über *Analysieren / Korrelationen / Bivariat* aufgerufen werden. Je nach Skalenniveau werden unterschiedliche Berechnungsgrundlagen gewählt.

Bedingt durch die Kausalitätsfrage wird im Allgemeinen auf eine zweiseitige Testung zurückgegriffen.

Unter **Optionen** kann bspw. eine Deskription der Daten angefordert werden.

Die Ausgabe zeigt alle Korrelationen der in die Analyse miteinbezogen Variablen an (bspw. könnten auch drei oder mehr Variablen analysiert werden).

Im Beispiel korrelieren Körpergröße und -gewicht zu $+0,259$, was einem kleinem Effekt entspricht ($r > 0,1$).

Korrelationen

		13 Wie groß sind Sie? (in cm)	14 Wieviel wiegen Sie? (in kg)
13 Wie groß sind Sie? (in cm)	Korrelation nach Pearson	1	,259*
	Signifikanz (2-seitig)		,036
	N	68	66
14 Wieviel wiegen Sie? (in kg)	Korrelation nach Pearson	,259*	1
	Signifikanz (2-seitig)	,036	
	N	66	66

*. Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

Die Körpergröße und das Körpergewicht weisen einen signifikanten linearen Zusammenhang mit schwacher Effektstärke auf ($r = .259, p = .036$). Je größer eine Person ist, desto schwerer ist Sie und vice versa.

Anm.: Da nach Cohen (1992) erst ab 0,3 von einem mittleren Effekt gesprochen wird, wurde hier ein schwacher Effekt berichtet. Je nach Rundungsgenauigkeit wäre hier u.U. auch der Bericht eines moderaten Effekts denkbar.

Aus r kann das Bestimmtheitsmaß r^2 (auch R^2) gebildet werden, welches für den Anteil der erklärten Varianz durch die beiden Variablen steht.

*Korrelation Größe*Gewicht.

CORRELATIONS

/VARIABLES=height weight
/PRINT=TWOTAIL NOSIG
/MISSING=PAIRWISE.

Regressionsanalysen

Mittels Regressionsverfahren können Messwerte einer (oder mehrerer) Variablen aus den Messwerten anderer Variablen vorhergesagt werden. Regressionsverfahren stellen somit im Grunde eine Erweiterung von Korrelationsanalysen dar und dienen der vertiefenden Analyse von Zusammenhängen.

„Regressieren“ steht dabei für ein Zurückbilden der Werte einer abhängigen Variablen (AV, *syn*: Kriterium, Regressand) auf die einer (oder mehrerer) unabhängigen Variablen (UV, *syn*: Prädiktor, Regressor). Man spricht daher auch von einer „Regression von Y auf X“. Regressionsanalysen setzen ein aus der Theorie abgeleitetes Kausalmodell voraus und eignen sich dementsprechend zur Bestimmung kausaler Zusammenhänge.

Lineare Regression (einfach/multipel)

Die lineare Regression geht von einem linearen Zusammenhang aus, sodass die prognostizierten (geschätzten) Werte anhand einer Geraden darstellbar sind. Der Zusammenhang zwischen einer abhängigen Variable Y und einer unabhängigen Variable X lässt sich somit auf Basis der allgemeinen linearen Geradengleichung beschreiben: $y_i = \beta_0 + \beta_1 \cdot x_i$

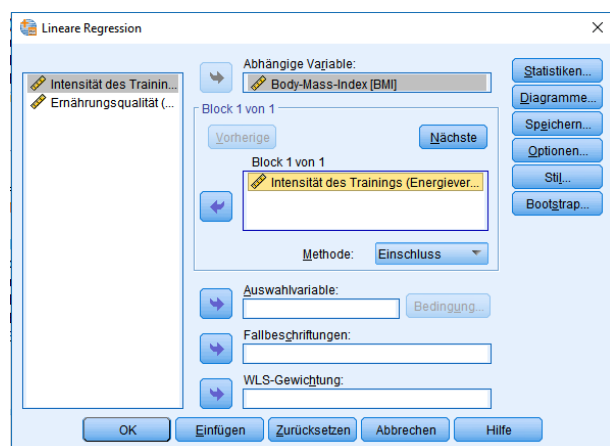
Der Y-Achsenabschnitt β_0 stellt dabei den Schnittpunkt von Regressionsgeraden und y -Achse dar (bei einer einfachen linearen Regression entspricht dies $x_i = 0$). Der Regressionskoeffizient β_1 entspricht der Geradensteigung und stellt die zu erwartende Veränderung von Y dar, wenn X um 1 Einheit steigt/fällt.

Ziel der linearen Regression ist die Vorhersage einer „optimalen“ Gerade, sodass die Differenz zwischen dem wahren Wert y_i und dem durch die Regressionsgleichung vorhergesagten Wert \hat{y}_i möglichst klein ist. Diese Differenz wird auch als Residuum e_i bezeichnet. Die bestmögliche Geradenschätzung ergibt sich nach der Methode der kleinsten Quadrate bei der kleinstmöglichen Quadratsumme der Residuen (QS_{res}).

Lineare Regressionsanalysen können über *Analyse* / *Regression* / *Linear* aufgerufen werden:

Zunächst müssen AV und UV in die jeweils dafür vorgesehenen Felder gebracht werden.

Unter **Statistiken** können u.a. Schätzungen, Konfidenzintervalle und deskriptive Statistiken für die Regressionskoeffizienten angefordert werden. Zusätzlich kann hier die Anpassungsgüte des Modells ausgegeben werden.



*Einfache lineare Regression – BMI, Trainingsintensität.

