

Titelblatt

Kurs 33208:SPSS

Autoren: Joachim Behnke

Nathalie Behnke

Copyright: 2007

Hinweise:

Bitte Polis-Signet auf das **Titelblatt**.

Inhaltsverzeichnis

Über die Autoren.....	4
1 Grundlagen des Arbeitens mit SPSS: Fenster und Dateitypen	5
1.1 Das Datenfenster: Fälle und Variablen.....	7
1.2 Das Syntax-Fenster: alternative Formen der Programmsteuerung.....	10
1.3 Das Ausgabe-Fenster: verschiedene Ansichten.....	16
1.4 Optionen	18
2 Bearbeiten eines Datensatzes	21
2.1 Erstellen eines Datensatzes.....	21
2.2 Öffnen, schließen und speichern.....	22
2.3 Teilen und Verkleinern eines Datensatzes.....	23
2.4 Vergrößern eines Datensatzes.....	31
3 Bearbeiten von Variablen	36
3.1 Variablendefinition	36
3.2 Fehlende Werte.....	38
3.3 Umcodierung bestehender Variablen	40
3.4 Bildung neuer Variablen.....	43
4 Deskriptive Statistiken	53
4.1 'FREQUENCIES'	53
4.2 'DESCRIPTIVES'	58
4.3 'CROSSTABS'	59
4.4 'CORRELATIONS'	64
5 Darstellungsformen	68
5.1 Grafiken.....	69
5.2 Tabellen	79
6 T-Tests	81
6.1 T-Test für zwei unabhängige Stichproben.....	81
6.2 T-Test für zwei abhängige Stichproben.....	88
7 Varianzanalyse.....	92
8 Der Chi²-Test.....	98
9 Bivariate lineare Regression	106

Über die Autoren

Joachim Behnke, Prof. Dr., geb. 1962; Studium der Kommunikationswissenschaft, Politikwissenschaft und Volkswirtschaftslehre an der Ludwig-Maximilians-Universität München. Zur Zeit Vertretung der Professur 'Empirische Politikforschung und Politikfeldanalyse' an der Ludwig-Maximilians-Universität München.

Forschungsschwerpunkte: Rational Choice, Social Choice, Spieltheorie, Wahlsysteme, Wissenschaftstheorie, Empirische und Statistische Methoden

Veröffentlichungen u.a.: mit Nina Baur und Nathalie Behnke (2006): *Empirische Methoden der Politikwissenschaft*. Paderborn: Schöningh (UTB); (2003) *Überhangmandate: Ein (behebbarer) Makel im institutionellen Design des Wahlsystems*. In: *Zeitschrift für Politikwissenschaft*, 13; (2002) *Colemans Theorie der Moderne*. In: Carsten Stark / Christian Lahusen (Hg.): *Theorien der Gesellschaft. Einführung in zentrale Paradigmen der soziologischen Gegenwartsanalyse*. München, 37-65; (2001) *Parteieneigung als Fakt oder Parteieneigung durch Fakten. Der Einfluss von Issues auf das Wahlverhalten*. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 53, 521-546;

Nathalie Behnke, Dr. phil., geb. 1973; Studium der Politikwissenschaft an den Universitäten Bamberg und Bologna. Wissenschaftliche Mitarbeiterin am Lehrgebiet Politikwissenschaft I der FernUniversität Hagen.

Forschungsschwerpunkte: vergleichende Institutionenanalyse, formale Demokratietheorie, Föderalismus, empirische Methoden, Ethik und Korruption

Veröffentlichungen u.a.: (2006) *Ethik-Maßnahmen für die öffentliche Verwaltung – Modeerscheinung oder Mauerblümchen?* In: Jörg Bogumil / Werner Jann / Frank Nullmeier (Hg.): *Politik und Verwaltung*. PVS-Sonderheft; (2004) *Ethik in Politik und Verwaltung. Entstehung und Funktionen ethischer Normen in Deutschland und den USA*. Baden-Baden. (2002) *A Nolan Committee for the German ethics infrastructure?* In: *European Journal of Political Research* 51 (5), 675-708

1 Grundlagen des Arbeitens mit SPSS: Fenster und Dateitypen

Für die statistischen Auswertungen größerer Datensätze, wie sie in der sozialwissenschaftlichen empirischen Forschung üblich sind, bietet es sich an, ein Programm zu verwenden, das genau hierauf spezialisiert ist. Zwar kann man viele Maßzahlen der deskriptiven Statistik, teilweise auch die Berechnung von Zusammenhangsmaßen, im Prinzip noch mit dem Taschenrechner bewältigen; auch ist es sinnvoll und hilfreich, wenn man weiß, wie diese Berechnungen anzustellen sind. Komfortabler, vor allem mit zunehmender Zahl an Variablen und Fällen, ist aber natürlich die Verwendung eines Computerprogramms. Dieser Kurs soll daher dazu dienen, eine Einführung und ein Grundverständnis davon zu vermitteln, wie man mit Hilfe des Statistikprogramms SPSS gängige statistische Verfahren und Auswertungen berechnen kann.

