

# Kurzeinführung in SPSS

(Einleitung)

## 1. Literatur:

A. BÜHL: *SPSS, Version 8*, Addison-Wesley-Verlag

W. KÄHLER: *SPSS für Windows*, Vieweg-Verlag

## 2. Internet:

<http://www.lrz-muenchen.de/services/schulung/unterlagen/einfuehrung>

## 3. Bemerkungen:

Wenn man SPSS startet, öffnet sich der sogenannte Dateneditor. Dort können Daten eingegeben und bearbeitet werden. Dabei stellt jede Zeile einen einzelnen Fall (eine Beobachtung) und jede Spalte eine Variable dar.

Nimmt man als Beispiel eine Umfrage, so ist jede befragte Person ein Fall (also eine Zeile) und jede gestellte Frage eine Variable (also eine Spalte).

Im Vergleich etwa zu Microsoft Excel können im Dateneditor keine Berechnungen ausgeführt oder Formeln eingegeben werden. Zum Berechnen neuer Werte und Variablen, sowie zum Umkodieren sollte das Menü **Transform** verwendet werden.

## 4. Tastenkombinationen:

- **Return-Taste**: eine Zelle tiefer innerhalb derselben Spalte
- **Tabulator-Taste**: eine Zelle nach rechts in derselben Zeile
- **Umschalt-Taste + Tabulator-Taste**: eine Zelle nach links in derselben Zeile
- **Strg + Cursor-Taste**: zur am weitesten entfernten Zelle in der gewählten Richtung
- **Strg + Pos1**: zur Zelle in der linken oberen Ecke der SPSS-Daten-Tabelle
- **Strg + Ende**: zur Zelle in der rechten unteren Ecke der SPSS-Daten-Tabelle
- **Umschalt-Taste + Entf**: Löschen des Inhalts der aktuellen Zelle
- **Strg + C**: Kopieren des Inhalts der aktuellen Zelle in den Zwischenspeicher
- **Strg + V**: Einfügen des Inhalts des Zwischenspeichers in die aktuelle Zelle

## 5. Editier-Hilfen:

- Einfügen einer neuen Tabellenspalte vor die aktuelle Spalte: **Data - Insert Variable**
- Markieren einer Tabellenspalte: Mit der Maus auf den Variablennamen klicken
- Löschen einer markierten Tabellenspalte: **Entf** betätigen
- Einfügen einer neuen Tabellenzeile oberhalb der aktiven Zeile: **Data - Insert Case**
- Markieren einer Tabellenzeile: Mit der Maus auf die Zeilennummer klicken
- Löschen einer markierten Tabellenzeile: **Entf** betätigen

## 6. Variablen definieren:

- **Variablenname**: Variablen dürfen aus bis zu 8 Zeichen bestehen, wobei das erste ein Buchstabe sein muß. Weitere erlaubt Zeichen danach: Buchstaben, Ziffern, "\_", "\$", "#", ".", "@". Der Dezimalpunkt "." darf nicht am Zeichenende stehen. Groß- und Kleinschreibung wird nicht als gesondert angesehen, d.h. "Alp" ist gleichbedeutend mit "aLp", ... Nicht als Variablennamen verwendet werden dürfen: **all and by eq ge gt le lt ne not or thru to with**. (**Data - Define Variable**)
- **Variablentyp**: Dieser ist vor der Eingabe des ersten Variablenwertes festzulegen. (**Data - Define Variable - Type**)
- **Variablenlabel**: Diese sind dazu da, um Analyseergebnisse leichter lesbar zu machen. (**Data - Define Variable - Labels**)

## 7. Anzeige von Werten: (Summary Report - Statistics)

- **Number of Cases:** Ausgabe der Anzahl der gültigen Fälle
- **Mean:** Ausgabe der arithmetischen Mittelwerte
- **Median:** Ausgabe der Medianwerte
- **Grouped Median:** Ausgabe der gruppierten Medianwerte
- **Std. Error of Mean:** Ausgabe der Standardfehler der arithmetischen Mittelwerte
- **Sum:** Ausgabe der Summenwerte
- **Minimum:** Ausgabe der kleinsten Werte
- **Maximum:** Ausgabe der größten Werte
- **Range:** Ausgabe der Spannweiten
- **First:** Ausgabe der jeweils ersten Werte
- **Last:** Ausgabe der jeweils letzten Werte
- **Standard Deviation:** Anzeige der Standardabweichungen
- **Variance:** Ausgabe der Varianzen
- **Kurtosis:** Ausgabe der Werte für die Wölbung
- **Std. Error of Kurtosis:** Ausgabe der Standardfehler für die Wölbung
- **Skewness:** Ausgabe der Werte für die Schiefe
- **Std. Error of Skewness:** Ausgabe der Standardfehler für die Schiefe
- **Harmonic Mean:** Ausgabe der harmonischen Mittelwerte
- **Geometric Mean:** Ausgabe der geometrischen Mittelwerte
- **Percent of Total Sum:** Ausgabe des prozentualen Anteils an der Gesamtsumme, bezogen auf jede einzelne Ausprägung der Gruppierungsdaten
- **Percent of Total N:** Ausgabe der prozentualen Fallzahl, bezogen auf jede einzelne Ausprägung der Gruppierungs-Variablen

