

Datenerfassung und Datenanalyse mit SPSS

Statistical Package for the Social Sciences

13.6.2012

Vera Heuser

Inhalt

- Aufrufen des Programms
- Die Variablenansicht
- Die Datenansicht
- Weitere SPSS-Fenster
- Univariate und Bivariate Häufigkeitsanalyse
- Korrelationen mit SPSS
- Literaturempfehlung

Aufrufen des Programms

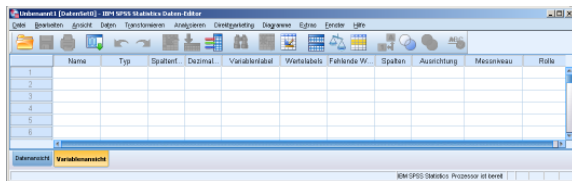
Nach Installation gemäß den Voreinstellungen:

Programme -> IBM SPSS Statistics -> IBM Statistics 19

Daten-Editor-Fenster und *Menü-Fenster* öffnen sich

Die Variablenansicht

- Für die betrachteten Merkmale einer statistischen Analyse müssen in SPSS *Variablen* definiert werden
- Für jede Variable steht eine Zeile zur Verfügung
- Die Spaltenbeschriftungen benennen die für die Variable festzulegenden Eigenschaften

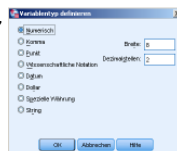


Name

- Der Name ist eine Zeichenkette, die zur eindeutigen Identifikation der Variable führt
 - kurz und prägnant
 - Ziffern und Sonderzeichen sind erlaubt, aber der Name muss mit einem Buchstaben beginnen (keine Leerzeichen!)
 - Zwischen Groß- und Kleinschreibung wird nicht unterschieden (Bsp.: Alter = alter)

Typ

- *numerisch* vs. *String-Datentyp*
- Voreinstellung: *numerischer Datentyp*
- Klickt man in das Feld, wird rechts neben dem Wert eine Schaltfläche mit drei Punkten [...] angezeigt, bei Mausklick hierauf öffnet sich ein kleines Fenster
- Bis auf den letzten Datentyp sind alle numerisch, lediglich die Darstellung und damit verbunden, die akzeptierten Werte bei der Eingabe variieren
- *String*: zur Eingabe expliziter Zeichenketten, die z.B. Buchstaben enthalten
- Empfehlung: numerische Datentypen (SPSS fordert bei der Auswertung, z.B. bei einer Gruppenbildung, meist numerische Datentypen)



Spaltenformat und Dezimalstellen

- Sind bereits voreingestellt, können aber jederzeit manuell verändert werden
- **Spaltenformat** steht für die Anzahl der Zeichen, die maximal verwendet werden sollen
- Die Spalte **Dezimalstellen** steht für sich
Bsp.: Die Voreinstellung **2** bewirkt, dass z.B. die eingegebenen Werte **5** oder **5,004** automatisch jeweils als **5,00** dargestellt werden

Variablenlabel

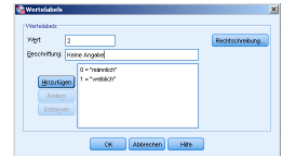
- dient der Beschreibung einer Variable

z.B. Beschreibung der Variable „Einkommen“
-> „monatliches Nettoeinkommen in Euro“

Achtung: Zwar sind theoretisch bis zu 256 Zeichen möglich, von sehr langen Variablenamen ist jedoch abzuraten, da diese per Voreinstellung in den Ergebnisausgaben statt des Namen verwendet werden

Wertelabel

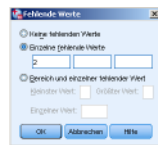
- Hier können den Variablen per **Codeplan** einzelne numerische Werte zugeordnet werden -> Klick auf **...**
- Im Dialogfenster gibt man den Wert und dessen Bedeutung (Beschriftung) ein und betätigt die Schaltfläche **Hinzufügen**
- Markierte Label können auch wieder gelöscht **Entfernen** oder korrigiert **Ändern** werden. Mit **OK** wird die Vergabe der Wertelabels abgeschlossen.



Fehlende Werte

- Hat man, wie im Beispiel der Variable Geschlecht, mit der 2 einen Wert festgelegt, der eigentlich einen fehlenden Wert darstellt, so kann man diesen explizit als solchen deklarieren
- fehlende Werte werden bei der Berechnung von Mittelwerten nicht berücksichtigt

Hinweis: Das gleiche Ergebnis erzielt man, wenn man in das Dateneingabefeld einfach keinen Wert einträgt!



