



Das statistische
Softwarepaket SPSS
(jetzt: IBM SPSS
Statistics)

Übersicht

Programmpaket SPSS

- Versionen und Verfügbarkeit
- Vor- und Nachteile
- Struktur: Basissystem mit Erweiterungsmodulen
- Fenster und Dateien
- Hilfesystem

Datenmanagement

- Variablenarten
- Variablendefinition
- Manuelle Dateneingabe
- Daten einlesen
- Fehlende Werte
- Variablenerzeugung und Variablentransformation
- Datentransformation
- Daten aggregieren

Kapitel 1

Das statistische Softwarepaket SPSS (jetzt: IBM SPSS Statistics)

Eine Kurzeinführung

Vorlesung *LVA Statistische Software*
Wintersemester 2015/2016

Christian Heumann, \LaTeX von Martin Dörr
Institut für Statistik
Ludwig-Maximilians-Universität München

1 ÜBERSICHT

2 PROGRAMMPAKET SPSS

- Versionen und Verfügbarkeit
- Vor- und Nachteile
- Struktur: Basissystem mit Erweiterungsmodulen
- Fenster und Dateien
- Hilfesystem

3 DATENMANAGEMENT

- Variablenarten
- Variablendefinition
- Manuelle Dateneingabe
- Daten einlesen
- Fehlende Werte
- Variablenerzeugung und Variablentransformation
- Datentransformation
- Daten aggregieren



Das statistische
Softwarepaket SPSS
(jetzt: IBM SPSS
Statistics)

Übersicht

Programmpaket SPSS

Versionen und
Verfügbarkeit
Vor- und Nachteile
Struktur: Basissystem mit
Erweiterungsmodulen
Fenster und Dateien
Hilfesystem

Datenmanagement

Variablenarten
Variablendefinition
Manuelle Dateneingabe
Daten einlesen
Fehlende Werte
Variablenerzeugung und
Variablentransformation
Datentransformation
Daten aggregieren

- Programmpaket SPSS
- Datenmanagement



**Das statistische
Softwarepaket SPSS
(jetzt: IBM SPSS
Statistics)**

Übersicht

Programmpaket SPSS

Versionen und
Verfügbarkeit

Vor- und Nachteile

Struktur: Basissystem mit
Erweiterungsmodulen

Fenster und Dateien

Hilfesystem

Datenmanagement

Variablenarten

Variablendefinition

Manuelle Dateneingabe

Daten einlesen

Fehlende Werte

Variablenerzeugung und
Variablentransformation

Datentransformation

Daten aggregieren



Das statistische
Softwarepaket SPSS
(jetzt: IBM SPSS
Statistics)

Übersicht

Programmpaket SPSS

Versionen und
Verfügbarkeit

Vor- und Nachteile

Struktur: Basissystem mit
Erweiterungsmodulen

Fenster und Dateien

Hilfesystem

Datenmanagement

Variablenarten

Variablendefinition

Manuelle Dateneingabe

Daten einlesen

Fehlende Werte

Variablenerzeugung und
Variablentransformation

Datentransformation

Daten aggregieren

- Entwicklung 1968 im akademischen Umfeld (Stanford University): Statistical Package for Social Sciences
- Anwendung in vielen Bereichen: öffentliche Verwaltungen, private Wirtschaft, Universitäten
- Weltweit weit verbreitetes Statistikprogramm
- Zwischenzeitlicher Name: Superior Performing Software Systems

- Aktuell: Version 22 für Linux, Mac (Intel Prozessor) und Windows, Version 22 LRZ
- Neue Features:
<http://www.ibm.com/software/analytics/spss/products/statistics/features.html>
- Campuslizenz beim LRZ erhältlich (Mitarbeiter und Studenten): EUR 72.-, Lizenzierung immer für ein Jahr



Das statistische
Softwarepaket SPSS
(jetzt: IBM SPSS
Statistics)

