

Umgang mit Statistikprogrammen

Variablen definieren Tanja Bossmann

In wissenschaftlichen Studien fallen jede Menge Daten an, die ausgewertet werden wollen. Physiotherapeuten, die gerade mit ihrer Bachelor- oder Masterarbeit beschäftigt sind, können davon ein Lied singen. Nachdem die sorgfältig geplante Studie dann endlich umgesetzt wurde und die Messungen erfolgreich verlaufen sind, fängt der Spaß jedoch erst richtig an. Die gesammelten Daten müssen geordnet, strukturiert und ausgewertet werden. SPSS ist ein bekanntes und in der Wissenschaft vielgenutztes Statistikprogramm, welches dabei ein unentbehrliches Hilfsmittel darstellt. Anhand eines Beispiels mit therapeutischem Bezug sollen zunächst grundlegende Funktionen vermittelt werden. Dieser Beitrag erklärt die Definition von Variablen – ein wichtiger Arbeitsschritt, bevor die Dateneingabe beginnen kann. Die beschriebenen Beispiele wurden dabei mit der Version 12.0 (SPSS Inc.) erarbeitet.


Ohne Struktur geht es nicht!

Die im Rahmen einer Studie erhobenen Daten müssen zur Analyse in eine Datendatei eingegeben werden, wobei es in SPSS eine eindeutig vorgegebene Struktur gibt, welche die Daten nach Fällen – Datensätzen – und Variablen anordnet. Doch vorher ist es elementar, die einzelnen Variablen genau zu definieren und Kodierungen vorzunehmen. Zur Erklärung dieser grundlegenden Schritte dient im Folgenden ein fiktiver Erhebungsbogen, wie er für eine Studie

denkbar wäre (Abb. 1). Nehmen wir an, dass dieser Bogen für 100 Probanden, deren Ausdauerleistungsfähigkeit mittels IPN-Test (📍_S. 952) untersucht wurde, ausgefüllt vorliegt. Die Messungen fanden einmal zu Beginn der Studie und noch einmal nach sechs Wochen statt, wobei 50 Probanden ihren normalen Aktivitäten nachgingen und die anderen ein intensives Trainingsprogramm absolvierten. Zur Charakterisierung der Stichprobe wurden zusätzlich verschiedene soziodemografische Daten,

Erhebungsbogen				Probanden-ID	
Alter... Jahre	Geschlecht	<input type="checkbox"/> männlich	<input type="checkbox"/> weiblich	Größe ... cm	
Familienstand					
<input type="checkbox"/> alleinstehen	<input type="checkbox"/> verheiratet	<input type="checkbox"/> Partnerschaft	<input type="checkbox"/> getrennt lebend	<input type="checkbox"/> geschieden	<input type="checkbox"/> verwitwet
Schulabschluss					
<input type="checkbox"/> kein Schulabschluss	<input type="checkbox"/> Hauptschule	<input type="checkbox"/> Realschule	<input type="checkbox"/> Gymnasium		
Ausbildung / Studium					
<input type="checkbox"/> keine Ausbildung	<input type="checkbox"/> Studium	<input type="checkbox"/> Lehre / Ausbildung			
Berufliche Situation					
<input type="checkbox"/> arbeitssuchend	<input type="checkbox"/> Angestellter	<input type="checkbox"/> leitender Angestellter	<input type="checkbox"/> selbstständig	<input type="checkbox"/> in Ausbildung / Studium	
<input type="checkbox"/> Beamter	<input type="checkbox"/> wehrpflichtig	<input type="checkbox"/> in Altersrente	<input type="checkbox"/> in Altersteilzeit	<input type="checkbox"/> erwerbsunfähig	
Gesundheitssportliche Aktivitäten (z. B. Walking, Gymnastik, etc.)					
kein Gesundheitssport		<input type="checkbox"/> 1-2x / Wo	<input type="checkbox"/> 3-4x / Wo	<input type="checkbox"/> >4x / Wo	
Nikotinkonsum			Alkoholkonsum		
<input type="checkbox"/> nie	<input type="checkbox"/> gelegentlich	<input type="checkbox"/> täglich	<input type="checkbox"/> nie	<input type="checkbox"/> gelegentlich	<input type="checkbox"/> täglich
Messzeitpunkt 1 am ...			Messzeitpunkt 2 am ...		
Gewicht ... kg	Blutdruck ... / ...	IPN-Test ... Watt	Gewicht ... kg	Blutdruck ... / ...	IPN-Test ... Watt