Es gibt eine ganze Reihe von guten Statistikprogrammen mit unterschiedlichen Schwerpunkten. Selbst Tabellenkalkulationsprogramme, die in Standard-Office-Paketen enthalten sind, sind mittlerweile für einfache statistische Berechnungen zu verwenden. STATA und SAS sind andere verbreitete Pakete, wovon insbesondere STATA in jüngerer Zeit in den Sozialwissenschaften großen Anklang findet. Wir haben uns hier aber aus zwei Gründen für eine Einführung in SPSS entschieden. Zum einen ist SPSS durch die an die Windows-Oberfläche angepasste Menüsteuerung schnell und intuitiv verwendbar, auch wenn man noch nicht mit den Tiefen der Programmiersprache vertraut ist. Dies macht es gerade für Anwender attraktiv, die statistische Auswertungen nicht als Lebensinhalt betreiben. Zum anderen ist SPSS in den USA wie in Europa nach wie vor in den Sozialwissenschaften, aber auch in der professionellen Marktforschung, mit am weitesten verbreitet. Wer also Grundkenntnisse in SPSS besitzt, verfügt damit über in höchstem Maße "marktgängige skills", die in den unterschiedlichsten Anwendungsbereichen nachgefragt werden.

SPSS wurde 1968 an der University of Stanford für die Politikwissenschaft entwickelt. Die Abkürzung stand damals für "Statistical Package for the Social Sciences"¹. Seit 1986 hat SPSS auch eine Niederlassung in Deutschland, genauer gesagt in München (<http://www.spss.com/de/>), und der Name der Abkürzung wurde abgewandelt in "Superior Performance Software System". Es besteht aus mehreren so genannten "Modulen", die im Baukastensystem verwendet werden können. Das Basismodul enthält bereits die meisten gängigen Auswertungsverfahren, kann aber durch viele Spezialmodule, etwa für komplexere Tabellen oder für Regressionsanalysen mit ordinalen Daten, ergänzt

1 Für einen kurzen historischen Abriss der Entstehung von SPSS vgl. Bühl/Zöfel 2004: 15.

werden. Eine Probeversion von SPSS kann bei der Münchner Niederlassung für 30 Tage kostenlos bezogen werden, sofern man nachweist, dass man Student ist.

Die Benutzeroberfläche von SPSS ist an Windows angepasst, das heißt, dass man über die Menü-Steuerung und in mehreren Fenstern gleichzeitig arbeiten kann. Wir gehen davon aus, dass alle Leser mit den grundsätzlichen Funktionen einer Windows-Oberfläche wie Dateien öffnen, speichern, verschieben, Text bearbeiten, markieren, kopieren, ausschneiden und einfügen, Fenster öffnen, vergrößern, verkleinern, überlappend anzeigen oder zwischen den Fenstern wechseln, vertraut sind. Hier besteht bei SPSS kein Unterschied zu anderen Windows-Anwendungen, etwa dem Office-Paket.

In SPSS arbeitet man üblicherweise mit drei verschiedenen Fenstern, dem *Datenfenster* oder Dateneditor, dem *Ausgabefenster* oder Viewer und dem *Syntax-Fenster*. Egal, in welchem der drei Fenster man sich befindet, die wesentlichen Menüs in der oberen Menüleiste bleiben immer erhalten. Ein bereits vorhandenes Daten-, Ausgabe- oder Syntaxfenster kann man über das Datei-Menü öffnen, indem man unter DATEI → ÖFFNEN jeweils "DATEN", "SYNTAX" oder "AUSGABE" wählt. Ebenso kann man über das Menü DATEI → NEU → ein neues Fenster des jeweiligen Typs anlegen und unter einem beliebigen Namen abspeichern. Je nachdem, welchen Typus von Fenster man geöffnet hat, speichert SPSS die Datei mit voreingestellten Datei-Endungen ab: Datendateien haben die Endung .sav, Syntaxdateien haben die Endung .sps und Ausgabedateien haben die Endung .spo. So kann man auch beim Öffnen schon vorhandener Dateien die einzelnen Dateitypen an den Endungen leicht unterscheiden, selbst wenn sie die gleichen Dateinamen haben.