REGRESSION

```
/DESCRIPTIVES MEAN STDDEV CORR SIG N  
/MISSING LISTWISE  
/STATISTICS COEFF OUTS CI(95) R ANOVA  
/CRITERIA=PIN(.05) POUT(.10)  
/NOORIGIN  
/DEPENDENT BMI  
/METHOD=ENTER Intensity.
```

Deskriptive Statistiken

	Mittelwert	Standardabweichung	N
Body-Mass-Index	22,44	3,504	9
Intensität des Trainings (Energieverbrauch in 100 KJ)	3,33	2,121	9

Korrelationen

		Body-Mass-Index	Intensität des Trainings (Energieverbrauch in 100 KJ)
Korrelation nach Pearson	Body-Mass-Index	1,000	-,813
	Intensität des Trainings (Energieverbrauch in 100 KJ)	-,813	1,000
Sig. (1-seitig)	Body-Mass-Index	.	,004
	Intensität des Trainings (Energieverbrauch in 100 KJ)	,004	.
N	Body-Mass-Index	9	9
	Intensität des Trainings (Energieverbrauch in 100 KJ)	9	9

Die Regressionsausgabe besteht aus insgesamt 6 Tabellen:

Die deskriptiven Statistiken geben einen Überblick über Mittelwert, Standardabweichung und Fallzahl der in das Modell aufgenommenen Variablen.

Anschließend wird die Korrelationsmatrix aller verwendeten Variablen aufgeführt.

Im Falle einer multiplen linearen Regression werden darauffolgend aus dem Modell entfernte und mit aufgenommene Variablen dargestellt (hier nicht enthalten), was v.a. für die Modellauswahl relevant ist.

In der Modellzusammenfassung wird der Anteil der Varianz wiedergegeben, der durch die Regressionsgleichung erklärt (Quadratsumme „Regression“) bzw. eben nicht erklärt werden kann (Quadratsumme „Residuen“). Der Determinationskoeffizient R-Quadrat (R^2) entspricht dem Anteil erklärter Varianz zur insgesamt zu erklärenden Varianz.

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,813 ^a	,661	,612	2,182

a. Einflußvariablen : (Konstante), Intensität des Trainings
(Energieverbrauch in 100 KJ)

Anm.: Nach Faustregel spricht man von einem besseren Modell, je mehr Varianz erklärt wird, bzw. je größer R^2 ist. Da R^2 zwischen 0 und 1 variieren kann, spricht man bei R^2 auch vom prozentualen Anteil der Varianz in Y, der auf die Varianz in X zurückgeführt werden kann. Ein Bestimmtheitsmaß von $R^2 = 0$ steht somit auch dafür, dass AV und UV unkorreliert sind bzw. $R^2 = 1$ dass AV und UV perfekt miteinander korrelieren.

Die folgende Ausgabe enthält die Ergebnisse einer Varianzanalyse, die auf dem Verhältnis der durch die Regression erklärten Varianz und der nicht erklärten Varianz (den sog. „Residuen“) des untersuchten Merkmals aufbaut.

ANOVA ^a

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	64,892	1	64,892	13,629	,008 ^b
	Nicht standardisierte Residuen	33,330	7	4,761		
	Gesamt	98,222	8			

a. Abhängige Variable: Body-Mass-Index

b. Einflußvariablen : (Konstante), Intensität des Trainings (Energieverbrauch in 100 KJ)

Mittels F-Test wird die Nullhypothese geprüft, dass zwischen der AV und der UV kein linearer Zusammenhang besteht. Bei einem signifikanten Testergebnis (wie hier der Fall) kann davon ausgegangen werden, dass die UV einen Anteil der Variation in der AV erklären kann.

Die Ausgabe der Koeffizienten zeigt die Regressionskoeffizienten b_0 („Konstante“) und b_1 („Intensität des Trainings“). Der Signifikanztest gibt Aufschluss darüber ob der Mehrwert einer Variable signifikant von Null abweicht, also „nützlich“ für das Modell ist. Eine signifikante Konstante bedeutet, dass der Y-Achsenabschnitt nicht 0 beträgt und damit die Regressionsgerade nicht durch den Ursprung geht. Ein signifikanter Koeffizient der UV bedeutet, dass die Variable einen signifikanten Einfluss auf die AV aufweist.

Koeffizienten ^a

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Sig.	95,0% Konfidenzintervalle für B	
		Regressionskoeffizient B	Standardfehler	Beta			Untergrenze	Obergrenze
1	(Konstante)	26,920	1,414		19,042	,000	23,577	30,263
	Intensität des Trainings (Energieverbrauch in 100 KJ)	-1,343	,364	-,813	-3,692	,008	-2,203	-,483

a. Abhängige Variable: Body-Mass-Index

Die Effektstärke einer (multiplen) linearen Regression kann mit f angegeben werden (mit $f = \sqrt{\frac{R^2}{1-R^2}}$).

Nach Cohen wird f ab 0,1, 0,25 und 0,40 als schwach, mittel und stark interpretiert.

Eine mögliche Zusammenfassung für ein [multiples] Regressionsmodell könnte folgendermaßen lauten:

Die [multiple] Regressionsanalyse zeigt, dass [sowohl] die Intensität eines standardisierten, täglichen Trainingsprogramms [als auch die Ernährungsqualität] einen Einfluss auf das Körpergewicht (BMI) von Frauen hat[/haben] ($F(1, 7) = 13,629$, $p = 0.008$, $n = 9$).

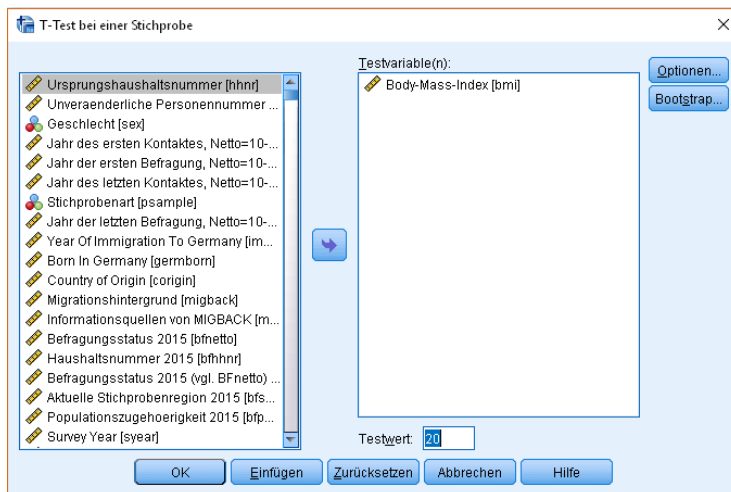
Steigt die Trainingsintensität um eine Einheit (hier: 100KJ), so sinkt der BMI um durchschnittlich 0,813 kg/m². [Steigt die Ernährungsqualität um eine Einheit (hier: Expertenrating), so sinkt der BMI um durchschnittlich 1,03 kg/m².]