Übersicht

Programmpaket SPSS

Versionen und
Verfügbarkeit

Vor- und Nachteile

Struktur: Basissystem mit
Erweiterungsmodulen

Fenster und Dateien

Hilfesystem

Datenmanagement

Variablenarten

Variablendefinition

Manuelle Dateneingabe

Daten einlesen

Fehlende Werte

Variablenerzeugung und
Variablentransformation

Datentransformation

Daten aggregieren

- Kommerziell, d. h. aktuelle statistische Methoden i. A. nicht sofort verfügbar
- + Komfortable GUI, anwenderfreundlich, 'intuitiv' bedienbar
- + Publikationsreife Ausgabe
- Graphiken nur begrenzt automatisch manipulierbar
- + Automatische Erstellung von Programmcode (Syntax), dadurch Reproduzierbarkeit der Ergebnisse, Kommunizierbarkeit des Vorgehens
- Unhandliches Programmierool: wenig entwicklerfreundlich
- + Umfassendes Hilfesystem



Das statistische
Softwarepaket SPSS
(jetzt: IBM SPSS
Statistics)

Übersicht

Programmpaket SPSS

Versionen und
Verfügbarkeit

Vor- und Nachteile

Struktur: Basissystem mit
Erweiterungsmodulen

Fenster und Dateien

Hilfesystem

Datenmanagement

Variablenarten

Variablendefinition

Manuelle Dateneingabe

Daten einlesen

Fehlende Werte

Variablenerzeugung und
Variablentransformation

Datentransformation

Daten aggregieren

STRUKTUR: BASISSYSTEM MIT ERWEITERUNGSMODULEN (OHNE ZUSATZ IBM SPSS)

Statistics Base
Bootstrapping
Complex Samples
Custom Tables
Data Preparation
Direct Marketing
Forecasting
Neural Networks
Programmability Extension

Advanced Statistics
Categories
Conjoint
Data Collection Data Entry
Decision Trees
Exact Tests
Missing Values
Regression
Amos

Heumann



Das statistische
Softwarepaket SPSS
(jetzt: IBM SPSS
Statistics)

Übersicht

Programmpaket SPSS

Versionen und
Verfügbarkeit

Vor- und Nachteile

Struktur: Basissystem mit
Erweiterungsmodulen

Fenster und Dateien
Hilfesystem

Datenmanagement

Variablenarten

Variablendefinition

Manuelle Dateneingabe

Daten einlesen

Fehlende Werte

Variablenerzeugung und
Variablentransformation

Datentransformation

Daten aggregieren



Das statistische
Softwarepaket SPSS
(jetzt: IBM SPSS
Statistics)

Übersicht

Programmpaket SPSS

Versionen und
Verfügbarkeit

Vor- und Nachteile

Struktur: Basissystem mit
Erweiterungsmodulen

Fenster und Dateien

Hilfesystem

Datenmanagement

Variablenarten

Variablendefinition

Manuelle Dateneingabe

Daten einlesen

Fehlende Werte

Variablenerzeugung und
Variablentransformation

Datentransformation

Daten aggregieren

■ Daten-Editor

- Daten-/Variablenansicht (Wechsel mit `Ctrl + T`)
- mehrere gleichzeitig geöffnete Datensätze (ab Version 15.0)
- Schließen des letzten Datenfensters beendet SPSS
- `<data>.sav` (→ `to save`)

■ Viewer/Ausgabe

- Output-Navigator: Baumstruktur
- Objekte bearbeitbar, z. B. Tabellen formatieren
- `<output>.spo` (→ `SPSS object`)

■ Syntax-Editor

- `<syntax>.sps` (→ `SPSS syntax`)

`Alt + Tab` schneller Wechsel des aktiven Fensters

- Menüpunkt `Hilfe > Themen` mit üblichen Registrierkarten `Inhalt`, `Index` und `Suchen`
- kontextbezogene Hilfe bei Prozeduraufruf über das Menü
- Ausgabefenster: Doppelklick auf Tabellenkopf
- Syntaxfenster: Hilfe zu markiertem Befehl mittels speziellem Icon oder über `Hilfe > Befehlssyntax`
- Handbücher
- Google group `comp.soft-sys.stat.spss`



Das statistische
Softwarepaket SPSS
(jetzt: IBM SPSS
Statistics)

Übersicht

Programmpaket SPSS

Versionen und
Verfügbarkeit
Vor- und Nachteile
Struktur: Basissystem mit
Erweiterungsmodulen
Fenster und Dateien

Hilfesystem

Datenmanagement

Variablenarten
Variablendefinition
Manuelle Dateneingabe
Daten einlesen
Fehlende Werte
Variablenerzeugung und
Variablentransformation
Datentransformation
Daten aggregieren



Das statistische
Softwarepaket SPSS
(jetzt: IBM SPSS
Statistics)

