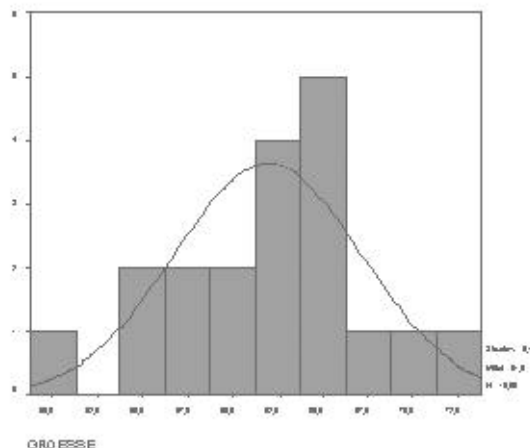




kompakt

© Copyright E.Grüner
Marburg 2001



Inhaltsübersicht

1	Überblick über SPSS	4
1.1	Einsatzzweck und Einordnung.....	4
1.2	Der Funktionsumfang von SPSS	5
1.3	Die wichtigsten SPSS-Prozeduren zur elementaren Statistik im Überblick.....	6
1.3.1	Prozeduren zur univariaten deskriptiven Statistik	6
1.3.2	Prozeduren zur bivariaten deskriptiven Statistik	7
1.3.3	Prozeduren zur Inferenzstatistik.....	7
2	Das Arbeiten mit SPSS unter Windows.....	8
2.1	Ablauf einer statistischen Auswertung mit SPSS	8
2.2	Dateien in SPSS	9
2.3	Die integrierte Entwicklungsumgebung	10
2.3.1	Die Elemente der graphischen Benutzeroberfläche	10
2.3.2	Die Entwicklungswerkzeuge.....	12
2.4	Arbeitsmodi in SPSS.....	13
2.5	Die Sprachelemente von SPSS und deren Syntax	14
2.6	Eingebaute Funktionen	17
2.7	Programmierung in SPSS	19
2.7.1	Praktische Hinweise für die Erstellung eines SPSS-Programms	19
2.7.2	Aufbau und Arbeitsweise eines SPSS-Programms	20
2.7.3	Die Syntax der SPSS-Befehlssprache	21
2.7.4	Kommentare in SPSS-Programmen (COMMENT).....	24
2.7.5	SPSS-Programm abbrechen (FINISH)	25

2.7.6	SPSS-Programmdatei einfügen (INCLUDE)	26
3	Datenmanagement in SPSS	27
3.1	Variablen in SPSS.....	27
3.1.1	Attribute von Variablen	27
	<i>Variablenamen</i>	28
	<i>Variablentyp</i>	28
	<i>Variablenlabels</i>	28
	<i>Wertelabels</i>	29
	<i>Fehlende Werte (Missing Werte)</i>	30
	<i>Spaltenformat</i>	30
	<i>Ausgabeformat</i>	31
	<i>Meßniveau</i>	32
3.1.2	Die SPSS-Datenformate.....	33
3.1.3	Systemvariablen	36
3.1.4	Arbeitsvariablen.....	36
3.1.5	Vektoren.....	37
3.1.6	Deklaration von Variablen und Vektoren	37
3.2	Ein- und Ausgabe.....	39
3.2.1	Der Dateneditor von SPSS für Windows	39
3.2.2	Ein- und Ausgabe von Rohdaten im SPSS-Format bzw. in Fremdformaten (GET, SAVE, GET TRANSLATE, SAVE TRANSLATE, ...)	40
3.2.3	Eingabe von ASCII-Daten (DATA LIST, BEGIN/END DATA).....	44
3.2.4	Eingabe von ASCII-Daten mit komplexer Struktur (INPUT PROGRAM RECORD TYPE,...).....	46
3.2.5	Erzeugung von Rohdaten (INPUT PROGRAM, CREATE)	46
3.2.6	Dateien zusammenfügen (ADD FILES, MATCH FILES).....	49
3.2.7	Löschen der Arbeitsdatei (NEW FILE)	52
3.2.8	Löschen einer externen Datei (ERASE)	53
3.2.9	Ausgabe von Daten im ASCII-Format (WRITE, PRINT..., LIST)	53
3.2.10	Definition von Kopfzeilen für die Ausgabe (TITLE, SUBTITLE)	57
3.3	Variablentransformationen.....	58
3.3.1	Umkodieren (RECODE, AUTORECODE)	58
3.3.2	Berechnen (COMPUTE)	62
3.3.3	Zählen (COUNT)	63
3.3.4	Bedingte Transformationen (IF, DO IF)	65
3.3.5	Temporäre Transformationen (TEMPORARY)	66
3.4	Aufbereitung der Analytestichprobe	67
3.4.1	Gewichten von Fällen (WEIGHT)	67
3.4.2	Sortieren von Fällen (SORT CASES)	68
3.4.3	Aufteilen der Fälle (SPLIT FILE)	69
3.4.4	Auswählen von Fällen (SELECT IF, SAMPLE, N OF CASES, FILTER, USE)	70
3.4.5	Arbeitsdatei transponieren (FLIP).....	75
3.5	Kontrollstrukturen in SPSS-Programmen.....	76

3.5.1	Offene Transformationen ausführen (EXECUTE)	76
3.5.2	Bedingte Blöcke (DO IF)	77
3.5.3	Schleifen (DO REPEAT, LOOP, BREAK)	78
4	Statistikprozeduren	82
4.1	Univariate deskriptive Statistik	82
4.1.1	Prozedur "Häufigkeiten" (FREQUENCIES)	83
4.1.2	Prozedur "Deskriptive Statistiken" (DESCRIPTIVES)	86
4.1.3	Prozedur "Explorative Datenanalyse" (EXAMINE)	88
4.1.4	Prozedur "Fälle zusammenfassen" (SUMMARIZE)	91
4.1.5	Prozedur "Aggregieren" (AGGREGATE)	94
4.2	Bivariate deskriptive Statistik	96
4.2.1	Prozedur "Kreuztabellen" (CROSSTABS)	96
4.2.2	Prozedur "Bivariate Korrelationen" (CORRELATIONS, NONPAR CORR)	99
4.2.3	Prozedur "Partielle Korrelation" (PARTIAL CORR)	102
4.2.4	Prozedur "Regression" (REGRESSION)	104
4.3	Inferenzstatistik	109
4.3.1	Prozedur "Nichtparametrische Tests" (NPAR TESTS)	109
4.3.2	Prozeduren für t-Tests (TTEST)	113
4.3.3	Prozedur "Mittelwerte" (MEANS)	116
4.3.4	Prozedur "Einfaktorielle Varianzanalyse" (ONEWAY)	118
4.3.5	Prozeduren für die mehrfaktorielle Varianzanalyse (UNIANOVA, ANOVA)	120
5	Graphik mit SPSS	126
5.1	Die SPSS-Standard-Graphikprozeduren	126
5.1.1	Balkendiagramme	127
5.1.2	Liniendiagramme	130
5.1.3	Flächendiagramme	133
5.1.4	Kreisdiagramme	135
5.1.5	Boxplots	137
5.1.6	Fehlerbalkendiagramme	138
5.1.7	Histogramme	140
5.1.8	Verteilungsplots	141
5.1.9	Streudiagramme	143
5.1.10	Die GRAPH-Prozedur	145
5.1.11	Die PLOT-Prozedur	148
5.2	Die Bearbeitung von Standard-Graphiken	150
5.3	Interaktive Graphiken	156
5.3.1	Die IGRAPH-Prozedur	156
6	Literaturhinweise	162
	Index	163

1 Überblick über SPSS

1.1 Einsatzzweck und Einordnung

- **Name und Zweck**

Ursprünglich war "SPSS" eine Abkürzung für "Statistical Package for the Social Sciences". Heute steht der Name angeblich für "Superior Performance Software System".

SPSS ist ein allgemeines und vielseitiges integriertes System zur statistischen Datenanalyse. Als solches bietet es viele Vorteile gegenüber einer losen Sammlung von einzelnen speziellen Statistikprogrammen, nämlich vor allem Einheitlichkeit in der Bedienung, leichtere Erlernbarkeit und Ökonomie.

- **Geschichte**

SPSS erschien erstmalig im Jahre 1970. In den folgenden Jahren gab es viele Versionen (Releases): SPSS,..., SPSS8, SPSS9. Anschließend kam SPSSX mit mehreren Versionen auf den Markt. Danach hieß das Produkt wieder SPSS. Die aktuelle Version für Windows 3.x ist die Version 6.1.3 und für Windows 95/98/NT die Version 10 (Stand: September 2000)¹.

- **Leistungsumfang**

SPSS enthält ein breites Spektrum an Routinen und Werkzeugen zur statistischen Analyse und graphischen Darstellung von Daten. In der Windows-Version verfügt es über eine graphische Benutzeroberfläche, eine integrierte Entwicklungsumgebung und eine Reihe von Online-Hilfen.

- **Verbreitung**

SPSS ist ein weltweit verbreitetes Produkt, das für verschiedene Rechnerplattformen verfügbar ist. SPSS wird in vielen Anwendungsbereichen eingesetzt. Vor allem an den Hochschulen wird es in verstärktem Maße in Lehre und Forschung verwendet.

- **Alternativen zu SPSS**

Zu SPSS gibt es auch etliche Konkurrenzprodukte wie SAS, SYSTAT, BMDP, STATISTICA, S-PLUS, STATGRAPHICS u.a. Diese Produkte unterscheiden sich von SPSS zum Teil erheblich in Bezug auf Leistungsumfang, Anwendungsschwerpunkt und Bedienung.

¹ Im folgenden wird SPSS Version 10 (deutsch) zugrunde gelegt.

1.2 Der Funktionsumfang von SPSS

SPSS besteht aus verschiedenen Programmpaketen, den sog. Moduln, die einzeln zu lizenzieren sind: Base, Advanced Models, Regression Models, Categories, Trends, Conjoint u.a. Im folgenden soll nur das Base System und von den darin enthaltenen Statistikprozeduren nur ein Teil berücksichtigt werden. Die anderen Moduln enthalten Verfahren zur multivariaten Statistik (z.B. Advanced Models und Regression Models) oder statistische Verfahren für spezielle Anwendungen (z.B. Trends).

Das Modul SPSS Base ist das Grundmodul, auf dem alle anderen Moduln aufbauen. Es stellt die Benutzungsoberfläche mit allen Tools und Funktionen zum Arbeiten mit SPSS bereit. Weiterhin enthält es alle notwendigen Hilfsmittel zur Definition von Rohdaten und ein reichhaltiges Inventar an Funktionen und Werkzeugen zum Datenmanagement. Außerdem sind darin noch eine Reihe "schlüsselfertiger" Programme (sog. Prozeduren) für verschiedene statistische Verfahren aus der deskriptiven Statistik und der Inferenzstatistik für den univariaten Fall enthalten. Schließlich umfaßt das Base Modul neben den erwähnten Statistikprozeduren noch die SPSS-Graphikprozeduren. Bei diesen unterscheidet man zwischen Standardgraphiken und interaktiven Graphiken.

Für die Definition der Rohdaten stehen dem Benutzer mehrere Möglichkeiten zur Verfügung: Eingabe von Daten über die Rechnerastatur mit Hilfe des SPSS-Dateneditors, Einlesen von Daten aus einer oder mehreren Dateien, Erzeugung von Daten innerhalb von SPSS. Die in SPSS vorliegenden Daten können auf eine Datei abgespeichert werden. In Normalfall erfolgt die Ein-/Ausgabe von Daten binär, d.h. im internen SPSS-Format. SPSS kann aber auch mit einigen Standardformaten wie ASCII, Excel, dBASE oder Lotus 1-2-3 umgehen. Außerdem werden auch Schnittstellen zu modernen Datenbanksystemen zur Verfügung gestellt.

In SPSS sind verschiedene Datentypen erlaubt: numerische Daten, Textdaten, Datumswerte u.a. Auch verschiedene Skalenniveaus werden in SPSS unterschieden: Nominaldaten, kategoriale Daten, metrische Daten. Bei unvollständigen Daten stehen dem Benutzer verschiedene Strategien zur Behandlung der fehlenden Werte zur Verfügung.

SPSS bietet eine Reihe von Möglichkeiten zum Datenmanagement wie das Sortieren, das Aufteilen und das Auswählen von Datensätzen, das Umkodieren von Variablen, das Berechnen neuer Variablen, das Zählen von Variablenwerten u.a.

Die graphische Benutzungsoberfläche von SPSS eröffnet dem Benutzer die Möglichkeit, eine statistische Auswertung ausschließlich interaktiv über Menüs und Dialogfenster vorzunehmen. Die meisten Funktionen und Prozeduren sind auf diese Weise erreichbar. Darüber hinaus bietet SPSS dem Benutzer die Möglichkeit, eine statistische Auswertung im "Batch-Betrieb" über ein SPSS-Syntaxprogramm durchzuführen. Eine solche programmgesteuerte Arbeitsweise bietet gegenüber dem rein interaktiven Vorgehen einige entscheidende Vorteile. Beide Vorgehensarten lassen sich auch beliebig mischen: man kann in der gleichen Sitzung einige Teile der Auswertung interaktiv und andere Teile programmgesteuert durchführen.

Neben der Verwendung von Syntaxprogrammen läßt sich in SPSS ein weitgehende Automatisierung der gesamten Auswertung im Sinne eines Produktionslaufs erreichen durch Einsatz von sog. Skripts. Die dafür zu verwendende BASIC-ähnliche Skriptsprache wird jedoch im folgenden nicht behandelt.

Abgerundet wird der Leistungsumfang von SPSS durch ein reichhaltiges Angebot an Online-Hilfen wie Direkthilfen, Howtos, Statistik-Assistent, Glossar, Tutorials und Syntaxhandbuch.

1.3 Die wichtigsten SPSS-Prozeduren zur elementaren Statistik im Überblick

1.3.1 Prozeduren zur univariaten deskriptiven Statistik

- **Zweck**

Beschreibung der Verteilung einer einzelnen Variablen; dies kann erfolgen:

- mit Hilfe einer empirischen Häufigkeitstabelle
- mit Hilfe von Maßzahlen (sog. "Statistiken")
 - Lagemaße
 - Streuungsmaße
 - andere Verteilungsmaße (z.B. Maße für Symmetrie und Breitigipfligkeit)
- mit Hilfe eines Diagramms.

- **Statistik-Prozeduren**

- Prozedur "Univariate Statistiken" (DESCRIPTIVES)
- Prozedur "Häufigkeiten" (FREQUENCIES)
- Prozedur "Explorative Datenanalyse" (EXAMINE)
- Prozedur "Fälle auflisten" (SUMMARIZE)

- **Standardgraphik-Prozeduren**

- Prozedur "Balkendiagramme" (GRAPH /BAR)
- Prozedur "Liniendiagramme" (GRAPH /LINE)
- Prozedur "Flächendiagramme" (GRAPH /AREA)
- Prozedur "Kreisdiagramme" (GRAPH /PIE)
- Prozedur "Boxplots" bzw. Prozedur "Explorative Datenanalyse" (EXAMINE)
- Prozedur "Fehlerbalkendiagramme" (GRAPH /ERRORBAR)
- Prozedur "Histogramme" (GRAPH /HISTOGRAM)
- Prozeduren "P-P-Diagramme" und "Q-Q-Diagramme" (PLOT)

- **Prozedur "Interaktive Graphiken"**

- Balkendiagramme (IGRAPH /BAR)
- Liniendiagramme (IGRAPH /LINE)
- Flächendiagramme (IGRAPH /AREA)
- Kreisdiagramme (IGRAPH /PIE)
- Boxplots (IGRAPH /BOX)
- Fehlerbalkendiagramme (IGRAPH /ERRORBAR)
- Histogramme (IGRAPH /HISTOGRAM)

1.3.2 Prozeduren zur bivariaten deskriptiven Statistik

- **Zweck**

Beschreibung der gemeinsamen empirischen Verteilung zweier Variablen; dies kann geschehen:

- mit Hilfe einer zweidimensionalen Häufigkeitstabelle
- mit Hilfe einzelner Maßzahlen (sog. "Statistiken")
- mit Hilfe eines Diagramms.

- **Verfügbare Statistiken**

- Anweichungsproduktsummen
- Kovarianzen
- Korrelationen

- **Statistik-Prozeduren**

- Prozedur "Kreuztabellen" (CROSSTABS)
- Prozedur "Korrelation" (CORRELATIONS, NONPAR CORR)
- Prozedur "Regression" (REGRESSION)

- **Standardgraphik-Prozedur**

- Prozedur "Streudiagramme" (GRAPH /SCATTERPLOT)

- **Prozedur "Interaktive Graphiken"**

- Streuungsdiagramme (IGRAPH /SCATTER)
- Anpassungsliniendiagramme (IGRAPH /FITLINE)

1.3.3 Prozeduren zur Inferenzstatistik

- **Zweck**

- Testen von Hypothesen

- **Im allgemeinen verfügbare Statistiken**

- Deskriptive Statistiken
- Teststatistiken mit P-Wert

- **Statistik-Prozeduren**

- Prozedur "Nichtparametrische Tests" (NONPAR TESTS)
- Prozedur "T-Tests" (T-TEST)
- Prozedur "Einfaktorielle ANOVA" (ONEWAY)
- Prozedur "Mittelwerte" (MEANS)
- Prozedur "GLM - Allgemein mehrfaktoriell" (ANOVA, UNIANOVA, GLM)

2 Das Arbeiten mit SPSS unter Windows

2.1 Ablauf einer statistischen Auswertung mit SPSS

In der Statistik betrachtet man das Auftreten von Merkmalen oder Variablen (z.B. Geschlecht, Alter, Schulbildung) bei einer Menge von Beobachtungseinheiten oder Fällen (z.B. Personen). Die erfaßten Datenwerte, die sog. Rohdaten, werden häufig in Rechteckform als Tabelle angeordnet, deren Zeilen standardmäßig die Fälle und deren Spalten die Variablen darstellen. Eine solche Datentabelle bezeichnet man üblicherweise als Datenmatrix.

Bei jeder statistischen Auswertung mit SPSS gilt es zunächst, eine Repräsentation der Daten aufzubauen. Diese bildet die Grundlage für alle nachfolgenden Schritte, in denen es z.B. um den Einsatz statistischer Verfahren geht. Bei der praktischen Arbeit mit SPSS lassen sich demnach wie bei vielen Vorgängen in der EDV grob drei Phasen unterscheiden: Eingabe, Verarbeitung, Ausgabe.

- **Dateneingabe**

Der erste Schritt bei der Arbeit mit SPSS ist die Definition der Rohdaten. Diese werden entweder "von Hand" direkt in SPSS eingegeben oder aus einer bereits vorhandenen Datendatei eingelesen. Diese Datei muß entweder im SPSS-Datenformat, als ASCII-Datei oder in einem von SPSS lesbaren Fremdformat (wie z.B. als Excel- oder dBASE-Datei) vorliegen. Eine weitere Möglichkeit zur Datendefinition, die vor allem bei Simulationsstudien Verwendung findet, besteht darin, die Daten durch SPSS (z.B. mit Hilfe eines Zufallszahlengenerators) erzeugen zu lassen.

Nachdem die Rohdaten vorliegen, können sie bei Bedarf noch editiert werden, d.h. es können weitere Daten hinzugefügt oder fehlerhafte Werte korrigiert werden. Nach einer Neueingabe der Daten und auch nach Editierarbeiten sollten die Daten in einer externen Datendatei abgespeichert werden, um sie in künftigen SPSS-Sitzungen wieder verfügbar zu haben.

- **Statistische Verarbeitung**

Erst wenn die Rohdaten in SPSS definiert sind, kann im zweiten Schritt mit Hilfe der von SPSS bereitgestellten Statistik- und Graphik-Prozeduren die gewünschte statistische Auswertung folgen.

Häufig ist es allerdings notwendig, die Rohdaten vorher noch einer Vorverarbeitung (z.B. irgendwelchen Transformationen) zu unterziehen. Beispielsweise möchte der Benutzer aus den erhobenen Originaldaten zunächst weitere Variablen erzeugen, etwa einen Test auswerten oder irgendwelche Indizes berechnen, um diese anschließend statistisch auszuwerten.

- **Ausgabe der Ergebnisse**

Die Ergebnisse liegen im allgemeinen in Form von Tabellen, Listings im Textformat oder Diagrammen vor. Diese können auf einem Drucker ausgegeben oder z.B. gleich in ein Textverarbeitungsprogramm übernommen werden.

2.2 Dateien in SPSS

Die Eingabe von Daten und die Ausgabe von Ergebnissen geschieht in SPSS generell über Dateien bzw. - wie bei graphischen Benutzungsoberflächen üblich - über Fenster. Auch die Steuerung des Auswertungsprozesses kann über ein Fenster bzw. über eine Datei erfolgen. Einige für das Arbeiten mit SPSS wichtige Dateien seien im folgenden aufgeführt:

- **Arbeitsdatei**

Die Arbeitsdatei ist die zentrale Datenablage von SPSS. Hier werden die eingegebenen, eingelesenen oder erzeugten Daten von SPSS intern gespeichert. Es ist dies eine temporäre Datei, die bei Beendigung von SPSS wieder gelöscht wird. Der Benutzer hat auf diese Datei außerhalb von SPSS keinen Zugriff. Er kann jedoch den Inhalt der Arbeitsdatei auf eine externe Datendatei abspeichern. Damit kann die Arbeitsdatei in einer späteren SPSS-Sitzung wieder restauriert und für weitere Auswertungsschritte verwendet werden.

Die Dimensionen der Arbeitsdatei werden von der Anzahl der Fälle und der Variablen bestimmt. Jeder Fall in der Arbeitsdatei enthält dieselben Variablen in derselben Reihenfolge. Die Reihenfolge der Variablen und der Fälle wird bestimmt durch die zeitliche Reihenfolge, in der sie eingegeben bzw. erzeugt wurden.

Bei SPSS für Windows ist die Arbeitsdatei im Datenfenster (Dateneditorfenster) repräsentiert.

- **Externe Datendatei**

Die Rohdaten werden üblicherweise von einer externen Datendatei eingelesen. Diese liegt entweder als binäre SPSS-Datendatei (der Name ist dann z.B. 'beispiel.sav') oder als ASCII-Datendatei (z.B. 'beispiel.dat') vor. Darüber hinaus kann SPSS auch mit einigen "Fremdformaten" (z.B. Excel, dBASE) umgehen.

Eine externe Datendatei kann auch fehlen. Beispielsweise können die Daten als Teil des SPSS-Programms vorliegen. Man spricht dann von Inline-Daten. Die Möglichkeit, ein SPSS-Programm zusammen mit den dazugehörigen Daten in einer einzigen Datei abzuspeichern, sollte allerdings nur bei überschaubaren Datenmengen (z.B. bei Demonstrationsbeispielen) verwendet werden. Weiterhin wird eine externe Datendatei normalerweise fehlen, wenn die Daten in SPSS generiert werden.

- **Ergebnisausgabe (Tabellen-/Text-/Graphikausgabe)**

Bei jeder SPSS-Sitzung wird normalerweise eine Ausgabe erzeugt. Neben einem Protokoll der Systemmeldungen erscheinen hier vor allem die Ausgaben der aufgerufenen Statistikprozeduren. Diese Ausgaben können entweder als Textausgabe oder in Tabellenform (in Form von sog. Pivot-Tabellen) vorliegen.

Bei SPSS für Windows erscheint die Ausgabe im Ausgabefenster. Auch die Ausgabe von Graphikprozeduren erscheint dort. Der Inhalt des Ausgabefensters kann mit den Ausgabe-Viewern betrachtet, in gewissem Umfang modifiziert, ausgedruckt oder auf eine externe Datei abgespeichert werden (z.B. unter dem Namen 'beispiel.spo').

- **Externe ASCII-Ausgabedateien**

Neben der Standard-Ergebnisausgabe kann der Benutzer noch weitere Ausgaben erzeugen, nämlich externe ASCII-Dateien. Eine solche Datei könnte dann als Eingabedatendatei für ein anderes Programm oder für SPSS selbst benutzt werden.

- **SPSS-Programm-Datei**

Für das programmgesteuerte Arbeiten werden SPSS-Programmdateien verwendet. Bei SPSS für Windows gibt der Benutzer sein SPSS-Programm in ein SPSS-Syntaxfenster ein oder liest es dort von einer externen Programmdatei ein. Dort kann es editiert und von dort gestartet werden. Im Normalfall wird man anschließend den Inhalt des Syntaxfensters als SPSS-Programmdatei (z.B. unter dem Namen 'beispiel.sps') abspeichern. SPSS-Programmdateien haben ASCII-Format, d.h. man könnte sie auch außerhalb von SPSS mit jedem beliebigen Texteditor bearbeiten.

2.3 Die integrierte Entwicklungsumgebung

Unter Windows steht dem SPSS-Benutzer eine integrierte Entwicklungsumgebung zur Verfügung. Integrierte Entwicklungsumgebungen sind heute vor allem bei den höheren Programmiersprachen weit verbreitet. Sie umfassen alle Werkzeuge, die für die Programmentwicklung benötigt werden, so daß der Programmierer die Entwicklungsumgebung nicht verlassen muß, um beispielsweise ein Programm einzugeben, zu starten oder zu korrigieren. Außerdem sind die Programmierwerkzeuge (z.B. der Editor) häufig auf die jeweilige Programmiersprache optimiert.

Bei SPSS für Windows sind Tätigkeiten wie Dateneingabe, Prozedurstart, Ausgabe von Listings und Diagrammen, Speichern, Drucken, sowie das Editieren und Starten von SPSS-Programmen innerhalb der integrierten Entwicklungsumgebung durchführbar. Die graphische Benutzungsoberfläche erlaubt eine weitgehend intuitive Arbeitsweise mit den bei Windows üblichen Techniken und Hilfsmitteln. So stehen dem Benutzer für die Kommunikation mit dem SPSS-System einerseits verschiedene graphische Bedienungselemente wie Fenster, Menüleisten, Symbolleisten, Statusleiste, Dialogfelder zur Verfügung, andererseits auch die nötigen Werkzeuge wie Dateneditor, Programmeditor, Viewer, Textviewer, Diagrammeditor u.a

2.3.1 Die Elemente der graphischen Benutzungsoberfläche

Fenster

- Datenfenster

Im Datenfenster wird der Inhalt der Arbeitsdatei angezeigt. Dieses Fenster wird zu Beginn einer SPSS-Sitzung automatisch geöffnet.

Es gibt nur ein Datenfenster, also nur eine Arbeitsdatei, d.h. man kann in SPSS z.B. nicht mehrere Datendateien gleichzeitig öffnen.

Das Datenfenster enthält zwei Registerkarten: die Datenansicht und die Variablenansicht. Beide sehen dem Anwendungsfenster eines Tabellenkalkulationsprogramms ähnlich.

Die Datenansicht stellt eine Tabelle dar, deren Zeilen die Fälle, deren Spalten die Variablen und deren Zellen² die Datenwerte darstellen. Alle Zeilen der Arbeitsdatei enthalten die gleiche Anzahl von Zellen. Die Dimensionen der Arbeitsdatei werden von der Anzahl der Fälle und der Anzahl der Variablen bestimmt. Bei numerischen Variablen gelten leere Datenwerte als sog. systemdefinierte fehlende Werte und werden durch ein alleinstehendes Dezimaltrennzeichen dargestellt. Bei String-Variablen gelten leere Felder als gültiger Wert.

Die Registerkarte Variablenansicht zeigt eine Tabelle, deren Zeilen die definierten Variablen und deren Spalten die Variablenattribute darstellen.

- Ausgabefenster

Alle statistischen Ergebnisse, Tabellen und Diagramme werden im Ausgabefenster angezeigt. Dieses Fenster wird automatisch geöffnet, wenn das erste Mal eine Prozedur aufgerufen wird, die eine Ausgabe erzeugt, oder wenn z.B. eine Fehlermeldung ausgegeben wird.

Das Ausgabefenster ist in zwei Fensterbereiche aufgeteilt: der linke Bereich enthält eine Gliederungsansicht des Inhalts, der rechte die eigentliche Ausgabe mit den Tabellen, Textausgaben und Diagrammen.

- Diagrammfenster

Dieses Fenster dient zur Bearbeitung von Standardgraphiken.

- Syntaxfenster

Dieses Fenster erlaubt die Eingabe, die Bearbeitung und den Start von SPSS-Programmen. Ein solches Fenster wird automatisch kreiert, wenn der Benutzer eine SPSS-Programmdatei öffnet. In einer SPSS-Sitzung können mehrere Syntaxfenster gleichzeitig geöffnet sein.

² Eine Zelle ist der Schnittpunkt einer Zeile mit einer Spalte.

- Skriptfenster

Dieses Fenster dient zur Eingabe und Bearbeitung von Skripten mit Hilfe des Skript-Editors.

Menüs

- Jedes der oben erwähnten Fenster besitzt eine spezielle Menüleiste, über die der Benutzer die Aktionen auswählen kann, die SPSS ausführen soll.
- Die Menüs "Analysieren" und "Grafiken" sind in allen Fenstern verfügbar. Dadurch muß der Benutzer für den Aufruf von Prozeduren nicht immer zwischen verschiedenen Fenstern umschalten.

Symbolleisten

- Jedes Fenster von SPSS verfügt über eine eigene Symbolleiste, mit der schnell und einfach auf häufig genutzte Funktionen zugegriffen werden kann. Einige Fenster verfügen über mehrere Symbolleisten.
- Als Hilfe wird zu den Symbolen in der Symbolleiste auch jeweils eine Quickinfo angezeigt.
- Der Benutzer hat die Möglichkeit, die Anzeige, das Aussehen und den Inhalt der Symbolleisten seinen Bedürfnissen anzupassen.

Statusleiste

- Die Statusleiste befindet sich am unteren Rand eines Fensters von SPSS.
- In ihr wird der Status des Systems angezeigt, z.B. Informationen über den im Moment aktiven Befehl. Außerdem kann man in der Statusleiste sehen, ob gerade eine Filterung, eine Gewichtung oder eine Aufteilung der Fälle eingeschaltet ist.

Dialogfelder

- Wenn der Benutzer eine Menüwahl trifft, wird in den meisten Fällen ein Dialogfeld (Dialogfenster) geöffnet. Darin sind die näheren Spezifikationen zu den ausgewählten Aktionen vorzunehmen (z.B. Auswählen von Variablen und Optionen für die Analyse).
- Die Hauptdialogfenster der SPSS-Prozeduren enthalten jeweils die folgenden Grundelemente:
 - Variablenliste: dies ist die Liste der Variablen in der Arbeitsdatei und dient zur Spezifikation der für die Analyse zu verwendenden Variablen (abhängige Variablen, unabhängige Variablen, usw.). In dieser Liste werden nur jene Variablen angezeigt, deren Typ für die ausgewählte Aktion zulässig ist (z.B. können String-Variablen in vielen Prozeduren nur eingeschränkt verwendet werden). Die Auswahl erfolgt jeweils durch Selektion der gewünschten Variable(n) aus der Variablenliste und anschließender Benutzung der geeigneten Pfeilschaltfläche.
 - Befehlsschaltflächen: mit diesen veranlaßt man SPSS beispielsweise, eine Prozedur auszuführen, die Online-Hilfe anzuzeigen oder ein untergeordnetes Dialogfenster für zusätzliche Angaben zu öffnen.
- Informationen zu einem Steuerelement in einem Dialogfenster erhält man, wenn man mit der Maus auf dieses Steuerelement zeigt und die rechte Maustaste drückt.

2.3.2 Die Entwicklungswerkzeuge

Dateneditor

- Der Dateneditor erlaubt die interaktive Eingabe und Bearbeitung der Arbeitsdatei im Datenfenster, sowie das Abspeichern des Inhalts der Arbeitsdatei auf eine externe Datendatei.
- In alle Zellen der Datentabelle können Daten eingegeben werden, allerdings nur Rohdaten und keine Formeln wie bei einem Tabellenkalkulationsprogramm. Wenn Daten in eine Zelle außerhalb der Grenzen der definierten Arbeitsdatei eingegeben werden, füllt SPSS die Tabelle entsprechend mit leeren Feldern bzw. systemdefinierten fehlenden Werten (SYSMIS, s.u.) auf.
- Neben der interaktiven Eingabe von Daten gibt es noch die Möglichkeit, sie von einer externen Datendatei einzulesen.

Viewer

- Der Viewer dient vor allem dazu, den Inhalt des Ausgabefensters anzusehen, zu bearbeiten, zu drucken und zu speichern.
- Zum leichteren Navigieren innerhalb der Ausgabe kann das Gliederungsfenster benutzt werden, um z.B. direkt zu bestimmten Tabellen oder Diagrammen zu springen. Mit Hilfe der Gliederung kann man auch ausgewählte Tabellen und Diagramme ein- oder ausblenden oder deren Anzeigereihenfolge ändern und Objekte zwischen dem Viewer und anderen Anwendungen verschieben.

Textviewer, Textausgabe-Editor

- Die Ausgabe von Statistikprozeduren kann entweder in Form von sog. Pivot-Tabellen oder als einfache Textausgabe erfolgen.
- Das Tool zum Anzeigen und Bearbeiten (z.B. Ändern der Schriftmerkmale) einer Textausgabe ist der Text-Viewer, bzw. der Textausgabe-Editor.

Diagrammeditor

- Mit Hilfe des Diagrammeditors kann eine Standardgraphik in einem Diagrammfenster bearbeitet werden.
- Beispielsweise können hier die Attribute von Texten, Achsen, Linien, Balken usw. geändert und sogar der Diagrammtyp nachträglich noch gewechselt werden.

Syntaxeditor

- Der Syntax-Editor wird für die Eingabe, die Bearbeitung und den Start von SPSS-Programmen verwendet. Außerdem kann ein SPSS-Programm in einer Datei gespeichert oder von einer Datei in ein Syntaxfenster geladen werden.

Skripteditor

- Mit dem Skripteditor kann man Skripts erstellen und bearbeiten. Solche Skripts können dazu verwendet werden, viele Aufgaben in SPSS an die eigenen Bedürfnisse anzupassen und zu automatisieren.

2.4 Arbeitsmodi in SPSS

- Prinzipiell kann eine statistische Auswertung mit SPSS unter Windows in zwei verschiedenen Modi durchgeführt werden:
 - Interaktiver Modus
Hier werden die unter Windows üblichen Elemente (Menüs, Symbolleisten, Dialogfenster) und Techniken (z.B. Maustechniken) verwendet, was eine weitgehend intuitive Benutzung erlaubt. Auch der Anfänger kommt nach kurzer Einarbeitungszeit zurecht, sofern er die nötigen Statistikkennnisse mitbringt.
 - Programmgesteuerter Modus
Hier werden die Anweisungen an SPSS in Form eines Programms formuliert, wobei die Syntax und Semantik der SPSS-Befehle zu beachten ist (hier ist im allgemeinen die Zuhilfenahme eines Handbuchs angebracht). Es gibt eine Reihe von Vorteilen, die für diese Arbeitsweise sprechen: ein Programm ist eine konzise Formulierung einer Problemlösung, es ist selbstdokumentierend und kann abgespeichert und wiederverwendet werden. Außerdem stehen hier dem Benutzer bei komplexeren Auswertungsproblemen wesentlich mehr Möglichkeiten zur Verfügung als bei der interaktiven Arbeitsweise.³
- Auf die meisten SPSS-Befehle kann man interaktiv, d.h. über Menüs und Dialogfenster, zugreifen. Einige Befehle und Optionen jedoch sind nur in der SPSS-Syntaxbefehlssprache verfügbar. Mit dieser verfügt man außerdem über die Möglichkeit, Jobs in einer Programmdatei zu speichern, um die Analyse zu einem späteren Zeitpunkt eventuell mit leicht geänderten Spezifikationen zu wiederholen oder diese automatisch im Produktionsmodus ausführen lassen.
- Auch wenn man mit SPSS rein interaktiv arbeitet, die Auswertung von Daten hat immer in mehreren logisch aufeinanderfolgenden Schritten zu erfolgen:
 - An erster Stelle steht immer die Definition der Daten, d.h. die Eingabe der Daten in die Arbeitsdatei. Dies geschieht entweder durch direkte Dateneingabe oder durch Übertragen der Daten aus einer gespeicherten Datendatei.
 - Erst wenn Daten definiert sind, kann über die Menüleiste des Datenfensters oder des Ausgabefensters die gewünschte Prozedur, Datentransformation usw. aufgerufen werden. Zu deren Ausführung sind im allgemeinen noch einige Spezifikationen erforderlich, die von SPSS in Dialogfenstern abgefragt werden. Um die Schreibarbeit beim Ausfüllen dieser Dialogfenster klein zu halten, gibt es einige Erleichterungen: z.B. erfolgt die Angabe der zu verwendenden Variablen im allgemeinen nicht durch Eingabe der Variablennamen über die Rechnertastatur, sondern komfortabel durch Auswahl aus der dargebotenen Variablenliste. Sobald eine Aufgabe hinreichend spezifiziert ist, kann durch die Befehlstaste [OK] deren Ausführung angestoßen werden.
 - Nach Ausführung der Prozedur kann man die Ergebnisse im Ausgabefenster betrachten, eventuell nachbearbeiten und am Ende beispielsweise ausdrucken oder in ein Textdokument übertragen.