Weitere Spalten in Kurzform

- Spalten**
Breite der Spalte in Zeichen in der Datenansicht; kann dort auch mit der Maus verändert werden
- Ausrichtung**
Ausrichtung einer Variable in der Datenansicht
- Messniveau**
Unterscheidung zwischen Nominalskala (Nominal), Ordinalskala (Ordinal) und metrischer Skala (Skala/Metrisch)

Datentyp	Definition	Beispiele
Nominal	Rein qualitative Merkmalsausprägungen ohne natürliche Ordnung	Geschlecht, Berufsstatus, dichotome Antwort vom Typ "ja/nein"
Ordinal	Qualitative Merkmalsausprägungen mit natürlicher Ordnung	Qualitätseinschätzung ("sehr gut", "gut", "mittel", "schlecht", "sehr schlecht")
Metrisch (auch: rational)	Merkmalsausprägungen, die in einer Zahl besteht und eine Dimension und einen Nullpunkt besitzt	Einkommen (in Euro), Alter (in Jahren), Leistung (in Stück pro Stunde, in km/h)

- Rolle**
Einstellung noch ohne große Konsequenzen, da neu in SPSS, daher für Einführung nicht relevant

Datenansicht

- Hier werden die erfassten Werte für die definierten Variablen eingegeben
- Wichtiger Unterschied zur Variablenansicht:
Eine Variable stellt nun jeweils eine Spalte dar!
Die Zeilen stehen nun für einzelne Merkmalsträger (Fälle)!

Dateneingabe

- Ein eingegebener Wert wird nach Verlassen der Zelle immer automatisch überprüft, ob er für diesen Typ der Variable gültig ist und ggfs. gemäß den Einstellungen formatiert

Bsp.: Aus 5 wird automatisch 5,00 wenn in der Voreinstellung zwei Dezimalstellen angegeben wurden

- Ungültige Werte werden nach Verlassen der Zelle als fehlender Wert durch einen Punkt dargestellt

Bsp.: 5.25 -> .
Da der Punkt kein gültiges Dezimaltrennzeichen ist

- Gespeichert werden die Daten über die Schaltfläche Speichern unter
- Man wählt als Dateierweiterung *sav*, wenn man die Daten weiterhin mit SPSS auswerten will
 - Es besteht aber auch die Möglichkeit, die Daten in Excel- oder ASCII-Format zu speichern

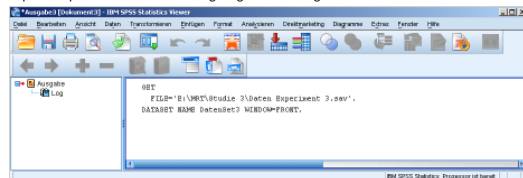
Weitere SPSS-Fenster

- **Das Syntax-Fenster**
dient als Editor, zum Ausführen der Syntax- Programme und zur Fehlersuche (debuggen)
- **Das Skript-Fenster**
eine spezielle Basic-Skriptsprache bietet die Möglichkeit, insbesondere die Ausgabe zu beeinflussen
- **Das Ausgabe- oder Viewer-Fenster**

Das Ausgabe- oder Viewer-Fenster

- Öffnet sich automatisch beim Öffnen des Daten-Editor-Fenster
- Muss nie, kann aber bei Bedarf gespeichert werden
- Erstellt ein Protokoll aller Aktionen, sprich alles, was auf der graphischen Benutzeroberfläche eingestellt wird, wird automatisch umgesetzt in Kommandos der SPSS-eigenen Syntax

Bsp.: Das Speichern der Daten erzeugt folgenden Eintrag:



Univariate Häufigkeitsanalyse

- Stellt geringe Anforderungen an Daten, Merkmale müssen lediglich nominalskaliert sein
- Klassische Ausgabe ist eine Häufigkeitstabelle mit den Merkmalsausprägungen und den zugehörigen absoluten und relativen Häufigkeiten
- Balken- und Kreisdiagramme können zur Veranschaulichung genutzt werden

Analysieren -> Deskriptive Statistiken -> Häufigkeiten

- Bsp.: Will man untersuchen wie viele Mädchen und wie viele Jungen an der Studie teilgenommen haben, wählt man die Variable Geschlecht und erhält dadurch eine Häufigkeitstabelle

Geschlecht		N	Gültig	218
Fehlend				18

		Geschlecht			
		Häufigkeit	Prozent	Gültige Prozente	Formatulierte Prozente
Gültig	Junge	109	49,2	50,0	50,0
	Mädchen	109	49,2	50,0	50,0
Gesamt		218	92,4	100,0	
Fehlend System		18	7,6		
Gesamt		236	100,0		

Es gab 236 Teilnehmer
109 Mädchen + 109 Jungen
+ 18 Teilnehmer, die keine Angabe zum Geschlecht gemacht haben

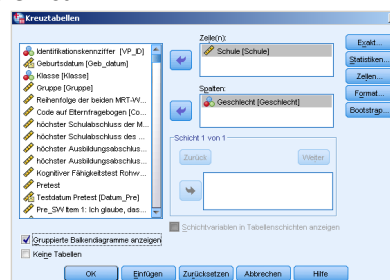
Bivariate Häufigkeitsanalyse - Zusammenhangsmaße für nominalskalierte Daten

- Zur gleichzeitigen Analyse von zwei Variablen
- Neben absoluten Häufigkeiten kann man in einer Kreuztabelle auch relative Häufigkeiten betrachten