Übersicht

Programmpaket SPSS

Versionen und
Verfügbarkeit
Vor- und Nachteile
Struktur: Basissystem mit
Erweiterungsmodulen
Fenster und Dateien
Hilfesystem

Datenmanagement

Variablenarten
Variablendefinition
Manuelle Dateneingabe
Daten einlesen
Fehlende Werte
Variablenerzeugung und
Variablentransformation
Datentransformation
Daten aggregieren

- Fälle, Untersuchungseinheiten
 - Menschen: Fragebögen, Messungen
 - Unternehmen
 - Objekte: Messungen
- Variablen, Untersuchungsmerkmale
 - demographisch: Alter, Geschlecht, Familienstand, Schulabschluss
 - klinisch: Blutzucker, Hormonkonzentration, Gewicht
 - soziologisch: Freizeitaktivität, Mediennutzung, Kaufverhalten
 - wirtschaftlich: Umsatz, Gewinn vor Steuer
 - technisch: Materialeigenschaften, experimentelles Design
- Datenmatrix
 - Zeile entspricht Untersuchungseinheit
 - Spalte entspricht Untersuchungsmerkmal

- 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.



Das statistische
Softwarepaket SPSS
(jetzt: IBM SPSS
Statistics)

Übersicht

Programmpaket SPSS

Versionen und
Verfügbarkeit
Vor- und Nachteile
Struktur: Basissystem mit
Erweiterungsmodulen
Fenster und Dateien
Hilfesystem

Datenmanagement

Variablenarten

Variablendefinition
Manuelle Dateneingabe
Daten einlesen
Fehlende Werte
Variablenerzeugung und
Variablentransformation
Datenanreicherung
Daten aggregieren

Variablenansicht des Daten-Editors erlaubt Spezifikation von 10 Attributen

- Name der Benutzervariablen:
 - max. 64 Zeichen
 - erstes Zeichen Buchstabe
 - beliebig Ziffern, Buchstaben, Sonderzeichen (@, #, \$, ., _), keine Leerzeichen, Vermeidung von Umlauten und "ß"
 - nicht . oder _ als Endung
 - keine reservierten Namen aus der SPSS-Syntax wie ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO, WITH, &, |
- Hilfsvariablen: temporär gültig, 'hash' als erstes Zeichen (z. B. #aux1)
- Systemvariablen: \$DATE, \$TIME, \$CASENUM, \$SYSMIS



Das statistische
Softwarepaket SPSS
(jetzt: IBM SPSS
Statistics)

Übersicht

Programmpaket SPSS

Versionen und
Verfügbarkeit
Vor- und Nachteile
Struktur: Basissystem mit
Erweiterungsmodulen
Fenster und Dateien
Hilfesystem

Datenmanagement

Variablenarten

Variablendefinition

Manuelle Dateneingabe
Daten einlesen
Fehlende Werte
Variablenerzeugung und
Variablentransformation
Datentransformation
Daten aggregieren

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



Das statistische
Softwarepaket SPSS
(jetzt: IBM SPSS
Statistics)

Übersicht

Programmpaket SPSS

Versionen und
Verfügbarkeit

Vor- und Nachteile

Struktur: Basissystem mit
Erweiterungsmodulen

Fenster und Dateien

Hilfesystem

Datenmanagement

Variablenarten

Variablendefinition

Manuelle Dateneingabe

Daten einlesen

Fehlende Werte

Variablenerzeugung und
Variablentransformation

Datentransformation

Daten aggregieren

- 1 Anlegen der Variablen
- 2 Eingeben der Daten im Dateneditor:
ENTER zeilenweises Füllen der Datenmatrix
Tab spaltenweise Füllen der Datenmatrix
- 3 Speichern der Datei
- 4 Tip: Sekundärdaten, z.B. Kontingenztafeln können durch Anlegen einer GewichtungsvARIABLE analysiert werden.