³ Praktische Hinweise für die Erstellung von SPSS-Programmen werden in Abschnitt 2.7 gegeben. Außerdem wird in den späteren Abschnitten neben der interaktiven Bedienung immer auch Syntax und Semantik der entsprechenden Programmbefehle behandelt.

2.5 Die Sprachelemente von SPSS und deren Syntax

Sowohl bei der programmgesteuerten Arbeitsweise als auch bei der interaktiven Arbeitsweise mit SPSS sind für einzelne Elemente wie Variablenamen, Werte oder Ausdrücke bestimmte syntaktische Regeln zu beachten⁴.

Variablenamen

- Ein Variablenname besteht aus einer Folge von 1 bis max. 8 Zeichen.
- Das 1. Zeichen muß ein Buchstabe oder eines der Sonderzeichen @ # \$ sein.
- Die restlichen Zeichen dürfen sein: Buchstaben, Ziffern, gewisse Sonderzeichen (. @ # \$ _).
- SPSS kennt noch spezielle Variablen, die in der SPSS-Steuersprache benutzt werden können:
 - Systemvariablen (siehe 3.1.3): deren Namen beginnen mit dem \$-Zeichen
 - Arbeitsvariablen (siehe 3.1.4): deren Namen beginnen mit dem #-Zeichen
- Manche Schlüsselwörter aus der SPSS-Steuersprache (z.B. "TO") sind reserviert und dürfen als Variablenamen nicht verwendet werden.

Variablenlisten

- Im einfachsten Fall ist eine Variablenliste ein einzelner Variablenname oder eine durch Kommata oder Leerzeichen getrennte Folge von einzelnen Variablenamen.
- Je nachdem, ob man Variablen erstmalig im Programm definiert oder auf schon definierte Variablen Bezug nimmt, unterscheidet man zwischen definierenden und reflexiven (referenzierenden) Variablenlisten.
- Bei definierenden Variablenlisten ist die Verwendung von numerierten Variablen in der Form "*aaamm to aaann*" möglich. *aaa* steht für eine Folge von einem oder mehreren Buchstaben und *mmm* bzw. *nnn* für ein- oder mehrstellige natürliche Zahlen mit *mmm < nnn*. SPSS erzeugt in diesem Falle die dazwischen liegenden Variablenamen automatisch. So bedeutet z.B. der Ausdruck FRAGE8 TO FRAGE11 die Variablenliste FRAGE8, FRAGE9, FRAGE10, FRAGE11.
- Bei reflexiven Variablenlisten ist immer auch das Schlüsselwort **to** möglich (dies gilt nicht nur für numerierte Variablen). Da bei SPSS die Variablen gemäß ihrem erstmaligen Auftreten geordnet sind, es also eine "natürliche" Reihenfolge der Variablen gibt, ist es klar, welche Variablen damit referenziert werden..
- Das Schlüsselwort **all** wird benutzt, wenn man alle bisher definierten Variablen referenzieren möchte.

Variablen-Labels

- Erlaubte Zeichen sind alle ASCII-Zeichen, auch Groß-/Kleinschreibung wird unterschieden.
- Die maximale Länge eines Variablenlabels ist 120 Zeichen.

Werte

- Einzelne numerische Werte sind ganze Zahlen oder Dezimalzahlen (z.B. 37, -17.02, 0.24, .24). Dabei beachte man, daß in arithmetischen Ausdrücken als Dezimaltrennzeichen immer der Dezimalpunkt zu verwenden ist, auch wenn bei der deutschen SPSS-Version für die Eingabe der Daten (im Dateneditor oder von einer ASCII-Datei) das Dezimalkomma verwendet werden muß.
- Numerische Intervalle werden angegeben durch *xu thru xo* (*xu* und *xo* sind zwei numerische Werte mit *xu < xo*). Für *xu* bzw. *xo* kann man auch die Schlüsselwörter **lo[west]** (=kleinste Zahl) bzw. **hi[higest]** (=größte Zahl) verwenden.

⁴ Hinweise zur verwendeten Metasprache sind in Abschnitt 2.7.3 zu finden.

- Zeichenketten werden in Hochkommata oder Gänsefüßchen gesetzt (Beispiele: 'Dies ist mein erster Versuch'; "Hoffentlich klappt's !" oder - mit zwei Hochkommata innerhalb der Zeichenkette - 'Hoffentlich klappt's !').
- Für Datum und Uhrzeit existieren eine Reihe verschiedener Datenformate, z.B. für das Datum das Format tt.mm.jjjj (siehe Abschnitt über SPSS-Datenformate in 3.1.1). Allerdings können diese nicht als Konstanten in Ausdrücken verwendet werden.
- Als spezielle Werte sind noch die folgenden Schlüsselwörter definiert:
 - **missing** benutzerdefinierter fehlender Wert (Missing-Wert)
 - **sysmis** systemdefinierter fehlender Wert (System-Missing)

Wertelisten

- Eine Werteliste ist eine durch Kommata oder Leerzeichen getrennte Aufzählung von einzelnen Werten bzw. Intervallen.

Werte-Labels

- Erlaubte Zeichen sind alle ASCII-Zeichen, auch Groß-/Kleinschreibung wird unterschieden.
- Die maximale Länge eines Wertelabels ist 60 Zeichen.

Arithmetische Ausdrücke

- Ein arithmetischer Ausdruck besteht aus beliebig vielen Bestandteilen. Diese können sein:
 - numerische Konstanten
 - Variablennamen
 - Funktionsaufrufe (siehe Abschnitt 2.6)
 - arithmetische Operatoren

+	Addition
-	Subtraktion
*	Multiplikation
/	Division
**	Exponentiation
 - runde Klammern
- Die Definition von arithmetischen Ausdrücken ist rekursiv: als Operanden von arithmetischen Operationen können Konstanten, Variablen, Funktionsaufrufe oder selbst wieder arithmetische Ausdrücke auftreten. Das Ergebnis eines arithmetischen Ausdrucks ist ein numerischer Wert.
- Da logische Werte in SPSS ebenfalls numerisch (mit 1 und 0) codiert sind (s.u.), können in arithmetischen Ausdrücken sogar Vergleichsausdrücke und logische Ausdrücke benutzt werden.
- Mit Hilfe von runden Klammern kann der Benutzer die Auswertungsreihenfolge der Teile des Ausdrucks immer festlegen. Für arithmetische Ausdrücke gilt die folgende Auswertungsreihenfolge:
 1. runde Klammern
 2. Funktionsaufrufe
 3. arithmetische Operatoren; dabei gilt wie in der Mathematik: "Punktrechnung" vor "Strichrechnung" (d.h. 1. Exponentiation, 2. Multiplikation und Division, 3. Addition und Subtraktion)
 4. logische Operatoren (hier gilt die Reihenfolge: 1. **and**, 2. **or**, 3. **not**)

Vergleichsausdrücke

- Der einfache Größenvergleich zweier numerischer Daten wird mit Hilfe eines Vergleichsausdrucks formuliert. Dieser hat die Form einer Vergleichsoperation.

- Die Operanden einer Vergleichsoperation sind arithmetische Ausdrücke, der Operator ist ein Vergleichsoperator und das Ergebnis ein logischer Wert.
- Vergleichsausdrücke bestehen also aus den folgenden Bestandteilen:
 - arithmetische Ausdrücke
 - Vergleichsoperatoren

eq	oder	=		gleich	
ne	oder	<>	oder	~=	ungleich
lt	oder	<		kleiner als	
gt	oder	>		größer als	
le	oder	<=		kleiner gleich	
ge	oder	>=		größer gleich	

Logische Ausdrücke

- Ein logischer Ausdruck besteht aus beliebig vielen Bestandteilen. Diese können sein:
 - logische Konstanten (1 oder 0)
 - Variablennamen
 - Vergleichsausdrücke
 - Funktionsaufrufe (siehe Abschnitt 2.6)
 - logische Operatoren

and	oder	&		logisches UND (Konjunktion)
or	oder			logisches ODER (Disjunktion)
not	oder	~		logisches NOT (Negation)
 - runde Klammern
- Die Definition von logischen Ausdrücken ist rekursiv: als Operanden von logischen Operationen können Konstanten, Variablen, Funktionsaufrufe oder selbst wieder logische Ausdrücke auftreten. Das Ergebnis eines logischen Ausdrucks ist ein logischer Wert.
- Logische Werte werden in SPSS numerisch codiert, u.zw. "wahr" durch 1 und "falsch" durch 0. Als logischen Ausdruck kann man somit jeden arithmetischen Ausdruck auffassen, der nur den Wert 0 oder 1 (oder SYSMIS) annimmt.
- Bei zusammengesetzten logischen Ausdrücken kann der Benutzer mit Hilfe von runden Klammern die Auswertungsreihenfolge immer festlegen. Für logische Ausdrücke gilt die folgende Auswertungsreihenfolge:
 1. runde Klammern
 2. Funktionsaufrufe
 3. Vergleichsausdrücke
 4. logische Operatoren (hier gilt die Reihenfolge: 1. **and**, 2. **or**, 3. **not**)
- Wahrheitstabellen

Seien a und b zwei logische Ausdrücke mit den logischen Werten 1, 0 oder fehlend (dargestellt durch ","), so ergibt sich der Wahrheitswert für die einzelnen logischen Operationen aus den folgenden Tabellen.

a	not a
1	0
0	1
,	,

a and b		b		
		1	0	,
a	1	1	0	,
	0	0	0	0
	,	,	0	,

a or b		b		
		1	0	,
a	1	1	1	1
	0	1	0	,
	,	1	,	,

2.6 Eingebaute Funktionen

SPSS stellt dem Benutzer eine Vielzahl spezieller Funktionen zur Verfügung. Für den Aufruf einer Funktion ist hinter dem Funktionsnamen in runde Klammern das Eingabeargument bzw. die Liste der Eingabeargumente zu setzen. Jede Funktion gibt einen Wert, nämlich den berechneten Funktionswert, zurück.

- Arithmetische Funktionen

abs (x)	Absolutbetrag von x
rnd (x)	gerundeter Wert von x ; bei 0,5 wird aufgerundet
trunc (x)	Wert von x nach Abschneiden der Nachkommastellen
mod (n,m)	Modulofunktion (Divisionsrest bei Division von n durch m)
sqrt (x)	Quadratwurzel von x
exp (x)	Exponentialfunktion von x (zur Basis e)
ln (x)	natürlicher Logarithmus von x
lg10 (x)	dekadischer Logarithmus von x
sin (x)	Sinus von x
cos (x)	Cosinus von x
arsin (x)	Arcussinus von x
artan (x)	Arcustangens von x

- Statistische Funktionen

sum [. n]($x_1,x_2[,...]$)	Summe
mean [. n]($x_1,x_2[,...]$)	Mittelwert
sd [. n]($x_1,x_2[,...]$)	Standardabweichung
var [. n]($x_1,x_2[,...]$)	Varianz
cfvar [. n]($x_1,x_2[,...]$)	Variationskoeffizient
min [. n]($x_1,x_2[,...]$)	Minimum
max [. n]($x_1,x_2[,...]$)	Maximum

Diese Funktionen liefern jeweils eine fallweise Statistik über die angegebenen Variablen. Die Anzahl der Eingabeargumente ist variabel. Als Eingabeargument ist auch eine Variablenliste in der Form x_1 to x_2 zugelassen. Der optionale Parameter n bewirkt, daß als Wert System-Missing (SYSMIS, s.u.) zurückgegeben wird, falls die Anzahl der Variablen mit gültigem Wert kleiner als n ist.

- Verteilungsfunktionen (Auswahl)

cdfnorm (z) , cdf.normal (z)	kumulative Standard-Normalverteilung: die Funktion berechnet die Wahrscheinlichkeit $P(Z \leq z)$
probit (p) , idf.normal (p)	Inverse der kumulativen Standard-NV: die Funktion berechnet den z -Wert, für den $P(Z \leq z) = p$ gilt

- Erzeugung von Zufallszahlen (Auswahl)

rv.uniform (a,b)	Zufallszahl aus einer stetigen Gleichverteilung zwischen a und b
uniform (b)	Zufallszahl aus einer stetigen Gleichverteilung zwischen 0 und b
rv.normal (m,s)	Zufallszahl aus einer Normalverteilung mit Mittelwert m und Standardabweichung s
normal (s)	Zufallszahl aus einer Normalverteilung mit Mittelwert 0 und Standardabweichung s

- Funktionen für fehlende Werte

missing (x)	ergibt 1, falls x Missing-Wert ist; 0 sonst
sysmis (x)	ergibt 1, falls x System-Missing ist; 0 sonst
nmiss ($x_1,x_2[,...]$)	Anzahl der Argumente, die Missing-Werte haben
nvalid ($x_1,x_2[,...]$)	Anzahl der Argumente, die gültige, nicht fehlende Werte haben
value (x)	Wert von x , d.h. Ignorieren einer Missing-Value-Definition

- Logische Funktionen

range ($x, u_1, o_1 [, u_2, o_2, \dots]$)	ergibt 1, falls x in einem der Intervalle $[u_1, o_1], [u_2, o_2], \dots$ liegt; 0 sonst
any ($x, x_1 [, x_2, \dots]$)	ergibt 1, falls x mit einem der Argumente x_1, x_2, \dots übereinstimmt; 0 sonst

- Funktionen für Strings (Auswahl)

concat ($s_1, s_2 [, \dots]$)	verkettet die Strings s_1, s_2, \dots
lpad ($s, l [, c]$)	String s bis auf Länge l mit Zeichen c links auffüllen
rpadd ($s, l [, c]$)	String s bis auf Länge l mit Zeichen c rechts auffüllen
ltrim ($s [, c]$)	Führende Zeichen c aus String s löschen
rtrim ($s [, c]$)	Abschließende Zeichen c aus String s löschen
substr ($s, n [, l]$)	ergibt den Teil des Strings s , der an der Position n beginnt und l Zeichen lang ist
length (s)	ergibt die Länge des Strings s
index ($s_1, s_2 [, l]$)	Absuchen des Strings s_1 von vorne nach einem Teilstring s_2 mit Länge l ; Rückgabe der Position
rindex ($s_1, s_2 [, l]$)	Absuchen des Strings s_1 von hinten nach Teilstrings s_2 mit Länge l ; Rückgabe der Position
string ($x, format$)	Konvertierung der Zahl x in einen String gemäß dem angegebenen Format
number ($s, format$)	Konvertierung des Strings s in eine Zahl gemäß dem angegebenen Format
lower (s)	Umwandlung von String s in Kleinschreibung
uppercase (s)	Umwandlung von String s in Großschreibung
min ($s_1, s_2 [, \dots]$)	alphabetisch kleinster Wert
max ($s_1, s_2 [, \dots]$)	alphabetisch größter Wert

Voreinstellung für das Zeichen c ist das Leerzeichen, für die Länge l die Zahl 1.

- Funktionen für Datum und Zeit (Auswahl)

date.dmy (t, m, j)	erzeugt ein Datum aus den Argumenten t (Tag), m (Monat) und j (Jahr)
ctime.days (x)	berechnet die Anzahl der Tage für die angegebene Zeitdauer x ; ist x ein Datum, so berechnet die Funktion die Anzahl der Tage vom 14. Oktober 1582 bis zu dem angegebenen Datum
yrmoda (j, m, t)	berechnet die Anzahl der Tage vom 14. Oktober 1582 bis zu dem durch die Argumente j (Jahr), m (Monat) und t (Tag) angegebenen Datum; ist j nur zweistellig, so wird 1900 hinzuaddiert (d.h. 99 entspricht 1999)
xdate.tday (x)	berechnet die Anzahl der Tage vom 14. Oktober 1582 bis zu dem angegebenen Datum x

- Sonstige Funktionen (Auswahl)

lag (var, n)	gibt den Wert der Variable var für jenen Fall zurück, der um n Fälle früher in der Arbeitsdatei steht; ergibt für die ersten n Fälle System-Missing (bei einer numerischen Variablen) bzw. einen leeren String (bei String-Variablen).
-------------------------	--

2.7 Programmierung in SPSS

2.7.1 Praktische Hinweise für die Erstellung eines SPSS-Programms

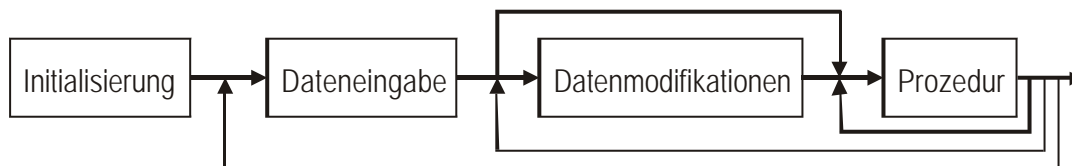
- Bei SPSS für Windows muß ein SPSS-Programm in einem Syntaxfenster eingegeben oder mit einem externen Texteditor erstellt werden. Ein Syntaxfenster wird automatisch geöffnet beim Laden eines SPSS-Programms ("Datei|Öffnen|Syntax...") bzw. beim Aufruf von "Datei|Neu|Syntax".
- Eine SPSS-Programmdatei ist eine einfache Textdatei, die SPSS-Syntaxbefehle enthält.
- Will man den gesamten Inhalt des Syntaxfensters oder nur einen Programmteil ausführen lassen, so selegiert man mit den üblichen Windows-Techniken zunächst die entsprechenden Programmbefehle. Für die Auswahl des gesamten Fensters dient auch die Tastenkombination [Strg]+[A]; ist kein Programmteil selegiert, so wird jener Programmbefehl ausgeführt, in dem der Cursor steht. Anschließend klickt man das Programmstartsymbol in der Symbolleiste an. Für den Programmstart gibt es auch eine Tastenkombination: [Strg]+[R]. Eventuelle Syntaxfehler werden im Ausgabefenster gemeldet, die Programmausgaben erscheinen wie üblich ebenfalls im Ausgabefenster. Dieses wird dabei im allgemeinen automatisch aktiviert.
- Der Benutzer tut gut daran, das Programm in einer SPSS-Programmdatei abzuspeichern. Dann ist es in Zukunft wieder verwendbar und es erfüllt außerdem die wichtige Aufgabe der Dokumentation der durchgeführten Auswertungsschritte.
- Wesentlich ist die Unterscheidung zwischen den Daten und dem SPSS-Programm. Syntaktische Regeln, so wie sie im folgenden formuliert werden, gelten nur für ein Programm, niemals für die Daten, auch wenn diese im ASCII-Format vorliegen. Meistens sind die Daten zwar getrennt vom SPSS-Programm auf einer eigenen Datei abgespeichert (nicht zuletzt aus Ökonomiegründen, da zu einem Datensatz im allgemeinen mehrere SPSS-Programme geschrieben werden), doch es ist auch möglich, die zu verarbeitenden Daten als Inline-Daten innerhalb des SPSS-Programms anzugeben (siehe BEGIN DATA/END DATA).
- Bei der Erstellung eines SPSS-Programms kann sich der Benutzer verschiedener Programmierhilfen bedienen:
 1. An erster Stelle ist hier der sog. Einfügemechanismus zu nennen. Sind beim interaktiven Aufruf eines Befehls (z.B. einer Datentransformation oder einer Statistikprozedur) alle Einstellungen und Optionen wunschgemäß spezifiziert, so kann man im Dialogfenster anstatt der Schaltfläche [OK] die Schaltfläche [Einfügen] wählen. Daraufhin wird der Befehl nicht ausgeführt, sondern lediglich der entsprechende SPSS-Syntaxbefehl bzw. die Syntaxbefehlsfolge automatisch ins Syntaxfenster geschrieben. Dadurch ist es möglich, durch interaktive Benutzung ein SPSS-Programm zu entwickeln, das man anschließend abspeichern und später wieder einlesen, verändern und starten kann. Dabei könnte man das Programm dann auch um jene Teile und Funktionen ergänzen, auf die über Menüs und Dialogfelder nicht zugegriffen werden kann.
 2. Weiterhin kann als Hilfe bei der Erstellung von SPSS-Programmen auch die Journaldatei verwendet und z.B. einzelne Teile daraus ins Syntaxfenster kopiert werden. Die Journaldatei ist eine ASCII-Datei, die ein fortlaufendes Protokoll aller Benutzereingaben in Form entsprechender Syntaxbefehle enthält, auch dann, wenn der Benutzer interaktiv gearbeitet hat.⁵
 3. Eine weitere Hilfe stellt das Befehlsprotokoll dar: ist es eingeschaltet, so können die Syntaxbefehle, die dann im Ausgabefenster protokolliert sind, ins Syntaxfenster kopiert und dort verwendet werden.⁶
 4. Als letzte Möglichkeit sei noch die Log-Tabelle im Ausgabefenster erwähnt, die beim Aufruf von Prozeduren jeweils erstellt wird. In dieser steht unter anderem auch der entsprechende Syntaxbefehl, den der Benutzer bei Bedarf kopieren und wiederverwenden könnte.

⁵ Einstellungen zur Journaldatei werden im Menü "Bearbeiten|Optionen...", Registerkarte "Allgemein" vorgenommen.

⁶ Das Befehlsprotokoll wird über das Menü "Bearbeiten|Optionen...", Registerkarte "Viewer" bzw. "Textviewer" ein- bzw. ausgeschaltet.

2.7.2 Aufbau und Arbeitsweise eines SPSS-Programms

- Ein SPSS-Programm besteht aus einer Folge von SPSS-Befehlen. Je nach ihrer Wirkung kann man verschiedene Typen von Befehlen unterscheiden:
 - deklarative Befehle,
 - status-ändernde Befehle,
 - ausführbare Befehle.
- Die Befehle werden SPSS normalerweise in der Reihenfolge interpretiert bzw. ausgeführt, in der sie im Programm stehen.⁷ Die richtige Wahl der Befehle und deren richtige Reihenfolge ist für die korrekte Arbeitsweise des Programms entscheidend. Die Reihenfolge der Befehle muß also so sein, wie logisch erforderlich. Beispielsweise muß eine Variable definiert sein, wenn auf sie Bezug genommen werden soll. Lediglich bei deklarativen Befehlen ist die Reihenfolge oft unkritisch.
- Bei einer statistischen Auswertung am Computer geht es im allgemeinen darum, irgendwelche Rohdaten einzulesen, an diesen Daten eventuell bestimmte Vorverarbeitungen oder Transformationen durchzuführen und schließlich irgendwelche statistischen Auswertungsverfahren anzuwenden. Ein SPSS-Programm beschreibt die Struktur und Codierung der Rohdaten, formuliert - wenn nötig - die erforderlichen Vorverarbeitungsschritte und spezifiziert schließlich die gewünschten statistischen und graphischen Auswertungsverfahren. Insgesamt unterscheidet SPSS vier verschiedene Programmstadien bzw. Programzustände:
 1. Initialisierung: Umgebung einstellen/initialisieren
 2. Dateneingabe: Arbeitsdatei definieren
 3. Datenmodifikationen: Daten aufbereiten/transformieren
 4. Prozedur: Daten mit Hilfe vorgefertigter Routinen auswerten.
 Ein SPSS-Programm bewegt sich durch diese Stadien, wobei das Stadium der Datenmodifikationen auch übersprungen werden kann. Bei komplexeren Programmen kommt es vor, daß bestimmte Stadien mehrfach durchlaufen werden.
 Eine Übersicht über die möglichen bzw. sinnvollen Abläufe gibt das folgende Diagramm:



⁷ Diese Form des Programmflusses, die sog. Sequenz, ist die am häufigsten vorkommende Kontrollstruktur, die in einem SPSS-Programm verwendet wird. Andere Kontrollstrukturen werden in Abschnitt 3.5 behandelt.

2.7.3 Die Syntax der SPSS-Befehlssprache

Metasprache zur Darstellung der Syntax von SPSS-Befehlen

In der EDV ist es üblich, die Syntax einer Programmiersprache in schematischer Form z.B. als Syntaxdiagramm oder als Menge von Ersetzungsregeln darzustellen. Dabei bedient man sich im allgemeinen einer Metasprache und erreicht damit eine knappe und präzise Definition der syntaktischen Regeln.

Im folgenden werden bei der Darstellung der Syntax der SPSS-Befehle immer folgende durchaus übliche metasprachlichen Elemente benutzt:

- Sprachkonstanten ("terminale Symbole"): sie müssen wörtlich im SPSS-Befehl auftauchen. Es sind dies:
 - Schlüsselwörter (sie werden in **Fettschrift** dargestellt)
 - Sonderzeichen außer den metasprachlichen Symbolen [] { } | ... (s.u.), d.h. beispielsweise auch die runden Klammern, das Gleichheitszeichen u.a.
- Metasprachliche Variablen (in *Kursivschrift* dargestellt): sie müssen durch entsprechende terminale Symbole ersetzt werden. Einige metasprachlichen Variablen haben eine spezielle Bedeutung, z.B.:

<i>var</i>	ein einzelner Variablenname
<i>varlist</i>	eine Variablenliste
<i>value</i>	ein numerischer Wert
<i>valuelist</i>	eine Werteliste
<i>n,m</i>	natürliche Zahlen
<i>string</i>	eine Zeichenkette
<i>filename</i>	ein Dateiname (incl. Pfadangabe).

Häufig wird der metasprachliche Name allerdings mehr im Hinblick auf den Inhalt bzw. Zweck des entsprechenden Befehlsteils und weniger auf dessen rein syntaktischen Aufbau gewählt.

- Optionale Teile eines SPSS-Befehls werden durch eckige Klammern ([]) gekennzeichnet: der entsprechende Teil des Befehls kann weggelassen werden; in diesem Fall treten evtl. Voreinstellungen in Kraft (in der Syntaxdarstellung unterstrichen); in manchen Prozeduren tauchen bei optionalen Spezifikationen zwei verschiedene Typen von Voreinstellungen auf: sog. "passive" (gekennzeichnet durch **) und "aktive" (nur unterstrichen) Voreinstellungen:
 - wird die betreffende Spezifikation ganz weggelassen, so tritt die passive Voreinstellung in Kraft;
 - wird bei der betreffenden Spezifikation lediglich der Name des Unterbefehls aufgeführt, ohne daß dazu weitere Angaben folgen, so tritt die aktive Voreinstellung in Kraft.
- Geschweifte Klammern sind Obligatklammern und treten zusammen mit dem Oder-Symbol (|) auf: genau eine der aufgeführten Alternativen muß verwendet werden.
- Fortsetzungspunkte (...) bedeuten, daß der SPSS-Befehl an dieser Stelle in analoger Weise beliebig lang fortgesetzt werden kann.

Syntax für SPSS-Befehle

- Ein SPSS-Programm besteht aus einer Folge von SPSS-Befehlen (engl. "commands").
- SPSS-Befehle beginnen mit einem Befehlswort. Dahinter stehen in den meisten Fällen Unterbefehle (engl. "subcommands"). Jeder Befehl kennt seine eigenen Unterbefehle, und diese haben jeweils eine bestimmte feste Bedeutung. Das Befehlswort gibt die Funktion des SPSS-Befehls an. Unterbefehle stellen nähere Spezifikationen zum SPSS-Befehl dar.
- Es gibt obligate und optionale Spezifikationen. Fehlt eine optionale Spezifikation, so tritt in vielen Fällen eine Voreinstellung in Kraft.
- Jeder Befehl muß auf einer neuen Zeile beginnen und mit einem Punkt (.) enden. Ein einzelner SPSS-Befehl kann sich über mehrere Programmzeilen erstrecken.

- Einzelne Unterbefehle werden meistens durch Schrägstriche (/) voneinander getrennt. Der Schrägstrich vor dem ersten Unterbefehl ist in der Regel optional.
- Das Befehlswort besteht aus einem oder manchmal auch mehreren fest vorgegebenen Schlüsselwörtern. Unterbefehle bestehen aus fest vorgegebenen Schlüsselwörtern, Variablenamen, Variablenlisten, Werten, Wertelisten und/oder Sonderzeichen (+ - * / ** () = u.a.).
- Schlüsselwörter müssen wortgetreu verwendet werden. Allerdings kann man sie bis auf die ersten drei Zeichen abkürzen, solange die Eindeutigkeit gewahrt ist. Variablenamen hingegen müssen vollständig ausgeschrieben werden. Variablenamen, die mit einem Punkt enden, können Fehler hervorrufen und sollten daher generell vermieden werden.
- Text in Apostrophen oder Anführungszeichen muß sich auf einer Zeile befinden.
- SPSS unterscheidet nicht zwischen Groß- und Kleinschreibung (Ausnahme: Strings). Manche kleingeschriebenen Wörter (z.B. kleingeschriebene Variablenamen) wandelt SPSS bei der Ausgabe automatisch in Großschreibung um.
- Trennzeichen zwischen Wörtern sind Leerzeichen (engl. "blank"), Komma, Zeilenumbruch (CR=Carriage Return), Sonderzeichen (/, =, runde Klammern u.a.). Überall im Programm können anstelle eines einzelnen Leerzeichens auch sog "white spaces" (d.h. eine beliebige Folge aus Leerzeichen, TAB-Zeichen und Zeilenumbrüchen) auftauchen. Durch Verwendung von Leerzeichen und Zeilenumbrüchen läßt sich ein SPSS-Programm strukturieren und damit für den Benutzer lesbarer machen.
- Syntaxdateien, die mit dem SPSS-Produktionsmodus oder dem Befehl INCLUDE ausgeführt werden, unterliegen etwas abweichenden Regeln:
 - Jeder Befehl muß in der ersten Spalte einer neuen Zeile beginnen.
 - Fortsetzungszeilen müssen um mindestens ein Leerzeichen eingerückt sein.
 - Der Punkt am Ende des Befehls ist optional.
- Wenn die Befehlsyntax mit der Schaltfläche [Einfügen] aus einem Dialogfeld in das Syntaxfenster übernommen wird, ist das Format für alle Betriebsmodi geeignet.

Als Beispiel soll der folgende SPSS-Befehl auf syntaktische Richtigkeit überprüft werden:

```
sort cases by fb(d) semester name. /* Hier wird sortiert! */
```

Die Syntax des SORT CASES-Befehls ist wie folgt:

```
sort cases [by] varlist [(a | d)]  
[varlist...].
```

- Die Folge der Schlüsselwörter "sort cases by" wurde wörtlich übernommen. Übrigens könnte das Schlüsselwort "by" auch fehlen.
- "fb" ist offenbar eine Variable, also auch eine Variablenliste.
- Danach folgt "(d)".
- "semester" und "name" müßten zwei weitere Variablen, also eine Variablenliste, sein.
- Nach dieser Variablenliste könnte noch ein geklammerter Ausdruck "(a)" oder "(d)" folgen. Dies ist jedoch optional und kann entfallen, da offenbar "(a)" gelten soll.
- Anschließend folgt ein Punkt.
- Die Leerstellen am Anfang und zwischen den Wörtern, sowie der Kommentar (siehe 2.6.4) am Ende sind syntaktisch zugelassen und werden überlesen.

Der obige Befehl ist also syntaktisch korrekt. Ob der Befehl auch inhaltlich richtig ist, ist eine andere Frage.

Einführendes Programmbeispiel

Das folgende Beispiel beschreibt die Lösung eines Auswertungsproblems in Form eines SPSS-Programms. Es beschäftigt sich nicht damit, wie dieses Programm konkret auf einem Rechner einzugeben und auszuführen ist (siehe dazu entsprechende Hinweise weiter unten).

Wir nehmen an, es lägen die Daten von 107 Personen in den folgenden acht Variablen vor: Geschlecht (m=männlich, w=weiblich), Alter, Körpergewicht in kg, Körpergröße in cm, Abiturnote, Schulbildung des Vaters und der Mutter (1=Hauptschule; 2=mittlere Reife; 3=Abitur; 4=Hochschulabschluß), PC-Besitz (0=nein, 1=ja). Die folgende Tabelle zeigt auszugsweise die Daten:

Geschlecht	Alter	Körpergewicht	Körpergröße	Abiturnote	Schulbildung des Vaters	Schulbildung der Mutter	PC-Besitz
1	20	48	163	1,70	4	1	1
1	19	56	166	1,70	3	4	1
2	20	78	189	3,20	4	4	1
1	19	61	172	3,20	4	4	1
1	20	63	175	1,30	2	2	0

Diese Daten seien bereits in SPSS mit dem Dateneditor eingegeben und auf Diskette im Verzeichnis "data" in der SPSS-Datei "beispiel.sav" abgespeichert worden. Bei der Eingabe der Daten seien unter Berücksichtigung der SPSS-Syntax die folgenden Variablennamen definiert worden: GESCHL, ALTER, GEWICHT, GROESSE, ABI, SCHULE_V, SCHULE_M, PC.

Zunächst sollen für die ersten 10 Personen die Werte in allen Variablen aufgelistet und anschließend die Häufigkeitsverteilung der Variablen Geschlecht, Alter, Körpergewicht, Körpergröße, Schulbildung des Vaters, Schulbildung der Mutter ermittelt werden.

Diese Aufgabe soll mit Hilfe eines SPSS-Programms gelöst werden. Dieses Programm sei als Programmdatei unter dem Namen "beispiel.sps" im Verzeichnis "programs" auf Diskette abgespeichert worden. Der Inhalt dieser Datei sei wie folgt:

```
get file = 'a:\data\beispiel.sav'.
title 'Mein erstes SPSS-Programm:'.
subtitle '- Auflisten von Datenwerten -'.
list variables = all /cases=1 to 10.
subtitle '- Häufigkeitsverteilungen -'.
frequencies variables = geschl to groesse, schule_v, schule_m.
```

Nachfolgend einige Erläuterungen zu diesem Programm:

- Die Daten liegen bereits im SPSS-Format vor und werden mit dem GET-Befehl von Diskette eingelesen.
- Durch TITLE- und SUBTITLE-Befehle werden Kopfzeilen für die Ausgabe definiert.
- Der LIST-Befehl bewirkt die Ausgabe der Rohwerte in allen Variablen für die ersten 10 Fälle.
- Die empirischen Häufigkeitsverteilungen für die Variablen Geschlecht bis Körpergröße, sowie Schulbildung des Vaters und Schulbildung der Mutter werden mit Hilfe des FREQUENCIES-Befehls erstellt.

2.7.4 Kommentare in SPSS-Programmen

Kommentare dienen der besseren Lesbarkeit von Programmen. In SPSS steht dem Benutzer dafür ein eigener Befehl, nämlich der COMMENT-Befehl, zur Verfügung. COMMENT-Befehle können an jeder Stelle eines SPSS-Programms auftauchen.

Neben dem COMMENT-Befehl hat der Benutzer noch die Möglichkeit, auch innerhalb eines SPSS-Befehls Kommentare einzustreuen, u.zw. überall dort, wo in dem Befehl Leerzeichen stehen dürfen.

DER COMMENT-BEFEHL UND EINGESTREUTE KOMMENTARE

Funktion

- Einstreuen von Kommentaren in ein SPSS-Programm

Syntax

Es gibt zwei verschiedene Formen des Kommentars:

a) als eigenständiger Befehl:

```
comment text.
```

oder:

```
* text.
```

b) eingestreut in einen beliebigen Befehl:

```
... /* text */ ...
```

Beispiel

```
comment   *** Beispiel eines SPSS-Programms (aus Schubö et al., 1991):
          *** Einlesen der Rohdaten aus einer ASCII-Datei: .
data list file='a:\data\beispiel.dat'
  /nummer 1-4
  geschl  6(a)      /* Geschlecht ist alphanumerisch codiert */
  alter   8-10
  gewicht 12-16(1) /* eine Dezimalstelle */
  beruf   18-39(a)
  tablette 40
  /* die folgenden Variablen sind die psychosomatischen Beschwerden */
  kopf    42
  foehn   43
  hals    44
  ohr     45
  schwind 46
  brechr  47
  magen   48
  herz    49.
title     'Medizinisch-soziologische Untersuchung'.
comment   *** Zunächst sollen die Daten nur aufgelistet werden.
subtitle  'Einlesen des Datenwerte, Auflistung'.
list      variables=geschl to tablette.
comment   *** Häufigkeitsverteilung für das Geschlecht.
subtitle  'Haeufigkeitsverteilung von Geschlecht'.
frequencies variables=geschl.
```


2.7.5 SPSS-Programm abbrechen (FINISH)

Will der Benutzer ein längeres SPSS-Programm vorzeitig beenden (z.B. um zu Testzwecken nur die ersten Befehle des Programms ausführen zu lassen), so kann er dafür den FINISH-Befehl einsetzen.

DER FINISH-BEFEHL

Funktion

- Abbruch der Abarbeitung von SPSS-Befehlen

Syntax

finish.

Hinweis

- SPSS-Befehle, die nach dem FINISH-Befehl folgen, werden nicht mehr interpretiert.

Beispiel

```
get file='a:\data\beispiel.sav'.  
correlations variables=groesse gewicht.  
finish.  
compute gindex=gewicht/alter.  
correlations variables=groesse gindex.
```

Die Daten werden eingelesen und die Korrelation zwischen Körpergröße und Körpergewicht berechnet. Anschließend beendet SPSS die Abarbeitung des Programms, d.h. der COMPUTE-Befehl und die nachfolgende Korrelationsberechnung werden nicht mehr ausgeführt.