Rund 61% der Streuung im Körpergewicht kann durch die [zwei] unabhängige[n] Variable[n] erklärt werden ($R^2 = 0.612$). Das entspricht nach Cohen (1992) einem starken Effekt ($f^2 = 2.34$).

Analysieren – Unterschiede

1-Stichproben-t-Test

Die einfachste Version des t-Tests, prüft den Mittelwert einer Stichprobe gegen einen vorher festgelegten Test(mittel)wert. In SPSS ist dieser Test unter *Analysieren / Mittelwerte vergleichen / T-Test bei einer Stichprobe* zu finden.



Im linken Abschnitt des Fensters sind wie gewohnt alle im Datensatz vorhandenen Variablen enthalten, im rechten Abschnitt, alle Variablen welche in die Analyse miteinbezogen werden sollen. Zusätzlich kann ein *Testwert* eingetragen werden, gegen den es zu prüfen gilt.

Unter **Optionen** kann das Signifikanzniveau spezifiziert werden (standardmäßig mit 5% hinterlegt).

Der Test führt automatisch eine zweiseitige Testung durch – sollte eine einseitige Testung von Interesse sein, muss der ausgegebene Signifikanzwert manuell halbiert werden!

*t-Test BMI vs. 20kg/m².

T-TEST

/TESTVAL = 20
/MISSING = ANALYSIS
/VARIABLES = bmi
/CRITERIA = CI(.95).

Die Ausgabe besteht aus zwei Teilen – einerseits der Übersicht der zu testenden Variable und andererseits der Teststatistik.

T-Test

Statistik bei einer Stichprobe

	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Body-Mass-Index	12094	25,8673	5,82745	,05299

Test bei einer Stichprobe

Testwert = 20

	T	df	Sig. (2-seitig)	Mittlere Differenz	95% Konfidenzintervall der Differenz	
					Untere	Obere
Body-Mass-Index	110,724	12093	,000	5,86728	5,7634	5,9711

Neben dem empirischen t-Wert, den Freiheitsgraden und der Test-Signifikanz können der Ausgabe auch die durchschnittliche Differenz der Werte und das Konfidenzintervall entnommen werden.

Eine Ergebnisdarstellung könnte hier bspw. so aussehen:

Der Mittelwert des BMIs der untersuchten Teilnehmer ($\bar{x} = 25,87$; $s = 5,83$) weist signifikante Unterschiede zum Testwert 20 auf ($t(12093) = 110,724, p < .001$).

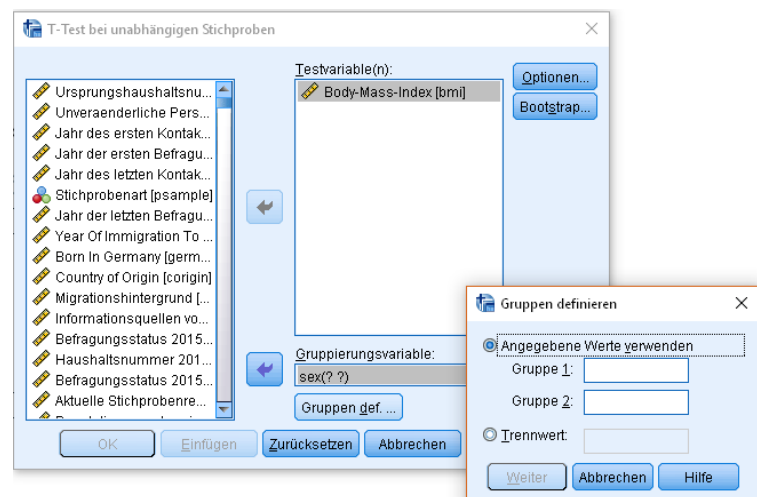
t-Test bei unabhängigen Stichproben

Der 2-Stichproben t-Test für unabhängige Stichproben kann über *Analysieren / Mittelwerte vergleichen / T-Test für unabhängige Stichproben* aufgerufen werden.

Der Aufbau unterscheidet sich kaum von der 1-Stichproben Variante, allerdings wird der Testwert hier durch eine Gruppierungsvariable ersetzt. Da der t-Test nur mit zwei (Teil)Stichproben durchgeführt werden kann, muss die Gruppierungsvariable bzw. deren Ausprägungen näher spezifiziert werden.

In Gruppe 1 bzw. Gruppe 2 werden die Codierungen der zu untersuchenden Gruppen eingetragen (hier: 1 und 2 für männlich und weiblich). Eine andere Möglichkeit liegt in der Angabe eines Trennwerts (alle TN kleiner als und größer gleich...).

Unter **Optionen** kann abermals das Signifikanzniveau angepasst werden (Standard 5%).



*t-Test BMI gruppiert nach Geschlecht.

T-TEST GROUPS=sex(1 2)
 /MISSING = ANALYSIS
 /VARIABLES = bmi
 /CRITERIA = CI(.95).

T-Test

Gruppenstatistiken					
	Geschlecht	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Body-Mass-Index	[1] männlich	5561	26,6593	4,95680	,06647
	[2] weiblich	6533	25,1931	6,40060	,07919

Test bei unabhängigen Stichproben									
		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit					
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz
Body-Mass-Index	Varianzen sind gleich	174,159	,000	13,900	12092	,000	1,46625	,10549	Untere: 1,25947 Obere: 1,67302
	Varianzen sind nicht gleich			14,182	11987,787	,000	1,46625	,10339	Untere: 1,26359 Obere: 1,66891

Neben den bereits bekannten Teilen der Ausgabe fügt SPSS automatisch einen Test auf Varianzhomogenität hinzu (→ Voraussetzungsprüfung).