Das statistische
Softwarepaket SPSS
(jetzt: IBM SPSS
Statistics)

Übersicht

Programmpaket SPSS

- Versionen und Verfügbarkeit
- Vor- und Nachteile
- Struktur: Basissystem mit Erweiterungsmodulen
- Fenster und Dateien
- Hilfesystem

Datenmanagement

- Variablenarten
- Variablendefinition

Manuelle Dateneingabe

- Daten einlesen
- Fehlende Werte
- Variablenerzeugung und Variablentransformation
- Datentransformation
- Daten aggregieren

Datei > Öffnen > Daten

- Textdateien: Tab-delimited oder cvs (comma separated values)
- 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
- Beispiel: Einlesen der Mietspiegeldaten 2003



Das statistische
Softwarepaket SPSS
(jetzt: IBM SPSS
Statistics)

Übersicht

Programmpaket SPSS

Versionen und
Verfügbarkeit

Vor- und Nachteile

Struktur: Basissystem mit
Erweiterungsmodulen

Fenster und Dateien

Hilfesystem

Datenmanagement

Variablenarten

Variablendefinition

Manuelle Dateneingabe

Daten einlesen

Fehlende Werte

Variablenerzeugung und
Variablentransformation

Datentransformation

Daten aggregieren

”freies” ASCII

- Format per Editor einsehen/kontrollieren beim Import/Export
- unsichtbare Steuer-/Formatierungszeichen: <TAB>, <CR>, <SPACE>
- leere Zelle kodiert vorzugsweise für fehlenden Wert
⇒ Stringerkennung unproblematisch
- variable Feldbreite
- Variablennamen in erster Zeile möglich
- SPSS Datenimport Assistent
z. B. 'mm/tt/jjjj'–Datumsformat für Variable
Geburtsdatum spezifizieren; ändern in der
SPSS Variablenansicht



Das statistische
Softwarepaket SPSS
(jetzt: IBM SPSS
Statistics)

Übersicht

Programmpaket SPSS

Versionen und
Verfügbarkeit
Vor- und Nachteile
Struktur: Basissystem mit
Erweiterungsmodulen
Fenster und Dateien
Hilfesystem

Datenmanagement

Variablenarten
Variablendefinition
Manuelle Dateneingabe

Daten einlesen

Fehlende Werte
Variablenerzeugung und
Variablentransformation
Datenanreicherung
Daten aggregieren

"freies" ASCII

- Lesen von Dezimalzahlen abhängig von Einstellung im Betriebssystem; Änderung des Dezimalzeichens in Windows (XP) unter Systemsteuerung > Regions- und Sprachoptionen > Regionale Einstellungen > Anpassen > Zahlen
- SPSS liest beim 1. Aufruf die Konventionen aus der Systemsteuerung
Vorsicht: deutsche Version von Excel interpretiert Komma stets als Dezimalzeichen



Das statistische
Softwarepaket SPSS
(jetzt: IBM SPSS
Statistics)

Übersicht

Programmpaket SPSS

Versionen und
Verfügbarkeit
Vor- und Nachteile
Struktur: Basissystem mit
Erweiterungsmodulen
Fenster und Dateien
Hilfesystem

Datenmanagement

Variablenarten
Variablendefinition
Manuelle Dateneingabe

Daten einlesen

Fehlende Werte
Variablenerzeugung und
Variablentransformation
Datentransformation
Daten aggregieren

systemdefiniert

- numerische Variablen:
 - keine Zahl oder Leerzeichen liefert ein `MISS`
 - Darstellung als `'.'`

- Stringvariablen: Interpretation als leerer String

benutzerdefiniert: Deklaration in der Variablenansicht

- Differenzierung der Ursachen eines fehlenden Wertes, etwa keine Antwort geben wollen oder können, da nichts zutrif.

212 "Ersatzwert" definieren, z. B. für Einkommen:

'-9' Verweigerung, '-99' Nicht angetroffen

- fehlender Wert vom Typ String: z.B. 'KA' kodiert für 'Keine Angabe'



Das statistische Softwarepaket SPSS (jetzt: IBM SPSS Statistics)

Übersicht

Programmpaket SPSS

- Versionen und Verfügbarkeit
- Vor- und Nachteile
- Struktur: Basissystem mit Erweiterungsmodulen
- Fenster und Dateien
- Hilfesystem

Datenmanagement

- Variablenarten
- Variablendefinition
- Manuelle Dateneingabe
- Daten einlesen

Fehlende Werte

- Variablenerzeugung und Variablentransformation
- Datentransformation
- Daten aggregieren

- 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)



Das statistische
Softwarepaket SPSS
(jetzt: IBM SPSS
Statistics)

Übersicht

Programmpaket SPSS

Versionen und
Verfügbarkeit
Vor- und Nachteile
Struktur: Basissystem mit
Erweiterungsmodulen
Fenster und Dateien
Hilfesystem

