

Einführungskurs in SPSS

Susanne Heim
Raum 210, Mi 15:00 - 16:00
<susanne.heim@stat.uni-muenchen.de>

Christian Heumann
Raum E06, Di 10:30 - 11:30
<chris@stat.uni-muenchen.de>

SS 2007

Organisatorisches

- Termine: 10 - 18 Uhr (**nach Absprache**); 29. Juni, 6./13./20. Juli
- vormittags 'Theorie', nachmittags 'Übung'
- Datensätze, Folien und Übungsblätter im Internet unter <http://www.statistik.lmu.de/institut/lehrstuhl/semwiso/spss>
- Teilnahmebestätigung ohne Leistungsnachweis
- Schein 'Einführung in statistische Programmpakete' (2 LP) bei Abgabe der Hausarbeit bis zum 17. August 2007

luK-Pool

- Kennung für Studierende der Statistik/BWL/VWL/Wipäd gegen Vorlage des Studenten- und Personalausweises in der WiWi-Bibliothek
- Übungsdateien im Ordner Trainingslaufwerk / k108 sind lokal zu kopieren
- langfristiges Speichern ist nur auf eigenem Netzlaufwerk möglich
- Laptops dürfen nicht ans Netzwerk angeschlossen werden; WLAN ist verfügbar
- Essen und Trinken ist nicht gestattet (Kamera!)

Inhalt

- Anmerkungen zum Programmpaket SPSS
- Datenmanagement (Eingabe und -bearbeitung, Verschmelzen und Aggregieren von Datendateien)
- Syntaxbefehle und Variablentransformation
- deskriptive, explorative und induktive Datenanalyse
- Kreuztabellen, Mittelwertvergleiche, t-Test, Korrelationsanalyse, Regressionsanalyse
- graphische Verfahren

SPSS: Allgemeines

- Entwicklung 1968 im akademischen Umfeld
 - Statistical Package for Social Sciences
- Universitäten, private Wirtschaft, öffentliche Verwaltungen
- Soziologie, Psychologie, BWL/VWL, Biologie, Medizin, u. a.
- weltweit am meisten verbreitetes Statistikprogramm
 - Superior Performing Software Systems
 - Statistical Product and Service Solutions
 - heute nunmehr das Akronym

SPSS: Versionen

- Version 13.0 für Mac OS X, ausschließlich englisch
- Version 15.0 für Windows ME/2000/XP, englisch und deutsch
- Kategoriale Regression (Base), GLMs/GEEs (Advanced Models); mehrere geöffnete Datensätze innerhalb einer Sitzung (> 14.0); Lesen/Schreiben des Stata-Formats, Ergebnisse als PDF exportierbar
- Demo-Version für 14 Tage von der Homepage: <http://www.spss.com>
- Campuslizenz beim LRZ erhältlich: EUR 45,- + EUR 5,- Datenträger; Gültigkeit: 01.06.2007 - 31.03.2008

SPSS: Vor- und Nachteile

- kommerziell, d. h. jüngere statistische Methoden i. A. nicht verfügbar
- +/- Windows-konform, keine UNIX/Linux Version
- + komfortable GUI, anwenderfreundlich
- +/- 'intuitiv' bedienbar
- +/- publikumsreifer Output
- Graphiken nur begrenzt automatisch manipulierbar
- + automatische Erstellung von Programmcode (Syntax)
- unhandliches Programmierool: wenig entwicklerfreundlich
- weniger statistische Verfahren als andere kommerzielle Software
- + umfassendes Hilfesystem

SPSS Struktur: Basissystem mit Erweiterungsmodulen

SPSS Base	SPSS Classification Trees
SPSS Advanced Models	SPSS Conjoint
SPSS Regression Models	SPSS Tables
SPSS Exact Tests	SPSS Complex Samples
SPSS Categories	SPSS Data Preparation
SPSS Trends	SPSS Missing Values Analysis
SPSS Programmability Extension	SPSS Maps

Zusatzprodukte: Sample Power, SPSS Data Entry, Amos, ...