In diesem Fall besteht keine Varianzhomogenität! Der Levene-Test zeigt eine signifikante Abweichung zwischen den Varianzen auf. In diesem Fall kann der t-Test nicht „normal“ interpretiert werden und es muss eine Anpassung berechnet werden – diese liefert SPSS allerdings automatisch in der zweiten Zeile. Diese Anpassung nach Welch (auch Welch-Test genannt) ist robuster gegen die Voraussetzungsverletzung der Varianzhomogenität und führt in diesem Fall zu einem „sichereren“ Ergebnis.

Das Ergebnis könnte wie folgt formuliert werden:

In der Gruppe der männlichen Teilnehmer wurden höhere BMI-Werte ($\bar{x} = 26,66 \pm 4,96$) beobachtet als in der Gruppe der weiblichen Teilnehmer ($\bar{x} = 25,19 \pm 6,40$). Dieser Unterschied konnte als signifikant nachgewiesen werden ($t(11987,78) = 14,18, p < .001$).

Zu berichten, dass es sich um eine Anpassung nach Welch bzw. den Welch-Test handelt, ist nicht unbedingt notwendig – ggf. wird dies auch an den dezimalen Freiheitsgraden sichtbar.

Hinsichtlich Fragstellungen mit Gruppenunterschieden, ist u.U. auch die durchschnittliche Differenz beider Gruppen berichtenswert. Diese findet sich unter „Mittlere Differenz“ und sollte mit einem Maß der Variabilität angegeben werden – bspw. dem Konfidenzintervall:

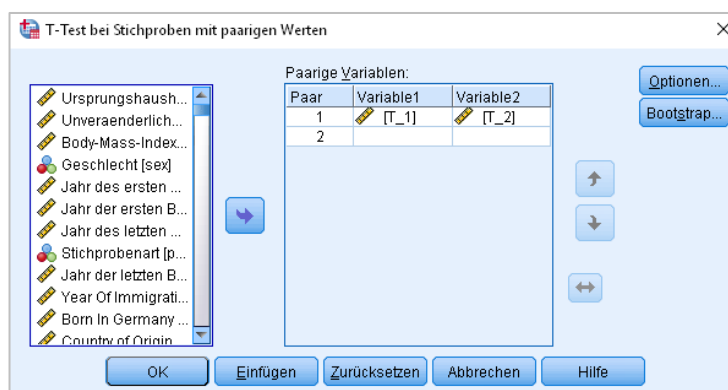
Der BMI liegt in der Gruppe der Männer (entgegen dem der Frauen) durchschnittlich 1,47 Punkte höher (95% – KI [1,26, 1,67]).

Eine Effektstärke kann bspw. mit $r = \sqrt{\frac{t^2}{t^2 + df}}$ angegeben werden – also hier:

Der untersuchte Unterschied weist einen schwachen Effekt auf ($r = 0,13$).

t-Test bei abhängigen Stichproben

Der t-Test bei abhängigen bzw. verbundenen Stichproben verhält sich entsprechend und wird über *Analysieren / Mittelwerte vergleichen / T-Test bei verbundenen Stichproben* aufgerufen



Da es sich um paarige Vergleiche handelt, erfordert die Analyse auch die Eingabe von zwei Variablen (bspw. Test_t1 und Test_t2 oder Test_Teilnehmer und Test_Partner). Beide Variablen werden in das entsprechende Feld überführt; das Signifikanzniveau kann über **Optionen** angepasst werden.

29

*t-Test Zeitpunkt1 vs. Zeitpunkt2.

T-TEST PAIRS=T_1 WITH T_2 (PAIRED)
/CRITERIA = CI(.95)
/MISSING = ANALYSIS

Eine mögliche Ergebnisdarstellung könnte folgendermaßen lauten:

Die Testwerte zum Zeitpunkt 2 waren nach der Intervention signifikant höher als zuvor,
 $t(df) = \dots, p = \dots$.

Bzw.: Die Intervention hatte keinen signifikanten Einfluss auf die Testwerte ($t(df) = \dots, p = \dots$).

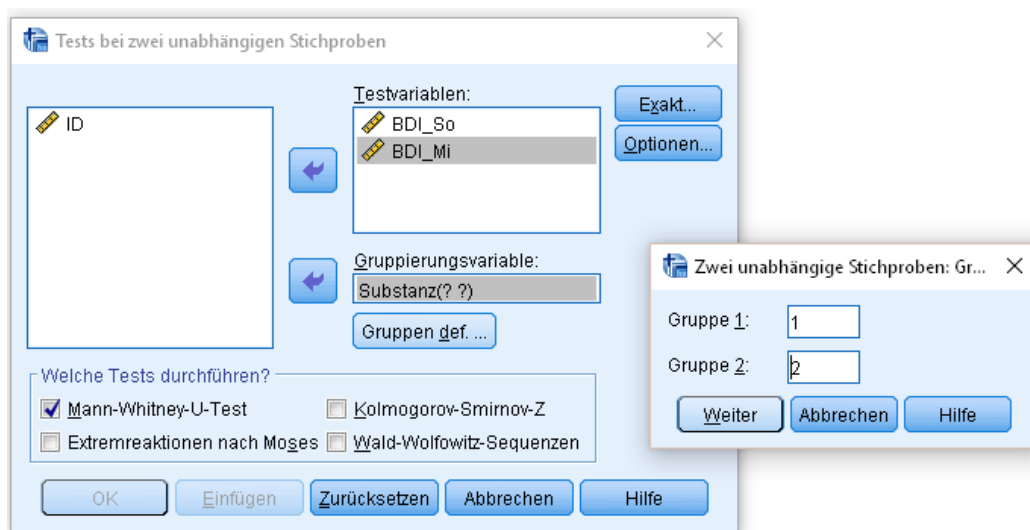
Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummentest

Der Mann-Whitney-U-Test ist das verteilungsfreie Äquivalent des t-Tests für zwei unabh. Stichproben.