2.7.6 SPSS-Programmdatei einfügen (INCLUDE)

Bei größeren Projekten werden in der Regel eine Reihe von verschiedenen Auswertungsprogrammen erstellt. Diese unterscheiden sich häufig nur in wenigen Befehlszeilen (z.B. in verschiedenen Prozeduraufrufen oder zusätzlichen Datentransformationen) und bestehen ansonsten aus einem identischen umfangreicheren Programmteil zur Dateneingabe und/oder zur Datenaufbereitung. Im Sinne der besseren Übersichtlichkeit kann nun dieser gemeinsame Programmteil auch in eine eigene Datei ausgliedert werden, die durch einen INCLUDE-Befehl an der gewünschten Stelle im Programm eingefügt wird.

Ein weiterer Einsatzzweck des INCLUDE-Befehls ist es, anderen Benutzern eine Bibliothek von fertigen Auswertungsroutinen zur Verfügung zu stellen (hier finden oft auch Makro-Definitionen⁸ Verwendung).

DER INCLUDE-BEFEHL

Funktion

- Einfügen einer Datei mit SPSS-Befehlen in das aktuelle Programm

Syntax

include [file=]'file'.

Hinweise

- Die Befehle in der INCLUDE-Datei müssen in Spalte 1 beginnen und Fortsetzungszeilen müssen eingerückt sein. Die Zeilen dürfen nicht über die Spalte 80 hinausgehen.
- Eine INCLUDE-Datei darf selbst wiederum INCLUDE-Befehle enthalten.
- Die INCLUDE-Datei wird an der Stelle eingefügt, an der der INCLUDE-Befehl steht.

Beispiel

Die Datei 'd:\diplom\vvdadef.sps' enthalte die folgenden Zeilen:

```
get file='d:\diplom\vvdadef.sps'.
recode geschl ('w'=1) ('m'=2) into grpvar1.
value labels grpvar1
  1 'weiblich'
  2 'maennlich'.
count rohwert=a1 to a23(1).
```

Das eigentliche SPSS-Programm sei wie folgt:

```
include file='d:\diplom\vvdadef.sps'.
t-test groups=grpvar variables=rohwert.
```

Die Wirkung ist dieselbe, als hätte man in dem eigentlichen SPSS-Programm geschrieben:

```
get file='d:\diplom\vvdadef.sps'.
recode geschl ('w'=1) ('m'=2) into grpvar1.
value labels grpvar1
  1 'weiblich'
  2 'maennlich'.
count rohwert=a1 to a23(1).
t-test groups=grpvar variables=rohwert.
```

⁸ MACRO-Definitionen werden hier nicht behandelt. Informationen dazu findet man im Handbuch "SPSS 10.0 Syntax Reference Guide" bzw. bei Schubö et al. (1991).

3 Datenmanagement in SPSS

3.1 Variablen in SPSS

3.1.1 Attribute von Variablen

SPSS kennt die folgenden Variablenattribute (in Klammern sind jeweils die Voreinstellungen angegeben, die der Benutzer je nach Bedarf definieren bzw. ändern kann):

- Variablenname (VAR00001, VAR00002,...)
- Typ (numerisch 8.2)
- Labels
 - Variablenlabel (keine)
 - Wertelabels (keine)
- Fehlende Werte (keine)
- Spaltenformat
 - Spaltenbreite (8)
 - Textausrichtung (rechts)
- Meßniveau (metrisch).

DEFINIEREN VON VARIABLEN

Die Definition neuer Variablen und die Bearbeitung sämtlicher Variablenattribute kann interaktiv in der Registerkarte "Variablenansicht" des Datenfensters vorgenommen werden. In dieser Tabelle bilden die Zeilen die einzelnen Variablen und die Spalten die Variablenattribute, nämlich:

- Name
- Typ
- Spaltenformat
- Dezimalstellen
- Variablenlabel
- Wertelabels
- Fehlende Werte
- Spalten
- Ausrichtung
- Meßniveau

Variablennamen

- Variablennamen dienen der Referenzierung von Variablen z.B. beim Aufruf einer Statistik- oder Graphikprozedur.
- Der Benutzer kann die Namen seiner Variablen frei wählen, allerdings unter Beachtung der Syntax und der Eindeutigkeit der Namen.
- Bei der Dateneingabe über den SPSS-Dateneditor sind die Variablennamen zunächst auf 'VAR00001', 'VAR00002', usw. voreingestellt.

Variablentyp

- In der Statistik kennt man verschiedene Datentypen. Diese werden hierbei meist mit dem Skalenniveau der Variablen gleichgesetzt.
- SPSS kennt ebenfalls verschiedene Datentypen. Diese entsprechen aber mehr den Datentypen, die in Programmiersprachen Verwendung finden. Der Datentyp einer Variablen gibt hier Auskunft über die Werte, die die Variable annehmen kann. Beispielsweise wird bei Programmiersprachen durchgängig unterschieden zwischen numerischen Daten (Zahlen) und alphanumerischen Daten (Text). Daneben gibt es noch den Datentyp boolesch für logische Werte (wahr/falsch). Bei manchen Datentypen gibt es oft noch Untertypen, so beispielsweise bei den numerischen Daten die Untertypen Integerzahlen, Realzahlen, u.a.
- In SPSS sind folgende Datentypen verfügbar:
 - numerisch $n.m$ (Zahl mit maximal n Stellen insgesamt und m Stellen hinter dem Dezimalkomma)
 - String n (Zeichenfolge aus n Zeichen)
 - Datum und Uhrzeit
 - Währung
- Beim Datentyp numerisch unterscheidet SPSS noch die Varianten Komma (z.B. 2,456,700.25), Punkt (z.B. 2.456.700,25) und wissenschaftliche Notation (z.B. 2,45670025E6). Diese Varianten haben lediglich mit dem Darstellungsformat der Zahl zu tun und nicht mit deren Wert. Für den Datentyp Datum gibt es ebenfalls verschiedene Darstellungsformate (z.B. '10.01.96' oder '10-JAN-96') und beim Datentyp Währung gibt es verschiedene Untertypen (Dollar, spezielle Währung).
- Bei Strings unterscheidet SPSS zwischen kurzen (Länge ≤ 8) und langen Strings (Länge > 8). Diese beiden Typen werden in SPSS-Prozeduren teilweise verschieden behandelt.
- Es gibt in SPSS keinen Datentyp logisch (boolesch). Die Wahrheitswerte 'falsch' und 'wahr' werden hier einfach durch die Zahlen 0 und 1 dargestellt.
- Man beachte, daß viele Berechnungen in Statistikprozeduren oder bei Datentransformationen nur mit bestimmten Datentypen verträglich sind.

Variablen-Labels

- Variablennamen sind wegen der restriktiven Syntax oft wenig aussagekräftig. Der Benutzer hat die Möglichkeit, Variablen durch Definition von Variablen-Labels näher zu kennzeichnen (zur Syntax von Variablenlabels siehe Abschnitt 2.5). Diese sind dann in der Ausgabe jeweils beim Variablennamen mit aufgeführt.
- Variablen-Labels dienen nicht der Referenzierung.

DER VARIABLE LABELS-BEFEHL

Funktion

- Vereinbarung von Variablen-Labels

Syntax

variable labels var 'label' [/var...].

Beispiel

Im folgenden Programmfragment werden für die 3 Variablen Geschlecht (GESCHL), Schulbildung des Vaters (SCHULE_V) und Schulbildung der Mutter (SCHULE_M) aussagekräftige Variablenlabels definiert.

```
get file='a:\data\frabo.sav'.
variable labels geschl 'Geschlecht'
  / schule_v 'Schulbildung des Vaters'
  / schule_m 'Schulbildung der Mutter'.
.....
```

Werte-Labels

- Mit Hilfe der Werte-Labels kann der Benutzer die Werte einer Variable näher kennzeichnen und damit z.B. bei codierten Variablen den jeweiligen Code näher beschreiben (zur Syntax von Wertelabels siehe Abschnitt 2.5). Werte-Labels erscheinen in der Ausgabe überall dort, wo Einzelwerte aufgelistet werden (z.B. in einer Häufigkeitstabelle).
- Werte-Labels sind nicht für die statistische Verarbeitung selbst vorgesehen.
- Für lange Stringvariablen können keine Wertelabels definiert werden.

DER VALUE LABELS-BEFEHL

Funktion

- Vereinbarung von Werte-Labels

Syntax

value labels varlist value 'label' [value..] [/varlist...].

Beispiel

Im folgenden Programmfragment werden für die alphanumerische Variable Geschlecht sowie für die numerischen Variablen Item 1 bis Item 48 (z.B. die Antworten von Personen in einem Fragebogen) aussagekräftige Wertelabels definiert:

```
get file='a:\data\items.sav'.
value labels
  geschl
    'M' 'männlich'
    'W' 'weiblich'
  /item1 to item48
    0 'falsch'
    1 'richtig'.
...
```

Fehlende Werte (Missing-Werte)

- In empirischen Untersuchungen treten aus den verschiedensten Gründen immer wieder fehlende Meßwerte auf, z.B. durch Fehler in der Apparatur oder dadurch, daß eine Person eine Antwort ausgelassen, die Antwort "weiß nicht" gegeben oder die Antwort verweigert hat. Solche fehlende Daten werden beim Eintragen in die Datentabelle mit dem SPSS-Dateneditor normalerweise einfach leer gelassen. Ein anderer, aber ähnlich gelagerter Fall liegt vor, wenn beim Lesen aus einer ASCII-Datendatei das Eingabefeld fehlerhaft ist oder nur Leerstellen enthält. Schließlich tauchen fehlende Werte oft auch bei Datentransformationen auf (z.B. bei Operationen mit undefinierten Operanden oder bei unerlaubten Argumenten in Funktionsaufrufen). In all diesen Fällen spricht man vom System-Missing-Wert, in SPSS auch mit SYSMIS bezeichnet (manchmal findet man auch die Bezeichnung "automatischer fehlender Wert"). Als Symbol für diesen Wert verwendet SPSS bei numerischen Datentypen das Dezimaltrennzeichen und sonst Leerzeichen).
- Neben diesem System-Missing-Wert gibt es in SPSS aber noch einen anderen Typ von fehlenden Werten, nämlich die benutzerdefiniert fehlenden Werte oder einfach Missing-Werte: der Benutzer kann nämlich für fehlende Werte auch einen bestimmten Code (z.B. eine bestimmte Zahl) verwenden und anschließend SPSS mitteilen, daß es sich bei diesem Code um einen fehlenden Wert handelt. Dies geschieht mit Hilfe der Deklaration von Missing-Werten.
- Es können nicht beliebig viele Missing-Werte deklariert werden, sondern lediglich:
 - max. 3 Einzelwerte oder
 - ein Werte-Intervall (unter Angabe der unteren und oberen Grenze) oder
 - ein Einzelwert und ein Intervall.
- SYSMIS-Werte werden immer aus den Berechnungen ausgeklammert.
- Normalerweise werden auch die Missing-Werte aus den statistischen Berechnungen ausgeschlossen, doch auf expliziten Wunsch des Benutzers können sie auch einbezogen, d.h. wie valide Daten behandelt werden.
- Bei der simultanen Analyse mehrerer Variablen (z.B. bei Korrelationsberechnungen) hat man oft die Wahl zwischen fallweisem vs. paarweisem Ausschluß: fehlt der Wert bei einem Fall in einer der Variablen, so wird dieser Fall beim fallweisen Ausschluß vollständig aus den Berechnungen ausgeschlossen, beim paarweisen Ausschluß hingegen nur dort, wo die Variable mit dem fehlenden Wert benutzt werden soll.

DER MISSING VALUES-BEFEHL

Funktion

- Deklaration von Missing-Werten

Syntax

```
missing values varlist({valuelist})  
[varlist...].
```

Hinweise

- Gibt es in einem SPSS-Programm zu einer Variable mehrere MISSING VALUES-Befehle, so gilt immer der letzte.
- Missing Value-Deklarationen werden für nachfolgende SPSS-Befehle gelöscht durch Angabe einer leeren Werteliste, d.h. eines leeren Klammerspaars: **missing values** *varlist*() .

Beispiel

```
....  
missing values alter(0) item1 to item48(9).  
missing values groesse (lo thru 50).  
....
```

Für die Variable ALTER wird der Wert 0 als fehlender Wert deklariert, für die Variablen ITEM1, ITEM2, ... ITEM48 der Wert 9 und für die Variable GROESSE alle Werte, die kleiner oder gleich 50 sind.

Spaltenformat

- Das Spaltenformat besteht aus:
 - Anzahl der Spalten (Voreinstellung: 8)
 - Ausrichtung (Voreinstellung: rechtsbündig)
- Das Spaltenformat betrifft nur die Darstellung der Variablenwerte im Dateneditorfenster und hat keinerlei Wirkung auf die Variablenwerte selbst. Diese werden immer mit der vollen Rechnergenauigkeit abgespeichert.

DER VARIABLE WIDTH-BEFEHL

Funktion

- Vereinbarung der Spaltenbreite für eine oder mehrere Variablen

Syntax

variable width *varlist* (*n*) [*/varlist...*].

Hinweis

- Dieser Befehl betrifft nur die Darstellung von Variablen im Dateneditorfenster, nicht deren Werte.

DER VARIABLE ALIGNMENT-BEFEHL

Funktion

- Vereinbarung der Ausrichtung für eine oder mehrere Variablen

Syntax

variable alignment *varlist* (**{right**|left|center}**) [*/varlist...*].

Hinweis

- Dieser Befehl betrifft nur die Darstellung von Variablen im Dateneditorfenster, nicht deren Werte.

Ausgabeformat

- In manchen Statistikprozeduren (z.B. bei LIST, FREQUENCIES, CROSSTABS) werden nicht nur Statistiken, sondern auch Werte einzelner Variablen ausgegeben. Dafür wird ein bestimmtes Ausgabeformat (z.B. Anzahl der Stellen hinter dem Komma) benutzt. Dieses kann mit Hilfe des FORMATS-Befehls definiert werden. Fehlt ein solcher Befehl, so wird das voreingestellte, d.h. dem Datentyp entsprechende Format benutzt.
- Das Ausgabeformat hat keine Wirkung auf die Variablenwerte selbst. Diese werden immer mit der vollen Rechnergenauigkeit abgespeichert und werden durch den FORMATS-Befehl nicht verändert.

DER PRINT FORMATS-, WRITE FORMATS-, FORMATS-BEFEHL

Funktion

- Definition des Ausgabeformats für Variablen

Syntax

print formats *varlist(format)* [*varlist ...*].

write formats *varlist(format)* [*varlist ...*].

formats *varlist(format)* [*varlist ...*].

Hinweise

- Mit dem PRINT FORMATS-Befehl wird das PRINT-Format, mit dem WRITE FORMATS-Befehl das WRITE-Format und mit dem FORMATS-Befehl beide Formate spezifiziert.
- Das WRITE-Format wird für die Ausgabe beim WRITE-Befehl benutzt, das PRINT-Format bei der Ausgabe von Einzelwerten in Prozeduren und beim PRINT-Befehl.
- Die Angabe zum Datenformat muß eine Formatspezifikation nach der FORTRAN-Syntax sein.

Beispiel

```
data list file='a:\data\formats.dat' /name gewicht (a8,f5.1).
formats gewicht(f3).
list variables=name gewicht.
```

Die Werte für die Variable GEWICHT werden zwar mit einer Dezimalstelle hinter dem Komma (d.h. im FORTRAN-Format f5.1) von der ASCII-Datei "a:\data\formats.dat" eingelesen, in der LIST-Prozedur jedoch auf ganze Zahlen gerundet (f3) ausgegeben.

Meßniveau

- Eines der "jüngeren" Attribute von Variablen bei SPSS ist das Meßniveau. SPSS kennt drei verschiedene Meßniveaus: nominal, ordinal und metrisch.
- Das Meßniveau einer Variable spielt eine Rolle innerhalb einiger Statistik- und Graphikprozeduren in dem Sinne, daß manche Berechnungen nur bei bestimmtem Meßniveau durchgeführt werden.

DER VARIABLE LEVEL-BEFEHL

Funktion

- Vereinbarung des Meßniveaus für eine oder mehrere Variablen

Syntax

variable level *varlist* (**{scale**|ordinal|nominal}**) [*/varlist...*].

Beispiel

```
get file='a:\data\frabo.sav'.
variable level religion(nominal).
...
```

Für die Variable RELIGION wird das Nominalskalenniveau festgelegt.

3.1.2 Die SPSS-Datenformate

DIE SPSS-DATENFORMATE

Funktion

Das Eingabe- bzw. Ausgabeformat kann in SPSS in Anlehnung an die Programmiersprache FORTRAN in Form einer FORTRAN-Formatliste angegeben werden. Eine SPSS-Datenformatliste ist eine Liste von sog. Umwandlungsschlüsseln ("Formatelementen"). Die einzelnen Umwandlungsschlüssel der Formatliste werden durch Kommata getrennt. Es folgen die gängigsten Umwandlungsschlüssel.

Zahlenformate

o Format F

fw[.d]

Dies ist das Standardformat für numerische Daten.

o Format E

ew[.d]

Dies ist das Zahlenformat für wissenschaftliche Notation mit Mantisse und Exponent.

o Format COMMA

commaw[.d]

Zahlen werden durch Kommata in die Tausender unterteilt; Dezimaltrennzeichen ist der Punkt.

o Format DOT

dotw[.d]

Zahlen werden durch Punkte in die Tausender unterteilt; Dezimaltrennzeichen ist das Komma.

o Format DOLLAR

dollarw[.d]

Ausgabeformat für die Dollarwährung: Ausgabe mit Dollarzeichen und Betrag.

o Format CCA

ccaw[.d]

Ausgabeformat für die aktuelle nationale Währung.

o Format PIBHEX

pibhexw

Hexadezimale Ausgabe von positiven ganzen Zahlen.

o Format RBHEX

rbhexw

Hexadezimale Ausgabe von reellen Zahlen.

Zeichenkettenformate

o Format A

aw

Das A-Format ist das Standardformat für Zeichenketten.

o Format AHEX

ahex_w

Hexadezimale Ausgabe von Zeichenketten.

Formate für Datum und Uhrzeit

o Format DATE

date_wAusgabe des Datums in der Form: "tt-*mmm*-*jj*" bzw. "tt-*mmm*-*jjjj*"

o Format ADATE

adate_wAusgabe des Datums in der Form: "*mm*/*tt*/*jj*" bzw. "*mm*/*tt*/*jjjj*"

o Format EDATE

edate_wAusgabe des Datums in der Form: "*tt*/*mm*/*jj*" bzw. "*tt*/*mm*/*jjjj*"

o Format JDATE

jdate_wAusgabe des Datums in der Form: "*jjtttt*" bzw. "*jjjjtttt*"

o Format SDATE

sdate_wAusgabe des Datums in der Form: "*jj*/*mm*/*tt*" bzw. "*jjjj*/*mm*/*tt*"

o Format TIME

time_w[.d]Ausgabe der Uhrzeit in der Form: "*hh*:*mm*" bzw. "*hh*:*mm*:*ss*.*s*"

o Format DTIME

dtime_w[.d]Ausgabe von Tag und Uhrzeit in der Form: "*tt* *hh*:*mm*" bzw. "*tt* *hh*:*mm*:*ss*.*s*"

o Format DATETIME

datetime_w[.d]

Ausgabe von Tag und Uhrzeit in der Form:

"*tt*-*mmm*-*jjjj* *hh*:*mm*" bzw. "*tt*-*mmm*-*jjjj* *hh*:*mm*.*ss*.*s*"Formatelemente zum Positionieren

o Formatelement X

nx

Das X-Formatelement dient zum Überspringen von Spalten.

o Formatelement T

tn

Dieses Formatelement ist der Tabulator, d.h. man kann damit auf eine bestimmte Spalte der aktuellen Zeile positionieren.

o Formatelement /
/

Dieses Formatelement dient zum Abbrechen oder zum Überspringen von Zeilen: ein "/" bewirkt die Beendigung der aktuellen Zeile und ein Springen zur nächsten Zeile.

Sonstiges

o Wiederholungsfaktor

r formatelement

oder:

r(liste-von-formatelementen)

Der Wiederholungsfaktor dient zum Wiederholen eines einzelnen Formatelements oder einer ganzen Liste von Formatelementen.

Bedeutung der Parameter

<i>w</i>	Anzahl der Spalten (beim T-Formatelement: Spaltennummer)
<i>d</i>	Anzahl der Nachkommastellen (Voreinstellung: 0)
<i>r</i>	Anzahl der Wiederholungen

Hinweise

- Ein FORTRAN-Format wird bei der Ein- und Ausgabe von Daten in folgender Weise abgearbeitet: das System arbeitet mit einem Zeilen- und Spaltenzeiger, d.h. es gibt jeweils eine aktuelle Zeile und eine aktuelle Spalte. Die Abarbeitung der Variablenliste ist synchronisiert mit der Abarbeitung der FORTRAN-Formatliste.
- Im DATA LIST- und im WRITE-Befehl können die verschiedenen Formatspezifikationen (d.h. Spaltenangaben, FORTRAN-Formatangaben) auch gemischt werden.

Beispiele

(a) Der folgende DATA LIST-Befehl:

```
data list file='a:\data\reaktion.dat'
  /vp 1-4 geschl 5(a) gruppe 7 beding 8
  rz1 10-12 rz2 16-18 rz3 22-24 rz4 28-30 rz5 34-36
  f1 13-14 f2 19-20 f3 25-26 f4 31-32 f5 37-38
  index 40-45(2).
```

läßt sich mit Hilfe des FORTRAN-Formats auch so schreiben:

```
data list file='a:\data\reaktion.dat'
  /vp geschl gruppe beding, rz1 to rz5, f1 to f5, index
  (f4,a1,1x,2f1,1x,5(f3,3x),t9,5(4x,f2),1x,f6.2).
```

(b) Dieselben Daten in einer anderen Reihenfolge gelesen:

```
data list file='a:\data\reaktion.dat'
  /vp 1-4 geschl 5(a) gruppe 7 beding 8
  rz1 10-12 f1 13-14
  rz2 16-18 f2 19-20
  rz3 22-24 f3 25-26
  rz4 28-30 f4 31-32
  rz5 34-36 f5 37-38
  index 40-45(2).
```

Formulierung mit Hilfe des FORTRAN-Formats:

```
data list file='a:\data\reaktion.dat'
  /vp geschl gruppe beding rz1 f1 rz2 f2 rz3 f3 rz4 f4 rz5 f5 index
  (f4,a1,1x,2f1,5(1x,f3,f2),1x,f6.2).
```

3.1.3 Systemvariablen

- In SPSS gibt es folgende Systemvariablen
 - \$CASENUM Laufende Fall-Nr.
 - \$SYSMIS automatischer fehlender Wert
 - \$JDATE gegenwärtiges Datum als Anzahl der Tage seit dem 15.10.1582
 - \$DATE gegenwärtiges Datum in der Form tt_mmm_jj (z.B. 31 MAR 96)
 - \$LENGTH Anzahl der Zeilen je Ausgabeseite im Listing
 - \$WIDTH Anzahl der Zeichen je Ausgabezeile im Listing
- Systemvariablen sind automatisch definiert und können in einem SPSS-Programm nicht verändert werden. Auch kann deren PRINT- und WRITE-Format nicht verändert werden.
- Systemvariablen können nicht in Prozeduren spezifiziert werden.
- In Datenmodifikationsbefehlen können Systemvariablen wie "normale" Variablen benutzt werden.

3.1.4 Arbeitsvariablen

- Arbeitsvariablen beginnen mit einem "#" - Zeichen.
- Arbeitsvariablen sind lediglich temporäre Variablen; sie werden nicht in die Arbeitsdatei abgespeichert und können in keiner Prozedur aufgerufen werden.
- Arbeitsvariablen können keine fehlenden Werte oder Etiketten zugewiesen werden.
- Arbeitsvariablen und andere Variablen können nicht zusammen mit dem Schlüsselwort **to** angesprochen werden.
- Bei Arbeitsvariablen wird der Wert der vorangehenden Beobachtung als Initialisierungswert verwendet⁹, bzw. bei der ersten Beobachtung die Zahl 0 bzw. Leerzeichen.
- Arbeitsvariablen werden vor Bearbeitung des nächsten Prozedur- oder TEMPORARY-Befehls gelöscht.

Beispiel

Im folgenden Programm wird aus dem Geburtsdatum einer Person ihr Alter in Jahren berechnet und eine Häufigkeitstabelle für diese Variable erstellt. Da das Geburtsdatum selbst nicht von Interesse ist, werden dafür die drei Arbeitsvariablen #GEBTAG, #GEBMON und #GEBJAHR benutzt. Im übrigen liefert die Systemvariablen \$DATE den Funktionswert von YRMODA für das heutige Datum.

```
data list file='a:\data\arbvar.dat'
  /nummer 1-4 geschl 6(a)
  #gebtag #gebmon #gebjahr 8-13.
compute alter=trunc($jdate-yrmoda(#gebjahr,#gebmon,#gebtag))/365.25.
frequencies variables=alter.
```

Nicht erlaubt wäre der folgende Prozeduraufruf (etwa um die Verteilung der Geburtsraten für die verschiedenen Monate des Jahres zu bestimmen):

```
frequencies variables=#gebmon.
```

Für diesen Zweck müßte man den Geburtsmonat als reguläre Variable definieren (z.B. als Variablennamen GEBMON benutzen).

⁹ Eine ähnliche Wirkung hat der LEAVE-Befehl.

3.1.5 Vektoren

Bei der programmgesteuerten Arbeitsweise hat der Benutzer die Möglichkeit, Vektoren von Variablen zu kreieren oder einzelne vorhandene Variablen zu einem Vektor zusammenzufassen. Ein Vektor ist eine spezielle Datenstruktur, nämlich die eines eindimensionalen Feldes ("Arrays"). Bei den höheren Programmiersprachen kennt man üblicherweise auch noch mehrdimensionale Felder sowie noch ganz andere Datenstrukturen. Wie dort werden die Elemente eines Vektors (d.h. hier: die einzelnen Variablen) über einen in runde Klammern gesetzten Index angesprochen.

Vektoren eröffnen dem Benutzer zusammen mit bestimmten Kontrollstrukturen eine Reihe sehr guter Möglichkeiten für die Lösung kniffliger Datentransformationsprobleme.

3.1.6 Deklaration von Variablen und Vektoren

Bei der programmgesteuerten Arbeitsweise hat der Benutzer die Möglichkeit, Variablen vor deren Benutzung zu deklarieren. Normalerweise werden Variablen durch SPSS implizit automatisch kreiert. Wenn aber durch Datentransformationen neue Zeichenkettenvariablen erzeugt werden sollen, so müssen diese mit Hilfe eines STRING-Befehls vorher erst deklariert werden. Bei Vektoren ist eine Deklaration immer notwendig.

DER NUMERIC-BEFEHL

Funktion

- Deklaration von numerischen Variablen

Syntax

numeric *varlist*(*{format}*) [/varlist...]

Hinweise

- Mit Hilfe dieses Befehls kann der Benutzer numerische Variablen deklarieren. Die Variablen werden in der angegebenen Reihenfolge in die Arbeitsdatei aufgenommen. In einem INPUT PROGRAM kann man damit z.B. vorneweg die Reihenfolge der Variablen abweichend von deren Reihenfolge im DATA LIST-Befehl festlegen.
- Hinter der Variablenliste kann jeweils in Klammern eine Formatangabe z.B. in der Form **fw**[*d*] angegeben werden (siehe S. 32).

DER STRING-BEFEHL

Funktion

- Deklaration von Zeichenkettenvariablen

Syntax

string *varlist*(*an*) [/varlist...]

Hinweise

- Mit Hilfe dieses Befehls kann der Benutzer Zeichenkettenvariablen deklarieren.
- Sollen durch Datentransformationen neue Zeichenkettenvariablen erzeugt werden, so müssen diese mit Hilfe eines STRING-Befehl erst deklariert werden.
- Hinter der Variablenliste muß in Klammern das FORTRAN-Format in der Form **an** angegeben werden (*n* ist die Länge der Variable in Anzahl der Zeichen).

DER VECTOR-BEFEHL

Funktion

- Definition eines Vektors von numerischen Variablen

Syntax

Es gibt zwei verschiedene Formen der VECTOR-Befehls:

a) Vektordefinition für existierende Variablen:

```
vector name=vara to varb [/name...].
```

b) Vektordefinition für neue Variablen:

```
vector vectorlist(n [format]) [/vectorlist...]
```

Hinweise

- In der ersten Form werden bereits definierte Variablen (*vara*, ..., *varb*) im nachhinein zu einem Vektor erklärt. Diese Variablen müssen in der Arbeitsdatei unmittelbar aufeinander folgen.
- In der zweiten Form werden neue numerische Variablen definiert und gleichzeitig als Vektor deklariert. Die Namen dieser neu definierten Variablen setzen sich zusammen aus dem angegebenen Vektornamen und einer laufenden Nummer (1,2,...,n). Hinter der Anzahl der Variablen (*n*) kann auch ein Variablenformat angegeben werden. Mit Hilfe dieses VECTOR-Befehls können auch gleich mehrere Vektoren definiert werden (*vectorlist* ist eine Liste von Vektornamen, im einfachsten Fall ein einzelner Name), dann werden die entsprechenden Variablen für alle angegebenen Vektornamen erzeugt.
- Der Vektorname wird nicht in die Arbeitsdatei abgespeichert; Vektoren können nicht in Prozedurbefehlen, wohl aber in Datenmodifikationsbefehlen benutzt werden.
- Außerhalb des VECTOR-Befehls muß der Vektorennamen immer mit einem Subskript (numerische Konstante, Variable, Ausdruck) in Klammern benutzt werden: '*vektor(subskript)*'.
- Eine Vektordefinition ist immer nur bis zum nächsten Prozedurbefehl gültig.

Beispiel

```
data list free file='a:\data\vector.dat'
  / vp nr1 nr2 nr3.
vector symptom(8).
do repeat symp=symptom1 to symptom8.
  compute symp=0.
end repeat.
compute symptom(nr1)=1.
compute symptom(nr2)=1.
compute symptom(nr3)=1.
frequencies variables=symptom1 to symptom8.
```

Für jede Person liegen die Nummern von drei Symptomen vor, unter denen sie leidet. Daraus werden die Werte für acht Variablen (SYMPTOM1,...,SYMPTOM8) erzeugt, wobei die drei Variablen mit den angegebenen Nummern den Wert 1 und die anderen den Wert 0 erhalten sollen.

3.2 Ein- und Ausgabe

3.2.1 Der Dateneditor von SPSS für Windows

Die manuelle Eingabe von Informationen (z.B. eines ASCII-Textes oder eines Programms in einer Programmiersprache) in den Rechner und deren Korrektur nennt man im Fachjargon 'Editieren'. Ein Programm zum Editieren ist ein 'Editor'. In der Windows-Version von SPSS steht ein Editor für die interaktive Eingabe und Korrektur von Rohdaten zur Verfügung. Der Editor präsentiert sich dem Benutzer in Form eines Fensters mit einer Datentabelle. Die Zeilen der Tabelle stellen die Beobachtungseinheiten (Fälle) und die Spalten die Variablen dar. In den Zellen der Tabelle stehen die Einzelwerte. In der Kopfzeile findet man die Variablenamen.

- **Aufruf des Dateneditors**

- Für die Neueingabe von Rohdaten wird ein leeres Dateneditorfenster benötigt. Ein leeres Dateneditorfenster (mit dem Titel "Unbenannt") wird geöffnet beim Start von SPSS und bei der Menüwahl von "Datei|Neu|Daten".
- Für die Aufnahme bereits gespeicherter SPSS-Daten: über "Datei|Öffnen..."; damit ist auch die Möglichkeit gegeben, die gespeicherten Daten im nachhinein zu korrigieren. Beim Öffnen einer externen Datendatei wird automatisch ein entsprechendes Datenfenster geöffnet. In SPSS für Windows gibt es immer nur ein Datenfenster, d.h. eine Arbeitsdatei. Wird eine Datendatei geöffnet, so wird das bisher vorhandene Datenfenster gelöscht (SPSS gibt aber eine Warnung aus, falls noch nicht abgespeichert wurde).

- **Praktisches Vorgehen bei der Dateneingabe in SPSS**

- Aktivieren des Editorfensters
- Variablen definieren und, wenn nötig, deren Attribute ändern (es gibt Voreinstellungen!)
- Eingabe von Datenwerten
 - beachten, welches die aktuelle Zelle ist
 - Datentyp beachten
 - Dezimaltrennzeichen beachten (hier: Komma)
- Navigieren über die Zellen, Zellen selektieren: mit der Maus, den Cursortasten, der [TAB]-Taste
- Korrektur von Werten: Zelle selektieren und Wert überschreiben
- Selektieren von Zeilen bzw. Spalten, z.B. zum Löschen oder Kopieren: mit Klick (bzw. Ziehen mit der Maus) in die Kopfspalte bzw. -zeile

3.2.2 Ein- und Ausgabe von Daten im SPSS-Format bzw. in Fremdformaten (GET, SAVE, GET TRANSLATE, SAVE TRANSLATE)

Prinzipiell basieren alle Auswertungsschritte in SPSS auf der internen Arbeitsdatei. Diese geht allerdings mit Verlassen von SPSS verloren. Der Benutzer hat aber die Möglichkeit, den Inhalt der Arbeitsdatei auf einer externen SPSS-Datendatei (auch Speicherdatei genannt) abzuspeichern. Wie durch einen 'Schnappschuß' wird damit der derzeitige Wissensstand konserviert. In einer späteren SPSS-Sitzung können wir durch Einlesen der Datendatei diesen Wissensstand wieder herstellen. Dies hat den Vorteil, daß wir dann die Variablen mit all ihren Attributen nicht mehr neu definieren und die Daten nicht mehr eingeben oder einlesen oder durch Datenmodifikationen neu erzeugen müssen.

In einer SPSS-Datendatei sind die folgenden Informationen festgehalten:

- Datum und Zeit der Erstellung, Dateietikett und Dokumentation der Datei¹⁰.
- Für jede Variable: Variablenname, Variablentyp, Ausgabeformat, Missing-Werte, Variablen- und Wertelabels, Meßniveau. Arbeitsvariablen und Systemvariablen werden allerdings nicht mit aufgenommen.
- Alle Datenwerte (alle Variablenwerte bei allen Fällen).

Eine SPSS-Datendatei ist eine Datei in einem speziellen Binärformat, das nur von SPSS (und einigen wenigen anderen Programmen) und dies sogar nur auf derselben Rechnerplattform lesbar ist.

Neben seinem eigenen Datenformat kann SPSS noch eine Reihe sogenannter Fremdformate lesen bzw. schreiben (z.B. Excel, dBase, Lotus). Die wichtigsten Programmbefehle dafür sind GET TRANSLATE und SAVE TRANSLATE. Neben diesen existieren noch eine Reihe weiterer Befehle, z.B. GET DATA (für Excel ab Version 5), GET BMDP, GET SAS. Auch ein Datenbankzugriff über die ODBC-Schnittstelle ist in SPSS möglich (Befehle GET CAPTURE und SAVE TRANSLATE).¹¹

Natürlich kann SPSS auch mit ASCII-Daten umgehen (siehe 3.2.3 bis 3.2.5).

ÖFFNEN UND SPEICHERN VON DATEN

Aufruf zum Öffnen von Daten

- 1 Menü
 - Datei
 - Öffnen
 - Daten ...
- 2 Dialogfenster "Datei öffnen"
 - Suchen in: [_____] [U]
 - Dateiname: [_____]
 - Dateityp: [SPSS (*.sav) _____] [U]
 - [SPSS/PC+ (*.sys)]
 - [Systat (*.syd)]
 - [Systat (*.sys)]
 - [SPSS portable (.por)]
 - [Excel (*.xls)]
 - [Lotus (*.w*)]
 - [SYLK (*.slk)]
 - [dBase (*.dbf)]
 - [Text(*.txt)]
 - [Daten (*.dat)]
 - [Alle Dateien (*.*)]

3a Dialogfenster "Datei öffnen: Optionen" (nur bei den Dateitypen Excel, Lotus, SYLK)

¹⁰ Informationen dazu erhält man über den FILE LABEL- bzw. den DOCUMENT-Befehl (hier nicht behandelt).

¹¹ Für weitere Informationen zu den hier nicht behandelten Befehlen siehe SPSS Syntax-Handbuch.

- Variablenamen einlesen *(falls die erste Zeile der Tabelle die Variablenamen enthält)*
 Bereich [_____] *(falls nur ein Teil der Tabelle eingelesen werden soll)*

3b Bei den Datentypen Text und Daten wird der "Assistent für Textimport" aufgerufen, der den Benutzer in mehreren Schritten durch die Definition der Daten führt.

Aufruf zum Speichern der Daten

1 Dateneditorfenster aktivieren

2 Menü


 Datei

 Speichern


 bzw.

 Speichern unter...

3 Dialogfenster "Daten speichern unter...."

Speichern in: [_____] []

Dateiname: [_____]

Dateityp: [SPSS (*.sav) _____] []

[SPSS 7.0 (*.sav)]

[SPSS/PC+ (*.sys)]

[SPSS portable (*.por)]

[Tabulator-getrennt (*.dat)]

[ASCII (feste Formate) (*.dat)]

[Excel (*.xls)]

[1-2-3 Ver 3.0 (*.wk3)]

[1-2-3 Ver 2.0 (*.wk1)]

[1-2-3 Ver 1.0 (*.wks)]

[SYLK (*.slk)]

[dBASE IV (*.dbf)]

[dBASE III (*.dbf)]

[dBASE II (*.dbf)]

Variablenamen im Arbeitsblatt speichern *(nur bei Tabulator-getrennt, Excel, 1-2-3, SYLK)*

DER GET- UND SAVE-BEFEHL

Funktion

- Einlesen einer SPSS-Datendatei bzw. Abspeichern der Arbeitsdatei als externe SPSS-Datendatei.