SPSS: Fenster und Dateien

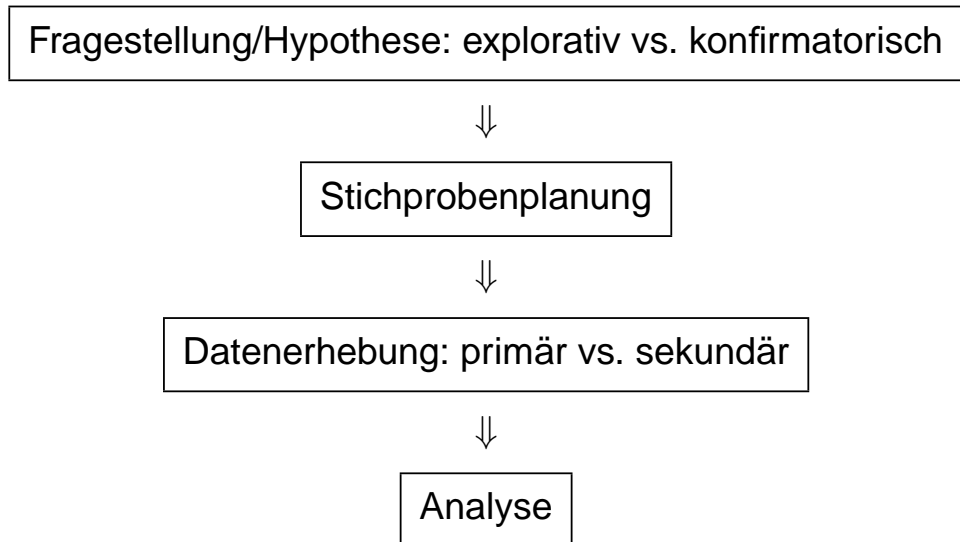
- Daten-Editor
 - Daten-/Variablenansicht (Wechsel mit `Ctrl + T`)
 - mehrere gleichzeitig geöffnete Datensätze (> 14.0)
 - Schließen des letzten Datenfensters beendet SPSS
 - `<data>.sav` (\rightarrow to save)
- Viewer/Ausgabe
 - Output-Navigator
 - Objekte bearbeitbar, z. B. Tabellen formatieren
 - `<output>.spo` (\rightarrow SPSS object)
- Syntax-Editor
 - `<syntax>.sps` (\rightarrow SPSS syntax)

`Alt + Tab` schneller Wechsel des aktiven Fensters

SPSS: Hilfesystem

- Menüpunkt `Hilfe > Themen` mit üblichen Registrierkarten Inhalt, Index und Suchen
- kontextbezogene Hilfe bei Prozeduraufruf übers Menü
- Ausgabefenster: Doppelklick auf Tabellenkopf
- Syntaxfenster: Hilfe zu markiertem Befehl mittels speziellem Icon oder über `Hilfe > Befehlssyntax`
- Menüpunkt `Hilfe > Algorithmen` für z. B. Teststatistiken
- `Hilfe > [Lernprogramm, Fallstudien, Statistics Coach]`
- Handbücher, wenig geeignet zum Einstieg in die Kommandosprache
- USENET Newsgroup `comp.soft-sys.stat.spss`

statistische Untersuchung/Studie



Datenstruktur

SPSS: Fälle, Untersuchungseinheiten

- Menschen: Fragebögen, Messungen
- Unternehmen
- Objekte: Messungen

SPSS: Variablen, Untersuchungsmerkmale

- demographisch: Alter, Geschlecht, Familienstand, Schulabschluss
- klinisch: Blutzucker, Hormonkonzentration, Gewicht
- soziologisch: Freizeitaktivität, Mediennutzung, Kaufverhalten
- wirtschaftlich: Umsatz, Gewinn vor Steuer
- technisch: Materialeigenschaften, experimentelles Design

SPSS: Datenmatrix, Relationale Datenstruktur

- Zeile = Untersuchungseinheit,
- Spalte = Untersuchungsmerkmal

Variablenarten

- stetige Variablen (Alter, Gewicht): reelle Zahlen
 - ordinale Variablen (Schulnoten, Klassifizierungen): größer/kleiner
- ⇒ Merkmalsausprägungen werden Zahlen zugeordnet
- nominale Variablen (Geschlecht, Farben): gleich/ungleich
- ⇒ Merkmalsausprägungen werden willkürlich Zahlen zugewiesen
- Datumsangaben
 - Text

Dies Skala legt die erlaubten Operationen fest.