Datenmanagement

Variablenarten
Variablendefinition
Manuelle Dateneingabe
Daten einlesen
Fehlende Werte

Variablenerzeugung und Variablentransformation

Datentransformation
Daten aggregieren

- 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)



Das statistische
Softwarepaket SPSS
(jetzt: IBM SPSS
Statistics)

Übersicht

Programmpaket SPSS

Versionen und
Verfügbarkeit
Vor- und Nachteile
Struktur: Basissystem mit
Erweiterungsmodulen
Fenster und Dateien
Hilfesystem

Datenmanagement

Variablenarten
Variablendefinition
Manuelle Dateneingabe
Daten einlesen
Fehlende Werte
Variablenerzeugung und
Variablentransformation

Datentransformation

Daten aggregieren



Das statistische
Softwarepaket SPSS
(jetzt: IBM SPSS
Statistics)

Übersicht

Programmpaket SPSS

Versionen und
Verfügbarkeit
Vor- und Nachteile
Struktur: Basissystem mit
Erweiterungsmodulen
Fenster und Dateien
Hilfesystem

Datenmanagement

Variablenarten
Variablendefinition
Manuelle Dateneingabe
Daten einlesen
Fehlende Werte
Variablenerzeugung und
Variablentransformation
Datentransformation

Daten aggregieren

Daten > Aggregieren ...

- Ziel:
 - Zusammenfassen von Beobachtungen
 - wiederholte Beobachtungen
 - Gruppenrepräsentant
- Break-Variable zur Unterteilung/Gruppierung
- Aggrrierungsfunktionen: Lage-/Streuungsmaße, prozentuale Anteile
- gruppenspezifische Fallzahlen
- Aggregieren in Kombination mit Zusammenfügen von Dateien zum Umbau von Dateien

DATEN AGGREGIEREN: ZUSAMMENFÜGEN VON DATEIEN, FÄLLE HINZUFÜGEN

Daten > Dateien zusammenfügen

- Fälle hinzufügen:
Variablen können in unterschiedlicher Reihenfolge vorliegen
- Variablen, die in beiden Dateien vorkommen, bleiben in der Arbeitsdatei unverändert
- bisherige Variablen sind mit (*) gekennzeichnet, hinzukommende Variablen mit (+)
- optionaler Herkunftsindikator: 0=Arbeitsdatei, 1=externe Datei



Das statistische
Softwarepaket SPSS
(jetzt: IBM SPSS
Statistics)

Übersicht

Programmpaket SPSS

Versionen und
Verfügbarkeit
Vor- und Nachteile
Struktur: Basissystem mit
Erweiterungsmodulen
Fenster und Dateien
Hilfesystem

Datenmanagement

Variablenarten
Variablendefinition
Manuelle Dateneingabe
Daten einlesen
Fehlende Werte
Variablenerzeugung und
Variablentransformation
Datentransformation

Daten aggregieren

DATEN AGGREGIEREN: ZUSAMMENFÜGEN VON DATEIEN, VARIABLEN HINZUFÜGEN

- Fall 1:** Beide Dateien haben gleiche Anzahl an Fällen und die gleichen Fälle. Verwendung einer Schlüsselvariablen optional.
- Fall 2:** Externe Datei ist größer als Arbeitsdatei, es sollen aber nur Variablen zu den Fällen der Arbeitsdatei hinzugefügt werden. Externe Datei enthält zu verwendenden Schlüssel.
- Fall 3:** Externe Datei ist kleiner Arbeitsdatei, es interessieren nur Fälle der externen Datei. Verwendeter Schlüssel ist in Arbeitsdatei.
- Fall 4:** In beiden Dateien kommen Beobachtungen vor; Vereinigungsmenge bilden (teilweise fehlende Werte).



Das statistische
Softwarepaket SPSS
(jetzt: IBM SPSS
Statistics)

Übersicht

Programmpaket SPSS

- Versionen und Verfügbarkeit
- Vor- und Nachteile
- Struktur: Basissystem mit Erweiterungsmodulen
- Fenster und Dateien
- Hilfesystem

Datenmanagement

- Variablenarten
- Variablendefinition
- Manuelle Dateneingabe
- Daten einlesen
- Fehlende Werte
- Variablenerzeugung und Variablentransformation
- Datentransformation

Daten aggregieren