Syntax

get file=file

 [/keep={**all**** | varlist}] [/drop=varlist]

 [/rename=(old_varlist=new_varlist) [(old_varlist=...)]

 [/map].

save outfile=file

 [/keep={**all**** | varlist}] [/drop=varlist]

 [/rename=(old_varlist=new_varlist) [(old_varlist=...)]

 [/map] [/{**compressed** | **uncompressed**}].

Hinweise

- Beim Unterbefehl **file=** ist der Name der SPSS-Datendatei anzugeben, eventuell incl. Laufwerk und Pfad (alles in Hochkommata setzen!).
- Mit dem Unterbefehlen **/keep** und **/drop** kann man für das Speichern bzw. Öffnen eine Teilmenge der Variablen auswählen. Damit ist es auch möglich, die Variablen umzusortieren.
- Mit dem Unterbefehl **/rename** kann man den Variablen neue Namen geben.
- Der Unterbefehl **/map** bewirkt die Ausgabe eine Übersicht über die Variablen in der Arbeitsdatei und den entsprechenden Namen in der SPSS-Datendatei.
- Der Unterbefehl **/compressed** bewirkt eine Ausgabe in komprimierter Form (kleine ganze Zahlen werden im 1-Byte-Format abgespeichert). Dadurch verbraucht die Datendatei weniger externen Speicher.
- Der SAVE-Befehl hat den Status eines Prozedurbefehls.
- Mithilfe des DISPLAY-Befehls kann man sich die Informationen über die abgespeicherten Variablen oder die Dokumentation der Datendatei ausgeben lassen: DISPLAY DICTIONARY bzw. DISPLAY DOCUMENTS.
- Neben dem GET-Befehl existieren noch einige komplexere Befehle zum Einlesen von SPSS-Dateien. Beispielsweise können mit Hilfe des ADD FILES- bzw. MATCH FILES-Befehls verschiedene SPSS-Datendateien zusammengefügt werden.

Beispiele

(a)

```
data list file='a:\data\beispiel.dat'
  /nummer 1-4 geschl 6(a) alter 8-10 gewicht 12-16(1) beruf 18-39(a)
  tablette 40 kopf foehn hals ohr schwind brech magen herz 42-49.
compute altgrup=trunc(alter/10).
count beschw=kopf to herz(1).
save outfile='a:\uebung\beispiel.sav'.
```

(b)

```
get file='a:\uebung\beispiel.sav'.
crosstabs tables=altgrup by tablette, beschw.
```

Bei (a) werden Rohdaten von einer ASCII-Datei gelesen, zwei neue Variablen (ALTGRUP und BESCHW) gebildet und am Ende als SPSS-Datendatei abgespeichert.

Bei (b) wird diese Datei eingelesen und zwei Kreuztabellen erstellt.

DER GET TRANSLATE- UND DER SAVE TRANSLATE-BEFEHL

Funktion

- Einlesen von Daten in einem Fremdformat bzw. Abspeichern der Arbeitsdatei auf eine Datei in einem Fremdformat.

Syntax

```
get translate file=file
  [/type={sys|wk|xls|dbf|tab|...}]
  [/fieldnames]
  [/range={rangename|start..stop|start:stop}]
  [/keep={all** | varlist}] [/drop=varlist]
  [/rename=(old_varlist=new_varlist) [(old_varlist=...)]
  [/map].
```

```
save translate outfile=file
  [/type={wk3|xls|db4|tab|por|txt|pc|slk|...}]
  [/fieldnames]
  [/keep={all**|varlist}] [/drop=varlist]
  [/rename=(old_varlist=new_varlist) [(old_varlist=...)]
```

```
[/map] [/{compressed|uncompressed}]
[/unselected={retain** | delete}]
```

Hinweise

- Beim Unterbefehl **file=** ist der Name der Datendatei anzugeben, eventuell incl. Laufwerk und Pfad (alles in Hochkommata setzen!).
- Der Unterbefehl **type=** definiert das Datenformat:

dbf	dBASE
db4	dBASE IV
pc	SPSS/PC+
por	SPSS portable
slk	Multiplan/Excel
sys	Systat
tab	ASCII tabulator-getrennt
txt	ASCII fester Länge
wk	Lotus 1-2-3
wk3	Lotus 1-2-3 Release 3
xls	Excel

Es gibt noch einige weitere Formate.

- Mit **/fieldnames** werden die Feldnamen (z.B. bei dBASE) bzw. die erste Tabellenzeile (z.B. bei Excel) als Variablennamen übernommen (bei GET TRANSLATE; SAVE TRANSLATE arbeitet analog).
- Der Unterbefehl **/range** dient dazu, aus einer Tabelle (z.B. bei Excel) einen Teilbereich auszuwählen. Fehlt die Angabe, so wird die gesamte Tabelle eingelesen.
- Mit dem Unterbefehlen **/keep** und **/drop** kann man für das Speichern bzw. Öffnen eine Teilmenge der Variablen auswählen.
- Mit dem Unterbefehl **/rename** kann man den Variablen neue Namen geben.
- Der Unterbefehl **/map** bewirkt die Ausgabe einer Übersicht über die Variablen in der Arbeitsdatei und die entsprechenden Namen in der Datendatei.
- Der Unterbefehl **/compressed** bewirkt eine Ausgabe in komprimierter Form. Dadurch verbraucht die Datendatei weniger externen Speicher.
- Für den Fall, daß eine Filterung der Daten eingeschaltet ist, kann mit dem Unterbefehl **/unselected** angegeben werden, ob alle Fälle ausgegeben werden sollen (**retain**) oder nur jene, die das durch FILTER oder USE definierte Kriterium erfüllen (**delete**).

Beispiel

```
get translate file='d:\data\tabelle.xls'
  /fieldnames /range=a2..d65.
save translate outfile='d:\data\tabelle.db4'
  /fieldnames.
```

Es wird eine Datei im Excel 4 Format eingelesen, allerdings nur die Teiltabelle der Spalten a bis d und Zeilen von 2 bis 65. Die Einträge in der ersten Tabellenzeile werden als Variablennamen übernommen. Anschließend werden die Daten im dBASE IV Format auf eine neue Datei geschrieben. Die Variablennamen werden als Namen der Felder übernommen.

3.2.3 Eingabe von ASCII-Daten (DATA LIST, BEGIN/END DATA)

Falls die Rohdaten nicht in Form einer SPSS-Datendatei, sondern als ASCII-Datei vorliegen, müssen die Variablen erst explizit definiert sowie die Datenwerte formatiert eingelesen und nach SPSS gewandelt werden.

EINLESEN VON ASCII-DATEN

Aufruf

1 Menü

 Datei

 Textdaten einlesen

2ff. Der "Assistent für Textimport" wird gestartet, mit dessen Hilfe der Benutzer in mehreren Schritten alle notwendigen Spezifikationen definieren kann.

DER DATA LIST-BEFEHL

Funktion

- Einlesen von Rohdaten aus einer ASCII-Datei

Syntax

```
data list file = dateiname [{fixed|free|list}] [records = {1|n}] [{table|notable}]
/[1|eingabezeile}] varlist startspalte [- endspalte] [(format)]
[varlist...]
/[2|eingabezeile}]...
[/...].
```

Bedeutung der Parameter

- **fixed**: festes Format (spaltengebunden)
- **free**: freies Format (Datenwerte nicht spaltengebunden, nur mit Blanks oder Kommas getrennt; auf einer Datenzeile auch mehr als ein Fall möglich)
- **list**: freies Format, aber jeder Fall beginnt mit einer neuen Datenzeile
- **records**: Anzahl der Datenzeilen pro Fall
- **table/notable**: Ausgabe der Formatliste ins Protokoll: ja/nein
- **format**: {n|a|datenformat}

n: Anzahl der Dezimalstellen hinter dem Komma (Voreinst.: 0)

a: alphanumerischer Typ

datenformat: SPSS-Datenformat (siehe Abschnitt 3.1.1)

Hinweis

- Der angegebene Spaltenbereich wird gleichmäßig auf die entsprechenden Variablen aufgeteilt, wenn nicht eine einzelne Variable, sondern eine Variablenliste vor den Spaltenangaben steht.

Beispiele

(a)

```
data list file='a:\data\psysomat.dat'
/ nummer 1-4 geschl 6(a) alter 8-10 gewicht 12-16(1)
```

```

beruf 18-39(a)  tablette 40
kopf foehn hals ohr schwind brechr magen herz 42-49.
compute altgrup=trunc(alter/10).
count beschw=kopf to herz(1).
crosstabs tables=altgrup by beschw.

```

Hier werden Rohdaten von einer ASCII-Datei gelesen, zwei neue Variablen (ALTGRUP und BESCHW) gebildet und eine Kreuztabelle erstellt. Die ASCII-Datei enthält pro Fall eine Datenzeile. Die Daten werden im festen Format gelesen (notwendig, da die letzten 8 Variablen einspaltig ohne Zwischenraum eingegeben sind). Die Variablen GESCHL und BERUF sind alphanumerisch (Strings), alle anderen sind numerische Variablen. Fehlt bei der Variablen GEWICHT das Dezimaltrennzeichen, so wird automatisch die letzte Ziffer (Spalte 16) als Zehntelziffer interpretiert, die Ziffer davor als Einerziffer usw. Die Variable BERUF ist ein langer String (länger als 8 Zeichen).

(b)

```

data list free file='a:\data\dlist.dat'
  / vp(a) gruppe geschl alter iq frage1 to frage15.
....

```

Hier werden ebenfalls Rohdaten von einer ASCII-Datei gelesen (insgesamt 20 Variablen: VP, GRUPPE, GESCHL, ALTER, IQ, FRAGE1, FRAGE2, ..., FRAGE15), diesmal allerdings im freien Format. Die einzelnen Variablenwerte sind offenbar durch Trennzeichen (z.B. Leerzeichen) voreinander getrennt. Die Variable VP ist übrigens alphanumerisch.

Neben der Eingabe von Rohdaten in eine externe ASCII-Datendatei hat der Benutzer auch die Möglichkeit, seine Daten im ASCII-Format innerhalb des SPSS-Programms unterzubringen (man spricht dann von Inline-Daten). Dies ist vor allem bei kleinen Datensätzen durchaus sinnvoll, vor allem wenn dieselben Daten nur einmal zur Auswertung anstehen.

BEGIN DATA - END DATA

Funktion

- Beginn und Ende von ASCII-Datenzeilen in einem SPSS-Programm

Syntax

```

begin data
  ASCII-datenzeilen
end data.

```

Hinweise

- Der gesamte BEGIN DATA - END DATA-Block muß vor dem ersten SPSS-Prozedurbefehl zu liegen kommen (am besten direkt hinter dem DATA LIST-Befehl).
- Zwischen **begin data** und **end data** dürfen keine SPSS-Befehle liegen.
- Die Datenzeilen dürfen nur max. 80 Spalten lang sein; sie dürfen nicht mit einem Punkt abgeschlossen werden.
- **end data** muß in Spalte 1 beginnen.

Beispiel (entnommen aus dem SPSS-Syntax-Handbuch)

```

data list /xvar 1 yvar zvar 3-12 cvar 14-22(a) jvar 24.
begin data
1  3424  274 ABU DHABI  2
2  39932   86 AMSTERDAM 4
3   8889  232 ATHENS
4   3424  294 BOGOTA    3
5  11323  332 HONG KONG 3
6   323   232 MANILA   1
7   3234  899 CHICAGO  4

```

```
8 78998 2344 VIENNA 3
9 8870 983 ZURICH 5
end data.
means xvar by jvar.
```

Bei diesem Beispiel sind die Daten nicht in einer externen Datei, sondern als Rohdaten im SPSS-Programm selbst (zwischen BEGIN DATA und END DATA) abgelegt. Da sie damit im ASCII-Format vorliegen, müssen sie mit Hilfe des DATA LIST-Befehl gelesen werden. Die Daten werden im festen Format gelesen, die anderen

Formate (FREE oder LIST) hätten wegen des fehlenden Wertes im 3. Datensatz und wegen des Leerzeichens im Wert "HONG KONG" zu fehlerhaften Daten geführt. Die Prozedur MEANS berechnet übrigens eine Varianzanalyse.

3.2.4 Eingabe von ASCII-Daten mit komplexer Datenstruktur

Die Datenstruktur, mit denen der Statistiker üblicherweise zu tun hat, ist die Rechtecksmatrix. Neben dieser gibt es aber noch andere Datenstrukturen wie z.B.

- hierarchische Datenstrukturen (z.B. Daten von Kliniken, Kliniks-Stationen und Patienten auf einer Datei)
- gruppierte Daten (z.B. für jeden Fall mehrere Eingabezeilen, die aber nicht geordnet sind o.ä.)
- Eingabesätze verschiedenen Typs (z.B. pro Fall nicht immer dieselben Variablen oder nicht dasselbe Datenformat)

Mit Hilfe des SPSS-Dateneditors (d.h. interaktiv) ist es lediglich möglich, eine rechteckige Datenmatrix einzulesen. Auch mit einem DATA LIST-Befehl allein kann man keine andere Datenstruktur verarbeiten. Will man eine andere Form von Daten verarbeiten, so gibt es in SPSS nur die Möglichkeit, den FILE TYPE-Befehl zu benutzen. Dieser Befehl taucht in drei Varianten auf:

- FILE TYPE MIXED: Verarbeitung von Eingabesätzen verschiedenen Typs
- FILE TYPE GROUPEd: Verarbeitung von gruppierten Eingabezeilen
- FILE TYPE NESTED: Einlesen einer hierarchischen Datenstruktur

Der FILE-TYPE-Befehl soll hier jedoch nicht weiter behandelt werden.

Genügen diese Möglichkeiten nicht, so kann man mit Hilfe des INPUT PROGRAM-Befehls (s.u.) ein Eingabeprogramm schreiben und damit das Einlesen der Rohdaten steuern.

3.2.5 Erzeugung von Rohdaten

In einem Eingabeprogramm ist es auch möglich, Daten zu erzeugen, ohne von einer Datei zu lesen (von dieser Möglichkeit wird z.B. in Simulationsstudien Gebrauch gemacht). Diese Möglichkeit steht allerdings nur im programmgesteuerten Modus zur Verfügung.

DER INPUT PROGRAM-BEFEHL

Funktion

- Allgemein: Definition eines INPUT Programms für das Einlesen oder Kreieren von Daten
- Hier: Definition eines Programms zur Erzeugung von Rohdaten mit Hilfe von Datentransformationsbefehlen

Syntax

input program.

befehle-zum-kreieren-von-daten

end input program.

Hinweise

- Der Block zwischen **input program.** und **end input program** bildet das Eingabeprogramm und dient zum Kreieren oder Definieren von Fällen. Als Befehle können dabei z.B. verwendet werden: DATA LIST, LOOP/END LOOP, DO IF/END IF, COMPUTE u.a.
Daneben gibt es noch einige spezielle Befehle, die häufig innerhalb eines Eingabeprogramms auftreten, wie z.B.:

- END CASE

Die Ausführung dieses Befehls definiert das Ende der Daten eines Falles, d.h. damit wird die Datendefinition für den aktuellen Fall abgeschlossen und ein neuer Fall begonnen, falls - direkt oder indirekt - weitere Datendefinitionsbefehle folgen. Beim Fehlen eines END CASE-Befehls wird ein Fall durch den DATA LIST-Befehl definiert.

- END FILE

Das Einlesen bzw. Kreieren von Daten wird abgeschlossen. Werden Daten durch einen DATA LIST-Befehl definiert, so steht der END FILE-Befehl im allgemeinen in einem DO IF-Block. Damit wird die Dateneingabe dann vorzeitig beendet.

- REREAD

Die zuletzt gelesene Eingabezeile steht zum nochmaligen Einlesen zur Verfügung.

Beispiele

(a)

* Beispiel nach Schubö et al., 1991, S.222.

```
input program.
  data list file='a:\data\medsoz.dat'
    / abt 51-65 (a).
  do if (abt='PRODUKTION').
    reread.
    data list
      /nummer 1-4 geschl 6 (a) alter 8-10 gewicht 12-16 (1).
    end case.
  else if (abt='ADMINISTRATION').
    reread.
    data list
      /nummer 1-4 geschl 6 (a) alter 8-10 gewicht 20-24 (1) belast 26.
    end case.
  end if.
end input program.
....
```

Zunächst wird von der angegebenen ASCII-Datei lediglich die Variable ABT (Abteilung) eingelesen. Abhängig davon, aus welcher Abteilung der aktuelle Fall stammt, werden verschiedene DATA LIST-Befehle ausgeführt, da offensichtlich die restlichen Variablen (NUMMER, GESCHL, usw.) z.T. in verschiedenem Format vorliegen. Außerdem ist die Variable BELAST offenbar nur für die Angehörigen der Administrationsabteilung erhoben worden.

(b)

```
input program.
  loop #i=1 to 500.
    compute testwert=rnd(50+normal(10)).
  end case.
end loop.
end file.
end input program.
execute.
```

Es werden die Daten von 500 Fällen erzeugt, u.zw. nur eine einzige Variable mit Namen TESTWERT. Diese Variable ist eine Zufallszahl aus einer Normalverteilung mit Mittelwert 50 und Standardabweichung 10. Sie nimmt allerdings nur ganzzahlige Werte an.

(c)

```
input program.
  vector test(9).
  loop #case=1 to 145.
    loop #var=1 to 9.
```

```

    compute test(#var)=rnd(50+normal(10)).
  end loop.
end case.
end loop.
end file.
end input program.
execute.

```

Mit diesem Programm werden die Daten von 145 Fällen erzeugt, u.zw. 9 Variablen TEST1,TEST2,...,TEST9. Diese Variablen sind unabhängig und normalverteilt mit Mittelwert 50 und Standardabweichung 10. Außerdem sollen sie nur ganzzahlige Werte annehmen.

DER CREATE-BEFEHL

Funktion

- Berechnung von Variablen unter Zugrundelegung von Variablenwerten bei anderen (z.B. unmittelbar vorhergehenden) Fällen; die Hauptanwendung ist die Zeitreihenanalyse

Syntax

```

create var=function(var[,parameter,...])
  [var=...].

```

Hinweise

- Es stehen verschiedene Funktionen zur Verfügung. Einige davon sind:
 - csum**(var) der Wert der Ergebnisvariable ist die kumulative Summe der angegebenen Variable
 - diff**(var,n) Differenz der Ordnung n in der angegebenen Variable
 - lag**(var,n) der Wert der Ergebnisvariable beim Fall *i* ist gleich dem Wert beim Fall *i-n* in der angegebenen Variable
 - lead**(var,n) der Wert der Ergebnisvariable beim Fall *i* ist gleich dem Wert beim Fall *i+n* in der angegebenen Variable

Beispiel

```

create v=csum(u).
create w=diff(u,1).
create x=diff(u,2).
create y=lag(u,1).
create z=lead(u,1).

```

Nehmen wir an, die Variable u hätte die Werte 1,3,7,12,9,20,27, so sähe die Arbeitsdatei anschließend so aus:

u	v	w	x	y	z
1	1	,	,	,	3
3	4	2	,	1	7
7	11	4	2	3	12
12	23	5	1	7	9
9	32	-3	-8	12	20
20	52	11	14	9	27
27	79	7	-4	20	,

3.2.6 Dateien zusammenfügen (ADD FILES, MATCH FILES)

Zuweilen sollen Daten aus verschiedenen Dateien gemeinsam ausgewertet werden. Für diesen Fall bietet SPSS die Möglichkeit, Dateien zusammenzufügen. Diese Dateien müssen SPSS-Datendateien sein (eine davon kann auch die Arbeitsdatei sein). Nach dem Zusammenfügen steht das Ergebnis in der Arbeitsdatei.

Für das Zusammenfügen gibt es verschiedene Möglichkeiten:

"Fälle hinzufügen"

Die verschiedenen Dateien enthalten im Normalfall dieselben ("gepaarten", d.h. mit identischem Namen und Datentyp definierten) Variablen, aber unterschiedliche Fälle: die Fälle werden jeweils in der Arbeitsdatei unten angefügt. Einzelne gepaarte Variablen können auch ausgeschlossen werden. Aufgenommene nicht gepaarte Variablen ergeben bei den Fällen, bei denen sie nicht definiert sind, den Wert SYSMIS. Es ist aber auch möglich, jeweils zwei nicht gepaarte Variablen explizit als zusammengehörig zu definieren und in die Liste aufzunehmen. Schließlich können Variablen auch umbenannt werden.

"Variablen hinzufügen"

Die einzelnen Dateien enthalten verschiedene Variablen zu denselben Fällen: die weiteren Variablen werden in der Arbeitsdatei rechts hinten angefügt. Die Fälle aus den einzelnen Dateien werden dabei entweder 1:1 der Reihe nach zugeordnet (Voraussetzung: die Fälle müssen in den verschiedenen Datendateien gleich sortiert sein und kein Fall darf fehlen) oder anhand übereinstimmender Werte in einer oder mehreren Schlüsselvariablen (Voraussetzung: die Namen der Schlüsselvariablen müssen in den Datendateien übereinstimmen und diese müssen nach den Schlüsselvariablen aufsteigend sortiert sein). Dabei kann auch eine Datei (bzw. die Arbeitsdatei) als Schlüsseltabelle oder Tabellenindexdatei dienen, d.h. als Datei, deren Daten für jeden Fall mehreren Fällen der anderen Datendateien zugeordnet werden können. Wenn aufgrund der Werte in den Schlüsselvariablen keine vollständige Zuordnung möglich ist (d.h. ein Datensatz in einer der Datendateien fehlt), so werden die fehlenden Variablen auf SYSMIS gesetzt. Haben mehrere Fälle identische Werte in den Schlüsselvariablen, so erfolgt die Zuordnung der entsprechenden Fälle in der Reihenfolge, in der sie in den Dateien stehen. Wenn in den verschiedenen Datendateien irgendwelche Variablenamen übereinstimmen, so wird nur jeweils der erste Wert übernommen (man kann aber "doppelte" Variablen umbenennen, um sie dann hinzuzunehmen). Generell ist es auch hier wiederum möglich, einzelne Variablen auszuschließen.

Beim Zusammenfügen von Daten gibt es bei der interaktiven Arbeitsweise eine Einschränkung: es können lediglich die Daten der aktuellen Arbeitsdatei mit den Daten aus einer einzelnen externen Datei zusammengefügt werden.

DATEIEN ZUSAMMENFÜGEN

Aufruf

1 Menü

Daten

Dateien zusammenfügen >

Fälle hinzufügen...

Variablen hinzufügen...

2a Dialogfenster "Fälle hinzufügen: Datei lesen"

Dateiname: [_____] (die Datei auswählen, die hinzugefügt werden soll)

2b Dialogfenster "Variablen hinzufügen: Datei lesen"

Dateiname: [_____] (die Datei auswählen, die hinzugefügt werden soll)

3a Dialogfenster "Fälle hinzufügen aus ..."

Variablen in der neuen Arbeitsdatei:

[_____]

(hier stehen schon die Namen aller gepaarten Variablen; aus dieser Liste können Variablen ausgeschlossen werden; weitere Variablen lassen sich noch hinzufügen u.zw. entweder einzeln oder

[Umbenennen]

Datei-Indikator als Variable:
[_Quelle01_]

jeweils zwei Variablen mit [Paar] als
zusammengehöriges Variablenpaar
(damit kann man nicht-gepaarte Variablen
umbenennen)
(damit wird eine neue logische Variable definiert,
die angibt, ob der Fall aus der hinzugefügten Datei
stammt)

3b Dialogfenster "Variablen hinzufügen aus ..."

Neue Arbeitsdatei:
[_____]

(hier stehen schon die Namen aller in Frage
kommenden Variablen, d.h. jener Variablen, die
nicht mehrfach vorhanden sind; aus dieser Liste
können Variablen ausgeschlossen werden; weitere
Variablen lassen sich noch hinzufügen, wenn sie
umbenannt werden)

[Umbenennen]

Fälle mittels Schlüsselvariablen verbinden:
 Beide Dateien liefern Fälle
 Externe Datei ist Schlüsseltabelle
 Arbeitsdatei ist Schlüsseltabelle
 Schlüsselvariablen:
 [_____]

(damit kann man ausgeschlossene Variablen
umbenennen)
(falls die Fälle nicht 1:1, sondern über die Werte
in Schlüsselvariablen verbunden werden sollen)
(falls eine Schlüsseltabelle verwendet
werden soll)
(hier sind die Schlüsselvariablen anzugeben)

Datei-Indikator als Variable:
[_Quelle01_]

(damit wird eine neue logische Variable definiert,
die angibt, ob der Fall aus der hinzugefügten Datei
stammt)

DER ADD FILES-BEFEHL

Funktion

- Zusammenfügen von Fällen aus verschiedenen SPSS-Datendateien

Syntax

```
add files
  file={file}*} [/rename=(old_varlist=new_varlist)] [/in=var]
  /file=... [/rename=...] [/in=...]
  [/file=... ]
  [/map]
  [/keep={all}*|varlist}] [/drop=varlist]
  [/by varlist] [/first=var] [/last=var].
```

Hinweise

- Mit **file=** sind die zusammenzufügenden Dateien anzugeben ('*' steht für die Arbeitsdatei). Dahinter können jeweils mit **rename=** Variablen umbenannt werden. Mit dem Unterbefehl **in=** wird eine neue Variable erzeugt, die den Wert 1 oder 0 erhält je nachdem, ob der jeweilige Fall aus dieser Datei stammt oder nicht.
- Mit **map=** wird eine Liste über die Variablen in der Ergebnisdatei erstellt. Dieser Unterbefehl darf auch mehrfach vorkommen.
- Mit **keep=** bzw. **drop=** kann man die zusammenzufügenden Variablen auswählen.
- Der Unterbefehl **by=** bewirkt, daß die Fälle in der Ergebnisdatei nach den angegebenen Variable sortiert sind. Mit **first=** bzw. **last=** kann dann noch je eine neue Variable angegeben werden, die den Wert 1 oder 0 erhält je nachdem, ob der aktuelle Fall jeweils der erste bzw. letzte Fall mit identischen Werten in den **by**-Variablen ist oder nicht.

Beispiel

```
add files
  /file='d:\diplom\gruppe1.sav' /rename test1=y /in=grp1
```

```

/file='d:\diplom\gruppe2.sav' /rename test2=y /in=grp2
/file='d:\diplom\gruppe3.sav' /rename test3=y /in=grp3
/by vpid.
execute.

```

Die Daten aus den drei SPSS-Dateien werden zusammengefügt und sortiert nach der Variable VPID in die Arbeitsdatei geschrieben. Die Variablen TEST1, TEST2 bzw. TEST3 werden in Y umbenannt. Zudem werden noch drei logische Variablen GRP1, GRP2 und GRP3 erzeugt, die angeben, ob der aktuelle Fall aus der jeweiligen Gruppe stammt.

DER MATCH FILES-BEFEHL

Funktion

- Zusammenfügen von Variablen zu gleichen Fällen aus verschiedenen SPSS-Datendateien

Syntax

match files

```

{file|table}=[{file|*}] [rename=(old_varlist=new_varlist)] [in=var]
/{file|table}=... [rename=...] [in=...]
[/{file|table}=...]
[map]
[keep={all|*|*|varlist}] [/drop=varlist]
[/by varlist] [/first=var] [/last=var].

```

Hinweise

- Mit **file=** sind die zusammenzufügenden Dateien anzugeben (*' steht für die Arbeitsdatei). Der Unterbefehl **table=** erlaubt die Angabe einer Datei, die als Schlüsseltabelle verwendet werden soll (erfordert dann auch den Unterbefehl **by**). Dahinter können jeweils mit **rename=** Variablen umbenannt werden. Mit dem Unterbefehl **in=** wird eine neue Variable erzeugt, die den Wert 1 oder 0 erhält je nachdem, ob der jeweilige Fall aus dieser Datei stammt oder nicht.
- Falls die Fallzuordnung nicht 1:1, sondern anhand der Werte von Schlüsselvariablen erfolgen soll, müssen diese im Unterbefehl **by=** aufgeführt werden.
- Mit **map=** wird eine Liste über die Variablen in der Ergebnisdatei erstellt. Dieser Unterbefehl darf auch mehrfach vorkommen.
- Mit **keep=** bzw. **drop=** kann man die zusammenzufügenden Variablen auswählen.
- Mit **first=** bzw. **last=** kann je eine neue Variable angegeben werden, die den Wert 1 oder 0 erhält je nachdem, ob der aktuelle Fall jeweils der erste bzw. letzte Fall mit identischen Werten in den **by**-Variablen ist oder nicht.

Beispiel

```

match files
/file='d:\diplom\vorher.sav' /rename y=yvor /in=vorher
/file='d:\diplom\nachher.sav' /rename y=ynach /in=nachher
/by vpid.
execute.

```

Die Daten aus den drei SPSS-Dateien werden anhand der Variable VPID zusammengefügt und in die Arbeitsdatei geschrieben. Die Variable Y wird in YVOR bzw. YNACH umbenannt. Zudem werden noch zwei logische Variablen VORHER und NACHHER erzeugt, die angeben, ob die Daten für den aktuellen Fall aus der erstgenannten und/oder aus der zweitgenannten Datei stammen.

3.2.7 Löschen der Arbeitsdatei (NEW FILE)

Zuweilen möchte der Benutzer anschließend an eine Auswertung in derselben SPSS-Sitzung ganz neue Daten eingeben und zu diesem Zwecke die Arbeitsdatei (d.h. alle Variablen, Labels, Werte usw.) löschen. Dazu dient der folgende Befehl.

LÖSCHEN DER ARBEITSDATEI

Aufruf

- 1 Menü
 - Datei
 - Neu ▷
 - Daten

DER NEW FILE-BEFEHL

Funktion

- Löschen der Arbeitsdatei

Syntax

new file.

Hinweise

- Vor einem Datendefinitionsbefehl wie GET, DATA LIST usw. ist kein NEW FILE-Befehl notwendig, da damit automatisch die Arbeitsdatei ersetzt wird.

Beispiel

```
get file='a:\data\beispiel.sav'.
frequencies variables=alter.
new file.
```

Zunächst werden Daten eingelesen, eine Häufigkeitstabelle erstellt und anschließend die Arbeitsdatei gelöscht. Nun hat man wiederum eine leere Arbeitsdatei und könnte z.B. eine gänzlich neue Auswertung beginnen, d.h. Variablen definieren, Daten eingeben, usw.

3.2.8 Löschen einer externen Datei (ERASE)

Zuweilen werden im Laufe einer umfangreicheren Auswertung temporäre Dateien angelegt, die anschließend wieder gelöscht werden sollen. Dies ist innerhalb von SPSS möglich unter Zuhilfenahme des ERASE-Befehls.

DER ERASE-BEFEHL

Funktion

- Löschen einer (externen) Datei

Syntax

erase file [=] 'file'.

Beispiel

```
get file='a:\data\demodat.sav'.
write outfile='c:\temp.dat' records=2
  / 'v' 1 groe_v 3-5
  / 'm' 1 groe_m 3-5.
execute
data list free file='c:\temp-dat'
  /geschl(a) groesse.
ttest groups=geschl('m' 'w')
  /variables=groesse.
erase file='c:\temp.dat'.
```

Die Körpergrößen für Vater und Mutter sollen verglichen werden. Dazu werden die beiden Variablen auf eine ASCII-Datei so ausgegeben, daß Vater und Mutter als zwei verschiedene Fälle eingelesen werden können. Die Datei c:\temp.dat wird anschließend nicht mehr gebraucht und deshalb gelöscht.

3.2.9 Ausgabe von Daten im ASCII-Format (WRITE, PRINT..., LIST)

SPSS bietet die Möglichkeit, individuelle Werte ins Listing oder in eine ASCII-Datei auszugeben. Die Ausgabe von Rohdaten ins Listing ist z.B. dann sinnvoll, wenn die Daten auf ihre Korrektheit überprüft werden sollen. Die Ausgabe von Rohdaten auf eine ASCII-Datei ist oft die einzige Möglichkeit des Datenaustausches zwischen verschiedenen Programmen oder Rechnersystemen. Damit können beispielsweise die in SPSS vorverarbeiteten und transformierten Daten als Eingabe für ein anderes Programm benutzt werden, auch wenn dieses das SPSS-Datenformat und auch die anderen von SPSS benutzbaren Ausgabeformate (z.B. dBase, Excel) nicht versteht, was häufig beim Wechsel der Rechnerplattform der Fall ist.

AUSGABE VON DATEN AUF EINE ASCII-DATEI

Aufruf

- 1 Dateneditorfenster aktivieren
- 2 Menü
Datei
Speichern unter...

- 3 Dialogfenster "Daten speichern unter..."
- Speichern in: [_____][👉]
- Dateiname: [_____]
- Dateityp: [ASCII (feste Formate) (*.dat)_____][👉]
 oder: [Tabulator-getrennt (*.dat)_____][👉]
- Variablennamen im Arbeitsblatt speichern (*nur bei Tabulator-getrennt!*)

Hinweise

- Das Ausgabeformat (z.B. Spaltenanzahl, Dezimalstellen) wird durch den jeweiligen Variablentyp definiert.
- Beim Dateityp "ASCII (feste Formate)" werden die einzelnen Datenwerte ohne Trennzeichen abgelegt, beim Dateityp "Tabulator-getrennt" hingegen mit Tabulatorzeichen voneinander getrennt.
- Beim Dateityp "Tabulator-getrennt" können die Variablennamen als erste Zeile mit auf die ASCII-Datei geschrieben werden (Voreinstellung).

DER WRITE-BEFEHL

Funktion

- Ausgabe von Rohdaten auf eine ASCII-Datei (oder ins Listing)

Syntax

```
write [outfile = dateiname] [records = {1 | n}] [{table | notable}]
/[{1 | ausgabezeile}] varlist [{startspalte [- endspalte] [(n|a)]} | (format) | *}]
[varlist...]
/[{2 | ausgabezeile}]...]
[/...].
```

Bedeutung der Parameter

- **records**: Anzahl der Datenzeilen pro Fall
- **table/notable**: Ausgabe der Formatliste ins Protokoll: ja/nein
- Angaben zum Datenformat

<i>n</i>	Anzahl der Dezimalstellen hinter dem Komma (Voreinst.: 0)
a	alphanumerischer Typ
<i>format</i>	Formatspezifikation nach der FORTRAN-Syntax

Hinweise

- Der WRITE-Befehl dient in erster Linie zur Ausgabe von Rohdaten auf eine ASCII-Datei. Ist keine Ausgabedatei angegeben, so erfolgt die Ausgabe ins Listing (d.h. ins Ausgabefenster).
- In der Variablenliste sind jene Variablen zu spezifizieren, deren Werte ausgegeben werden sollen.
- Nach der Variablenliste kann jeweils festgelegt werden, in welche Spalten und in welchem Format die einzelnen Werte geschrieben werden sollen (diese Angaben sind analog zu jenen im DATA LIST-Befehl). Das Ausgabeformat kann jedoch auch in einem getrennten FORMATS- bzw. WRITE FORMATS-Befehl spezifiziert werden. Bei Fehlen jedwelcher Angaben zum Ausgabeformat wird ein voreingestelltes Format (z.B. das Format F8.2) benutzt.
- Neben Variablenwerten kann man auch Zeichenketten-Konstanten in die Ausgabedatei schreiben (Hochkommata nicht vergessen!).
- Eine mit dem WRITE-Befehl erzeugte Datei kann wieder als Eingabedatei für ein anderes Programm benutzt werden.
- Der WRITE-Befehl hat nicht den Status eines Prozedurbefehls, d.h. es ist bei Fehlen eines Prozeduraufrufs nach dem WRITE-Befehl der EXECUTE-Befehl auszuführen.

Beispiele

(a)

```
data list file='a:\data\items.dat'
  / vp 1-2(a)  item1 to item40 4-43.
count rw=item1 to item40(1).
write outfile='a:\data\rohwerter.dat'
  / vp 1-2(a)  rw 4-6.
execute.
```

Hier werden von einer ASCII-Datei jeweils ein Vp-Code, sowie die Einzelergebnisse in den 40 Aufgaben eines Leistungstests eingelesen und jeweils der Rohwert, d.h. die Anzahl der richtigen Antworten (1=richtig, 0=falsch) berechnet. Dieser Rohwert wird anschließend zusammen mit dem VP-Code auf eine ASCII-Datei ausgegeben.

(b)

```
data list file='a:\data\einzeil.dat'
  / vp 1-2 (a)  test1 to test3 3-11.
write outfile='a:\data\dreizeil.dat' records=3
  / vp 1-2 (a)  '1' 3  test1 5-7
  / vp 1-2 (a)  '2' 3  test2 5-7
  / vp 1-2 (a)  '3' 3  test3 5-7.
execute.
```

Die drei Variablen TEST1, TEST2 und TEST3, die auf der externen Datei jeweils auf einer Datenzeile hinter dem Vp-Code abgelegt sind, werden durch den obigen WRITE-Befehl auf drei hintereinander liegende Datenzeilen verteilt: in den ersten beiden Spalten steht in jeder Datenzeile der Vp-Code, in Spalte 3 als Konstante (1, 2 bzw. 3) noch die Nummer der jeweiligen Datenzeile; Spalte 4 bleibt frei; in den Spalten 5 bis 7 schließlich steht der Wert der jeweiligen Testvariable (TEST1, TEST2 bzw. TEST3).