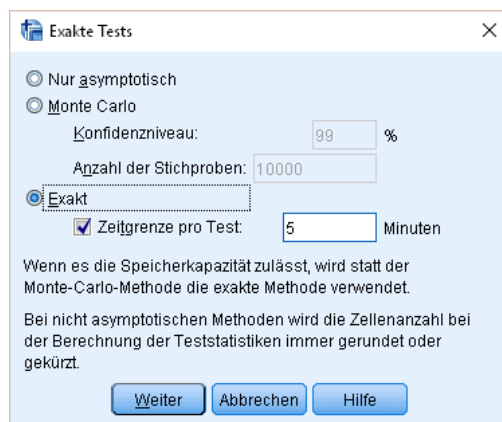
Auch wenn das Verfahren weniger Voraussetzungen hat, als der entsprechende t-Test, sollten dennoch vorab zur Abklärung dessen Voraussetzungen geprüft werden (Normalverteilung, Varianzhomogenität).

Anm.: Auch, wenn Varianzhomogenität nicht per se separat aufgerufen werden kann, ist der Levene-Test ein möglicher Teil der explorativen Datenanalyse, aufrufbar über Diagramme im entsprechenden Menü (Analysieren/Deskriptive Statistik/Explorative Datenanalyse) über Diagramme und nicht-transformierte Streubreite vs. Mittleres Niveau.

In SPSS ist der Mann-Whitney-U-Test unter *Analysieren/Nicht parametrische Tests/Alte Dialogfelder/ 2 unabhängige Stichproben* zu finden.



Die Testvariablen beinhalten die zu prüfenden UVs; innerhalb der Gruppierungsvariable werden die Kodierungen der zu unterscheidenden Gruppen hinterlegt. Unter **Optionen** kann die Anzeige der deskriptiven Statistik angewählt werden, welche jedoch nicht gruppenspezifisch angezeigt wird und hier somit wenig relevant ist.



Unter **Exakt** kann anschließend die Berechnung der Signifikanz spezifiziert werden. Die Standardanwahl ist „nur asymptotisch“ welche v.a. bei großen Stichprobengrößen genutzt werden sollte.

Im hier behandelten Beispiel sollte aufgrund der Fallzahl (n=20) eher auf die exakte Methode zurückgegriffen werden. Eine exakte Berechnung verlangt mehr Arbeitsspeicher des Computers und somit mehr Zeit, ist aber v.a. bei kleinen Stichproben und „schlecht“ verteilten Daten (Ausreißer etc.) ist die exakte Methode vorzuziehen.

Die Monte Carlo Methode ist etwas weniger anspruchsvoll und kann für große Stichproben herangezogen werden.

*Mann-Whitney-U-Test: Unterschiede im BDI nach Substanzkonsum, Sonntag und Mittwoch.

NPARTESTS

/M-W = BDI_So BDI_Mi BY Substanz(1 2)

/MISSING ANALYSIS

/METHOD=EXACT TIMER(5).

Ränge				
	Substanz	N	Mittlerer Rang	Rangsumme
BDI_So	Ecstasy	10	11,95	119,50
	Alkohol	10	9,05	90,50
	Gesamt	20		
BDI_Mi	Ecstasy	10	15,10	151,00
	Alkohol	10	5,90	59,00
	Gesamt	20		

Statistik für Test ^a		
	BDI_So	BDI_Mi
Mann-Whitney-U	35,500	4,000
Wilcoxon-W	90,500	59,000
Z	-1,105	-3,484
Asymptotische Signifikanz (2-seitig)	,269	,000
Exakte Signifikanz [2*(1-seitige Sig.)]	,280 ^b	,000 ^b
Exakte Signifikanz (2-seitig)	,288	,000
Exakte Signifikanz (1-seitig)	,144	,000
Punkt-Wahrscheinlichkeit	,013	,000

a. Gruppenvariable: Substanz

b. Nicht für Bindungen korrigiert.

Die Ausgabe des Tests ist vergleichsweise übersichtlich. Die erste Tabelle zeigt (v.a.) die Rangsummen der untersuchten Variablen in beiden Gruppen an. Kleine Ränge stehen für kleine Werte, also kann hieraus abgeleitet werden, dass (hier) die Gruppe der Alkohol-Konsumenten sowohl am Sonntag als auch am Mittwoch kleinere BDI Werte aufgewiesen hat.

Die zweite Tabelle zeigt die Teststatistik(en). Es werden sowohl U als auch W-Statistik angezeigt (welche wie äquivalent sind). Anschließend werden z-Wert und die gewählte(n) Signifikanz(en) angezeigt. Wenn nicht die exakte Testung angewählt wurde, fällt diese Tabelle kleiner aus und nur der Signifikanzwert der 2-seitigen Testung wird angegeben.

Anm.: Bei einseitig formulierter Hypothese wird auch hier der p-Wert manuell halbiert!

Mithilfe der Ausgabe kann ebenfalls manuell die Effektstärke r berechnet werden – mithilfe der angegebenen z-Werte und der Stichprobengröße ($r = \left| \frac{z}{\sqrt{n}} \right|$). Hier also $r_{\text{Sonntag}} = 0,25$ und $r_{\text{Mittwoch}} = 0,78$.

Die Ergebnisse sprechen dafür, dass am Tag nach der Substanzeinnahme (Sonntag) kein Unterschied im BDI-Wert zwischen den Gruppen Ecstasy und Alkohol vorliegt – allerdings in der Messung am Mittwoch!