## 8. Häufigkeitsverteilungen und Statistiken: (Analyze-Descriptive Statistics-Frequencies)

- **Auswahl Display frequency tables:** Häufigkeitstabellen anzeigen
  - **Auswahl Statistics**
    - **Quartiles:** Ausgabe der Quartilwerte.
    - **Cut points for:** Der Bereich von 0% bis 100% soll in  $n$  Teile mit jeweils annähernd gleichen Ausprägungshäufigkeiten gegliedert werden. Beispielsweise führt der Eintrag "4" zur Ausgabe der Quartilwerte.
    - **Percentile(s):** Bei einem Perzentilwert von  $p\%$  handelt es sich um denjenigen Wert, unterhalb dem  $p\%$  aller beobachteten Merkmalsausprägungen liegen.
    - **Mode:** Modalwert
    - **S.E.mean:** Standardfehler für den Mittelwert
    - **Values are group midpoints:** Auswählen, falls mit gruppierten Daten gerechnet wird.
- Ein 95%-Konfidenzintervall für die zentrale Tendenz (Mittelwert, Median bzw. Modalwert) ist jenes Intervall, welches das arithmetische Mittel als Mittelpunkt enthält, und dessen halbe Breite der Größe entspricht, die sich aus der Multiplikation des Standardfehlers mit dem Faktor 1,96 ergibt.
- **Auswahl Charts** Anzeige von Diagrammen
    - **Bar charts:** Balkendiagramme
    - **Pie charts:** Kreisdiagramme
    - **Histograms:** Histogramme, **With normal curve:** Normalverteilungskurve in das Histogramm einblenden

- **Frequencies:** Häufigkeiten verwenden
- **Percentages:** Prozente verwenden
- **Auswahl Format**
  - **Ascending values:** Die Einträge in der Tabelle sind aufsteigend nach Variablenwerten geordnet (Voreinstellung!).
  - **Descending values:** Die Einträge in der Tabelle sind absteigend nach Variablenwerten geordnet.
  - **Ascending counts:** Die Einträge in der Tabelle sind aufsteigend nach den Häufigkeiten der Variablenwerter geordnet.
  - **Descending counts:** Die Einträge in der Tabelle sind absteigend nach den Häufigkeiten der Variablenwerter geordnet.
  - **Compare variables:** Sofern eine Häufigkeitsauszählung für mehr als eine Variable abgerufen wird, werden die Ergebnisse innerhalb einer Gesamttabelle angezeigt.
  - **Organize output by variables:** Sofern eine Häufigkeitsauszählung für mehr als eine Variable angefordert wird, werden die Ergebnisse für jede Variable getrennt ausgegeben.
  - **Suppress tables with more than  $n$  categories:** Die Ausgabe erfolgt nur für diejenigen Tabellen, die höchstens die innerhalb des zugehörigen Textfeldes eingetragene Anzahl von Merkmalsausprägungen halten.

## 9. Statistiken von kontinuierlichen Merkmalen:

(Analyze - Descriptive Statistics - Descriptives)

Kann bei Merkmalen theoretisch jeder Wert eines bestimmten Intervalls als Meßwert auftreten, dann ist es nicht sinnvoll, Häufigkeitstabellen auszugeben. Hier sollte man geeignete Statistiken verwenden, wie etwa den Mittelwert (**Mean**), die Varianz (**Variance**), ...

Sollen einzelne Merkmalsträger bezüglich verschiedener Merkmale, bei denen sich z.B. die Maßeinheiten unterscheiden, miteinander verglichen werden, so ist das Kontrollkästchen **Save standardized values as variables** zu aktivieren.

## 10. Explorative Datenanalyse (Analyze - Descriptive Statistics - Explore)

- **Auswahl Statistics:**
  - Durch Wahl von **Descriptives** ist die Ausgabe folgender Statistiken festgelegt: arithmetisches Mittel, Median, um 5% getrimmtes arithm. Mittel (d.h. die 5% kleinsten bzw. größten Werte werden bei der Berechnung nicht berücksichtigt), Standardfehler, Varianz, Standardabweichung, Minimum, Maximum, Spannweite, IQR, Schiefe, Wölbung, sowie Standardfehler der letzten beiden und ein Konfidenzintervall für den Mittelwert.
  - **M-estimator:** Das arithmetische Mittel als Schätzung der zentralen Tendenz kann durch Ausreißer beeinflusst werden, daher berechnet man die M-Schätzer, wo Ausreißer schwächer als zentrale Werte gewichtet werden.
  - Durch Wahl von **Outliers** werden die Ausreißer untersucht.
  - **Percentiles:** Perzentilwerte für die gesamte Konstellation der Prozentwerte 5, 10, 25, 50, 75, 90 und 95.
- **Auswahl Plots:** Um Diagramme anzeigen zu können, muß im Kontrollfeld **Display** das Optionfeld **Both** oder das Optionfeld **Plots** aktiviert sein.
  - **Boxplot:** Damit dieser in Form einer Grafik aufgebaut wird, **Dependents together** wählen, für mehrere Grafiken: **Factor levels together**.
  - **Histogram:** Gleichzeitig **Histogramm** und bei **Boxplots** das Feld **None** wählen.
- **Überprüfung auf Normalverteilung:** Um dies graphisch zu überprüfen, muß das Kontrollkästchen **Normality plots with test** aktiviert werden. Als Ausgabe erhält man einen QQ-Plot. Ist das Merkmal normalverteilt, so liegen die Wertepaare auf einer Geraden. Wählt man bei **Boxplots** die Auswahl **None** so ergibt sich die gewünschte Ausgabe als Kolmogorov-Smirnov-Test. Um die Annahme zu akzeptieren, muß der ausgegebene Signifikanz-Wert größer als das vorgegebene Testniveau sein.