DER PRINT- UND PRINT EJECT-BEFEHL

Funktion

- Ausgabe von Rohdaten ins Listing (oder auf eine ASCII-Datei)

Syntax

```
print [outfile=filename] [records={1n}] [notable|table].
  /[{1|ausgabezeile}] varlist [{startspalte [-endspalte] [(n|a)]} | (format) | *}]
  [varlist...]
  /[{2|ausgabezeile}]...
  [/...].
```

```
print eject [outfile=filename] [records={1n}] [notable|table].
  /[{1|ausgabezeile}] varlist [{startspalte [-endspalte] [(n|a)]} | (format) | *}]
  [varlist...]
  /[{2|ausgabezeile}]...
  [/...].
```

Bedeutung der Parameter

- **records**: Anzahl der Datenzeilen pro Fall
- **table/notable**: Ausgabe der Formatliste ins Protokoll: ja/nein
- Angaben zum Datenformat

<i>n</i>	Anzahl der Dezimalstellen hinter dem Komma (Voreinst.: 0)
<i>a</i>	alphanumerischer Typ
<i>format</i>	Formatspezifikation nach der FORTRAN-Syntax

Hinweise

- Der PRINT- bzw. PRINT EJECT-Befehl dient in erster Linie zur Ausgabe von Rohdaten ins Listing (d.h. ins Ausgabefenster). Ist jedoch eine Ausgabedatei angegeben, so werden die entsprechenden Daten im ASCII-Format in diese Datei geschrieben.
- Beim PRINT EJECT-Befehl erfolgt vor der eigentlichen Ausgabe ein Seitenvorschub.
- Der PRINT-Befehl entspricht weitgehend dem WRITE-Befehl. Im Unterschied zum WRITE-Befehl wird beim PRINT-Befehl zwischen die auszugebenden Werte automatisch jeweils ein Leerzeichen gesetzt, falls keine Formatangaben vorliegen. Außerdem werden Druckersteuerzeichen in die Ausgabe geschrieben.
- In der Variablenliste sind jene Variablen zu spezifizieren, deren Werte ausgegeben werden sollen.
- Nach der Variablenliste kann jeweils festgelegt werden, in welche Spalten und in welchem Format die einzelnen Werte geschrieben werden sollen (diese Angaben sind analog zu jenen im DATA LIST-Befehl). Das Ausgabeformat kann jedoch auch in einem getrennten FORMATS- bzw. PRINT FORMATS-Befehl spezifiziert werden. Bei Fehlen jedwelcher Angaben zum Ausgabeformat wird ein voreingestelltes Format (z.B. das Format F8.2) benutzt.
- Neben Variablenwerten kann man auch Zeichenketten-Konstanten in die Ausgabedatei schreiben (Hochkommata nicht vergessen!).
- Der PRINT- bzw. PRINT EJECT-Befehl hat nicht den Status eines Prozedurbefehls, d.h. es ist bei Fehlen eines Prozeduraufrufs der EXECUTE-Befehl auszuführen.

Beispiel

```
get file='a:\data\beispiel.sav'.
do if $casenum eq 1.
  print eject /'Die größten Leute'
             /'====='.
end if.
do if groesse ge 180.
  print /'Fall Nr.' $casenum (f3) groesse (f6) ' cm'.
end if.
execute.
```

Zunächst wird nach einem Seitenvorschub eine Überschrift ins Ausgabefenster geschrieben und anschließend werden alle Fälle aufgelistet, deren Körpergröße mindestens 180 cm beträgt.

DER PRINT SPACE-BEFEHL*Funktion*

- Ausgabe von Leerzeilen ins Listing (oder auf in eine ASCII-Datei)

Syntax

```
print space [outfile=filename] [{numerischer-ausdruck}]
```

Hinweise

- Ist eine Ausgabedatei angegeben, so erfolgt die Ausgabe nicht ins Listing, sondern in diese Datei.
- Die Anzahl der auszugebenden Leerzeilen ergibt sich aus dem Wert des numerischen Ausdrucks.

Beispiel

```
get file='a:\data\beispiel.sav'.
print /'Fall Nr.' $casenum (f3) groesse (f6) ' cm'.
do if mod($casenum,10) eq 0.
  print space.
end if.
execute.
```

Es wird die Fallnummer und die Körpergröße ausgegeben und nach jeweils 10 Fällen zusätzlich eine Leerzeile.

Mit Hilfe des folgenden Befehls ist es möglich, Rohdaten in Tabellenform ins Listing zu schreiben.

DER LIST-BEFEHL

Funktion

- Ausgabe von Rohdaten ins Listing

Syntax

list

```
[[variables]={all**|varlist}]
[/format={wrap**|single}] [{unnumbered**|numbered}]
[/cases={from {1**|n}] [to {eof**|n}] [by {1**|n}]
```

Hinweise

- Die Ausgabe erfolgt im Textformat.
- Überlange Zeilen werden bei der Ausgabe umgebrochen.
- Im Gegensatz zum PRINT-Befehl ist der LIST-Befehl eine Prozedur.

Beispiel

```
get file='a:\data\beispiel.sav'.
list cases=from 1 to 5.
```

Für die ersten 5 Fälle werden die Werte aller Variablen ausgegeben.

3.2.10 Definition von Kopfzeilen für die Ausgabe (TITLE, SUBTITLE)

Zu Dokumentationszwecken kann der Benutzer für die Ausgabe eine oder zwei Kopfzeilen definieren. Diese werden dann auf jeder Ausgabeseite als Seitenüberschriftszeilen ausgegeben. Wird diese Option nicht benutzt, so bleiben die Kopfzeilen leer.

Bei der interaktiven Arbeitsweise werden Kopfzeilen mit Hilfe des Viewers definiert. Dies hat im Ausgabefenster über das Menü "Datei|Optionen|Seite einrichten..." zu erfolgen.

DER TITLE- UND SUBTITLE-BEFEHL

Funktion

- Definition einer ersten bzw. zweiten Seitenüberschriftszeile

Syntax

title []text[].

subtitle []text[].

Hinweise

- TITLE- und/oder SUBTITLE-Befehle können an jeder Stelle im SPSS-Programm stehen und wirken sich ab dort aus.
- Die Hochkommata kann man weglassen, wenn der Titel keine Sonderzeichen (auch keine Blanks) enthält.

Beispiel

```
title 'Mein erstes SPSS-Programm'.
get file='a:\data\beispiel.sav'.
descriptives variables=all.
```

Auf jeder Ausgabeseite erscheint als Überschrift der Text "Mein erstes SPSS-Programm".

3.3 Variablentransformationen

In der Praxis kommt es häufig vor, daß die Rohdaten nicht direkt in die statistische Analyse eingehen, sondern erst noch transformiert werden oder aus ihnen neue, d.h. die letztlich interessierenden Variablen erzeugt werden. Beispiele hierfür sind:

- Berechnung des Therapieerfolgs, gemessen als Differenz eines Symptomwerts vor und nach einer Behandlung
- Einteilung einer Gruppe in Altersklassen, wenn das Alter der Personen vorliegt
- Auszählung der richtigen Antworten bei einem Leistungstest, wenn nur die Einzelantworten vorliegen.

Allgemein ist zu sagen daß Datentransformationen zeilenweise abgearbeitet werden. Dies bedeutet, daß eine angegebene Datentransformation getrennt für jede Zeile der Arbeitsdatei, also intraindividuell, ausgeführt wird. Im Gegensatz dazu arbeiten die Statistik- und Graphikprozeduren spaltenweise, d.h. sie führen ihre Berechnungen für Gruppen von Fällen durch. Der Unterschied in der Arbeitsweise von SPSS bei Datentransformationen und bei Prozeduren soll an einem Beispiel verdeutlicht werden: angenommen, die Arbeitsdatei enthalte die Daten von 60 Personen mit 10 Variablen, die 10 aufeinanderfolgende Reaktionszeiten bei einem Reaktionsversuch darstellen, und man möchte für jede Person die mittlere Reaktionszeit berechnen, so benutzt man die Datentransformation "Berechnen" (den Datentransformationsbefehl COMPUTE) unter Zuhilfenahme der Funktion MEAN und speichert diesen Wert auf eine neue Variable ab; möchte man hingegen die Gruppenmittelwerte der 10 Reaktionszeiten sowie eventuell der mittleren Reaktionszeit berechnen, so hat dies mit Hilfe einer Statistikprozedur zu geschehen (z.B. mit der Prozedur FREQUENCIES).

Es gibt verschiedene Klassen von Datentransformationen:

- Umkodieren
- Berechnen
- Zählen.

Daneben unterscheidet man noch bedingte von unbedingten Datentransformationen. Unbedingte Datentransformationen werden für jede Beobachtungseinheit ausgeführt, bedingte nur dann, wenn eine anzugebende Bedingung für die jeweilige Beobachtungseinheit erfüllt ist.

3.3.1 Umkodieren (RECODE, AUTORECODE)

Mit Hilfe dieser Datentransformation können beliebige Werte einer oder mehrerer Variablen durch neue Werte ersetzt werden.

Die Umkodierung kann innerhalb der Originalvariable oder auf eine neue Variable vorgenommen werden, wobei dann die Originalvariable bestehen bleiben kann. Wenn die Umkodierung allerdings mit einem Typwechsel verbunden ist (z.B. wenn beim Geschlecht aus 'm' eine 1 und aus 'w' eine 2 werden soll), so ist die Umkodierung nur auf eine neue Variable möglich.

Normalerweise muß für das Umkodieren angegeben werden, welche neuen Werte den ursprünglichen Werten zugeordnet werden sollen. Eine Ausnahme bildet hier das automatische Umkodieren. Bei diesem ist die Zuordnungsvorschrift bereits implizit definiert. Zudem kann hierbei das Umkodieren nur auf neue Variablen erfolgen.

Es folgen einige Anwendungsbeispiele, bei denen eine Umkodierung von Nutzen ist:

- Auswertung eines Multiple-Choice-Tests durch Umkodierung der richtigen Antwortalternative(n) in die Zahl 1 und der falschen in die Zahl 0
- Umpolung von Antworten auf einzelne Items eines Fragebogens (d.h. aus 0 wird 1 und aus 1 wird 0)
- Nachträgliche Klasseneinteilung einer Variable
- Ersetzen der ursprünglichen alphanumerischen Datenwerte durch die Zahlenwerte 1, 2, 3, usw. (beispielsweise müssen bei manchen Prozeduren die unabhängigen Variablen vom numerischen Typ sein)

UMKODIEREN

Aufruf

1 Menü

Transformieren

Umkodieren ▷

In dieselben Variablen...

In andere Variablen...

2a Dialogfenster "Umkodieren in dieselben Variablen"

Variablen: [_____]

(aus der Variableliste jene Variablen auswählen, die umkodiert werden sollen)

[Falls...]

(nur für bedingtes Umkodieren)

[Alte und neue Werte...]

2b Dialogfenster "Umkodieren in andere Variablen"

Eingabevar. -> Ausgabevar.:

[_____]

(aus der Variablenliste jene Variablen auswählen, die umkodiert werden sollen)

Ausgabevariable:

Name [_____]

(für die neue Variable Angaben eintragen und [Ändern] wählen; damit wird das Fragezeichen in der Variablen-Zuordnungsliste durch diesen Variablennamen ersetzt)

Label [_____]

[Falls...]

(nur für bedingtes Umkodieren)

[Alte und neue Werte...]

Mit Hilfe des daraufhin erscheinenden Dialogfensters muß nun die Liste der Zuordnungsregeln für das Umkodieren erstellt werden.

3a Dialogfenster "Umkodieren in dieselben Variablen: Alte und neue Werte"

Alter Wert:

Wert [_____]

Systemdefiniert fehlend

System- oder benutzerdefinierte fehlende Werte

Bereich: [_____] bis [_____]

Bereich: Kleinster Wert bis [_____]

Bereich: [_____] bis größter Wert

Alle anderen Werte

(damit kann man alle bisher noch nicht aufgeführten Werte ansprechen)

Neuer Wert:

Wert [_____]

Systemdefiniert fehlend

(nachdem alter und neuer Wert spezifiziert sind, wird mit [Hinzufügen] der Eintrag in die Zuordnungsliste "Alt -> Neu" aufgenommen)

3b Dialogfenster "Umkodieren in andere Variablen: Alte und neue Werte"

Alter Wert:

Wert [_____]

Systemdefiniert fehlend

System- oder benutzerdefinierte fehlende Werte

Bereich: [_____] bis [_____]

Bereich: Kleinster Wert bis [_____]

Bereich: [_____] bis größter Wert

Alle anderen Werte

(damit kann man alle bisher noch nicht aufgeführten Werte ansprechen)

Neuer Wert:

Wert [_____]

Systemdefiniert fehlend

(nachdem alter und neuer Wert spezifiziert sind, wird mit [Hinzufügen] der Eintrag in die Zuordnungsliste "Alt -> Neu" aufgenommen)

- Alte Werte kopieren
- Ausgabe der Variablen als Strings Breite: [____]
- Num. Strings in Zahlen umwandeln ('5'->5)

Hinweis

- Der Schalter "Falls..." dient der Durchführung bedingter Umkodierungen (wird später behandelt).

DER RECODE-BEFEHL*Funktion*

- Umkodieren von Variablen

Syntax

```
recode varlist(value_list=value) [(valu_list=value)...] [into varlist]
  [/varlist...].
```

Hinweise

- Die Variablen müssen bereits definiert sein.
- Der RECODE-Befehl wird von links nach rechts abgearbeitet. Dabei wird der Wert einer Beobachtung nur einmal recodiert. Nicht gefundene Werte bleiben unverändert.
- Stringwerte sind in Hochkommata zu setzen.
- Die Werteliste kann bei numerischen Variablen auch Wertebereiche enthalten (Schlüsselwörter **thru**, **lo[west]**, **hi[ghest]**).
- Alle bisher nicht aufgeführten Werte können mit dem Schlüsselwort **else** angesprochen werden.
- In der Werteliste können auch die Schlüsselwörter **missing** und **sysmis** auftauchen, als neuer Wert auch **sysmis**.
- Wird in einem RECODE-Befehl das Schlüsselwort **into** benutzt, so wird das Ergebnis der Rekodierung einer (oder mehreren) neuen Variablen zugewiesen, und die Originalvariablen bleiben erhalten. Dabei kann als neuer Wert auch das Schlüsselwort **copy** auftreten (d.h. die in der Werteliste aufgeführten Variablenwerte sollen unverändert übernommen werden). Achtung bei Missing-Werten: der Missing-Status wird nicht mitkopiert!
- Beim Umkodieren auf eine neue Variable ist es auch möglich, numerische Strings in numerische Werte umwandeln (z.B. '5' -> 5). Für diesen Zweck gibt es auch die Funktionen CONVERT, NUMBER, STRING.

Beispiele

```
/* Umpolung der Items 5, 12 und 23: */
recode item5,item12,item23 (0=1) (1=0).

/* Auswertung eines Multiple Choice Tests mit jeweils */
/* einer richtigen Antwortalternative: */
recode item3,item4,item18 (1=1) (else=0)
  /item1,item8,item14,item15 (2=1) (else=0)
  /item2,item5 to item7 (3=1) (else=0)
  /item9,item11 to item13,item17 (4=1) (else=0)
  /item10,item16,item19,item20 (5=1) (else=0).

/* Dichotomisierung: */
recode tablette (1,2=1) (3 thru 5=2) (else=sysmis).

/* Typänderung; INTO ist notwendig! */
recode gruppe ('K'=1)('E'=2) into treat.

/* Kopieren der ursprünglichen Variablen in einen Vektor */
recode kopf,foehn,hals,ohr,schwind,brechr,magen,herz(0=2)
  (else=copy) into symptom1 to symptom8.
```

```

/* Die ursprünglichen Variablen hatten die Werte 0..15, sind aber */
/* alphanumerisch codiert mit 0,1,...9,A,B,C,D,E,F: */
recode fragel to frage24 (convert)('A'=10) ('B'=11)
 ('C'=12) ('D'=13) ('E'=14) ('F'=15) (ELSE=SYSMIS)
into nfragel to nfrage24.

```

AUTOMATISCH UMKODIEREN

Aufruf

1 Menü

Transformieren
Automatisch umkodieren...

2 Dialogfenster "Automatisch umkodieren"

Variable -> neuer Name

[_____]
[_____]

Neuer Name [_____]

Umkodierung beginnen bei

Kleinstem Wert Größtem Wert

(aus der Variableliste jeweils jene Variable auswählen, die umkodiert werden soll, den Namen für die umkodierte Variable eingeben sowie angeben, ob die Werte aufsteigend oder absteigend zugeordnet werden sollen)

DER AUTORECODE-BEFEHL

Funktion

- Automatisches Umkodieren von Variablen in neue Variablen

Syntax

```

autorecode [variables]=varlist /into varlist
  [/descending]
  [/print].

```

Hinweise

- Mit **variables=** werden die Variablen angegeben, die umkodiert werden sollen, und mit dem Unterbefehl **/into** die zu kreierenden neuen Variablen. Diese dürfen noch nicht existieren.
- Die Werte in den Ausgangsvariablen werden der Größe nach durchnummeriert. Der numerisch bzw. alphabetisch kleinste Werte erhält im Normalfall die Zahl 1 zugeordnet, der nächstgrößere die Zahl 2 usw. Ist der Unterbefehl **/descending** angegeben, so wird umgekehrt dem größten Wert die Zahl 1 zugeordnet, dem zweitgrößten die Zahl 2 usw.
- Für die neuen Variablen werden außerdem automatisch auch Wertelabels definiert.
- Der Unterbefehl **/print** bewirkt, daß eine Zuordnungstabelle ausgegeben wird.

Beispiel

```

get file='a:\data\studold.sav'.
autorecode kurs /into nkurs.
oneway studzufr by nkurs.

```

3.3.2 Berechnen (COMPUTE)

Mit dieser Datentransformation ist es möglich, einen beliebig komplexen arithmetischen Ausdruck zu berechnen und diesen einer Ergebnisvariable, d.h. einer neuen oder bereits definierten Variable, zuzuweisen. Existiert die Ergebnisvariable bereits, so wird deren alter Wert überschrieben.

BERECHNEN	
<i>Aufruf</i>	
1	Menü Transformieren Berechnen...
2	Dialogfenster "Variable berechnen" Zielvariable: [_____] = Numerischer Ausdruck: [_____] [Typ und Label...] <i>(die Zielvariable ist aus der Variablenliste auszuwählen oder über die Rechartastatur einzugeben; der numerische Ausdruck kann entweder direkt über die Rechartastatur eingegeben werden oder mit Hilfe der Maus durch Übernahme der benötigten Elemente aus der Variablenliste, der Funktionenliste und der abgebildeten Tastatur)</i> [Falls...] <i>(nur für bedingtes Berechnen)</i>
3	Dialogfenster "Variable berechnen: Typ und Label" Label <input checked="" type="radio"/> Label: [_____] <input type="radio"/> Ausdruck als Label einsetzen Typ <input checked="" type="radio"/> Numerisch <input type="radio"/> String Breite [__8__]
<i>Hinweise</i>	
<ul style="list-style-type: none"> • Ist die angegebene Zielvariable neu, so wird sie kreiert; ist sie schon vorhanden, wird sie mit neuen Werten überschrieben. • Falls SPSS den Wert eines Ausdrucks nicht berechnen kann, wird SYSMIS als Wert benutzt. • Der Schalter "<u>Falls...</u>" dient der Durchführung <u>bedingter</u> Berechnungen (wird später behandelt). 	

DER COMPUTE-BEFEHL	
<i>Funktion</i>	
<ul style="list-style-type: none"> • Wertzuweisung 	
<i>Syntax</i>	
compute var=expression.	
<i>Hinweise</i>	
<ul style="list-style-type: none"> • Die Variable links vom Gleichheitszeichen ist die Ergebnisvariable, ihr wird der Wert des Ausdrucks zugewiesen. • Die Ergebnisvariable kann schon definiert sein oder sie wird durch den COMPUTE-Befehl neu definiert. • Kann SPSS den Ausdruck nicht berechnen, so wird der System-Missing (SYSMIS) als Wert eingesetzt. 	

Beispiele

```

compute s=0. /* Die Variable s wird auf 0 gesetzt */
compute summe=x1+x2+x3. /* Individuelle Mittelwertsberechnung */
compute xquer=summe/3. /* - Fortsetzung - */
compute xquer=(x1+x2+x3)/3. /* Indiv. Mittelwert.ohne Zw.rechnung */
compute diff=xnach-xvor. /* Differenz Nachher minus Vorher */
compute adiff=abs(xnach-xvor). /* Berechnung der absoluten Differenz */
compute n=n+1. /* n wird um 1 erhöht */
compute jalter=jahre+monate/12. /* Genaues Alter */
compute rzmax=max(rz1 to rz60). /* Maximale Reaktionszeit */
compute rzq=mean.10(rz1 to rz60). /* Mittlere Reaktionszeit;
/* aber: mind. 10 Reaktionszeiten
/* müssen valide sein!
compute eziff=mod(x,10). /* Isolierung der Einerstelle
compute zziff=mod(trunc(x/10),10). /* Isolierung der Zehnerstelle
compute frac=zahl-trunc(zahl). /* Die Stellen hinter dem Komma

```

3.3.3 Zählen (COUNT)

Mit dieser Datentransformation ist es möglich, die Häufigkeit eines angegebenen Ereignisses zu zählen. Das Zählergebnis wird einer anzugebenden Ergebnisvariable zugewiesen.

Mögliche Anwendungsfälle für das Zählen sind:

- die Anzahl der richtigen Antworten bei einem Test, falls sie z.B. mit 1 codiert sind;
- die Anzahl der verlangsamten Reaktionen bei einem aus mehreren Durchgängen bestehenden Reaktionsversuch;
- die Anzahl der gültigen Variablenwerte bei einer Vp.

ZÄHLEN**Aufruf****1 Menü**

Transformieren
Zählen...

2 Dialogfenster "Häufigkeiten von Werten in Fällen zählen"

Zielvariable: [_____] Label: [_____]

Variablen: [_____] (aus der Variablenliste auswählen!)

[Werte definieren...] (aufrufen und die zu zählenden Werte eingeben)

[Falls...] (nur für bedingtes Zählen)

3 Dialogfenster "Werte in Fällen zählen: Welche Werte?"

Wert: Zu zählende Werte:

 Wert [_____] [_____]

 Systemdefiniert fehlend

 System- oder benutzerdefinierte fehlende Werte

 Bereich: [_____] bis [_____] (mit [Hinzufügen] wird jeweils der Eintrag in die Liste

 Bereich: Kleinster Wert bis [_____] der zu zählenden Werte aufgenommen)

 Bereich: [_____] bis größter Wert

Hinweise

- Bei der interaktiven Arbeitsweise ist es nicht möglich, für die verschiedenen aufgeführten Variablen auch verschiedene Wertelisten für das Zählen anzugeben (diese Möglichkeit bräuchte man z.B. für die Auswertung eines Leistungstests in Multiple-Choice-Form, denn dabei wechselt die richtige Alternative üblicherweise von Aufgabe zu Aufgabe).
- Der Schalter "Falls..." dient der Durchführung bedingten Zählens (wird später behandelt).

DER COUNT-BEFEHL*Funktion*

- Auszählung von bestimmten Werten bei Variablen

Syntax

```
count var = varlist(value_list) [varlist...]  
      [var...].
```

Hinweise

- Die Variable vor dem Gleichheitszeichen ist die Ergebnisvariable.
- Für jede in der Variablenliste bzw. den Variablenlisten aufgeführte Variable wird geprüft, ob der Wert des aktuellen Falles mit einem der Werte der entsprechenden Werteliste übereinstimmt. Die Anzahl dieser Übereinstimmungen wird ermittelt und der Ergebnisvariable zugewiesen.
- Bei numerischen Variablen sind in der Werteliste die Schlüsselwörter **lo[west]**, **hi[ghest]**, **thru**, **missing** und **sysmis** erlaubt.

Beispiele

```
/* Testauswertung, z.B. Anzahl der richtigen Antworten: */  
count rohwert=item1 to item48(1).  
/* Testauswertung, z.B. bei mehreren richtigen Antwortalternativen: */  
count erg=var1,var4 to var10(1) var2,var3(1,2).  
/* Fragebogenauswertung (Anzahl der Ja-Antworten): */  
count jazahl=frage1 to frage50('J').  
/* Anzahl Missing-Werte: */  
count fehl=kopf to herz(missing).  
/* Anzahl überlanger Reaktionszeiten: */  
count extrem=rz1 to rz120 (500 thru hi).
```


3.3.4 Bedingte Transformationen (IF, DO IF)

Sollen irgendwelche Datentransformationen nicht unbedingt, d.h. für alle Fälle der Arbeitsdatei durchgeführt werden, sondern nur für bestimmte, so kann der Benutzer eine Bedingung formulieren, unter der sie ausgeführt werden sollen. Als Bedingung kann ein beliebig komplexer logischer Ausdruck angegeben werden. Zunächst sollen die Möglichkeiten betrachtet werden, die der Benutzer in SPSS für die Formulierung eines logischen Ausdrucks hat.

BEDINGTE TRANSFORMATIONEN

Aufruf

- 1 Im Dialogfenster des jeweiligen Transformationsmenüs [Falls...] wählen

In dem nun folgenden Dialogfenster muß die gewünschte Bedingung in Form eines logischen Ausdrucks formuliert werden.

- 2 Dialogfenster "....: Falls Bedingung erfüllt ist"

Alle Fälle einschließen

Fall einschließen, wenn Bedingung erfüllt ist: [_____]

(einen logischen Ausdruck eingeben: entweder über die Rechnertastatur oder mit der Maus durch Übernahme der benötigten Elemente aus der Variablenliste, der Funktionenliste und der abgebildeten Tastatur)

Der interaktive Aufruf des bedingten Berechnens entspricht dem Syntaxbefehl IF. Die anderen Typen von bedingten Transformationen (Umkodieren, Zählen,...) entsprechen dem jeweiligen Befehl für die unbedingte Transformation innerhalb eines DO IF-Blocks, in dem die Bedingung formuliert ist. (Der DO IF-Befehl wird in einem späteren Kapitel behandelt.)

DER IF-BEFEHL

Funktion

- Bedingte Berechnung von Variablen

Syntax

if [(|logical_expression|)] var=expression.

Hinweise

- Die Wertzuweisung erfolgt nur, wenn die Bedingung erfüllt ist (man beachte, daß bei der Wertzuweisung das Befehlswort **compute** nicht geschrieben werden darf).

Beispiele

(a)

```
if (tablette le 2) konsum=1.
```

```
if (tablette gt 2) konsum=2.
```

Dichotomisierung der Stichprobe nach dem Tablettenkonsum (geht auch mit dem RECODE-Befehl).

(b)

```
if (alter lt 22) agruppe=1.
```

```
if (alter ge 22 and alter lt 26) agruppe=2.
```

```
if (alter ge 26) agruppe=3.
```

Bildung von Altersgruppen (geht ebenfalls mit dem RECODE-Befehl).

(c)

```
compute belast=0.
if (kopf) belast=belast+2.
if (foehn) belast=belast+1.
if (hals) belast=belast+1.
    usw.
if (herz) belast=belast+3.
```

Berechnung eines Belastungswerts als gewichtete Summe von Symptomen.

3.3.5 Temporäre Transformationen (TEMPORARY)

Man unterscheidet permanente und temporäre Datenmodifikationen:

- permanente Datenmodifikationen gelten bis zum Ende des SPSS-Programms
- temporäre Datenmodifikationen gelten nur für die nächstfolgende Prozedur

Falls in einem SPSS-Programm der TEMPORARY-Befehl fehlt, sind die im Programm auftauchenden Datentransformationen permanent.

DER TEMPORARY-BEFEHL

Funktion

- Kennzeichnung von nachfolgenden Datenmodifikationen als temporär

Syntax

```
temporary.
transformation_commands
```

Beispiel

```
.....
count beschw=kopf to herz(1).
temporary.
recode tablette (1,2=1) (3 thru 5=2).
crosstabs tables = tablette by beschw.
frequencies variables = tablette.
```

In der Prozedur CROSSTABS wird die recodierte Version der Variable TABLETTE benutzt, in der darauffolgenden Prozedur FREQUENCIES hingegen die Originalversion.

3.4 Aufbereitung der Analysestichprobe

3.4.1 Gewichten von Fällen (WEIGHT)

Im Normalfall repräsentiert jede Zeile der Datenmatrix eine Beobachtungseinheit und jede Beobachtungseinheit ist "gleich wichtig" (d.h. hat das Gewicht 1). In SPSS hat man aber auch die Möglichkeit, Fälle unterschiedlich zu gewichten. Auf diesem Wege könnte man z.B. eine nicht repräsentative Stichprobenziehung im nachhinein korrigieren. Eine weitere Situation, bei der das Gewichten von Fällen notwendig ist, liegt vor, wenn die Daten nicht als Rohdaten, sondern in Form einer Häufigkeitstabelle eingegeben werden. Dann muß neben der bzw. den ursprünglichen Variablen noch eine weitere Variable eingegeben werden, die der jeweiligen Häufigkeit entspricht. Die "Fälle" stellen dann z.B. nicht mehr Einzelpersonen, sondern Gruppen von Personen dar, die entsprechend der Häufigkeit gewichtet werden müssen. In allen nachfolgenden Prozeduren werden die Fälle entsprechend der angegebenen Variable gewichtet.

GEWICHTEN VON FÄLLEN

Aufruf

- 1 Menü
 - Daten
 - Fälle gewichten...
- 2 Dialogfenster "Fälle gewichten"
 - Fälle nicht gewichten
 - Fälle gewichten mit Häufigkeitsvariable:
 - [_____] (aus der Variablenliste auswählen!)

DER WEIGHT-BEFEHL

Funktion

- Gewichtung von Fällen

Syntax

weight {by var | off}.

Beispiel

```
data list free file='a:\data\htab.dat'
  / punkte n.
weight by n.
frequencies variables=punkte /statistics.
```

Offenbar sind in der ASCII-Datendatei nicht die Rohdaten (d.h. pro Person eine Datenzeile mit ihrem Punktwert) abgelegt, sondern bereits die Häufigkeitstabelle für die gesamte Stichprobe (d.h. in jeder Datenzeile ein Punktwert und dahinter die Anzahl der Personen, die diesen Punktwert erreicht haben). Durch den FREQUENCIES-Befehl erhält man neben den angeforderten Statistiken auch die Häufigkeitstabelle, die dem Inhalt der Arbeitsdatei entsprechen müßte. Ohne den WEIGHT-Befehl wäre diese Häufigkeitstabelle falsch, denn dann wäre die Häufigkeit für jeden Punktwert gleich 1.

3.4.2 Sortieren von Fällen (SORT CASES)

In der Arbeitsdatei liegen die Fälle normalerweise in der Reihenfolge vor, die bei der Dateneingabe festgelegt wurde. SPSS bietet jedoch auch die Möglichkeit, die Fälle nach den Werten in einer oder mehreren Variablen neu sortieren zu lassen. Dies ist z.B. wünschenswert, wenn man die individuellen Werte sortiert auflisten möchte (Prozedur LIST, s.u.). Notwendig ist eine Neusortierung der Fälle im allgemeinen dann, wenn man die Gesamtgruppe der Fälle in Untergruppen aufteilen möchte (SPLIT FILE-Befehl, s.u.).

SORTIEREN VON FÄLLEN

Aufruf

1 Menü

Daten
Fälle sortieren...

2 Dialogfenster "Fälle sortieren"

Sortieren nach: [_____]

Sortierreihenfolge:

- aufsteigend
 absteigend

(Sortiervariable aus der Variablenliste auswählen und Sortierreihenfolge spezifizieren; soll nach mehreren Variablen sortiert werden, so sind sie nach fallendem Hierarchiegrad anzugeben)

Hinweise

- Die Reihenfolge der Sortiervariablen gibt die Hierarchie der Sortierreihenfolge an: die erstgenannte Variable ist die hierarchisch höchste und die zuletzt genannte die hierarchisch niedrigste.
- Haben mehrere Fälle identische Werte in den Sortiervariablen, so bleibt ihre relative Reihenfolge erhalten.

DER SORT CASES-BEFEHL

Funktion

- Sortieren der Fälle

Syntax

```
sort cases [by] varlist [(a | d)]  
[varlist...].
```

Hinweise

- Jede permanente Variable kann als Sortierungsvariable benutzt werden.
- Die Reihenfolge der Sortiervariablen gibt die Hierarchie der Sortierreihenfolge an: die erstgenannte Variable ist die hierarchisch höchste und die zuletzt genannte die hierarchisch niedrigste.
- Haben mehrere Fälle identische Werte in den Sortiervariablen, so bleibt ihre relative Reihenfolge erhalten.
- Wird hinter der Variablenliste ein '**d**' angegeben, so wird absteigend sortiert, sonst aufsteigend.
- SORT CASES ist eine Prozedur, d.h. es ist kein EXECUTE-Befehl erforderlich.

Beispiel

```
get file='a:\data\studis.sav'.  
sort cases by fb semester.  
list variables=all.
```

Die Daten werden nach Fachbereich und innerhalb eines Fachbereichs nach Semesterzahl sortiert aufgelistet.

3.4.3 Aufteilen der Fälle (SPLIT FILE)

Will man in SPSS eine statistische Analyse nicht für die Gesamtgruppe der Fälle durchführen, sondern getrennt für jede Untergruppe, so bietet sich die Möglichkeit, die Fälle in die gewünschten Untergruppen aufzuteilen und anschließend die entsprechende Statistikprozedur aufzurufen. Diese wird dann automatisch für alle Untergruppen durchgeführt, und die Ergebnisse werden gruppenweise getrennt ausgegeben.

Für die Aufteilung der Fälle müssen eine oder mehrere Variablen angegeben werden, deren Werte bzw. Wertekombinationen die einzelnen Gruppen definieren. Konkret bedeutet dies, daß es so viele Gruppen gibt, wie viele verschiedenen Werte bzw. Wertekombinationen die Gruppenvariablen besitzen. Genauer gesagt: wechselt der Wert einer der Gruppenvariablen von einer Beobachtung zur nächsten, so definiert SPSS eine neue Untergruppe. Daher sollte man sinnvollerweise die Daten erst nach den Gruppenvariablen sortieren (SORT-Befehl!), bevor man die Fälle aufteilt (diese Vorgehensweise ist bei der interaktiven Benutzung voreingestellt). Eine Aufteilung der Fälle kann später auch wieder rückgängig gemacht werden.

AUFTEILEN DER FÄLLE

Aufruf

- 1 Menü
 - Daten
 - Datei aufteilen...

- 2 Dialogfenster "Datei aufteilen"
 - Alle Fälle analysieren, keine Gruppen bilden *(damit kann man z.B. eine bestehende Aufteilung rückgängig machen)*
 - Gruppen vergleichen *(Ausgabe der Ergebnisse für die Untergruppen in einer gemeinsamen Tabelle)*
 - Ausgabe nach Gruppen aufteilen *(Ausgabe der Ergebnisse einzeln für jede Untergruppe)*

 - Gruppen basierend auf
 [_____] *(aus der Variablenliste auswählen!)*

 - Datei nach Gruppenvariablen sortieren
 - Datei ist sortiert

DER SPLIT FILE-BEFEHL

Funktion

- Aufteilung der Fälle

Syntax

split file {[**layered** | **separate**]} **by** varlist | off}

Hinweise

- In der Variablenliste sind jene Variablen zu spezifizieren, nach denen Untergruppen gebildet werden sollen.
- Die Beobachtungen sollten sinnvollerweise nach diesen Variablen sortiert sein.
- Wechselt der Wert einer der Variablen in der Variablenliste von einer Beobachtung zur nächsten, so beginnt SPSS eine neue Untergruppe.
- Beim Modus '**layered**' (Voreinstellung) sind die Ergebnisse für die einzelnen Untergruppen zwar getrennt ausgewiesen, jedoch in einer gemeinsamen Tabelle angeordnet.

- Beim Modus **'separate'** werden die nachfolgenden Prozeduren getrennt für jede Untergruppen durchgeführt und die Ergebnisse einzeln für jede Untergruppe ausgegeben.
- Tritt in einem SPSS-Programm nach einem SPLIT FILE-Befehl ein weiterer auf, so setzt er die Wirkung des vorherigen außer Kraft.
- Durch **split file off** kann man eine bestehende Untergruppenbildung ganz aufheben. Auch ein SORT CASES-Befehl hebt eine bestehende Untergruppenbildung auf.
- Durch einen vorausgehenden TEMPORARY-Befehl kann man einen SPLIT FILE-Befehl auf den nächsten Prozeduraufruf beschränken.