Abb. 1_Fiktiver Erhebungsbogen für eine Studie



INTERNET

Informationen zur neuesten SPSS-Version
www.spss.com/de/spss17/index.htm

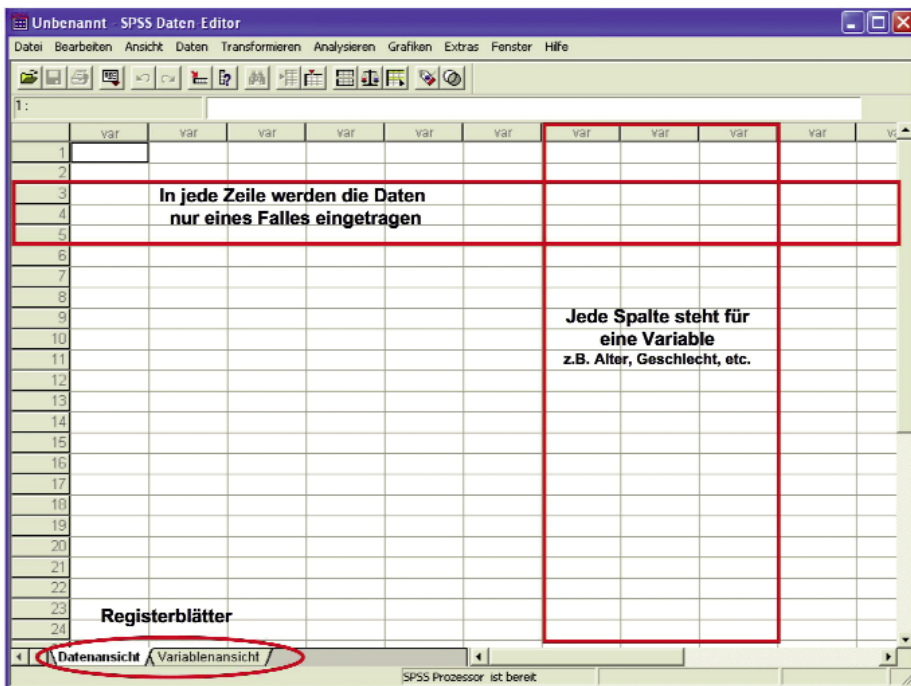


Abb. 2_Arbeitsoberfläche von SPSS für Windows

Variablen brauchen einen Namen und Eigenschaften

Im Registerblatt Variablenansicht bezieht sich nun jede Zeile auf eine Variable und deren Eigenschaften. Zur Definition der Variablen sind die zwei Eigenschaften Name und Datentyp zwingend notwendig, denn es ist wichtig festzulegen, welche Werte zulässig sind und vor allem welche Eingaben man nicht machen darf.

Jede Variable bekommt also zunächst einen frei wählbaren Namen. Der Datentyp legt dann die Art der Daten fest – das entsprechende Dialogfeld kann man über die graue Schaltfläche am rechten Rand des Feldes öffnen. SPSS unterscheidet hier insgesamt acht verschiedene Möglichkeiten, zum Beispiel »Numerisch« oder »String«. Der numerische Variablentyp eignet sich dabei für alle Arten von Zahlen, der Variablentyp »String« ist für alle Eingaben geeignet, die Text enthalten.

Das Spaltenformat legt fest, wie viele Zeichen insgesamt eingegeben werden können, die Eigenschaft Dezimalstellen spezifiziert zusätzlich die vorge- >>>

wie zum Beispiel Alter, Geschlecht, Familien- und Bildungsstand dokumentiert. Da es hier um die Anwendung von SPSS geht, werden die genauen Studien-

und jede Zeile für einen einzelnen Datensatz. Bevor man nun aber Daten eingeben oder importieren kann, müssen die einzelnen Variablen definiert werden und dies ist im Registerblatt Variablenansicht möglich.



Abb. 3_Dialogfeld »Wertelabels definieren« – dieses Beispiel zeigt die Kodierung für die Variable »gesundheitsportliche Aktivität«

vorgänge, wie sie in einem Studienprotokoll (pt 6_2008) dargelegt werden müssen, an dieser Stelle jedoch nicht weiter berücksichtigt.

Die Arbeitsoberfläche des Programms gliedert sich grundsätzlich in zwei Registerblätter: Datenansicht und Variablenansicht (Abb. 2). In der Datenansicht steht jede Spalte für eine andere Variable