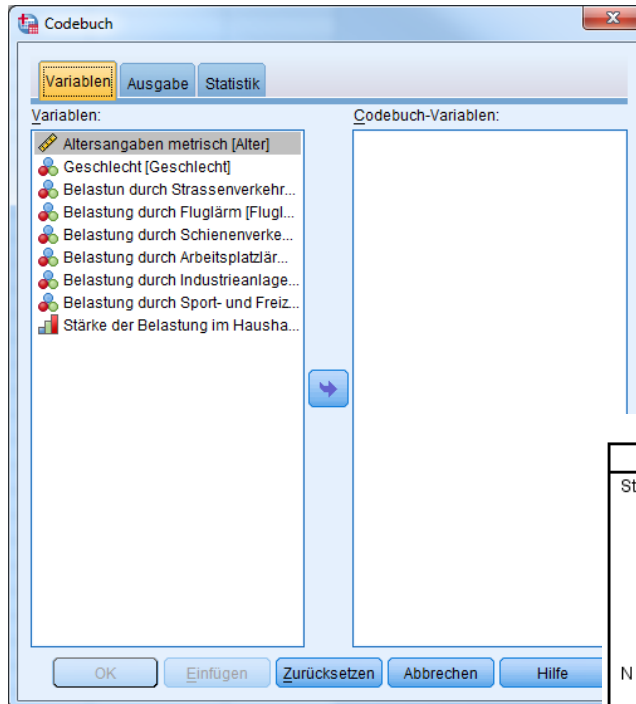
Ein Ergebnis könnte folgendermaßen dargestellt werden:

Es zeigen sich keine signifikanten Unterschiede in der Schwere der depressiven Symptomatik von Ecstasy-Konsumenten (Median=17,5) und Alkoholkonsumenten (Median=16,0) am Tag nach der Einnahme der jeweiligen Substanz ($U = 35,5$; $z = -1,11$; ns ; $r = 0,25$).

Zur Wochenmitte weisen Ecstasy-Konsumenten allerdings signifikant höhere BDI-Werte auf als Alkoholkonsumenten (Median 33,5 zu 7,5; $U = 4,0$; $z = -3,48$; $p < 0,05$; $r = ,78$).

Dokumentation

Codebuch



Um einen Datensatz dokumentieren zu können bzw. das Kodierungsschema transparent zu machen kann unter *Analysieren/Berichte* ein *Codebuch* ausgegeben werden.

Sowohl die Variablen an sich, deren statistische Werte sowie die Ausgabe deren Information kann individuell gewählt und angepasst werden.

Standardmäßig sähe die Ausgabe bspw. so aus:

Alter		Wert
Standardattribute	Position	1
	Beschriftung	Altersangaben metrisch
	Typ	Numerisch
	Format	F11
	Messung	Skala
	Rolle	Eingabe
N	Gültig	1674
	Fehlend	4
Lagemaße und Streuung	Mittelwert	51,69
	Standardabweichung	15,791
	25. Perzentil	40,00
	50. Perzentil	54,00
	75. Perzentil	63,00

SPSS-Webbericht

Vor allem wenn es darum geht, andere Personen an den Ergebnissen und Berechnungen teilhaben zu lassen – also die Ausgabedatei zu publizieren – wird man schnell mit dem Problem der „SPSS-Verfügbarkeit“ konfrontiert werden. U.a. aufgrund des Preises des Programms, kann nicht davon ausgegangen werden, dass die Praxis Zugang zu SPSS hat und somit eine SPSS-Ausgabedatei öffnen kann.

Eine weitere Möglichkeit die Ausgabe zu publizieren, besteht im *Speichern unter* als *SPSS-Webbericht*. Hierbei wird der Inhalt der Ausgabe in einer html-Datei abgelegt, welche von einem Internetbrowser geöffnet werden kann (was standardmäßig auf allen Betriebssystemen vorhanden ist). Die Ausgabe ist hierbei sogar interaktiv – ähnlich der „normalen“ Ausgabe wird die Gliederung der einzelnen Berechnungen übernommen und kann mittels Klick zur schnelleren Navigation verwendet werden.

Zusatzmaterial

Syntax: Operatoren

Neben den jeweiligen Befehlswörtern und Variablennamen können auch einzelne Operatoren zum Tragen kommen, welche bspw. eine bestimmte Bedingung für die Analyse voraussetzen oder die Analyse für diese Bedingung ausschließen.

Relationale Operatoren	Abkürzung/Zeichen			Definition
		EQ	=	Ist gleich
	NE	\neq	\neq <>	Ist ungleich
		LT	<	Kleiner als (less than)
		LE	<=	Kleiner gleich (less than or equal to)
		GT	>	Größer als (greater than)
Logische Operatoren		GE	>=	Größer gleich (greater than or equal to)
	Abkürzung/Zeichen			Definition
		AND	&	Beide Bedingungen müssen erfüllt sein
		OR		Eine der Bedingungen muss erfüllt sein
		NOT		Alle Bedingungen außer...

Effektgrößen

Je nach Verfahren können unterschiedliche Effektgrößen/-stärken berechnet werden welche je nach Verfahren oder Wissenschaftszweig unterschiedlich gängig und/oder berichtenswert sind. Zudem werden Effektstärken im Allgemeinen in drei Größen eingeteilt bzw. dementsprechend interpretiert (klein, mittel und groß). Die meisten Effektstärken können ineinander überführt werden.

Effektgröße/-stärke	Verwendung (u.a.)	Effektinterpretation		
		klein	mittel	groß
r / ρ (Rho) / ϕ (Phi)	Korrelation	0,1	0,3	0,5
η^2 / η_p^2 (Eta)	(M)ANOVA	0,01	0,06	0,14
ω^2 (Omega)	ANOVA	0,01	0,06	0,14
Cohen's f	An(c)ova/Regression	0,1	0,25	0,4
Cohen's f^2	Multiple Regression	0,02	0,15	0,35
Cohen's d	t-Test	0,2	0,5	0,8
Cohen's ω	Chi ² -Test	0,1	0,3	0,5
Cohen's Δ	ANOVA	0,25	0,75	1,25
Odds Ratio	4-Felder-Tafel	1,5	3,5	9,0

Literatur

Primärliteratur

Backhaus, Klaus; Erichson, Bernd; Plinke, Wulff; Weiber, Rolf (2016): Multivariate Analysemethoden. Eine anwendungsorientierte Einführung. 14., überarbeitete und aktualisierte Auflage. Berlin, Heidelberg: Springer Gabler (Lehrbuch).

Bortz, Jürgen; Schuster, Christof (2010): Statistik für Human- und Sozialwissenschaftler. 7., vollst. überarb. u. erw. Aufl. Berlin: Springer (Springer-Lehrbuch).

Eid, Michael; Gollwitzer, Mario; Schmitt, Manfred; Hohmann, Louisa; Lawes, Maria; Wang, Tianqi (2016): Formelsammlung Statistik und Forschungsmethoden. Weinheim, Basel: Beltz.