Beispiel

```
get file='a:\data\symptome.sav.'
recode alter(lo thru 18) (18 thru 30) (30 thru 50) (50 thru 65)
      (65 thru hi) into altgrup.
sort cases by altgrup.
split file by altgrup.
frequencies variables = beschw.
```

Aus der Variable ALTER werden zunächst Altersgruppen (Variable ALTGRUP) gebildet, und die Daten nach dieser Variable aufgeteilt. Für jede Altersgruppen getrennt wird dann die Häufigkeitsverteilung für die Variable BESCHW ermittelt. Vor Aufteilung der Fälle muß erst nach der Altergruppe sortiert werden.

3.4.4 Auswählen von Fällen (SELECT IF, SAMPLE, N OF CASES, FILTER, USE)

Zuweilen ist es wünschenswert, eine statistische Analyse lediglich für eine bestimmte Untergruppe z.B. von Personen durchzuführen, d.h. bestimmte Fälle aus der Gesamtstichprobe auszuwählen. In SPSS stehen dafür folgende Auswahlstrategien zur Verfügung:

1. Jene Fälle auswählen, für die eine bestimmte Bedingung zutrifft.
2. Einen zusammenhängenden Bereich von Fällen (von Fall Nr. ... bis Fall Nr. ...) auswählen.
3. Eine Zufallsauswahl von Fällen vorzunehmen (dabei ist die Anzahl oder der ungefähre Prozentsatz der auszuwählenden Fälle anzugeben).
4. Jene Fälle auswählen, für die eine anzugebende numerische Variable ("Filtervariable") einen Wert verschieden von 0 hat.

Für die nicht ausgewählten Fälle muß dann noch entschieden werden,

1. ob sie auch aus der Arbeitsdatei gelöscht werden sollen oder
2. ob sie lediglich ausgefiltert, d.h. zwar aus der statistischen Analyse ausgeschlossen, aber nicht aus der Arbeitsdatei entfernt werden sollen; hat man dabei gleichzeitig eine der obigen Alternativen (1) oder (3) gewählt, so erzeugt SPSS automatisch eine Variable FILTER_\$ (die man im übrigen auch im Dateneditor bearbeiten und in einem nächsten Schritt als Filtervariable für die Alternative (4) benutzen könnte); das Ergebnis der Filterung ist auch in der Kopfspalte der Dateneditortabelle zu ersehen.

AUSWÄHLEN VON FÄLLEN

Aufruf

- 1 Menü
Daten
Fälle auswählen...

<p>2 Dialogfenster "Fälle auswählen"</p> <p>Auswahl:</p> <ul style="list-style-type: none"> <input checked="" type="radio"/> Alle Fälle <input type="radio"/> Falls Bedingung zutrifft [Falls...] <input type="radio"/> Zufallsstichprobe [Stichprobe...] <input type="radio"/> Nach Zeit- oder Fallbereich [Bereich...] <input type="radio"/> Filtervariable verwenden: [_____] <p>Nicht ausgewählte Fälle:</p> <ul style="list-style-type: none"> <input checked="" type="radio"/> Filtern <input type="radio"/> Löschen <p>3a Dialogfenster "Fälle auswählen: Falls"</p> <p>[_____]</p> <p>3b Dialogfenster "Fälle auswählen: Zufallsstichprobe"</p> <p>Größe der Stichprobe</p> <ul style="list-style-type: none"> <input checked="" type="radio"/> Ungefähr [_____] % aller Fälle <input type="radio"/> Exakt [_____] aus den ersten [_____] Fällen <p>3c Dialogfenster "Fälle auswählen: Bereich"</p> <p>Beobachtung: Erster Fall [_____] Letzter Fall [_____]</p>	<p><i>(aus der Variablenliste auswählen!)</i></p> <p><i>(einen logischen Ausdruck eingeben: entweder über die Rechnerastatur oder mit der Maus durch Übernahme der benötigten Elemente aus der Variablenliste, der Funktionenliste und der abgebildeten Tastatur)</i></p>
--	---

Die Arbeitsweise von SPSS bei der interaktiven Benutzung der Auswahl von Fällen nach den Auswahlstrategien bedingte Auswahl bzw. Zufallsstichprobe entspricht der folgenden Befehlsfolge:

```
use all.
compute filter_$=.... .
variable label filter_$ "....".
value labels filter_$ 0 'Nicht ausgewählt' 1 'Ausgewählt'.
format filter_$ (f1.0).
filter by filter_$.
execute.
```

Daraus ersieht man, daß SPSS

1. eine Filtervariable berechnet (COMPUTE) und ihr automatisch passende Labels und ein Format zuweist und
2. diese Variable in einem FILTER-Befehl benutzt.

Für die Auswahl gemäß einer Bedingung lautet der COMPUTE-Befehl:

```
compute filter_$=bedingung.
```

und für die Zufallsauswahl mit vorgegebenem Anteil:

```
compute filter_$(uniform(1)<=anteil).
```

Neben dem FILTER- und dem USE-Befehl kennt SPSS aber noch einige ältere Befehle für das Auswählen von Fällen: SELECT IF, SAMPLE, N OF CASES. Die Wirkung dieser Befehle kann man aber auch mit Hilfe von COMPUTE, FILTER und USE erreichen.

DER SELECT IF-BEFEHL

Funktion

- Auswahl von Fällen in Abhängigkeit von einer Bedingung

Syntax

```
select if [(|logical_expression|)].
```

Hinweise

- Der Fall wird nur dann in die Arbeitsdatei übernommen, falls die Bedingung wahr ist.
- Spezifiziert man in einem SPSS-Programm mehrere SELECT IF-Befehle, so wirken sie wie ein einzelner SELECT IF-Befehl mit AND-Verknüpfung der Bedingungen.
- Durch einen vorausgehenden TEMPORARY-Befehl kann man die Wirkung des SELECT IF-Befehls auf die nächste Prozedur beschränken.

Beispiel

```
....
select if (fb=4 and geschl='W').
....
```

Die weitere Analyse wird lediglich für die Psychologie-Studentinnen durchgeführt.

DER SAMPLE-BEFEHL

Funktion

- Zufallsauswahl von Beobachtungen

Syntax

```
sample {decimal_value | n from m}
```

Hinweise

- Wird ein Dezimalwert (zwischen 0 und 1) angegeben, so entspricht dieser dem auszuwählenden Anteil; allerdings stimmt der tatsächlich ausgewählte Anteil nur ungefähr mit dem angegebenen Wert überein.
- Wird 'n from m' angegeben, so werden n Beobachtungen zufällig gezogen, falls die tatsächliche Anzahl der Beobachtungen größer oder gleich m ist; im anderen Fall wird ein proportional kleinerer Anteil ausgewählt.
- Ein SAMPLE-Befehl wird jeweils auf der eventuell reduzierten Basis ausgeführt (zu beachten bei mehreren SAMPLE-Befehlen oder bei zusätzlichem SELECT IF).

Beispiele

(a)

```
....
sample 0.10.
....
```

Für die weitere Analyse wird eine Zufallsstichprobe von ca. 10% gezogen.

(b)

```
....
do if (geschl='W').
  sample 0.30.
else if (geschl='M').
  sample 0.25.
end if.
frequencies variables=punkte.
```


Von den weiblichen Personen werden ca. 30% zufällig ausgewählt und von den männlichen ca. 25%.

DER N OF CASES-BEFEHL

Funktion

- Auswahl der ersten n Beobachtungen

Syntax

n [of cases] n

Hinweise

- Es werden nur die ersten n Beobachtungen aus der Arbeitsdatei ausgewählt (n bezieht sich auf die Beobachtungen in der Arbeitsdatei, nicht auf die Rohdaten-Eingabesätze!).
- Stehen im SPSS-Programm außer dem N OF CASES-Befehl noch SELECT IF- und/oder SAMPLE-Befehle, so werden diese solange ausgeführt, bis n Beobachtungen ausgewählt sind oder das Ende der Arbeitsdatei erreicht ist.

Beispiel

```
get file='a:\data\testdat.sav'.
n of cases 50.
....
```

Die weitere Analyse wird nur mit den ersten 50 Fällen durchgeführt.

DER FILTER-BEFEHL

Funktion

- Auswahl von Fällen, ohne sie aus der Arbeitsdatei zu entfernen

Syntax

filter {by var | off}

Hinweise

- Die angegebene Variable muß numerisch sein.
- Fälle, die der angegebenen Variable einen Wert 0 oder fehlend haben, werden von der weiteren Analyse ausgeschlossen.
- Durch einen vorausgehenden TEMPORARY-Befehl kann man die Wirkung des FILTER-Befehls auf die nächste Prozedur beschränken.

Beispiel

```
....
filter by geschl.
frequencies punkte.
```

Ist die Variable GESCHL das Geschlecht der Personen mit dem Wert 0 für männlich und 1 für weiblich, so wird die Häufigkeitsverteilung der Variable PUNKTE nur für die weiblichen Personen ausgegeben.

DER USE-BEFEHL

Funktion

- Auswahl der Fälle für die weitere Analyse

Syntax

use {**all** | [{*start_case* | **first**}] [**thru** {*end_case* | **last**}]}

Hinweise

- Bei dem USE-Befehl wird ein Bereich angegeben, aus dem die Fälle ausgewählt werden sollen. Diese Angabe besteht entweder aus einem Startwert (z.B. einer Fallnummer) oder aus einem Endwert oder beidem.
- Mit **use all** werden alle Fälle ausgewählt, d.h. eine bisher bestehende Auswahl außer Kraft gesetzt.
- Stellen die Daten in der Arbeitsdatei eine Zeitreihe dar, so kann der USE-Befehl dazu verwendet werden, Anfang und Ende der zu analysierenden Zeitreihe zu definieren. In diesem Fall können als Startwert bzw. Endwert auch Datumsangaben benutzt werden.

Beispiel

```
.....  
use 51 thru last.  
.....
```

Für die weitere Analyse werden nur die Fälle 51ff. zugrundegelegt.

3.4.5 Arbeitsdatei transponieren (FLIP)

Es ist in SPSS möglich, die Arbeitsdatei durch ihre Transponierte zu ersetzen, d.h. die Zeilen und Spalten zu vertauschen (die Fälle werden zu Variablen und die Variablen zu Fällen). Hierbei ist es auch möglich, nur einzelne Variablen auszuwählen (die nicht ausgewählten Variablen gehen dann verloren).

Die Variablennamen der untransponierten Arbeitsdatei werden in der transponierten Matrix zu den Werten einer Stringvariable namens CASE_LBL. Die Variablennamen der transponierten Matrix werden entweder von SPSS automatisch erzeugt oder vom Benutzer durch Spezifizierung einer Namensvariable vorgegeben, wobei SPSS nötigenfalls noch Änderungen vornimmt, um die Namen syntaktisch korrekt und eindeutig zu machen.

TRANSPONIEREN

Aufruf

- 1 Menü
 - Daten
 - Transponieren auswählen...
- 2 Dialogfenster "Transponieren"
 - Variable(n):
[_____] *(aus der Variablenliste auswählen!)*
 - Namensvariable:
[_____] *(aus der Variablenliste auswählen!)*

DER FLIP-BEFEHL

Funktion

- Transponieren der Arbeitsdatei

Syntax

```
flip [[variables=]{all**|varlist}]  
[/newnames=var].
```

Hinweise

- Mit **variables=** kann man die Spalten der Arbeitsdatei auswählen, die transponiert werden sollen. Die anderen Variablen gehen verloren.
- Mit **newnames=** kann man die Namen der Variablen in der transponierten Matrix vorgeben. Fehlt dieser Unterbefehl, so werden die Variablennamen von SPSS generiert (es sei denn, es existiert eine Variable CASE_LBL: dann werden deren Werte als Variablennamen verwendet).
- SPSS erzeugt zusätzlich eine Stringvariable CASE_LBL als Fall-ID.
- Nach dem Transponieren ist der Datentyp aller Variablen numerisch mit dem Standardformat F8.2 (eine Ausnahme bildet hier nur die Variable CASE_LBL).
- Variablen- und Wertelabels, sowie die Gewichtung von Fällen gehen durch das Transponieren verloren. Missing Werte werden nach dem Transponieren zu SYSMIS, ebenso Stringvariablen.

Beispiel

```
get file='d:\data\mydata.sav'.  
flip /newnames=vpid.  
compute xquer=mean(vp1 to vp72). execute.
```

Nach Ausführung des Programms enthält die Variable XQUER die Mittelwerte der ursprünglichen Variablen.

3.5 Kontrollstrukturen in SPSS-Programmen

Im allgemeinen werden die Programmbefehle der Reihe nach abgearbeitet, so wie sie im Programm stehen, d.h. der normale Ablauf eines Programms ist die Sequenz. Zuweilen möchte man aber von diesem Ablauf abweichen, und sog. Kontrollstrukturen zu Hilfe nehmen. In allen höheren Programmiersprachen gibt es eine Reihe von Kontrollstrukturen, um den Ablauf eines Programms zu steuern, wie z.B. unbedingte und bedingte Sprünge, Fallunterscheidungen, Blöcke, Schleifen, Rekursion, u.a. Einige dieser Kontrollstrukturen sind auch in SPSS verfügbar, allerdings nicht bei der interaktiven Arbeitsweise.

3.5.1 Offene Transformationen ausführen (EXECUTE)

Beim interaktiven Arbeiten wird jeder über Menüs veranlaßte Auswertungsschritt sofort ausgeführt. Anders bei der syntaxgesteuerten Arbeitsweise: hier wird ein Programm erst bei Aufruf einer Statistik- oder Graphikprozedur zur Ausführung gebracht (dies betrifft z.B. das Einlesen der Daten oder die Ausführung von Datentransformationen). Fehlt in einem SPSS-Programm ein Prozedur-Befehl, so muß im Programm ein EXECUTE-Befehl stehen, wenn das Programm ausgeführt werden soll.

OFFENE TRANSFORMATIONEN AUSFÜHREN

Aufruf

- 1 Menü
 - Transformieren
 - Offene Transformationen ausführen...

DER EXECUTE-BEFEHL

Funktion

- Programm zur Ausführung bringen

Syntax

execute.

Beispiel

```
data list file='a:\data\execin.dat'
  / vp 1-2(a) item1 to item40 4-43.
count rw=item1 to item40(1).
write outfile='a:\data\execout.dat'
  / vp 1-2(a) rw 4-6.
execute.
```

In dem Programm werden zunächst Rohdaten (Einzelantworten in einem aus 40 Aufgaben zusammengesetzten Test) aus einer ASCII-Datei eingelesen (DATA LIST), der Rohwert gebildet (COUNT) und dieser zusammen mit der VP-Kennung auf eine ASCII-Datei ausgegeben (WRITE). Da ein Prozedur-Befehl fehlt, wird für die Programmausführung ein EXECUTE-Befehl benötigt.

3.5.2 Bedingte Blöcke (DO IF)

DER DO IF-BEFEHL

Funktion

- Definition eines Block-IF zur bedingten Ausführung von mehreren Datenmodifikationen

Syntax

```

do if [(logical_expression)].
    transformation_commands
[else if [(logical_expression)].
    transformation_commands]
[else if...].
[else.
    transformation_commands]
end if.

```

Hinweise

- Die Programmbefehle zwischen den DO IF-, ELSE IF-, ELSE- und END IF-Befehlen bilden jeweils einen Block und können beliebig viele Transformationsbefehle enthalten.
- Zunächst wird die Bedingung für den ersten, dem DO IF-Block, geprüft. Falls sie wahr (1) ist, werden die Befehle dieses Blocks ausgeführt und anschließend mit dem ersten Befehl nach dem END IF fortgefahren, d.h. dann ist die Abarbeitung der gesamten DO IF/END IF Struktur beendet. Ist die Bedingung nicht wahr, so wird die Bedingung des nächsten Blocks geprüft und die Befehle dieses Blocks gegebenenfalls ausgeführt usw. Ein ELSE-Block (falls vorhanden) wird ausgeführt, wenn alle vorhergehenden Blöcke wegen Nichtzutreffen der jeweiligen Bedingung übersprungen wurden.
- DO IF / END IF- Strukturen können beliebig oft geschachtelt werden.
- Im Syntaxprotokoll sind die Befehle in der DO IF / END IF-Struktur speziell gekennzeichnet, und zwar durch eine Zahl größer als 0 (=Schachteltiefe) hinter der Befehlsnummer.

Beispiele

(a)

```

do if (tablette le 2).
    compute konsum=1.
else if (tablette gt 2).
    compute konsum=2.
end if.

```

Dichotomisierung der Stichprobe nach dem Tablettenkonsum (siehe Beispiel (a) beim IF-Befehl).

(b)

```

do if (alter lt 22).
    compute agruppe=1.
else if (alter lt 26).
    compute agruppe=2.
else.
    compute agruppe=3.
end if.

```

Bildung von Altersgruppen (siehe Beispiel (b) beim IF-Befehl).

(c)

```

do if (abt=1).
    do if (tablette<=2).
        compute belast=3.
    else if (tablette>2).
        compute belast=4.
    end if.
else if (abt=2 and tablette<=2).

```

```

        compute belast=1.
else.
        compute belast=2.
end if.

```

Berechnung eines Belastungswerts in Abhängigkeit von zwei Variablen.

3.5.3 Schleifen (DO REPEAT, LOOP, BREAK)

Bei den höheren Programmiersprachen ist die Schleife eine der wichtigsten Kontrollstrukturen. Schleifen werden benötigt für iterative Wiederholungen von Programmbefehlen. In SPSS sind aber nicht alle Typen von Schleifen realisiert, die man von den höheren Programmiersprachen her kennt. Es gibt hier lediglich eine "Listenschleife" (DO REPEAT) und eine "Zählschleife" (LOOP), die allerdings auch als Bedingungsschleife (while/do, do/until) formuliert werden kann.

DER DO REPEAT-BEFEHL

Funktion

- Definition einer Schleife, z.B. zum Ausführen von gleichen Datenmodifikationen für verschiedene Variablen

Syntax

```

do repeat stand-in_var = {varlist | value_list}
           [stand-in_var...].
           transformation_commands
end repeat [print].

```

Hinweise

- Die Befehle, die zwischen DO REPEAT und END REPEAT stehen, werden für jede Variable bzw. jeden Wert hinter '*stand-in_var*='genau einmal ausgeführt.
- Die *stand-in_var* ist eine sog. Stellvertretervariable. Sie ist genaugenommen keine Variable, sondern lediglich ein Platzhalter für die jeweilige Variable bzw. den jeweiligen Wert und ist nur innerhalb der DO REPEAT / END REPEAT-Schleife definiert. Bei den Transformationsbefehlen ist anstelle der Variablen der Variablenliste bzw. der Werte der Werteliste die entsprechende Stellvertretervariable zu benutzen.
- Die Variablenlisten können auch eine definierende Variablenlisten sein.
- Alle Variablen- bzw. Wertelisten einer DO REPEAT/END REPEAT-Schleife müssen gleich viele Elemente enthalten.
- Wird hinter END REPEAT das Schlüsselwort PRINT angegeben, so werden die von SPSS generierten Befehle der Schleife protokolliert (gekennzeichnet mit "+").

Beispiele

(a)

```

compute belast=0.
do repeat symptom=kopf to herz
    /gewicht=2,1,1,1,2,3,2,3.
    if (symptom) belast=belast+gewicht.
end repeat.

```

Berechnung eines Belastungswerts als gewichtete Summe von Symptomen (siehe Beispiel (c) beim IF-Befehl).

(b)

```

do repeat symp=symptom1 to symptom8
    /var=kopf to herz.
    compute symp=var.
end repeat print.

```

Die Variablen KOPF, FOEHN, usw. werden auf neue Variablen SYMPTOM1, SYMPTOM2, usw. kopiert. Übrigens bewirkt das Schlüsselwort **print** im END REPEAT-Befehl folgende Ausgabe im Syntaxprotokoll:

```
+COMPUTE SYMPTOM1=KOPF.
+COMPUTE SYMPTOM2=FOEHN.
+COMPUTE SYMPTOM3=HALS.
+COMPUTE SYMPTOM4=OHR.
+COMPUTE SYMPTOM5=SCHWIND.
+COMPUTE SYMPTOM6=BRECH.
+COMPUTE SYMPTOM7=MAGEN.
+COMPUTE SYMPTOM8=HERZ.
```

(c)

```
compute n=0.
compute s=0.
do repeat x=var1 to var48.
  do if (not missing(x)).
    compute n=n+1.
    compute s=s+x.
  end if.
end repeat.
do if (n>0).
  compute mitte=s/n.
else.
  compute mitte=SYSMIS.
end if.
```

Hier wird der intraindividuelle Mittelwert der Variablen VAR1 bis VAR48 gebildet, incl. Berücksichtigung von fehlenden Werten. Die Wirkung dieses Programmfragments entspräche also der folgenden Programmzeile:

```
compute mitte=mean(var1 to var48).
```

DER LOOP-BEFEHL

Funktion

- Definition einer Schleife zur iterativen Abarbeitung von Datenmodifikationsbefehlen

Syntax

```
loop [var=n to m [by {1 | k}]] [if [(logical_expression)].
  transformation_commands...
end loop [if [(logical_expression) ]].
```

Hinweise

- Ist nur **loop** ... **end loop** angegeben, so wird die Schleife sooft abgearbeitet, wie der Systemparameter MXLOOPS angibt; diesen Systemparameter kann man mit dem Befehl: SET MXLOOPS=n setzen.
- Mithilfe von 'var=' spezifiziert man die Anzahl der Durchläufe durch die Schleife. Die aufgeführte Variable ist die indizierende Variable, die alle Werte von n bis m mit Schrittweite k (oder 1) durchläuft. Sie ist eine ganz 'normale' Variable, es sei denn, sie beginnt mit dem '#'- Zeichen (dann ist sie eine Arbeitsvariable). Ihr Wert kann innerhalb der Schleife benutzt, aber nicht verändert werden.
- Der Wert von n, m und k muß numerisch sein, braucht aber nicht ganzzahlig oder positiv zu sein. Ist n gleich m, so wird die Schleife einmal durchlaufen. Die Schleife wird nicht durchlaufen, wenn bei positivem k der Wert von n größer wird als der Wert von m oder wenn bei negativem k der Wert von n kleiner ist als der Wert von m oder wenn der Wert von k gleich 0 ist oder wenn einer der Werte von n, m, k fehlend ist.
- Es besteht auch die Möglichkeit, die Schleife durch Angabe einer Bedingung abbrechen zu lassen. Steht das Schlüsselwort **if** mit einer Bedingung nach dem **loop** (d.h. vor dem Beginn der Schleife), so wird die Bedingung geprüft, bevor die Schleife abgearbeitet wird: die Schleife wird abgearbeitet, falls die Bedingung wahr (1) ist, und nicht, falls die Bedingung falsch (0) oder fehlend ist. Steht das Schlüsselwort **if** mit einer Bedingung nach dem **end loop**, so wird die Schleife erneut abgearbeitet, falls die Bedingung falsch (0) ist, und nicht abgearbeitet, falls die Bedingung wahr (1) oder fehlend ist.

Beispiele

(a)

```
vector sym=kopf to herz.
compute belast=0.
loop #i=1 to 8.
  compute belast=belast+sym(#i).
end loop.
```

Eine weitere Möglichkeit, eine Summe von Variablen zu bilden.

(b)

```
data list free file='a:\data\symptom.dat'
  /vp nr1 nr2 nr3.
vector symptom(8).
loop #i=1 to 8.
  compute symptom(#i)=0.
end loop.
compute symptom(nr1)=1.
compute symptom(nr2)=1.
compute symptom(nr3)=1.
frequencies variables=symptom1 to symptom8
```

Für jede Person liegen die Nummern von drei Symptomen vor, unter denen sie leidet. Daraus werden die Werte für acht Variablen (SYMPTOM1,...,SYMPTOM8) erzeugt, wobei die drei Variablen mit den angegebenen Nummern den Wert 1 und die anderen den Wert 0 erhalten sollen.

(c)

```
data list free file='a:\data\zigarett.dat'
  /x1 to x182.
vector x=x1 to x182.
vector ssum(7).
loop #i=1 to 7.
  compute ssum(#i)=0.
  loop #j=#i to 182 by 7.
    compute ssum(#i)=ssum(#i)+x(#j).
  end loop.
end loop.
frequencies variables=ssum1 to ssum7.
```

Für jede Person liegen 182 Variablen vor, die eine Matrix von 26 Zeilen und 7 Spalten darstellen (z.B. die Anzahl gerauchter Zigaretten an den 7 Tagen der Woche während der Beobachtungszeit von 26 Wochen). Es sollen die Spaltensummen berechnet werden, d.h. die Gesamtzahl der gerauchten Zigaretten pro Wochentag.

(d)

```
data list free file='a:\data\zigarett.dat'.
  /x1 to x182.
vector x=x1 to x182.
vector zsum(26).
compute #je=0.
loop #i=1 to 26.
  compute zsum(#i)=0.
  compute #ja=#je+1.
  compute #je=#je+7.
  loop #j=#ja to #je.
    compute zsum(#i)=zsum(#i)+x(#j).
  end loop
end loop.
frequencies variables=zsum1 to zsum26.
```

Wie im vorigen Beispiel, jedoch sollen diesmal die Zeilensummen berechnet werden, d.h. die Gesamtzahl der gerauchten Zigaretten pro Woche.

DER BREAK-BEFEHL

Funktion

- Abbruch der Abarbeitung einer LOOP-Schleife

Syntax

break.

Hinweise

- Der BREAK-Befehl ist innerhalb einer LOOP/END LOOP-Struktur zu spezifizieren, und zwar dort innerhalb einer DO IF/END IF-Struktur (da ja sonst die Schleife bereits beim ersten Durchlauf abgebrochen würde).
- Sind LOOP/END LOOP-Strukturen ineinander verschachtelt, so wird durch BREAK nur die innerste Schleife abgebrochen, innerhalb derer sie spezifiziert ist.

4 SPSS-Statistikprozeduren

4.1 Univariate deskriptive Statistik

Übersicht über die verfügbaren Statistiken

	Prozedur		
	Häufigkeiten	Deskriptive Statistiken	Explorative Datenanalyse
Häufigkeitstabelle	+		
n	+	+	+
Mittelwert	+	+	+
Median	+		+
Modalwert	+		
Summe	+	+	
Varianz	+	+	+
Standardabweichung	+	+	+
Quartile	+		+
Perzentile	+		+
Interquartilabstand			+
Minimum	+	+	+
Maximum	+	+	+
Spannweite	+	+	+
Schiefe	+	+	+
Exzeß	+	+	+
Extremwerte			+
Getrimmter Mittelwert			+
Std.fehler des Mittelwerts	+	+	+
Std.fehler der Schiefe	+	+	+
Std.fehler des Exzesses	+	+	+
KI für den Mittelwert			+
M-Schätzer			+

4.1.1 Prozedur "Häufigkeiten" (FREQUENCIES)

Mit Hilfe dieser Prozedur können eindimensionale empirische Häufigkeitsverteilungen für numerische und Zeichenkettenvariablen incl. einiger Statistiken (Lagemaße, Streuungsmaße, Perzentilwerte, Verteilungsmaße) und Diagramme (Balken- und Kreisdiagramme, Histogramme, evtl. mit Normalverteilungskurve) angefordert werden.

DIE PROZEDUR "HÄUFIGKEITEN"	
<i>Aufruf</i>	
1	Menü Analysieren Deskriptive Statistiken > Häufigkeiten...
2	Dialogfenster "Häufigkeiten" Variable(n): [_____] <input checked="" type="checkbox"/> Häufigkeitstabellen anzeigen [Statistik...] [Diagramme...] [Format...]
	<i>(die abhängigen Variablen aus der vorgegebenen Variablenliste auswählen) (damit werden für die abhängigen Variablen eindimensionale Häufigkeitstabellen angefordert; in diesem Falle kann man über [Format...] deren Aussehen beeinflussen)</i>
3	Dialogfenster "Häufigkeiten: Statistik" Perzentilwerte: <input type="checkbox"/> Quartile <input type="checkbox"/> Trennen in [_10_] gleiche Gruppen <input type="checkbox"/> Perzentile: [_____] Streuung: <input type="checkbox"/> Std.-Abweichung <input type="checkbox"/> Varianz <input type="checkbox"/> Spannweite <input type="checkbox"/> Minimum <input type="checkbox"/> Maximum <input type="checkbox"/> Std.-Fehler Lagemaße: <input type="checkbox"/> Mittelwert <input type="checkbox"/> Median <input type="checkbox"/> Modalwert <input type="checkbox"/> Summe <input type="checkbox"/> Werte sind Gruppenmittelpunkte Verteilung: <input type="checkbox"/> Schiefe <input type="checkbox"/> Kurtosis
	<i>(gewünschte Werte eingeben; mit [Hinzufügen] werden sie übernommen)</i>
4	Dialogfenster "Häufigkeiten: Diagramme" Diagrammtyp: <input checked="" type="radio"/> Keiner <input type="radio"/> Balkendiagramme <input type="radio"/> Kreisdiagramme <input type="radio"/> Histogramme <input type="checkbox"/> Mit Normalverteilungskurve Diagrammwerte: <input checked="" type="radio"/> Häufigkeiten <input type="radio"/> Prozente
5	Dialogfenster "Häufigkeiten: Format"

Sortieren nach:

- Aufsteigenden Werten
- Absteigenden Werte
- Aufst. Häufigkeiten
- Abst. Häufigkeiten

Mehrere Variablen:

- Variablen vergleichen
- Ausgabe nach Variablen ordnen
- Keine Tabellen mit mehr als [_10_] Kategorien

DIE FREQUENCIES-PROZEDUR

Funktion

- Berechnung eindimensionaler empirischer Häufigkeitsverteilungen für numerische und Zeichenkettenvariablen
- Berechnung von Statistiken: Lagemaße, Streuungsmaße, Perzentilwerte, Verteilungsmaße
- Erstellung von Diagrammen: Balken- und Kreisdiagramme, Histogramme, evtl. mit NV-Kurve

Syntax

frequencies

```
[variables=]varlist
[/format=[{notable | limit(n)}] [{avalue | dvalue | afreq | dfreq}]
[/missing=include]
[/barchart=[minimum(n)] [maximum(n)] [{freq(n) | percent(n)}]]
[/piechart=[minimum(n)] [maximum(n)] [{freq(n) | percent(n)}]]
[/histogram=[min(n)] [max(n)] [{freq(n) | percent(n)}] [{nonnormal | normal}] [increment(n)]]
[/grouped=varlist[{(width) | (boundary list)}]]
[/ntiles=n]
[/percentiles=value list]
[/statistics=[default] [mean] [stddev] [minimum] [maximum] [sum] [median] [mode] [variance]
[range] [skewness] [kurtosis] [semean] [seskew] [sekurt] [all] [none**]]
[/order={analysis | variable}]
```

Bedeutung der Spezifikationen

- Bei **format**:

avalue	Kategorien aufsteigend nach dem Wert sortieren
dvalue	Kategorien absteigend nach dem Wert sortieren
afreq	Kategorien aufsteigend nach der Häufigkeit sortieren
dfreq	Kategorien absteigend nach der Häufigkeit sortieren
notable	alle Häufigkeitstabellen unterdrücken
limit(n)	Häufigkeitstabellen mit mehr als <i>n</i> Kategorien unterdrücken
- Bei **missing**:

include	Einbeziehung von Missing-Werten
----------------	---------------------------------
- Bei **barchart, piechart, histogram**:

min,max	zu berücksichtigender Wertebereich für den Plot der Verteilung(en)
freq(n)	Achsenkalierung in Häufigkeiten (mit <i>n</i> als Maximum)
percent(n)	Achsenkalierung in Prozent (mit <i>n</i> als Maximum)
normal	Überlagern der empirischen Verteilung mit einer Normalverteilungskurve
nonnormal	keine Normalverteilungskurve
increment(n)	Vorgabe einer Intervallbreite
- Bei **statistics**:

default	wie mean, stddev, minimum, maximum
semean	Std.fehler des Mittelwerts

seskew	Std.fehler der Schiefe
sekurt	Stdfehler des Exzesses
all	alle verfügbaren Statistiken
none	keine Statistik

Hinweise

- Obligatorisch ist nur der Unterbefehl **variables=**.
- Normalerweise arbeitet SPSS im "Allgemeinen Modus": sowohl numerische als auch Stringvariablen sind hier zugelassen; jeder vorkommende distinkte Datenwert stellt eine eigene Kategorie dar; es gibt keine leeren Kategorien (z.B. bei Lücken in den Datenwerten).
- Ist beim Unterbefehl **variables=** hinter der Variablenliste ein Wertebereich (*min,max*) angegeben, so arbeitet SPSS im "Integermodus": hierbei sind nur numerische Variablen zugelassen; eventuelle Dezimalstellen werden abgeschnitten; es wird nur der angegebene Wertebereich berücksichtigt.
- Mit **/ntiles=** werden jene Quantile ausgegeben, die die Verteilung in *n* gleiche Teile teilen.
- Mit Hilfe von **/percentiles=** werden die angegebenen Percentile berechnet.
- Mit dem Unterbefehl **/grouped=** kann man SPSS mitteilen, daß die Werte in den angegebenen Variablen als Klassenmitten aufzufassen sind. Damit werden dann alle Quantilmaße für den Fall von gruppierten Daten korrekt berechnet (d.h. es wird linear interpoliert). In Klammern kann man bei Bedarf noch die Klassenbreite bzw. die Klassengrenzen angeben.
- Mit **/barchart** fordert man ein Balkendiagramm, mit **/piechart** ein Kreisdiagramm und mit **/histogram** ein Histogramm an.

Beispiel

Im folgenden Programm werden für die 3 Variablen Geschlecht, Schulbildung des Vaters und Schulbildung der

Mutter die empirische Häufigkeitstabelle und die Default-Statistiken (arithmetischer Mittelwert, Standardabweichung; Minimum und Maximum) berechnet sowie ein Balkendiagramm erstellt.

```
get file='a:\data\frabo.sav'.
frequencies variables=geschl schule_v schule_m
/statistics /barchart.
```

4.1.2 Prozedur "Deskriptive Statistiken" (DESCRIPTIVES)

Mit dieser Prozedur können alle gebräuchlichen eindimensionalen deskriptiven Statistiken für numerische Variablen berechnet werden, allerdings keine Quantile. Weiterhin erlaubt die Prozedur die Berechnung von z-Werten.

DIE PROZEDUR "DESKRIPTIVE STATISTIKEN"

Aufruf

- 1 Menü
 Analysieren
 Deskriptive Statistiken >
 Deskriptive Statistiken...

- 2 Dialogfenster "Deskriptive Statistik"
 Variable(n): [_____] *(die abhängigen Variablen aus der vorgegebenen Variablenliste auswählen)*
 Standardisierte Werte in Variablen speichern *(Berechnung von z-Werten)*
 [Optionen...]

- 3 Dialogfenster " Deskriptive Statistik Optionen" *(Angabe der gewünschten Statistiken und Ausgabespezifikationen)*
 Mittelwert
 Summe
 Streuung:
 Std.-Abweichung
 Varianz
 Spannweite
 Minimum
 Maximum
 Std.-Fehler
 Verteilung:
 Kurtosis
 Schiefe
 Anzeigereihenfolge:
 Variablenliste
 Alphabetisch
 Aufsteigende Mittelwerte
 Absteigende Mittelwerte

Hinweise

- Falls Standardisierung der abhängigen Variablen angefordert wird, werden deren z-Werte als neue Variablen in die Arbeitsdatei aufgenommen. Die Namen dieser Variablen sind die ursprünglichen Variablennamen mit einem 'Z' davor. Zusätzlich wird jeweils als Variablenlabel automatisch generiert: "Zscore(var)".
- Die Option Anzeigereihenfolge bezieht sich auf die Reihenfolge der abhängigen Variablen bei der Ausgabe.

DIE DESCRIPTIVES-PROZEDUR

Funktion

- Berechnung eindimensionaler deskriptiver Statistiken für numerische Variablen (allerdings keine Quantile!)
- Berechnung von z-Werten

Syntax

descriptives

```
[variables=]var[(z-var)] [var...]
[/missing={variable** | listwise} [include]]
[/save]
[/statistics={default** | mean** | stddev** | min** | max**} [sum] [variance] [range]
[skewness] [kurtosis] [semean] [all]]
[/sort={[mean | semean | stddev | variance | kurtosis | skewness | range | min | max | sum | name]
[(a) | (d)]}]
```

Bedeutung der Spezifikationen

- Bei **missing**:

variable	variablenweiser Ausschluß fehlender Werte
listwise	fallweiser Ausschluß fehlender Werte
include	Einbeziehung von Missing-Werten
- Bei **statistics**:

default	wie mean , stddev , min , max
semean	Standardfehler des Mittelwerts
all	alle verfügbaren Statistiken
- Bei **sort**:

mean , ...	Sortieren nach den Mittelwerten, ...
name	Sortieren nach dem Variablennamen
a	aufsteigend
d	absteigend

Hinweise

- Obligatorisch ist nur der Unterbefehl **variables = ...**
- Ist **/save** angegeben, so werden die angegebenen Variablen z-transformiert und als neue Variablen in die Arbeitsdatei aufgenommen. Die Namen dieser neuen Variablen sind *z-var*,..., bzw. gleich dem ursprünglichen Variablennamen mit einem 'Z' davor, falls *z-var* nicht angegeben ist. Zusätzlich wird jeweils als Variablenlabel automatisch generiert: "Zscore(*var*)".
- Es werden nur jene Statistiken ausgegeben, die hinter dem Unterbefehl **/statistics=** aufgeführt sind. Fehlt dieser völlig, so werden die mit ****** gekennzeichneten Statistiken ausgegeben.
- Mit Hilfe von **/sort=** kann man die Reihenfolge der Variablen bei der Ausgabe steuern: sie bestimmt sich nach der Größe der angegebenen Statistik. Fehlt die Spezifikation hinter **/sort**, so erfolgt die Ausgabe aufsteigend sortiert nach dem Mittelwert. Fehlt der Unterbefehl ganz, so ist die Reihenfolge der Variablen im Unterbefehl **variables=** maßgeblich.