Wir verwenden für unsere Rechnungen hier einen Beispieldatensatz. Es handelt sich hierbei um einen Auszug aus einer Umfrage, die nach der Bundestagswahl 1998 vom MZES (Mannheimer Zentrum für Europäische Sozialforschung), dem WZB (Wissenschaftszentrum Berlin), dem ZA (Zentralarchiv für Sozialforschung in Köln) und dem ZUMA (Zentrum für Umfrage- und Meinungsforschung in Mannheim) durchgeführt wurde. In Anlehnung an die Kursnummer dieses Studienbriefs haben wir den auszugsweisen Datensatz 'k3208' benannt. Diese Bezeichnung taucht in den Beispielen immer wieder auf. Die originale Umfrage ist im ZA in Köln unter der Studiennummer s3073 archiviert und dort bei Bedarf auch zu beziehen, falls man Rechnungen nachvollziehen oder Details über die Erhebung und das Codebuch herausfinden möchte.

Die Beispielrechnungen hier im Kurs beziehen sich auf die entsprechenden Kapitel im Kurs 33207. Dort wird etwa in Kap. 24 der T-Test eingeführt. Die Ausführungen in Kap. 6 des vorliegenden Kurses nehmen darauf unmittelbaren Bezug, und die Beispiele sind parallelisiert. Alle Berechnungen hier wurden mit SPSS Version 11 erstellt. Mittlerweile sind neuere Versionen von SPSS

verfügbar, so dass das Layout, etwa von Tabellen, etwas von den Darstellungen hier abweichen kann. Die Befehls-Syntax ist jedoch unverändert, und unseres Wissens wurde auch Anordnung der Befehle in der Menü-Steuerung bislang nicht verändert.

1.1 Das Datenfenster: Fälle und Variablen

Das Datenfenster zeigt in einer Datenblattansicht die Codierung der einzelnen Fälle an. SPSS folgt hier der üblichen Konvention zum Aufbau eines Datensatzes. In den Zeilen werden die einzelnen erhobenen Fälle abgetragen, in den Spalten die Variablen. Eine Zelle in einem Datenblatt ist also immer so zu lesen, dass der Fall der *i*-ten Zeile für die *j*-te Variable die Ausprägung *x* hat. Um ein Beispiel zu geben: In unserem Datensatz *k3208.sav* enthält die 4. Zeile in der 5. Spalte den Wert "2". Dies bedeutet, dass der 4. Befragte bei der Variablen 5 (wie zufrieden ist man mit der Art und Weise, wie die Demokratie in Deutschland funktioniert) die Antwort "ziemlich zufrieden" gegeben hat, die im Datensatz mit der Zahl "2" codiert wurde.

Unser Datensatz *k3208.sav* hat insgesamt 2019 Fälle und 266 Variablen, das heißt, er umfasst 2019 Zeilen und 266 Spalten. Da das Fenster nicht groß genug ist, um alle diese Fälle und Variablen anzuzeigen, muss man gegebenenfalls die Scroll-Balken bewegen. Dennoch ist es sehr unübersichtlich, in einem so großen Datenblatt zu arbeiten, weshalb gerade bei der Eingabe von Fällen aus Fragebögen große Sorgfalt aufgewendet werden muss, damit man sich nicht in Zeilen oder Spalten irrt. Diese 2019x266-Matrix enthält also auf den ersten Blick einen recht verwirrenden Wust an Zahlen. Um zu wissen, was diese Zahlen bedeuten (etwa dass die Zahlen der fünften Spalte etwas über die Demokratiezufriedenheit des Befragten aussagen), gibt es verschiedene Wege. Üblicherweise ist jeder Datensatz dokumentiert. Handelt es sich, wie in unserem Beispiel, um Daten, die durch eine standardisierte Umfrage erhoben wurden, ist dem Datensatz ein so genanntes *Codebuch* beigelegt. Diesem Codebuch kann man entnehmen, wie die Fragen im Fragebogen ursprünglich formuliert waren, welche Antwortmöglichkeiten es auf die Fragen gab und welche Zahlenwerte den einzelnen Antwortmöglichkeiten zugeordnet wurden. Die Fragen des Fragebogens bilden also die Variablen des Datensatzes, die Antwortmöglichkeiten zu den Fragen die Ausprägungen oder Kategorien der einzelnen Variablen und die Zahlenwerte die den Antwortmöglichkeiten zugeordnet werden, die Werte der Variablen.