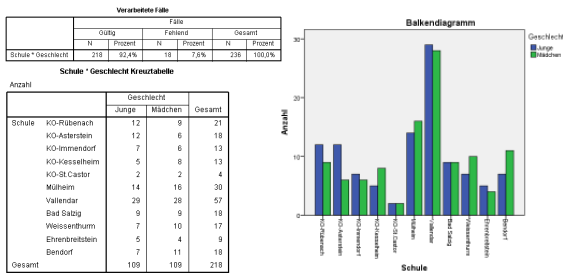
Analysieren -> Deskriptive Statistiken -> Kreuztabellen

- Man erhält ein Dialogfenster, indem man Variablen auswählen kann – diesmal speziell in Felder *Zeile(n)* und *Spalte(n)*
- Außerdem hat man die Option neben Tabellen auch *Balkendiagramme* erstellen zu können

- Bsp.: Man möchte untersuchen wie die Verteilung von Mädchen und Jungen an unterschiedlichen Schulen ist



- Im **Ausgabe-Fenster** werden die verarbeiteten Fälle, eine Kreuztabelle der Variablen und ein Balkendiagramm angezeigt

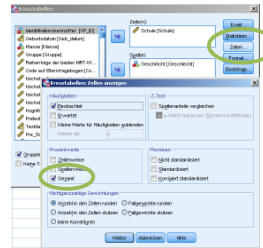


Hier lässt sich die Anzahl der Schüler und Schülerinnen an den Schulen ablesen

Bsp.: In den vierten Klassen der GS Vallendar gibt es 29 Mädchen und 28 Jungen

Über die Schaltfläche Zellen lässt sich noch mehr als die absolute Häufigkeit abfragen.

Z.B. kann man sich den prozentualen Gesamtwert angeben lassen etc.



Schule	Anzahl	Geschlecht		Gesamt
		Junge	Mädchen	
KO-Rübenach	21	12	9	21
KO-Asterstein	18	12	6	18
KO-Immenhof	13	7	6	13
KO-Kesselheim	13	5	8	13
KO-St.Caspar	4	2	2	4
Mühlern	30	14	16	30
Vallendar	57	29	28	57
Bad Säckig	18	9	9	18
Weissenburg	17	7	10	17
Ehrenstettlin	9	5	4	9
Bendorf	18	7	11	18
Gesamt	218	109	109	218

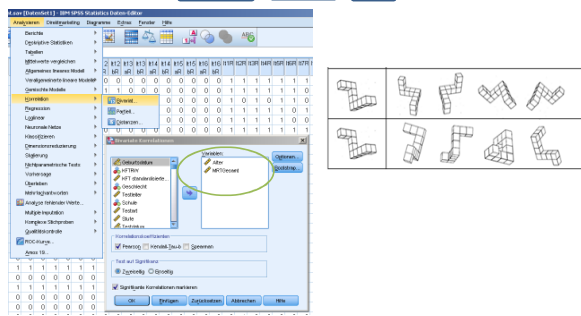
Korrelationen mit SPSS

- Der **Korrelationskoeffizient** gibt Stärke und Richtung eines **linearen Zusammenhangs** von zwei metrischen oder ordinalen Variablen wieder
- Der Korrelationskoeffizient liegt immer zwischen **-1** und **+1**
- Der Wert **-1** gibt an, dass eine vollständig negative Korrelation vorliegt.
- Der Wert **+1** gibt an, dass eine vollständig positive Korrelation vorliegt.
- Beim Wert **0** besteht kein nachweisbarer Zusammenhang zwischen zwei Variablen

Korrelationsberechnung

- Bsp.: Es soll untersucht werden, ob ein Zusammenhang zwischen dem Alter und dem Gesamtergebnis des MRT-Würfeltests (Stichprobe von N=432)

Analysieren -> Korrelation -> Bivariat



Signifikanz

		Alter	MRTGesamt
Alter	Korrelation nach Pearson	1	,287**
	Signifikanz (2-seitig)		,000
	N	432	432
MRTGesamt	Korrelation nach Pearson	,287**	1
	Signifikanz (2-seitig)	,000	
	N	432	432

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

- Der Wert des empirischen **Korrelationskoeffizienten** beträgt 0,287
-> Positiv, daher lässt sich daraus schließen, dass sich mit steigendem Alter, das Gesamtergebnis verbessert.
- Die Korrelation ist auf dem Niveau von 0,01 signifikant.
- Ein Wert unter **.05** bezeugt **Signifikanz**:
-> 0.01 zeugt also von hoher Signifikanz
- Signifikanz= Ein gemessener Zusammenhang zwischen zwei Variablen tritt in der Stichprobe nicht einfach zufällig auf, sondern trifft auch für die Grundgesamtheit zu. Je kleiner der Signifikanzwert, desto unwahrscheinlicher ist es, dass die Korrelation durch Zufall entstanden ist.

Literaturempfehlung

- Reiß, S. & Sarris, V. (2012). *Experimentelle Psychologie – Von der Theorie zur Praxis*. München: Pearson Studium.
- Bühl, A. (2006). *SPSS 14 – Einführung in die moderne Datenanalyse* (10. überarb. & erw. Aufl.). München: Pearson Studium.