Beispiel

Im folgenden Programm werden für die Variablen Mathe- und Deutschpunkte sowie Abiturnote alle verfügbaren Statistiken sowie z-Werte berechnet; die neuen Variablen mit den z-Werten sollen ZDEU, ZMAT und ZABI heißen.

```
get file='a:\data\frabo.sav'.
descriptives variables=pkte_deu(zdeu) pkte_mat(zmat) abi-note(zabi)
  /missing=listwise /save /statistics=all.
```

4.1.3 Prozedur "Explorative Datenanalyse" (EXAMINE)

Dies ist eine universelle, sehr leistungsfähige Prozedur für univariate deskriptive Statistik. Folgende Werte können berechnet werden: Lage- und Streuungsmaße, Schiefe und Exzeß, Perzentile, Ausreißer (die 5 größten und die 5 kleinsten Werte) u.a. Außerdem könne auch Diagramme angefordert werden, nämlich Boxplots, Stem-and-Leaf-Diagramme Histogramme (mit oder ohne Normalverteilungskurve).

DIE PROZEDUR "EXPLORATIVE DATENANALYSE"

Aufruf

1 Menü

Analysieren

Deskriptive Statistiken >

Explorative Datenanalyse...

2 Dialogfenster "Explorative Datenanalyse"

Abhängige Variablen:

[_____]

(die abhängigen Variablen aus der vorgegebenen Variablenliste auswählen)

Faktorenliste:

[_____]

(optional; die entsprechenden Variablen aus der vorgegebenen Variablenliste auswählen)

Fallbeschriftung:

[_____]

(optional; eine Variable angeben, deren Werte zur Kennzeichnung individueller Datenwerte, z.B. für Ausreißer benutzt werden sollen)

Anzeigen:

Beide Statistik Diagramme

[Statistik...]

[Diagramme...]

[Optionen...]

3 Dialogfenster "Explorative Datenanalyse: Statistik"

Deskriptive Statistik

Konfidenzintervall für den Mittelwert [_95_] %

M-Schätzer

Ausreißer

Perzentile

4 Dialogfenster "Explorative Datenanalyse: Diagramme"

Boxplots

Faktorstufen zusammen

Abhängige Variablen zusammen

Keiner

Deskriptiv

Stengel-Blatt

Histogramm

Normalverteilungsdiagramm mit Tests

Streubreite vs. mittleres Niveau mit Levene Test:

Keiner

Exponentenschätzung

Transformiert

Exponent: [Natürl. Log

[1/Quadratwurzel]

[Reziprok]

[Quadratwurzel]

[Quadratisch]

[Kubisch]

Nicht transformiert

5 Dialogfenster "Explorative Datenanalyse: Optionen"

Fehlende Werte:

- Listenweiser Fallausschluß
- Paarweiser Fallausschluß
- Werte einbeziehen

Hinweise

- Faktorvariablen werden angegeben, wenn die Berechnungen für Untergruppen getrennt durchgeführt werden sollen. Jeder Wert bzw. jede Wertekombination in diesen Variablen erzeugt eine eigene Untergruppe.
- Bei den angeforderten Statistiken bedeuten: Univariate Statistiken: Lage- und Streuungsmaße, Schiefe und Exzeß, u.a. M-Schätzer: robuste Maximum-Likelihood-Schätzer für die mittlere Tendenz Ausreißer: die 5 größten und kleinsten Werte Perzentile: 5%, 10%, 25%, 50%, 75%, 90%, 95%

DIE EXAMINE-PROZEDUR

Funktion

- Universelle Prozedur mit einem vielfältigen Angebot an Statistiken und Diagrammen für eindimensionale deskriptive Statistik

Syntax

examine

```
variables=varlist [by varlist [by ...]]
[/compare={groups** | variables}]
[/total** | nototal]
[/id={case number** | variable}]
[/percentiles (({5,10,25,50,75,90,95 | value list})=[{haverage | waverage | round | aempirical |
empirical}] [none**])]
[/plot={stemleaf**} [boxplot**] [npplot] [spreadlevel(n)] [histogram] [{all | none}]]
[/statistics={descriptives**} [extreme({n | n})] [{all | none}]]
[/cinterval {95** | n}]
[/mestimator={none** | all}] [huber({1.339 | c})] [andrew({1.34 | c})] [hampel({1.7,3,4,8,5 | a,b,c})]
[tukey({4.685 | c})]
[/missing={listwise** | pairwise} [{exclude** | include}] [{noreport** | report}]].
```

Bedeutung der Spezifikationen

- Bei **missing**:

listwise	fallweiser Ausschluß fehlender Werte
pairwise	paarweiser Ausschluß fehlender Werte
exclude	Ausschluß von Missing-Werten
include	Einbeziehung von Missing-Werten
noreport	fehlende und Missing-Werte bei Faktorvariablen werden ausgeschlossen
report	fehlende und Missing-Werte bei Faktorvariablen gelten als eigene Kategorien
- Bei **plot**:

spreadlevel(n)	Spread-und-Level-Plots
-----------------------	------------------------
- Bei **statistics**:

descriptives	univariate Statistiken: Mittelwert, Median, 5% -getrimmter Mittelwert, Standardfehler, Varianz, Standardabweichung, Minimum, Maximum, Spannweite, Interquartilabstand, Schiefe, Std.fehler der Schiefe, Exzeß, Std.fehler des Exzesses
extreme(n)	die Fälle mit den n niedrigsten und den n höchsten Werten
all	alle verfügbaren Statistiken

Hinweise

- Obligatorisch ist nur der Unterbefehl **variables =**. Als erstes sind hier die abhängigen Variablen anzugeben und - falls gewünscht - nach dem Schlüsselwort **by** die Faktorvariablen zur Bildung von Zellen (Gruppen). Für jede der Faktorvariablen ergeben sich so viele Zellen, wie die Variable Werte hat, die Zellen werden also

getrennt für jede Faktorvariable gebildet. Sollen die Zellen aus den Wertekombinationen mehrerer Faktorvariablen gebildet werden, so führt man die entsprechenden Variablen auf, getrennt ebenfalls mit dem Schlüsselwort **by**.

- Der Unterbefehl **/compare**= bezieht sich auf die Anordnung der Boxplots bei Vorliegen mehrerer Gruppen und mehrerer abhängiger Variablen: entweder liegen pro abhängige Variable die Boxplots für die Gruppen hintereinander (**groups**) oder umgekehrt pro Gruppe die Boxplots für die angegebenen abhängigen Variablen (**variables**).
- Sind Faktorvariablen angegeben, so erzeugt **/total** Statistiken und Plots auch für die Gesamtgruppe.
- Mit **/id**= kann man eine Variable angeben, deren Werte zur Kennzeichnung individueller Datenwerte, z.B. für Ausreißer benutzt werden sollen.
- Mit dem Unterbefehl **/percentiles**= kann man die Perzentile angeben, die berechnet werden sollen, und auch die Methode, nach der die Berechnung erfolgen soll im Falle, daß das entsprechende Perzentil nicht exakt auf einen vorhandenen Wert fällt: z.B. durch lineare Interpolation oder durch Runden.
(Achtung: Rangbindungen werden bei keinem der Verfahren berücksichtigt! Beispielsweise wäre der Median von {1,1,2} bei allen Verfahren gleich 1).
- Mit dem Unterbefehl **/plot**= kann man Diagramme anfordern. Die Stem-and-Leaf-Diagramme erscheinen im Textformat, die anderen Diagramme als hochauflösende Graphik. Die Spezifikation **spreadlevel(n)** erzeugt sog. Spread-und-Level-Plots: dabei werden für alle Zellen die Interquartilabstände gegen die Mediane abgetragen. Entsprechend dem Parameter n werden die Daten aber vorher transformiert: ein $n > 0$ entspricht einer Potenztransformation, $n = 0$ einer logarithmischen Transformation. Fehlt der Parameter n völlig, so werden die natürlichen Logarithmen der Interquartilabstände der natürlichen Logarithmen der Mediane abgetragen.
- Der Unterbefehl **/mestimator**= fordert M-Schätzer (das sind robuste Maximum-Likelihood-Schätzer für die Lage einer Verteilung) an; vier Methoden stehen zur Auswahl.

Beispiel

Im folgenden Programm werden für die Variablen Körpergröße und Körpergewicht getrennt für Männer und Frauen deskriptive Statistiken und die Default-Perzentile berechnet sowie Normalverteilungsplots und Histogramme erstellt. Für die Diagramme wird dabei eine gemeinsame Skalierung verwendet.

```
get file='a:\data\frabo.sav'.
examine variables=groesse gewicht by geschl
  /percentiles
  /plot=npplot histogram
  /scale=uniform.
```

4.1.4 Prozedur "Fälle zusammenfassen" (SUMMARIZE)

Mit Hilfe dieser Prozedur können Rohdaten aufgelistet sowie verschiedene Statistiken berechnet werden. Dabei kann die Gesamtgruppe aufgrund der Werte in einer oder mehreren Gruppierungsvariablen in Untergruppen aufgeteilt werden.

DIE PROZEDUR "FÄLLE ZUSAMMENFASSEN"	
<i>Aufruf</i>	
1	Menü Analysieren Berichte ▷ Fälle zusammenfassen...
2	Dialogfenster "Fälle zusammenfassen" Variable(n): [_____] Gruppenvariable(n): [_____] <input checked="" type="checkbox"/> Fälle anzeigen <input checked="" type="checkbox"/> Fälle beschränken auf die ersten [_100_] <input checked="" type="checkbox"/> Nur gültige Fälle anzeigen <input type="checkbox"/> Fallnummern anzeigen [Statistik...] [Optionen...]
	<i>(die abhängigen Variablen aus der vorgegebenen Variablenliste auswählen)</i> <i>(optional; die Gruppenvariablen aus der vorgegebenen Variablenliste auswählen)</i> <i>(falls die Rohdaten aufgelistet werden sollen)</i> <i>(Rohdaten nur für diese Fälle auflisten)</i> <i>(falls die Rohdaten aufgelistet werden sollen)</i> <i>(bei Auflistung der Rohdaten: Fallnummern anzeigen)</i>
3	Dialogfenster "Zusammenfassung: Statistik" Auswahlliste: Anzahl der Fälle (ist bereits ausgewählt) Mittelwert Median Gruppiertes Median Standardfehler des Mittelwerts Summe Minimum Maximum Spannweite Erster Wert Letzter Wert Standardabweichung Varianz Kurtosis Standardfehler der Kurtosis Schiefe Standardfehler der Schiefe Harmonisches Geometrisches Mittel Prozent der Gesamtsumme Prozent der Gesamtanzahl

4 Dialogfenster "Optionen"

- Titel: [_____] (Überschrift)
 Erklärung: [_____] (Fußnote)
 Zwischentitel für Gesamtergebnisse
 Listenweiser Ausschluß bei fehlenden Werten
 Fehlende Statistiken erscheinen als: [_____] (Kennzeichnung für fehlende Werte)

DIE SUMMARIZE-PROZEDUR

Funktion

- Auflistung der Rohdaten
- Berechnung von Statistiken: Lagemaße, Streuungsmaße, Verteilungsmaße
- Einfaktorielle Varianzanalyse, Linearitätstest
- Die Gesamtgruppe kann für die Analyse in Untergruppen aufgeteilt werden.

Syntax

summarize

```

[tables={varlist | all} [by varlist] [by ...] [/varlist ...]
[/title='string']
[/footnote='string']
[/cells=[count] [mean] [stddev] [median] [gmedian] [semean] [sum] [min] [max] [range]
[variance] [kurt] [sekurt] [skew] [seskew] [first] [last] [ncpt] [spct] [npct(var)] [spct(var)]
[harmonic] [geometric] [default]**] [all] [none]]
[/missing=[{exclude** | include | dependent}] [{variable** | table}]]
[/format=[{nolist** | list | validlist}] [{casenum | nocasenum}] [{total** | nototal}] [limit=n]
[missing='string']]
[/statistics=[anova] [{linearity | all}] [none**]]

```

Bedeutung der Spezifikationen

- Bei **cells**:

gmedian	gruppierter Median
first	erster Wert
last	letzter Wert
npct	Prozent an der Gesamtzahl der Fälle
spct	Prozent an der Gesamtsumme
ncpt(var)	Prozent an der Gesamtzahl der Fälle in der angegebenen Gruppenvariable
spct(var)	Prozent an der Gesamtsumme in der angegebenen Gruppenvariable
harmonic	harmonischer Mittelwert
geometric	geometrischer Mittelwert
default	Mittelwert, Standardabweichung, Anzahl
- Bei **missing**:

exclude	Ausschluß von Missing-Werten
include	Einbeziehung von Missing-Werten
dependent	Ausschluß von Fällen mit fehlenden Werten in den abhängigen Variablen
variable	Ausschluß von Fällen mit fehlenden Werten in allen beteiligten Variablen
table	listenweiser Ausschluß bei fehlenden Werten

- Bei **format**:
 - nolist** keine Auflistung von Fällen
 - list** Auflistung von Fällen
 - validlist** Auflistung nur der gültigen Fälle
 - casenum** Auflistung mit Fallnummern
 - nocasenum** keine Fallnummern bei der Ausgabe der Rohdaten
 - total** Ergebnisse der Gesamtgruppe mit dem Label "Insgesamt" ausgeben
 - nototal** Ergebnisse der Gesamtgruppe ohne Label ausgeben
 - limit=*n*** es sollen nur die ersten *n* Fälle verwendet werden
 - missing='string'** Kennzeichnung für fehlende Werte
- Bei **statistics**:
 - anova** einfaktorielle Varianzanalyse; falls mehrere Gruppierungsvariablen angegeben sind, erfolgt die Untergruppenbildung nur nach der ersten
 - linearity** zusätzlich Test auf Linearität
 - all** wie **linearity**
 - none** keine Statistik

Hinweise

- Im Unterbefehl **/tables** werden die abhängigen Variablen angegeben und - falls gewünscht - nach dem Schlüsselwort **by** die Kontrollvariablen für die Untergruppenbildung.
- Die Unterbefehle **/title** und **/footnote** dienen zur Definition einer Überschrift bzw. Fußnote. Die Strings können sich über mehrere Zeilen erstrecken, die Maximallänge ist 255.
- Mit dem Unterbefehl **/cells** können die verschiedensten Ausgaben angefordert werden. Bei Untergruppenbildung werden diese Werte pro Zelle (d.h. Gruppe) ausgegeben.

Beispiel

Im folgenden Programm werden die Abiturnoten getrennt nach den Bundesländern aufgelistet und jeweils Mittelwert, Gruppengröße und Standardabweichung ausgegeben.

```
get file='a:\data\demodat.sav'.
summarize tables=abi_note by land
/format=list.
```

4.1.5 Prozedur "Aggregieren" (AGGREGATE)

Mit Hilfe dieser Prozedur lassen sich die Daten von Gruppen so zusammenfassen, daß jede Gruppe anschließend nur noch einen einzelnen Fall darstellt. Die Zusammenfassung der Daten der verschiedenen Fälle einer Gruppe zu einem einzigen Wert erfolgt mit Hilfe einer statistischen Funktion (z.B. des arithmetischen Mittelwerts). Die berechneten Gruppenstatistiken (z.B. die Gruppenmittelwerte) stellen anschließend Rohdaten dar, die der Benutzer für weitere Auswertungen verwenden kann.

DIE PROZEDUR "AGGREGIEREN"	
<i>Aufruf</i>	
1	Menü Daten Aggregieren...
2	Dialogfenster "Daten aggregieren " Break-Variable(n): [_____] Variablen aggregieren: [_____] [Name & Label...] [Funktion...] <input type="checkbox"/> Anzahl der Fälle in der Breakgruppe speichern: [_N_BREAK_] <i>(die Gruppierungsvariablen aus der vorgegebenen Variablenliste auswählen)</i> <input checked="" type="radio"/> Neue Datendatei anlegen [Datei...] <i>(die zu aggregierenden abhängigen Variablen aus der vorgegebenen Variablenliste auswählen)</i> <input type="radio"/> Arbeitsdatei ersetzen
3	Dialogfenster "Daten aggregieren: Variablenname und -label" Name: [_____] Label: [_____]
4	Dialogfenster "Daten aggregieren: Aggregierungsfunktion" Auswertungsfunktion für ausgewählte Variablen: <input checked="" type="radio"/> Mittelwert <input type="radio"/> Erster Wert <input type="radio"/> Letzter Wert <input type="radio"/> Anzahl Fälle <input type="checkbox"/> fehlend <input type="checkbox"/> ungewichtet <input type="radio"/> Standardabweichung <input type="radio"/> Minimalwert <input type="radio"/> Maximalwert <input type="radio"/> Summe Wert: [_____] <input type="radio"/> Prozentsatz oberhalb <input type="radio"/> Prozentsatz unterhalb <input type="radio"/> Anteil oberhalb <input type="radio"/> Anteil unterhalb Kleinster Wert: [_____] Größter Wert: [_____] <input type="radio"/> Prozentsatz innerhalb <input type="radio"/> Prozentsatz außerhalb <input type="radio"/> Anteil innerhalb <input type="radio"/> Anteil außerhalb

DIE AGGREGATE-PROZEDUR

Funktion

- Berechnung von deskriptiven Statistiken für Gruppen und Erzeugung einer SPSS-Datei mit den Gruppenstatistiken als Rohdaten.

Syntax

```

aggregate
  outfile={file|*}
  [/missing=columnwise]
  [/document]
  [/presorted]
  /break=varlist[({a|d})] [varlist ...]
  /aggvar [label'] [aggvar ...] = aggfunction(argument [, argument ...]) [/aggvar ...]

```

Hinweise

- Die Unterbefehle müssen in der oben aufgeführten Reihenfolge verwendet werden.
- Ist **outfile**=* angegeben, so werden die aggregierten Daten in die Arbeitsdatei geschrieben. Im anderen Fall wird die Ausgabe in die angegebene Datei geleitet, und die Arbeitsdatei bleibt unverändert.
- Der Unterbefehl **/missing=columnwise** bewirkt, daß die Zielvariable den Wert SYSMIS erhält, falls bei der entsprechenden Gruppe mindestens ein Fall in der Quellvariablen einen fehlenden Wert aufweist.
- Die Bildung von Gruppen erfolgt beim AGGREGATE-Befehl in analoger Weise wie beim SPLIT FILE-Befehl, d.h. falls sich von einem Fall zum nächsten eine der Gruppierungsvariablen ändert, so wird eine neue Gruppe begonnen. Daher werden die Daten von SPSS vor der Aggregation normalerweise erst sortiert. Ist **/presorted** angegeben, so entfällt dieser Schritt.
- Mit dem Unterbefehl **/break** werden die Gruppierungsvariablen incl. der Sortierrichtung angegeben. Dabei bedeutet **a** aufsteigend oder **d** fallend.
- Zum Schluß sind dann die zu kreierenden Zielvariablen anzugeben (eventuell mit Label) und nach dem Gleichheitszeichen die Aggregierungsfunktion mit den Quellvariablen als Argumente. Als Aggregierungsfunktionen stehen zu Verfügung:

sum	Summe
mean	Mittelwert
sd	Standardabweichung
max	Maximum
min	Minimum
first	Erster nichtfehlender Wert
last	Letzter nichtfehlender Wert
pgt	Prozent größer als der Wert
plt	Prozent kleiner als der Wert
pin	Prozent innerhalb des Intervalls
pout	Prozent außerhalb des Intervalls
fgt	Anteil größer als der Wert
flt	Anteil kleiner als der Wert
fin	Anteil innerhalb des Intervalls
fout	Anteil außerhalb des Intervalls
n	gewichteter Gruppenumfang
nu	ungewichteter Gruppenumfang
nmiss	gewichtete Anzahl fehlender Werte
numiss	ungewichtete Anzahl fehlender Werte

Beispiel

```

get file='a:\data\demodat.sav'.
aggregate outfile=* /break=land /xqabi=mean(abi_note) /nabi=n(abi_note).
sort cases by xqabi.
print / 'Rang', $casenum, ': ', land, ' Anzahl:', nabi, ' Abitur:', xqabi.
execute.

```

4.2 Bivariate deskriptive Statistik

4.2.1 Prozedur "Kreuztabellen" (CROSSTABS)

Mit dieser Prozedur können zwei- oder mehrdimensionale Kreuztabellen erstellt werden. Es stehen auch eine Reihe von Statistiken zur Verfügung: Chi-Quadrat Phi, Kontingenzkoeffizient, Lambda, Cramer's V, Kendall's τ , Produkt-Moment-Korrelation und Spearman's Rangkorrelation u.a. Zusätzlich zu den beobachteten absoluten Häufigkeiten in den Zellen können noch ausgegeben werden: erwartete absolute Häufigkeiten, Prozentwerte (zeilenweise, spaltenweise, gesamt), Residuen (unstandardisiert, standardisiert).

DIE PROZEDUR "KREUZTABELLEN"

Aufruf

1 Menü

Analysieren
 Deskriptive Statistiken >
 Kreuztabellen...

2 Dialogfenster "Kreuztabellen"

Zeilen:
 [_____]

(Zeilenvariable(n) für die Tabelle(n) angeben, d.h. die entsprechenden Variablen aus der vorgegebenen Variablenliste auswählen)

Spalten:
 [_____]

(Spaltenvariable(n) für die Tabelle(n) angeben, d.h. die entsprechenden Variablen aus der vorgegebenen Variablenliste auswählen)

Schicht __1__ von __1__:
 [_____]

(hier kann man weitere Variablen angeben, für drei- oder mehrdimensionale Tabellen; bei mehr als 3 Dimensionen kann mit [Weiter] eine neue Schicht definiert werden)

Gruppierte Balkendiagramme anzeigen
 Keine Tabellen
 [Statistik...]
 [Zellen...]
 [Format...]

3 Dialogfenster "Kreuztabellen: Statistik"

Chi-Quadrat
 Korrelationen
 Nominal:
 Kontingenzkoeffizient
 Phi und Cramer-V
 Lambda
 Unsicherheitskoeffizient

Ordinal:
 Gamma
 Sommers-d
 Kendall-Tau-b
 Kendall-Tau-c

Nominal bezüglich Intervall:
 Eta
 Kappa
 Risiko
 McNemar

- Cochran und Mantel-Haenszel-Statistik
Gemeinsames Quoten-Verhältnis [_1_]

4 Dialogfenster "Kreuztabellen: Zellen"

Häufigkeiten:

- Beobachtet
 Erwartet

Prozentwerte:

- Zeilenweise
 Spaltenweise
 Gesamt

Residuen:

- Nicht standardisiert
 Standardisiert
 Korrigiert standardisiert

5 Dialogfenster "Kreuztabellen: Format"

Zeilenfolge

- Aufsteigend:
 Absteigend

Hinweis

- Wird als Zeilen- oder Spaltenvariable nicht eine einzelne Variable, sondern eine Variablenliste angegeben, so werden die Berechnungen für jede Kombination der Variablen durchgeführt. Dies gilt auch für die Schichtvariable(n).

DIE CROSSTABS-PROZEDUR

Funktion

- Berechnung zwei- oder mehrdimensionaler Häufigkeitsverteilungen für numerische und Zeichenkettenvariablen

Syntax

- "Allgemeiner Modus":

crosstabs

```
[tables=]varlist by varlist [by...] [/varlist...]  
[/missing={table** | include}]  
[/write[={none** | cells}]]  
[/format=[{avalue** | dvalue}] [tables** | notables]  
[/cells=[{count** | none}] [row] [column] [total] [expected] [resid] [sresid] [asresid] [all]  
[/statistics=[chisq] [phi] [cc] [lambda] [uc] [bttau] [ctau] [risk] [gamma] [d] [kappa] [eta] [corr]  
[mcnemar] [all] [none].
```

- "Integermodus":

crosstabs

```
variables=varlist(min,max) [varlist...]  
/tables=varlist by varlist [by...] [/varlist...]  
[/missing={table** | include | report}]  
[/write[={none** | cells | all}]]  
[/format=[{avalue** | dvalue}] [tables** | notables]  
[/cells=[{count** | none}] [row] [column] [total] [expected] [resid] [sresid] [asresid] [all]  
[/statistics=[chisq] [phi] [cc] [lambda] [uc] [bttau] [ctau] [risk] [gamma] [d] [kappa] [eta] [corr]  
[mcnemar] [all] [none]  
[/method={mc[cin({99.0|value})] [samples({10000|value})] | exakt[timer({5|value})]}].
```

Bedeutung der Spezifikationen

- Bei **missing**:

table	Ausschluß fehlender Werte tabellenweise
include	Einbeziehung von Missing-Werten
- Bei **write**:

cells	Ausgabe der Häufigkeiten nur für besetzte Zellen
all	Ausgabe der Häufigkeiten für alle Zellen
- Bei **format**:

tables	Ausgabe von Tabellen
avalue	Kategorien aufsteigend sortiert ausgeben
dvalue	Kategorien absteigend sortiert ausgeben
- Bei **cells**:

count	absolute Häufigkeiten
row	Zeilenprozentwerte
column	Spaltenprozentwerte
total	Prozentwerte total
expected	erwartete Häufigkeiten
resid	Residuen (Differenzen zwischen beobachteten und erwarteten Häufigkeiten)
sresid	standardisierte Residuen
asresid	angepaßte standardisierte Residuen
- Bei **statistics**:

chisq	Chiquadrat (Pearson, Likelihood Ratio)
phi	Phi-Koeffizient und Cramer's V
cc	Kontingenzkoeffizient
lambda	Lambda (symmetrisch und asymmetrisch) und Goodman&Kruskal Tau
uc	Unsicherheitskoeffizient (symmetrisch und asymmetrisch)
btau	Kendall's tau-b
ctau	Kendall's tau-c
gamma	Gamma
d	Sommer's d (symm. und asymm.)
eta	Eta
corr	Produkt-Moment- und Rangkorrelationskoeffizient (nur für num. Daten)
kappa	Kappa Koeffizient
risk	relatives Risiko
mcnemar	McNemar-Test
all	alle verfügbaren Statistiken
none	keine Statistiken

Hinweise

- Obligatorisch ist nur der Unterbefehl **tables=**. Die Variablenliste nach **tables=** definiert die Zeilenvariable(n), jene nach dem ersten **by** die Spaltenvariable(n) und danach folgen evtl. (nach weiteren **by**'s) die Kontrollvariable(n).
- Fehlt der Unterbefehl **variables=**, so arbeitet SPSS im "Allgemeinen Modus": sowohl numerische als auch Stringvariablen sind hier zugelassen; jeder vorkommende distinkte Datenwert stellt eine eigene Kategorie dar; es gibt keine leeren Kategorien (z.B. bei Lücken in den Datenwerten).
- Ist der Unterbefehl **variables=** mit einer Variablenliste und einem Wertebereich (*min,max*) angegeben, so arbeitet SPSS im "Integermodus": hierbei sind nur numerische Variablen zugelassen; eventuelle Dezimalstellen werden abgeschnitten; es wird nur der angegebene Wertebereich berücksichtigt.
- Mit **/cells=** gibt man an, welche Werte für die Zellen ausgegeben werden sollen.
- Mit **/format=** steuert man verschiedene Parameter für die Ausgabe der Tabellen.
- Der Unterbefehl **/write=** bewirkt die Ausgabe der Zellenhäufigkeiten auf eine eigene Datei, deren Namen man mit einem PROCEDURE OUTPUT-Befehl spezifizieren muß
- Mit Hilfe des Unterbefehls **/method=** kann man die Methode der Schätzung der Erwartungswerte bestimmen (exakte Berechnung oder mit der Monte Carlo-Methode).

Beispiel

Im folgenden Programm werden Kreuztabellen für die Variable Geschlecht einerseits und die 9 Matheaufgaben andererseits erstellt und anschließend eine Tabelle für die Variablen Schulbildung der Mutter vs. Schulbildung des Vaters (insgesamt also 10 Kreuztabellen). Zusätzlich wird noch das Chiquadrat berechnet. In den Zellen stehen die beobachteten und die erwarteten Häufigkeiten sowie die standardisierten Residuen.

Mit dem zweiten CROSSTABS-Befehl wird eine dreidimensionale Tabelle für die Variablen Geschlecht, Rauchen und Religion erstellt und die Chiquadrat-Statistik angefordert.

```
get file='a:\data\frabo.sav'.
crosstabs tables=geschl by ma1 to ma9 /schule_m by schule_v
  /statistics=chisq
  /cells=count expected sresid.
crosstabs tables=geschl by rauchen by religion /statistics=chisq.
```

4.2.2 Prozedur "Bivariate Korrelationen" (CORRELATIONS, NONPAR CORR)

Mit diesen Prozeduren werden Produkt-Moment- und Rangkorrelationen berechnet incl. Signifikanzniveau für den ein- oder zweiseitigen Test. Außerdem können noch Mittelwerte und Standardabweichungen sowie Kovarianzen und Kreuzproduktabweichungen berechnet werden.

DIE PROZEDUR "BIVARIATE KORRELATIONEN"

Aufruf

1 Menü

Analysieren
Korrelation ▷
Bivariat...

2 Dialogfenster "Bivariate Korrelationen"

Variablen: _____ *(die zu korrelierenden Variablen aus der Variablenliste auswählen)*

Korrelationskoeffizienten:

- Pearson
 Kendalls tau-b
 Spearman

Test auf Signifikanz:

- Zweiseitig
 Einseitig

Signifikante Korrelationen markieren
[Optionen...]

3 Dialogfenster "Optionen"

Statistik:

- Mittelwerte und Standardabweichungen
 Kreuzproduktabweichungen und Kovarianzen

Fehlende Werte:

- Paarweiser Fallausschluß
 Listenweiser Fallausschluß

DIE CORRELATIONS-PROZEDUR

Funktion

- Berechnung von Produkt-Moment-Korrelationskoeffizienten (incl. Signifikanzprüfung), Abweichungsproduktsummen, Kovarianzen, Mittelwerten und Standardabweichungen
- Ausgabe der Korrelationsmatrix auf Datei möglich

Syntax

correlations

```
[variables=]varlist [with varlist] [/varlist...]
[/missing={pairwise** | listwise} [{include | exclude}]]
[/print={twotail** | onetail} {sig** | nosig}]
[/matrix=out(* | file)]
[/statistics=[descriptives] [xprod] [all]]
```

Bedeutung der Spezifikationen

- Bei **missing**:

pairwise	paarweiser Ausschluß fehlender Werte
listwise	fallweiser Ausschluß fehlender Werte
include	Einbeziehung von Missing-Werten
- Bei **print**:

sig	Test der Korrelationen auf Null
nosig	keine Signifikanztests
onetail	einseitiger Test
twotail	zweiseitiger Test
- Bei **statistics**:

descriptives	Mittelwerte, Standardabweichungen, n
xprod	Summen der Abweichungsprodukte, Kovarianzen
all	alle verfügbaren Statistiken

Hinweise

- Obligatorisch ist nur die Spezifikation **variables=**.
- Mit Hilfe des Unterbefehls **/matrix=** kann man die Ausgabe (Mittelwerte, Standardabweichungen, Anzahl der Fälle, Korrelationen) auf eine Matrix-Datendatei bzw. in die Arbeitsdatei (*) schreiben, um sie für andere Statistikprozeduren als Eingabe zu benutzen.

Beispiel

Im folgenden Programm werden alle Korrelationen zwischen den Variablen Körpergewicht und Körpergröße der Person sowie seiner Eltern incl. deskriptiver Statistiken berechnet. Die Signifikanzprüfung der Koeffizienten soll einseitig erfolgen. Bei fehlenden Werten in einer der beteiligten Variablen soll der jeweilige Fall nicht berücksichtigt werden.

```
get file='a:\data\frabo.sav'.
correlations variables=gewicht gew_m gew_v groesse groe_m groe_v
  /statistics /print=onetail /missing=listwise.
```

DIE NONPAR CORR-PROZEDUR

Funktion

- Berechnung von Rangkorrelationskoeffizienten nach Spearman und Kendall (mit Signifikanzprüfung)
- Ausgabe der Korrelationsmatrix auf Datei möglich

Syntax

```

nonpar corr
  [variables=varlist [with varlist] [/varlist...]]
  [/print={twotail** | onetail} {sig** | nosig} {spearman** | kendall | both}]
  [/sample]
  [/missing={pairwise** | listwise} [include]]
  [/matrix=out({* | file})]

```

Bedeutung der Spezifikationen

- Bei **missing**:

pairwise	paarweiser Ausschluß fehlender Werte
listwise	fallweiser Ausschluß fehlender Werte
include	Einbeziehung von Missing-Werten
- Bei **print**:

sig	Test der Korrelationen auf Null
nosig	keine Signifikanztests
onetail	einseitiger Test
twotail	zweiseitiger Test
spearman	Spearman's rho
kendall	Kendall's tau-b
both	beide Rangkorrelationskoeffizienten

Hinweise

- Obligatorisch ist nur die Spezifikation **variables=**.
- Mit Hilfe des Unterbefehls **/matrix=** kann man die Ausgabe (Anzahl der Fälle, Korrelationen) auf eine Matrix-Datendatei bzw. in die Arbeitsdatei (*) schreiben, um sie für andere Statistikprozeduren als Eingabe zu benutzen.
- Falls zu wenig Speicher zur Verfügung steht, um die Matrizen zu bilden, kann man mit **/sample** aus der Gesamtzahl aller Fälle eine Zufallsstichprobe ziehen lassen.

Beispiel

Im folgenden Programm werden die Spearman'schen Rangkorrelationskoeffizienten zwischen den schulischen Leistungen berechnet. Die Signifikanzprüfung der Koeffizienten soll einseitig erfolgen. Bei fehlenden Werten in einer der beteiligten Variablen soll der jeweilige Fall nicht berücksichtigt werden.

```

get file='a:\data\frabo.sav'.
nonpar corr variables=note_deu to abi_note
  /missing=listwise
  /print=onetail.

```

4.2.3 Prozedur "Partielle Korrelationen" (PARTIAL CORR)

Mit dieser Prozeduren können partielle Korrelationen incl. Signifikanzniveau für den ein- oder zweiseitigen Test und zusätzlich Mittelwerte und Standardabweichungen sowie die Korrelationen nullter Ordnung berechnet werden.

DIE PROZEDUR "PARTIELLE KORRELATIONEN"

Aufruf

1 Menü

Analysieren
Korrelation ▷
Partiell...

2 Dialogfenster "Partielle Korrelationen"

Variablen: [_____] *(die zu korrelierenden Variablen aus der Variablenliste auswählen)*
Kontrollvariablen: [_____] *(die Kontrollvariablen aus der Variablenliste auswählen)*
Test auf Signifikanz:
 Zweiseitig
 Einseitig
 Tatsächliches Signifikanzniveau anzeigen
[Optionen...]

3 Dialogfenster "Optionen"

Statistik:
 Mittelwerte und Standardabweichungen
 Korrelationen nullter Ordnung
Fehlende Werte:
 Listenweiser Fallausschluß
 Paarweiser Fallausschluß

DIE PARTIAL CORR-PROZEDUR

Funktion

- Berechnung von partiellen Produkt-Moment-Korrelationen sowie Mittelwerten, Standardabweichungen und Korrelationen nullter Ordnung aus Rohdaten oder aus einer Korrelationsmatrix (nullter Ordnung)
- Ausgabe der Korrelationsmatrix in die Arbeitsdatei bzw. auf eine externe Datei möglich

Syntax

partial corr

```
[variables=]varlist [with varlist] by varlist[(valuelist)] [/varlist...]  
[/significance={twotail** | onetail}]  
[/statistics={none**} [corr] [descriptives] [badcorr] [all]]  
[/format={matrix** | serial | condensed}]  
[/missing={listwise** | analysis} [{exclude** | include}]]  
[/matrix={none**} [in({* |file})] [out({* |file})]]
```

Bedeutung der Spezifikationen

- Bei **missing**:

analysis	in Analogie zum paarweisen Ausschluß (siehe CORRELATIONS-Prozedur) wird hier bei jeder einzelnen partiellen Korrelation geprüft, ob bei einer der beteiligten Variablen Werte fehlen und damit der Fall auszuschließen ist
listwise	fallweiser Ausschluß bei fehlenden Werten
exclude	Ausschluß von Missing-Werten
include	Einbeziehung von Missing-Werten
• Bei format:	
matrix	vollständige Ausgabe im Matrizenform
serial	nur die nichtredundenten Koeffizienten ausgeben
condensed	Ausgabe ohne Freiheitsgrade und Signifikanzwerte
• Bei statistics:	
corr	Korrelationen nullter Ordnung mit Freiheitsgraden und p-Wert
descriptives	Mittelwerte, Standardabweichungen, n
badcorr	Korrelationen nullter Ordnung, falls partielle Korrelationen nicht berechnet werden können
all	alle verfügbaren zusätzlichen Statistiken
none	keine zusätzlichen Statistiken

Hinweise

- Obligatorisch ist nur die Spezifikation **variables=**. Nach **by** sind die Kontrollvariablen anzugeben. Falls mehrere Kontrollvariablen vorliegen, ist in Klammern anzugeben, welche Ordnung die partiellen Korrelationen haben sollen (Voreinstellung ist 1, d.h. jede Kontrollvariable wird einzeln auspartialisiert).
- Der Unterbefehl **/matrix=in** ist zu verwenden, wenn als Eingabe für die Berechnung der partiellen Korrelationen eine Matrix von Korrelationen nullter Ordnung verwendet werden soll. In Klammern ist hierbei entweder eine Matrix-Datendatei anzugeben oder ein "*", falls die Matrixdaten in der Arbeitsdatei stehen.
- Mit dem Unterbefehl **/matrix=out** kann die Ausgabe (Anzahl der Fälle, Mittelwerte, Standardabweichungen, Korrelationen nullter Ordnung) auf eine Matrix-Datendatei bzw. in die Arbeitsdatei (*) geschrieben werden, um sie beispielsweise anderen Statistikprozeduren als Eingabe zur Verfügung zu stellen.