Field, Andy (2013): Discovering statistics using IBM SPSS statistics. And sex and drugs and rock 'n' roll. 4th edition. Los Angeles, London, New Delhi, Singapore, Washington DC: Sage (Mobile-Study).

Weiterführende Literatur

Backhaus, Klaus; Erichson, Bernd; Weiber, Rolf (2015): Fortgeschrittene Multivariate Analysemethoden. Eine anwendungsorientierte Einführung. 3., überarb. u. aktualisierte Aufl. 2015.

Bortz, Jürgen; Lienert, Gustav A.; Boehnke, Klaus (2008): Verteilungsfreie Methoden in der Biostatistik. Mit 247 Tabellen. 3., korrigierte Aufl. Heidelberg: Springer (Springer-Lehrbuch Bachelor, Master).

Benesch, Thomas (2013): Schlüsselkonzepte zur Statistik. Die wichtigsten Methoden, Verteilungen, Tests anschaulich erklärt. 1. Aufl. Heidelberg: Spektrum Akademischer Verlag (SpringerLink : Bücher).

Döring, Nicola; Bortz, Jürgen (2016): Forschungsmethoden und Evaluation. Für Human- und Sozialwissenschaften; 5., überarb. Aufl., Heidelberg: Springer-Medizin-Verl. (Springer-Lehrbuch Bachelor, Master).

Eid, Michael; Gollwitzer, Mario; Schmitt, Manfred (2011): Statistik und Forschungsmethoden. Lehrbuch. 3. Aufl., deutsche Erstausg. Weinheim: Beltz.

Häder, Michael (2015): Empirische Sozialforschung. Eine Einführung. 3. Aufl. 2015. Wiesbaden: Springer Fachmedien Wiesbaden.

Hilgers, Ralf-Dieter; Bauer, Peter; Scheiber, Viktor (2007): Einführung in die medizinische Statistik. 2., verb. und überarb. Aufl. Berlin [u.a.]: Springer (Statistik und ihre Anwendungen).

Rasch, Björn (2014a): Quantitative Methoden 1. Einführung in die Statistik für Psychologen und Sozialwissenschaftler. 4., überarb. Aufl. Wiesbaden: Springer (Springer-Lehrbuch).

Rasch, Björn (2014b): Quantitative Methoden 2. Einführung in die Statistik für Psychologen und Sozialwissenschaftler. 4., überarb. Aufl. Berlin: Springer (Springer-Lehrbuch).

Schendera, Christian (2008a): Clusteranalyse mit SPSS. München: Oldenbourg.

Schendera, Christian (2008b): Regressionsanalyse mit SPSS. München: Oldenbourg.

Weiber, Rolf; Mühlhaus, Daniel (2014): Strukturgleichungsmodellierung. Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS. 2., erw. und korr. Aufl. 2014. Berlin, Heidelberg: Springer Gabler (Springer-Lehrbuch).

Serie zur Bewertung wissenschaftlicher Publikationen

Editorial: Du Prel, Jean-Baptist; Röhrig, Bernd; Blettner, Maria (2009b): Biometrische Methoden in der medizinischen Forschung. Editorial zum Start der Serie zur Bewertung wissenschaftlicher Publikationen. In: Deutsches Ärzteblatt international 106 (7), S. 99.

Teil 1: Du Prel, Jean-Baptist; Röhrig, Bernd; Blettner, Maria (2009c): Kritisches Lesen wissenschaftlicher Artikel. Teil 1 der Serie zur Bewertung wissenschaftlicher Publikationen. In: Deutsches Ärzteblatt international 106 (7), S. 100–105. DOI: 10.3238/arztebl.2009.0100.

Teil 2: Röhrig, Bernd; Du Prel, Jean-Baptist; Blettner, Maria (2009a): Studiendesign in der medizinischen Forschung. Teil 2 der Serie zur Bewertung wissenschaftlicher Publikationen. In: Deutsches Ärzteblatt international 106 (11), S. 184–189. DOI: 10.3238/arztebl.2009.0184.

Teil 3: Röhrig, Bernd; Du Prel, Jean-Baptist; Wachtlin, Daniel; Blettner, Maria (2009b): Studientypen in der medizinischen Forschung. Teil 3 der Serie zur Bewertung wissenschaftlicher Publikationen. In: Deutsches Ärzteblatt international 106 (15), S. 262–268. DOI: 10.3238/arztebl.2009.0262.

Teil 4: Du Prel, Jean-Baptist; Hommel, Gerhard; Röhrig, Bernd; Blettner, Maria (2009a): Konfidenzintervall oder p-Wert? Teil 4 der Serie zur Bewertung wissenschaftlicher Publikationen. In: Deutsches Ärzteblatt international 106 (19), S. 335–339. DOI: 10.3238/arztebl.2009.0335.

Teil 5: Bausch, Wilfried (2009): Anforderungen und Bewertung der Ergebnisse von Laboruntersuchungen. Teil 5 der Serie zur Bewertung wissenschaftlicher Publikationen. In: Deutsches Ärzteblatt international 106 (24), S. 403–406. DOI: 10.3238/arztebl.2009.0403.

Teil 6: Rössing, Meike; Blettner, Maria; Klug, Stefanie J. (2009): Systematische Übersichtsarbeiten und Metaanalysen. Teil 6 der Serie zur Bewertung wissenschaftlicher Publikationen. In: Deutsches Ärzteblatt international 106 (27), S. 456–463. DOI: 10.3238/arztebl.2009.0456.

Teil 7: Spriestersbach, Albert; Röhrig, Bernd; Du Prel, Jean-Baptist; Gerhold-Ay, Aslihan; Blettner, Maria (2009): Deskriptive Statistik. Teil 7 der Serie zur Bewertung wissenschaftlicher Publikationen. Angabe statistischer Maßzahlen und ihre Darstellung in Tabellen und Grafiken. In: Deutsches Ärzteblatt international 106 (36), S. 578–583. DOI: 10.3238/arztebl.2009.0578.