	Name	Typ	S/D	De	Variablenlabel	Wertelabels	F	S	A	Meßniveau
1	id	Numerisch	3	0	Probanden ID	Kein	K	7	Re	Metrisch
2	alter	Numerisch	3	0	Alter	Kein	K	7	Re	Metrisch
3	geschl	Numerisch	1	0	Geschlecht	{1, Männlich}...	K	7	Re	Nominal
4	familienst	Numerisch	1	0	Familienstand	{1, alleinstehend}...	K	7	Re	Nominal
5	schulab	Numerisch	1	0	Schulabschluss	{1, kein Schulabschluss}...	K	7	Re	Nominal
6	ausbild	Numerisch	1	0	Ausbildung/Studium	{1, keine Ausbildung}...	K	7	Re	Nominal
7	berufsitu	Numerisch	2	0	Berufliche Situation	{1, arbeitssuchend}...	K	7	Re	Nominal
8	sport	Numerisch	1	0	Gesundheitssportliche Aktivitäten	{1, kein Gesundheitssport}	K	7	Re	Ordinal
9	nikotin	Numerisch	1	0	Nikotinkonsum	{1, nie}...	K	7	Re	Ordinal
10	alkohol	Numerisch	1	0	Alkoholkonsum	{1, nie}...	K	7	Re	Ordinal
11	größe	Numerisch	3	0	Größe in cm	Kein	K	7	Re	Metrisch
12	t1gewicht	Numerisch	4	1	Gewicht in kg zu Messzeitpunkt 1	Kein	K	7	Re	Metrisch
13	t1rsys	Numerisch	3	0	systolischer Blutdruck zu Messzeitpunkt 1	Kein	K	7	Re	Metrisch
14	t1rdias	Numerisch	3	0	diastolischer Blutdruck zu Messzeitpunkt 1	Kein	K	7	Re	Metrisch
15	t1ipn	Numerisch	3	0	Ergebnis IPN-Test zu Messzeitpunkt 1	Kein	K	7	Re	Metrisch
16	t2gewicht	Numerisch	4	1	Gewicht in kg zu Messzeitpunkt 2	Kein	K	7	Re	Metrisch
17	t2rsys	Numerisch	3	0	systolischer Blutdruck zu Messzeitpunkt 2	Kein	K	7	Re	Metrisch
18	t2rdias	Numerisch	3	0	diastolischer Blutdruck zu Messzeitpunkt 2	Kein	K	7	Re	Metrisch
19	t2ipn	Numerisch	3	0	Ergebnis IPN-Test zu Messzeitpunkt 2	Kein	K	7	Re	Metrisch
20	gruppe	Numerisch	1	0	Zuordnung Interventions- bzw Kontrollgruppe	{1, Interventionsgruppe}...	K	7	Re	Nominal
21										

Abb. 4_Variablenansicht für unseren Erhebungsbogen, welche die einzelnen Variablen genau definiert – zum besseren Verständnis sind die Variablenlabels in dieser Abbildung deutlich ersichtlich

sehenen Nachkommastellen. Im Variablenlabel ist es darüber hinaus möglich, ausführlichere Beschreibungen zu den einzelnen Variablen zu hinterlegen.

Arbeitet man mit Kodierungen, ist zusätzlich die Festlegung von Wertelabels ratsam, denn so kann die inhaltliche Bedeutung der eingegebenen Zahlen später schnell nachvollzogen werden (Abb. 3). Der Erhebungsbogen in unserem Beispiel beinhaltet zahlreiche kategoriale Variablen, bei denen diese Vorgehensweise angebracht ist, denn ohne die numerische Kodierung sind statistische Analysen mit diesen Daten nicht möglich.

Da wissenschaftliche Datensätze oftmals sehr umfangreich sind, kommt es häufig vor, dass Werte fehlen. Zur expliziten Kennzeichnung fehlender Werte kann man Kodierungen als Platzhalter für Datenlücken festlegen.

Die Definition von Spaltenbreite und Ausrichtung ist möglich, hat jedoch keine inhaltliche Bedeutung.

Die korrekte Festlegung des Messniveaus in der Variablenansicht ist zwar für die weitere Arbeit mit SPSS nur bedingt relevant, muss jedoch unbedingt bei der Anwendung verschiedener statistischer Verfahren bedacht werden. Das Messniveau – nominal, ordinal oder metrisch (pt 2_2008) – ist für die Wahl des geeigneten Analyseverfahrens entscheidend.

Für unseren Erhebungsbogen ist somit die Festlegung von insgesamt 20 verschiedenen Variablen nötig, denn um später einen Vergleich zwischen den beiden Gruppen – Alltagsaktivität vs. Trainingsprogramm – vornehmen zu können, muss in SPSS zusätzlich eingegeben werden, ob der Proband zur Interventions- oder Kontrollgruppe gehörte. Abbildung 4 zeigt die in SPSS festgelegte Struktur für unsere Beispielstudie in

der Variablenansicht.

Nach dieser wichtigen Vorarbeit kann nun die Dateneingabe beginnen! =

LITERATUR

Untersteiner H. 2007. *Statistik – Datenauswertung mit Excel und SPSS*. Wien: Facultas.wuv Universitätsverlag