- ⇒ Variablenkodierung, d. h. zulässige Skalenabbildung
(Dummykodierung, numerische Repräsentation der Antwort)

Kleine Studie im Kurs

- ca. 30 Teilnehmer
- 5 Variablen: Alter, Geschlecht, Gewicht, Größe, Salatkonsum

Geburtsdatum	_____	dd-mm-jjjj
Geschlecht	<input type="checkbox"/> männlich	<input type="checkbox"/> weiblich
Gewicht	_____	kg
Körpergröße	_____	cm
Ich esse Salat	_____	– mal pro Woche.

- Fragestellungen/Hypothesen:
 - Gibt es einen Zusammenhang zw. Geschlecht und Salatkonsum?
 - Wiegen Salatliebhaber signifikant weniger als Salatverweigerer?
 - Welchen Einfluss hat der Salatkonsum auf das Gewicht in Abhängigkeit vom Geschlecht und der Körpergröße?

Variablendefinition

Variablenansicht des Daten-Editors erlaubt Spezifikation von 10 Attributen

- Name der Benutzervariablen:
 - max. 64 Zeichen
 - erstes Zeichen Buchstabe (Case-Sensitivity)
 - beliebig Ziffern, Buchstaben, Sonderzeichen (@, #, \$, ., -), keine Leerzeichen, Vermeidung von Umlauten und "ß"
 - nicht . oder _ als Endung
 - keine SPSS-Syntax wie ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO, WITH

Hilfsvariablen: temporär gültig, nicht speicherbar/auswertbar, 'hash' als erstes Zeichen (z. B. #aux1)

Systemvariablen: \$DATE, \$TIME, \$CASENUM, \$SYSMIS

Variablendefinition

- Typ (Numerisch, String, Datum, Scientific, ...)
- Format (Dezimalstellen, Spaltenformat)
- Labels für Namen und Werte
- Fehlende Werte (systemdefiniert, benutzerdefiniert)
- Spalten und Ausrichtung
- Messniveau (metrisch, ordinal, nominal)

Manuelle Dateneingabe

1. Füllen Sie den ausgeteilten Fragebogen aus.
2. Erzeugen Sie eine SPSS–Datei zur Eingabe der Fragebögen.
3. Geben Sie 5 Fragebögen ein (ausgefüllte Fragebögen austauschen).
4. Speichern Sie die Datei in Ihrem Verzeichnis ab.

ENTER zeilenweises Füllen der Datenmatrix
Tab spaltenweise Füllen der Datenmatrix

Daten einlesen

Datei > Öffnen > Daten

- Tab–delimited, Semikolon-delimited
- dBase
- Lotus, Excel
- SQL Abfragen, ODBC
- Systat, SAS
- andere SPSS Versionen

Datei > Textdaten lesen

- ASCII frei, fest
- unter Verwendung eines Wizards
- Format für zukünftige Verwendung als *.tpf–Datei speicherbar

Variablentransformation

- kopieren, verschieben (markieren und drag & drop), löschen
- Variable vervielfältigen/Attribute übertragen
- Transformieren > Variable berechnen
für numerischen Ausdruck, u. a. verfügbare Funktionengruppen sind
Datumsarithmetik und Zufallszahlen
- Transformieren > Umkodieren unter Nebenbedingungen möglich
- professionell: Rohdatendatei (<daten>r.sav) → Transformationsprogramm (<daten>t.sps) → Fertigdatendatei (<daten>.sav)

Datentransformation

- Daten > Fälle sortieren
auf-/absteigend sortieren nach einer oder mehreren Variablen
- Daten > Transponieren
- Daten > Fälle auswählen
filtern nach einer oder mehreren Variablen
- Daten > Umstrukturieren (über Assistenten)

Übung

1. Lesen Sie den Datensatz `salat.dat` ein.
2. Spezifizieren Sie die Variablenattribute soweit sinnvoll.
3. Fügen Sie eine Variable zur Fallidentifikation ein.
4. Simulieren Sie jeweils das Körpergewicht nach zweimaliger Gewichts**zunahme** als weitere Variablen des Datensatzes. Achten Sie dabei auf Erhalt der Variablenattribute Gewicht.
5. Organisieren Sie den Datensatz mit Hilfe des Umstrukturierungsassistenten, sodass die wiederholten Messwerte in einer einzigen Variablen erfasst werden und eine Hilfsvariable für die Messungen kodiert.