Teil 8: Hammer, Gaël P.; Du Prel, Jean-Baptist; Blettner, Maria (2009): Vermeidung verzerrter Ergebnisse in Beobachtungsstudien. Teil 8 der Serie zur Bewertung wissenschaftlicher Publikationen. In: Deutsches Ärzteblatt international 106 (41), S. 664–668. DOI: 10.3238/arztebl.2009.0664.

Teil 9: Sauerbrei, Wilhelm; Blettner, Maria (2009): Interpretation der Ergebnisse von 2x2-Tafeln. Teil 9 der Serie zur Bewertung wissenschaftlicher Publikationen. In: Deutsches Ärzteblatt international 106 (48), S. 795–800. DOI: 10.3238/arztebl.2009.0795.

Teil 10: Victor, Anja; Elsässer, Amelie; Hommel, Gerhard; Blettner, Maria (2010): Wie bewertet man die p-Wert-Flut? Teil 10 der Serie zur Bewertung wissenschaftlicher Publikationen. Hinweise zum Umgang mit dem multiplen Testen. In: Deutsches Ärzteblatt international 107 (4), S. 50–56.

Teil 11: Rössing, Meike; Blettner, Maria; Klug, Stefanie J. (2010): Auswertung epidemiologischer Studien. Teil 11 der Serie zur Bewertung wissenschaftlicher Publikationen. In: Deutsches Ärzteblatt international 107 (11), S. 187–192. DOI: 10.3238/arztebl.2010.0187.

Teil 12: Du Prel, Jean-Baptist; Röhrig, Bernd; Hommel, Gerhard; Blettner, Maria (2010): Auswahl statistischer Testverfahren. Teil 12 der Serie zur Bewertung wissenschaftlicher Publikationen. In: Deutsches Ärzteblatt international 107 (19), S. 343–348. DOI: 10.3238/arztebl.2010.0343.

Teil 13: Röhrig, Bernd; Du Prel, Jean-Baptist; Wachtlin, Daniel; Kwiczen, Robert; Blettner, Maria (2010): Fallzahlplanung in klinischen Studien. Teil 13 der Serie zur Bewertung wissenschaftlicher Publikationen. In: Deutsches Ärzteblatt international 107 (31-32), S. 552–556. DOI: 10.3238/arztebl.2010.0552.

Teil 14: Schneider, Astrid; Hommel, Gerhard; Blettner, Maria (2010): Lineare Regressionsanalyse. Teil 14 der Serie zur Bewertung wissenschaftlicher Publikationen. In: Deutsches Ärzteblatt international 107 (44), S. 776–782. DOI: 10.3238/arztebl.2010.0776.

Teil 15: Zwiener, Isabella; Blettner, Maria; Hommel, Gerhard (2011): Überlebenszeitanalyse. Teil 15 der Serie zur Bewertung wissenschaftlicher Publikationen. In: Deutsches Ärzteblatt international 108 (10), S. 163–169.

Teil 16: Kwiecien, Robert; Kopp-Schneider, Annette; Blettner, Maria (2011): Konkordanzanalyse. Teil 16 der Serie zur Bewertung wissenschaftlicher Publikationen. In: Deutsches Ärzteblatt international 108 (30), S. 515–521. DOI: 10.3238/arztebl.2011.0515.

Teil 17: Kabisch, Maria; Ruckes, Christian; Seibert-Grafe, Monika; Blettner, Maria (2011): Randomisierte kontrollierte Studien. Teil 17 der Serie zur Bewertung wissenschaftlicher Publikationen. In: Deutsches Ärzteblatt international 108 (39), S. 663–668. DOI: 10.3238/arztebl.2011.0663.

Teil 18: Wellek, Stefan; Blettner, Maria (2012b): Vom richtigen Umgang mit dem Crossover-Design in klinischen Studien. Teil 18 der Serie zur Bewertung wissenschaftlicher Publikationen. In: Deutsches Ärzteblatt international 109 (15), S. 276–281. DOI: 10.3238/arztebl.2012.0276.

Teil 19: Spix, Claudia; Blettner, Maria (2012): Screening. Teil 19 der Serie zur Bewertung wissenschaftlicher Publikationen. In: Deutsches Ärzteblatt international 109 (21), S. 385–390. DOI: 10.3238/arztebl.2012.0385.

Teil 20: Wellek, Stefan; Blettner, Maria (2012a): Klinische Studien zum Nachweis von Äquivalenz oder Nichtunterlegenheit. Teil 20 der Serie zur Bewertung wissenschaftlicher Publikationen. In: Deutsches Ärzteblatt international 109 (41), S. 674–679. DOI: 10.3238/arztebl.2012.0674.

Teil 21: Binder, Harald; Blettner, Maria (2015): Big Data in der wissenschaftlichen Medizin – eine biostatistische Perspektive. Teil 21 der Serie zur Bewertung wissenschaftlicher Publikationen. In: Deutsches Ärzteblatt international 112 (9), S. 137–142. DOI: 10.3238/arztebl.2015.0137.

Teil 22: Kiefer, Corinna; Sturtz, Sibylle; Bender, Ralf (2015): Indirect Comparisons and Network Meta-Analyses. Teil 22 der Serie zur Bewertung wissenschaftlicher Publikationen. In: Deutsches Ärzteblatt international 112 (47), S. 803–808. DOI: 10.3238/arztebl.2015.0803.

Teil 23: Kuss, Oliver; Blettner, Maria; Borgermann, Jochem (2016): Propensity Score: an Alternative Method of Analyzing Treatment Effects. Teil 23 der Serie zur Bewertung wissenschaftlicher Publikationen. In: Deutsches Ärzteblatt international 113 (35-36), S. 597–603. DOI: 10.3238/arztebl.2016.0597.

Teil 24: Lange, Stefan; Sauerland, Stefan; Lauterberg, Jörg; Windeler, Jürgen (2017): The Range and Scientific Value of Randomized Trials. Teil 24 der Serie zur Bewertung wissenschaftlicher Publikationen. In: Deutsches Ärzteblatt international 114 (38), S. 635–640. DOI: 10.3238/arztebl.2017.0635.