Beispiel

Im folgenden Programm wird die partielle Korrelation zwischen Körpergewicht und Körpergröße unter Auspartialisierung des Alters berechnet. Zusätzlich werden die Korrelationen nullter Ordnung zwischen den drei Variablen berechnet.

```
get file='a:\data\beispiel.sav'.
partial corr variables=gewicht with groesse by alter
/statistics=corr.
```

4.2.4 Prozedur "Regression" (REGRESSION)

Mit Hilfe dieser Prozedur kann eine einfache oder multiple lineare Regressionsanalyse durchgeführt werden. Es gibt eine Vielzahl von Möglichkeiten für die Auswahl des Regressionsmodells, die Berechnung von Statistiken, die Erstellung von Diagrammen und die Analyse der Residuen. Es können auch individuelle Werte (z.B. Residuen, vorhergesagte Werte u.a.) berechnet und an die Arbeitsdatei angefügt werden.

DIE PROZEDUR "REGRESSION"

Aufruf

1 Menü

Analysieren
Regression ▷
Linear...

2 Dialogfenster "Lineare Regression"

Abhängige Variable:
[_____]

(die Kriteriumsvariable, d.h. jene Variable, die vorhergesagt werden soll, aus der vorgegebenen Variablenliste auswählen)

Block __1__ von __1__:
Unabhängige Variable(n):
[_____]

(die Prädiktorvariablen, d.h. jene Variablen, die für die Vorhersage benutzt werden soll, aus der vorgegebenen Variablenliste auswählen; wird mehr als eine Prädiktorvariable angegeben, so wird eine multiple Regression gerechnet)

Methode: [Einschluß]
[Schrittweise]
[Ausschluß]
[Rückwärts]
[Vorwärts]

(die Methode betrifft die Reihenfolge der Auswahl der Prädiktorvariablen bei einer multiplen Regression; für die einfache Regression ist die Methode unerheblich)

Auswahlvariable:
[_____]
[Bedingung...]

(falls die Analyse auf eine Untergruppe beschränkt werden soll, hier die Gruppenvariable angeben (hiermit wird ein Dialogfenster aufgerufen, in dem der Wert anzugeben ist, der die Untergruppe definiert)

Fallbeschriftungen:
[_____]

(die Variable angeben, deren Werte als Fallbeschriftung dienen sollen)

[WLS >>]
WLS-Gewichtung:
[_____]

(falls ein Weighted Least-Squares Modell zugrunde gelegt werden soll und diese Schaltfläche gewählt wird, ist hier Gewichtsvariable anzugeben, d.h. aus der vorgegebenen Variablenliste auszuwählen))

[Statistik...]
[Diagramme...]
[Speichern...]
[Optionen...]

3 Dialogfenster "Lineare Regression: Statistiken"

Regressionskoeffizienten:
 Schätzer
 Konfidenzintervalle
 Kovarianzmatrix
 Anpassungsgüte des Modells
 Änderung in R-Quadrat
 Deskriptive Statistik
 Teil- und partielle Korrelationen
 Kollinearitätsdiagnose
Residuen:
 Durbin-Watson

- Fallweise Diagnose
 - Ausreißer außerhalb [3] Standardabweichungen
 - Alle Fälle

5 Dialogfenster "Lineare Regression: Diagramme" (*hiermit können verschiedene Diagramme angefordert werden, die im wesentlichen der differenzierten Beurteilung der Modellanpassung dienen*)

Streudiagramm 1 von 1 :

Y: [_____]

X: [_____]

(zur Auswahl stehen hier die Kriteriumsvariable und verschiedene bei der Regression berechnete Variablen wie z.B. die Residuen; man wählt jene Variablen aus, die gegeneinander abgetragen werden sollen)

Diagramme der standardisierten Residuen:

- Histogramm
- Normalverteilungsdiagramm
- Alle partiellen Diagramme erzeugen

6 Dialogfenster "Lineare Regression: Speichern" (*falls z.B. Residuen oder andere berechnete Variablen in die Arbeitsdatei übernommen werden sollen, hier angeben*)

Vorhergesagte Werte:

- Nicht standardisiert
- Standardisiert
- Korrigiert
- Standardfehler des Mittelwerts

Residuen:

- Nicht standardisiert
- Standardisiert
- Studentisiert
- Ausgeschlossen
- Studentisiert, ausgeschl.

Distanzen:

- Mahalanobis
- nach Cook
- Hebelwerte

Einflußstatistiken:

- DfBeta
- Standardisierte(s) DfBeta
- DfFit
- Standardisiertes DfFit
- Kovarianzverhältnis

Vorhersageintervalle:

- Mittelwert Individuell

Konfidenzintervall: [95]%

In neuer Datei speichern:

- Koeffizientenstatistik [Datei...]

Modellinformation in XML-Datei exportieren:

[_____] [Durchsuchen]

7 Dialogfenster "Optionen"

Kriterien für schrittweise Methode:

- F-Wahrscheinlichkeit verwenden
 - Aufnahme: [,05] Ausschluß: [,10]
- F-Wert verwenden
 - Aufnahme: [3,84] Ausschluß: [2,71]

Konstante in Gleichung einschließen

Fehlende Werte:

(Behandlung von fehlenden Werten)

- Listenweiser Ausschluß
- Paarweiser Ausschluß
- Durch Mittelwert ersetzen

DIE REGRESSION-PROZEDUR

Funktion

- Leistungsfähige Prozedur zur multiplen linearen Regression

Syntax

regression

```
[matrix={in({file | *})}][out({file | *})]
[/variables={varlist | collect** | all}]
[/descriptives={defaults} [mean] [stddev] [corr] [cov] [variance] [xprod] [sig] [n] [badcorr] [all]
  [none**]]
[/select=var relation value]
[/missing={({listwise** | pairwise | meansubstitution})} [include]]
[/regwt=var]
[/statistics={defaults**} [r**] [coeff**] [anova**] [outs**] [zpp] [label] [cha] [ci]
  [f] [bcov] [ses] [xtx] [collin] [end] [tol] [selection] [all]]
[/criteria={defaults**} [tolerance({0.0001** | value})] [maxsteps(n)] [pin]({0.05** | value})]
  [pout]({0.10** | value})] [fin]({3.84 | value})] [fout]({2.71 | value})] [cin]({95** | value})]
[/noorigin** | origin]]
/dependent=varlist
{[/method=] stepwise [varlist] | forward [varlist] | backward [varlist] | enter [varlist] |
  remove varlist | test (varlist) (varlist) ...} [...] [/...]}
[/residuals={defaults} [durbin] [outliers ({zresid | tempvars})] [id(var)]
  [normprob ({zresid | tempvars})] [histogram ({zresid | tempvars})] [size ({small | large})]
  [({separate | pooled})]
[/casewise={defaults} [({outliers ({3 | value}) | all})] [plot ({zresid | tempvar})]
  [({dependent pred resid | tempvars})]
[/scatterplot=(var,var)]
[/partialplot={all | varlist}]
[/outfile={covb(filename) | corb(filename)varlist}]
[/save= tempvar(newname) [tempvar(newname)...] [fits]].
```

Bedeutung der Spezifikationen

- Bei **variables**:
 - collect** alle Variablen benutzen, die bei **dependent** und **method** aufgeführt sind
- Bei **descriptives**:
 - defaults** wie **mean**, **stddev**, **corr**
 - mean** Mittelwerte
 - stddev** Standardabweichungen
 - corr** Korrelationsmatrix
 - cov** Kovarianzmatrix
 - variance** Varianzen
 - xprod** Matrix der Kreuzprodukte der Abweichungen vom Mittelwert
 - sig** Signifikanzen für den einseitigen Test der Korrelationen
 - n** Anzahl der Fälle
 - badcorr** Korrelationsmatrix, falls manche Korrelationen nicht berechenbar waren
 - all** alle verfügbaren deskriptiven Statistiken
- Bei **select**:
 - relation** Relationsoperator (EQ, NE, LT, LE, GT, GE) für den logischen Ausdruck
- Bei **missing**:
 - meansubstitution** Ersetzung von fehlenden Werten durch den Mittelwert
 - include** Einbeziehung von Missing-Werten
- Bei **statistics**:
 - defaults** wie **r**, **coeff**, **anova**, **outs**
 - r** R , R^2 , R^2_{adj} , $SE(R)$
 - coeff** Statistiken für die Variablen in der Regression: b , $SE(b)$, b^* , t , p -Wert (zweiseitig)

anova	Varianzanalyse
outs	Statistiken für die Variablen außerhalb der Regression: b (falls die Variable hinzukäme), t, p-Wert, partielle Korrelation
zpp	normale ("zero order"), semipartielle und partielle Korrelationen mit dem Kriterium
cha	R^2 für Änderung zwischen den Steps, zugehöriger F-Wert und p-Wert
ci	95%-Konfidenzintervall für b
f	F-Werte anstelle der t-Werte bei coeff und outs
bcov	Varianz-Kovarianz-Matrix der b
ses	Standardfehler von b^*
xtx	Produktsummenmatrix
collin	Kollinearitäts-Diagnostik
tol	Toleranz, d.h. der durch die anderen Prädiktoren nicht aufgeklärte Varianzanteil der jeweiligen Prädiktorvariablen: $1-R^2_{i,1,2,\dots,k}$
selection	Selektionsstatistiken (z.B. Akaike's Informationskriterium)
all	alle verfügbaren Statistiken außer f , line , end

- Bei **criteria**:
 - defaults** entspricht: `pin(0.05) pout(0.10) tolerance(0.0001)`
 - tolerance(value)** Toleranz, d.h. der durch die anderen Prädiktoren nicht aufgeklärte Varianzanteil eines Prädiktors
 - maxsteps (n)** maximale Anzahl von Schritten
 - pin(value)** p-Wert für F-to-Enter
 - pout(value)** p-Wert für F-to-Remove
 - fin(value)** F-to-Enter
 - fout(value)** F-to-Remove
- Bei **method**:
 - stepwise** schrittweise optimales Hinzunehmen bzw. Entfernen
 - forward** schrittweise aufsteigend (schrittweise optimales Hinzunehmen)
 - backward** schrittweise absteigend (schrittweise optimales Entfernen)
 - enter varlist** forciertes Hinzunehmen der genannten Variablen zu den bisherigen Variablen
 - remove varlist** forciertes Entfernen der genannten Variablen aus den bisherigen Variablen
 - test(varlist)** forcierte Benutzung der genannten Variablen
- Bei **residuals**:
 - defaults** entspricht **histogram(zresid) normplot(zresid) outliers(zresid)**
 - histogram(vars)** Histogramm der angegebenen Variablen (Default: zresid)
 - normplot(vars)** Normalverteilungsplot der angegebenen Variablen (Default: zresid)
 - outliers(vars)** Ausgabe der Werte für die 10 extremsten Ausreißer (Default: zresid)
- Bei **casewise**:
 - defaults** entspricht **outliers(3) plot(zresid) dependent pred resid**
 - outliers(value)** Plot von Ausreißern gemäß dem angegebenen Wert
 - all** Plot aller Fälle
 - plot(var)** Plot der angegebenen Variable
 - varlist** Auflistung der angegebenen Variablen

Hinweise

- Die einzelnen Unterbefehle müssen in einer bestimmten Reihenfolge auftreten:
 - (1) `matrix=`
 - (2) `/variables= /descriptives= /select= /missing= /width= /regwgt=`
 - (3) `/statistics= /criteria= /origin bzw. /norigin`
 - (4) `/dependent=`
 - (5) `/method=`
 - (6) `/residuals= /save= /casewise= /scatterplot= /partialplot=`

Für die nebeneinander aufgeführten Unterbefehle ist die Reihenfolge beliebig. Die Unterbefehle bei (1) und (2) dürfen nur einmal auftreten. Die Teile (3) bis (6) bilden einen "Gleichungsblock" und (5) einen "Methodenblock". Es darf nur einen Gleichungsblock geben, allerdings dürfen innerhalb des Gleichungsblocks mehrere Methodenblöcke auftreten. In diesem Fall bezieht sich Block (6) lediglich auf die letzte Regressionsgleichung.

- Mit Hilfe des Unterbefehls **matrix=** kann man eine SPSS-Matrix-Datendatei einlesen (**in**) bzw. ausgeben (**out**). Steht in Klammern anstelle eines Dateinamens ein *, so wird die Arbeitsdatei benutzt.

- Der Unterbefehl **variables=** darf nur einmal auftreten.
- Obligatorisch sind nur die Unterbefehle **dependent=** und **method=**.
- Mit **/origin** wird ein Modell ohne additive Konstante geschätzt, mit **/noorigin** eines mit additiver Konstante.
- Mit Hilfe von **/select=** kann man eine Untermenge der Fälle für die Analyse (Berechnung der Korrelationen und Regressionsstatistiken) auswählen. Die Analyse der Residuen erfolgt sowohl für die selektierten als auch die unselektierten Fälle.
- Der Unterbefehl **/regwgt=** dient der Spezifikation der Gewichtsvariable bei Schätzung der Parameter nach dem Weighted Least-Squares Modell.
- Die Unterbefehle **/residuals=**, **/casewise=**, **/scatterplot=**, **/partialplot=** und **/save=** dienen der Analyse der Residuen. Die folgenden temporären Variablen (u.a.) können berechnet und bei der Analyse der Residuen verwendet werden:
 - **pred** vorhergesagte Werte
 - **zpred** standardisierte vorhergesagte Werte
 - **sepred** Standardfehler der vorhergesagten Werte
 - **resid** unstandardisierte Residuen
 - **zresid** standardisierte Residuen
 - **sresid** studentisierte Residuen
 - **adjpred** angepaßte vorhergesagte Werte
 - **dresid** unstandardisierte angepaßte Residuen ("deleted residuals")
 - **sdresid** studentisierte angepaßte Residuen
 - **mahal** Mahalanobis Distanzen
 - **cook** Cook Distanzen
 - **lever** Leverage-Werte
 - **dfbeta** Änderung des Regressionskoeffizienten, wenn der aktuelle Fall entfernt würde
 - **sdbeta** standardisiertes dfbeta
 - **dffit** Änderung des vorhergesagten Werts, wenn der aktuelle Fall entfernt würde
 - **sdfit** standardisiertes dffit
 - **covratio** Quotient der Determinanten der beiden Kovarianzmatrizen, wenn zum einen der aktuelle Fall entfernt würde und zum anderen wenn alle Fälle einbezogen sind
 - **mcin** untere (lmcin) und obere (umcin) Grenze des KI für den Mittelwert der Vorhersage
 - **icin** untere (licin) und obere (uicin) Grenze des KI für die Vorhersage eines Einzelwerts
- Mit **/save=** kann man eine oder mehrere dieser Variablen in die Arbeitsdatei übernehmen. Dabei kann man die Variablen auch umbenennen (neuer Name in Klammern dahinter).
- Mit **/scatterplot=** wird ein Scatterdiagramm der angegebenen Variablen angefertigt; den Namen von temporären Variablen muß jeweils ein * vorangestellt werden (z.B. *resid).
- Mit **/partialplot=** werden partielle Plots für die angegebenen unabhängigen Variablen erstellt.

Beispiel

Im folgenden Programm wird die Regression der Körpergröße der Person auf die Körpergröße der Mutter zusammen mit den Default-Verteilungsstatistiken berechnet und eine Analyse der Residuen angefordert.

```
get file='a:\data\frabo.sav'.
regression /descriptives
  /dependent=groesse /method=enter groe_m
  /residuals.
```

4.3 Inferenzstatistik

4.3.1 Prozedur „Nichtparametrische Tests“ (NPAR TESTS)

In dieser Prozedur sind alle gängigen nichtparametrischen Testverfahren zusammengefasst. Im einzelnen sind das Tests für:

- eine Stichprobe:
Binomialtest, Chi-Quadrat-Anpassungstest, Kolmogorov-Smirnov-Test für 1 Stichprobe, Runstest
- zwei unabhängige Stichproben:
Mann-Whitney U-Test, Kolmogorov-Smirnov-Test für 2 Stichproben, Wald-Wolfowitz-Test, Moses-Test
- zwei abhängige Stichproben:
McNemar-Test, Vorzeichentest, Wilcoxon-Test
- k unabhängige Stichproben:
Kruskal-Wallis H-Test, Mediantest
- k abhängige Stichproben:
Cochran Q-Test, Friedman-Test, Kendall W-Test

Auf Wunsch werden dabei neben den üblichen Teststatistiken incl. p-Wert auch deskriptive Statistien berechnet.

DIE PROZEDUR "NICHTPARAMETRISCHE TESTS"

Aufruf

1 Menü

Analysieren

Nichtparam. Tests ▷

- Chi-Quadrat...
- Binomial...
- Sequenzen...
- K-S bei einer Stichprobe...
- Zwei unabhängige Stichproben...
- K unabhängige Stichproben...
- Zwei verbundene Stichproben...
- K verbundene Stichproben...

2a Dialogfenster "Chi-Quadrat-Test" (*Chi-Quadrat-Anpassungstest: v.a. Test auf Gleichverteilung*)

Testvariablen: _____ (die abhängigen Variablen aus der Variablenliste auswählen)

Erwarteter Bereich: _____ (den zu verwendenden Wertebereich angeben)

Aus den Daten

Angegebenen Bereich verwenden:

Minimum: [_____]

Maximum: [_____]

Erwartete Werte:

Alle Kategorien gleich

Werte: [_____]

(die erwarteten Häufigkeiten der Reihe nach eingeben und in die Liste [Hinzufügen]; es müssen soviele Werte eingegeben werden, wie es Kategorien bzw. Klassen gibt!)

[Optionen...]

2b Dialogfenster "Test auf Binomialverteilung" (*Binomialtest*)

Testvariablen: _____ (die abhängigen Variablen aus der Variablenliste auswählen)

Dichotomie definieren:

- Aus den Daten *(es dürfen nur 2 verschiedene Werte vorkommen!)*
 Trennwert: [____] *(Trennwert für Dichotomisierung angeben)*
 Testanteil [__,50_] *(Wert für den Parameter p eingeben)*
 [Optionen...]

2c Dialogfenster "Sequenzentest" (*Sequenzanalyse*)

- Testvariablen: *(die abhängigen Variablen aus der Variablenliste auswählen)*
 [_____
 Trennwert: *(für jeden ausgewählten Trennwert wird der Test gerechnet)*
 Median
 Modalwert
 Mittelwert
 Anderer: [____]
 [Optionen...]

2d Dialogfenster "Kolmogorov-Smirnov-Test bei einer Stichprobe"

- Testvariablen: *(die abhängigen Variablen aus der Variablenliste auswählen)*
 [_____
 Testverteilung:
 Normal *(Normalverteilung)*
 Gleichvert. *(Gleichverteilung)*
 Poisson *(Poissonverteilung)*
 Exponentiell *(Exponentialverteilung)*
 [Optionen...]

2e Dialogfenster "Tests bei zwei unabhängigen Stichproben"

- Testvariablen: *(die abhängigen Variablen aus der Variablenliste auswählen)*
 [_____
 Gruppenvariable: *(die unabhängige Variable aus der Variablenliste auswählen)*
 [_____
 [Gruppen definieren...] *(die zu vergleichenden Gruppen durch Angabe der beiden Werte in der Gruppenvariable spezifizieren; dies erfolgt in einem eigenen Dialogfenster)*

 Welche Tests durchführen?:
 Mann-Whitney U-Test
 Kolmogorov-Smirnov Z
 Extremreaktionen nach Moses
 Wald-Wolfowitz-Sequenzen
 [Optionen...]

2f Dialogfenster "Tests bei mehreren unabhängigen Stichproben"

- Testvariablen: *(die abhängigen Variablen aus der Variablenliste auswählen)*
 [_____
 Gruppenvariable: *(die unabhängige Variable aus der Variablenliste auswählen)*
 [_____
 [Bereich definieren...] *(die zu vergleichenden Gruppen durch Angabe des Wertebereichs, d.h. von Minimum und Maximum in der Gruppenvariable, spezifizieren; dies erfolgt in einem eigenen Dialogfenster)*

 Welche Tests durchführen?:
 Kruskal-Wallis-H
 Median
 [Optionen...]

2g Dialogfenster "Tests bei zwei verbundenen Stichproben"

- Testvariablen: *(die abhängigen Variablen aus der Variablenliste auswählen)*
 [_____
 Ausgewählte Variablenpaare: *(jeweils nacheinander die beiden verbundenen Messungen aus der Variablenliste auswählen; die ausgewählten Variablen erscheinen bei "Aktuelle*
 [_____
aktuelle")

Auswahl" und können in die Liste der ausgewählten Variablenpaare übernommen werden)

Welche Tests durchführen?:

- Wilcoxon
- Vorzeichen
- McNemar

[Optionen...]

2h Dialogfenster "Tests bei mehreren verbundenen Stichproben"

Testvariablen:

[_____]

(die verbundenen Messungen aus der Variablenliste auswählen)

Welche Tests durchführen?:

- Friedman
- Kendall-W
- Cochran-Q

[Statistik...]

3a Dialogfenster "...: Optionen"

Statistik

- Deskriptive Statistik
- Quartile

Fehlende Werte

- Fallausschluß Test für Test
- Listenweiser Fallausschluß

3b Dialogfenster "Mehrere verbundene Stichproben: Statistiken"

Statistik

- Deskriptive Statistik
- Quartile

DIE NPAR TESTS-PROZEDUR

Funktion

- Durchführung der gängigen nichtparametrischen Tests

Syntax (allgemein)

npar tests

Unterbefehl für den Test

`[/missing={analysis**|listwise} [include]]`

`[/sample]`

`[/statistics={descriptives} [quartiles] [all]]`

`[/method={mc[cin({99.0**|wert})] | exact[timer({5|wert})] [samples({10000**|wert})]]`

o Einstichproben-Tests

- Binomial-Test

npar tests

`binomial[({0.5|p})]=varlist[({value1,value2|value})]`

...

- Chiquadrat-Anpassungstest

npar tests

`chisquare=varlist[(lo,hi)] [expected={equal**|f1,f2,...fn}]`

...

- Kolmogorov-Smirnov-Test für 1 Stichprobe

- npar tests**
 - k-s**(**{uniform**[*lo,hi*] | **normal**[*m,sd*] | **poisson**[*m*])=*varlist*
 - ...
- Runs-Test
 - npar tests**
 - runs**(**{mean** | **median** | **mode** | *value*)=*varlist*
 - ...
- Tests für zwei unabhängige Stichproben
 - Mann-Whitney U-Test
 - npar tests**
 - m-w**=*varlist* **by** *var* (*value1,value2*)
 - ...
 - Kolmogorov-Smirnov-Test für 2 Stichproben
 - npar tests**
 - k-s**=*varlist* **by** *var* (*value1,value2*)
 - ...
 - Wald-Wolfowitz-Test
 - npar tests**
 - w-w**=*varlist* **by** *var* (*value1,value2*)
 - ...
 - Moses-Test
 - npar tests**
 - moses**[*(n)*]=*varlist* **by** *var* (*value1,value2*)
 - ...
- Tests für zwei abhängige Stichproben
 - McNemar-Test
 - npar tests**
 - mcnemar**=*varlist* [**with** *varlist* [(**paired**)]]
 - ...
 - Vorzeichentest
 - npar tests**
 - sign**=*varlist* [**with** *varlist* [(**paired**)]]
 - ...
 - Wilcoxon-Test
 - npar tests**
 - wilcoxon**=*varlist* [**with** *varlist* [(**paired**)]]
 - ...
- Tests für k unabhängige Stichproben
 - Kruskal-Wallis H-Test
 - npar tests**
 - k-w**=*varlist* **by** *var* (*value1,value2*)
 - ...
 - Mediantest
 - npar tests**
 - median**[*(value)*]=*varlist* **by** *var* (*value1,value2*)
 - ...
- Tests für k abhängige Stichproben
 - Cochran Q-Test


```

npar tests
  cochran=varlist
  ...
• Friedman-Test
  npar tests
  friedman=varlist
  ...
• Kendall W-Test
  npar tests
  kendall=varlist
  ...

```

Beispiel

Das folgende Programm führt mit Hilfe eines U-Tests nach Mann-Whitney einen nichtparametrischen Mittelwertsvergleich der männlichen und weiblichen Personen hinsichtlich der Körpergröße durch.

```

get file='a:\data\frabo.sav'.
npar tests
  /m-w=groesse by geschl(1 2).

```

4.3.2 Prozeduren für t-Tests

Hiermit können die verschiedenen t-Tests durchgeführt werden:

- t-Test bei einer Stichprobe
- t-Test für zwei unabhängige Stichproben (t_{hom} und t_{het})
- t-Test für zwei abhängige Stichproben (t_{corr})

Neben den Teststatistiken incl. p-Wert werden auch deskriptive Statistiken sowie Konfidenzintervalle ausgegeben. Beim t-Test für zwei unabhängige Stichproben wird zusätzlich der Levene-Test auf Varianzhomogenität durchgeführt.

DIE PROZEDUR "T-TEST BEI EINER STICHPROBE"

Aufruf

- 1 Menü
 - Analysieren
 - Mittelwerte vergleichen ▷
 - T-Test bei einer Stichprobe...
- 2 Dialogfenster "T-Test bei einer Stichprobe"

Testvariable(n):	<i>(die abhängigen Variablen aus der Variablenliste auswählen)</i>
[_____]	
Testwert: [_0_]	<i>(den Populationsmittelwert angeben, gegen den getestet werden soll)</i>
[Optionen...]	
- 3 Dialogfenster "T-Test bei einer Stichprobe: Optionen"

Konfidenzintervall: [_95_] %
Fehlende Werte:
<input checked="" type="radio"/> Fallausschluß Test für Test
<input type="radio"/> Listenweiser Fallausschluß

DIE PROZEDUR "T-TEST BEI UNABHÄNGIGEN STICHPROBEN"

Aufruf

- 1 Menü
Analysieren
Mittelwerte vergleichen ▷
T-Test bei unabhängigen Stichproben...
- 2 Dialogfenster "T-Test bei unabhängigen Stichproben"
Testvariable(n): [] *(die abhängigen Variablen aus der Variablenliste auswählen)*
Gruppenvariable: [] *(die unabhängige Variable aus der Variablenliste auswählen)*
[Gruppen def. ...]
[Optionen...]
- 3 Dialogfenster "Gruppen definieren"
 Angegebene Werte verwenden:
Gruppe 1: [] *(Werte in der Gruppenvariable angeben, die die beiden Gruppen definieren)*
Gruppe 2: []
 Trennwert: [] *(Cutoff-Point für die Gruppenvariable angeben)*
- 4 Dialogfenster "T-Test bei unabhängigen Stichproben: Optionen"
Konfidenzintervall: [_95_] %
Fehlende Werte:
 Fallausschluß Test für Test
 Listenweiser Fallausschluß

DIE PROZEDUR "T-TEST BEI GEPAARTEN STICHPROBEN"

Aufruf

- 1 Menü
Analysieren
Mittelwerte vergleichen ▷
T-Test bei gepaarten Stichproben...
- 2 Dialogfenster "T-Test bei gepaarten Stichproben"
Gepaarte Variablen: [] *(jeweils nacheinander die beiden verbundenen Messungen aus der Variablenliste auswählen; die ausgewählten Variablen erscheinen bei "Aktuelle Auswahl" und können in die Liste der ausgewählten Variablenpaare übernommen werden)*
[Optionen...]
- 3 Dialogfenster "T-Test bei gepaarten Stichproben: Optionen"
Konfidenzintervall: [_95_] %
Fehlende Werte:
 Fallausschluß Test für Test
 Listenweiser Fallausschluß

DIE TTEST-PROZEDUR

Funktion

- Durchführung von t-Tests:
 - Einstichproben t-Test
 - t-Test für zwei unabhängige Stichproben mit homogenen bzw. heterogenen Varianzen
 - t-Test für abhängige Stichproben

Syntax

o Einstichproben-t-Test

t-test

```
/variables=varlist
[/missing={analysis**|listwise} [include]]
[/format={labels**|nolabels}]
```

o t-Test für unabhängige Stichproben

t-test

```
groups=var({1,2**|value|value1,value2})
/variables=varlist
[/missing={analysis**|listwise} [include]]
[/format={labels**|nolabels}]
```

o t-Test für abhängige Stichproben

t-test

```
pairs=varlist [with varlist [(paired)]] [/varlist...]
[/missing={analysis**|listwise} [include]]
[/format={labels**|nolabels}]
```

Hinweise

- Bei **groups**= darf nur eine Variable angegeben werden, nämlich Gruppenvariable. Dahinter stehen in Klammern die beiden Werte, die die Gruppen definieren. Wenn nur ein Wert angegeben ist, so ist dieser als Cutoff-Wert aufzufassen: alle Fälle mit Werten größer oder gleich dem angegebenen Wert kommen in die erste Gruppe und alle anderen in die zweite Gruppe.
- Bei **pairs**= wird jede Variable der ersten Variablenliste mit jeder Variable der zweiten Variablenliste (nach dem Schlüsselwort **with**) verglichen. Ist das Schlüsselwort **with** nicht angegeben, so wird jede Variable der Variablenliste mit jeder anderen verglichen. Bei der Angabe von **(paired)** zusammen mit **with** wird die erste Variable mit der ersten, die zweite Variable mit der zweiten usw. verglichen.
- Mit **/variables**= werden die abhängigen Variablen spezifiziert.
- **/missing=analysis** bewirkt, daß Fälle mit Missing-Werten nur dann ausgeschlossen werden, wenn sie für den jeweiligen Test einen Missing-Wert aufweist.

Beispiel

Im folgenden Programm wird das Körpergewicht des Vaters mit jenem der Mutter hinsichtlich des Mittelwerts verglichen. Es handelt sich hier um abhängige Gruppen.

```
get file='a:\data\frabo.sav'.
t-test pairs=gew_m with gew_v (paired).
```

4.3.3 Prozedur "Mittelwerte" (MEANS)

Mit dieser Prozedur können eine Vielzahl von deskriptiven Statistiken für unabhängige Gruppen berechnet sowie eine einfaktorielle Varianzanalyse und ein Test auf Linearität durchgeführt werden.

DIE PROZEDUR "MITTELWERTE"

Aufruf

1 Menü

Analysieren

Mittelwerte vergleichen ▷

Mittelwerte...

2 Dialogfenster "Mittelwerte"

Abhängige Variable(n):

[_____]

(die abhängigen Variablen aus der vorgegebenen Variablenliste auswählen)

Unabhängige Variablen:

[_____]

(optional; die Gruppenvariablen aus der vorgegebenen Variablenliste auswählen)

Schicht __1__ von __1__:

(für mehrdimensionale Tabellen sind mit [Weiter] jeweils die verschiedenen Schichten zu definieren)

Statistik für erste Schicht:

ANOVA-Tabelle und Eta

Linearitätstest

[Optionen...]

3 Dialogfenster "Mittelwerte: Optionen"

Auswahlliste:

Mittelwert (ist bereits ausgewählt)

Anzahl der Fälle (ist bereits ausgewählt)

Standardabweichung (ist bereits ausgewählt)

Median

Gruppiertes Median

Standardfehler des Mittelwerts

Summe

Minimum

Maximum

Spannweite

Erster Wert

Letzter Wert

Varianz

Kurtosis

Standardfehler der Kurtosis

Schiefe

Standardfehler der Schiefe

Harmonisches

Geometrisches Mittel

Prozent der Gesamtsumme

Prozent der Gesamtanzahl

DIE MEANS-PROZEDUR

Funktion

- Berechnung verschiedener deskriptiver Statistiken für unabhängige Gruppen sowie einfaktorielle Varianzanalyse und Test auf Linearität

Syntax

means

```
[tables={varlist | all} by varlist [by ...] [/varlist...]  
[/missing={table** | include | dependent }]]  
[/cells=[mean**] [count**] [stddev**] [median] [gmedian] [semean] [sum] [min] [max] [range]  
[variance] [kurt] [sekurt] [skew] [seskew] [first] [last] [ncpt] [scpt] [ncpt(var)] [scpt(var)]  
[harmonic] [geometric] [default] [all] [none]]  
[/statistics=[anova] [linearity | all] [none**]]
```

Bedeutung der Spezifikationen

- Bei **cells**:

gmedian	gruppierter Median
first	erster Wert
last	letzter Wert
npct	Prozent an der Gesamtzahl der Fälle
spct	Prozent an der Gesamtsumme
ncpt(var)	Prozent an der Gesamtzahl der Fälle in der angegebenen Gruppenvariable
scpt(var)	Prozent an der Gesamtsumme in der angegebenen Gruppenvariable
harmonic	harmonischer Mittelwert
geometric	geometrischer Mittelwert
default	Mittelwert, Standardabweichung, Anzahl
- Bei **missing**:

table	tabellenweiser Ausschluß bei fehlenden Werten
include	Einbeziehung von Missing-Werten
dependent	Ausschluß von Fällen mit Missing-Werten in den abhängigen Variablen
- Bei **statistics**:

anova	einfaktorielle Varianzanalyse sowie Eta und Eta-Quadrat; falls mehrere Gruppierungsvariablen angegeben sind, erfolgt die Untergruppenbildung nur nach der ersten
linearity	zusätzlich Test auf Linearität
all	wie linearity
none	keine Statistik

Hinweise

- Beim Unterbefehl **/tables=** ist als erstes die Liste der abhängigen Variablen anzugeben, gefolgt von dem Schlüsselwort **by** und der bzw. den unabhängigen Variablen. Werden hier mehrere unabhängige Variablen angegeben, so wird jede getrennt ausgewertet. Wird das Schlüsselwort **by** bei der Tabellenspezifikation mehrfach verwendet, so werden entsprechend höherdimensionale Tabellen gebildet.
- Die Prozedur MEANS hieß früher BREAKDOWN. Dieser Name darf immer noch verwendet werden.

Beispiel

Im folgenden Programm werden Mittelwert, Gruppengröße und Standardabweichung der Abiturnoten - aufgeteilt nach der Schulbildung des Vaters - sowie eine einfaktorielle Varianzanalyse berechnet.

```
get file='a:\data\demodat.sav'.  
means tables=abi_note by schule_v  
/statistics.
```


[Erste]
Test:	
<input checked="" type="radio"/> Zweiseitig	
<input type="radio"/> < Kontrolle	
<input type="radio"/> > Kontrolle	
Keine Varianz-Gleichheit angenommen:	
<input type="checkbox"/> Tamhane-T2	
<input type="checkbox"/> Dunnett-T3	
<input type="checkbox"/> Games-Howell	
<input type="checkbox"/> Dunnett-C	
Signifikanzniveau:	[__,05__]
5 Dialogfenster "Einfaktorielle ANOVA Optionen"	
Statistik:	
<input type="checkbox"/> Deskriptive Statistik	
<input type="checkbox"/> Homogenität der Varianzen	
<input type="checkbox"/> Diagramm der Mittelwerte	
Fehlende Werte:	
<input checked="" type="radio"/> Fallausschluß Test für Test	
<input type="radio"/> Listenweiser Fallausschluß	

DIE ONEWAY-PROZEDUR

Funktion

- Durchführung einer einfachen Varianzanalyse für unabhängige Gruppen und von multiplen Mittelwertsvergleichen

Syntax

oneway

```
varlist by var(min,max)
[/polynomial=n]
[/contrast=coefficient list] [/contrast=...]
[/ranges={ls|duncan|snk|tukeyb|tukey|mod|sdscheffe|range values}({0.05|alpha}) [/ranges=...]
[/harmonic={none**|pair**|all}]
[/format={nolabels**|labels}]
[/statistics={none**} [descriptives] [effects] [homogeneity] [all]]
[/missing={analysis**|listwise} [{exclude**|include}]]
[/matrix={in(*file)}] [out(*file)] [none]
```

Bedeutung der Spezifikationen

- Bei statistics:

descriptives	deskriptive Statistiken für die Gruppen: Gruppenumfang, Mittelwert, Standardabweichung, Standardfehler, Minimum, Maximum, 95% -KI
effects	Effekt-Statistiken: Standardabweichung, Standardfehler und 95% -KI für das Modell mit fixen Effekte sowie Standardfehler, 95% -KI und Schätzung der Varianzkomponenten für das Modell mit Random Effekten
homogeneity	Cochran's C, Bartlett-Box F, Hartley's F_{\max}

Hinweise

- Als erstes ist die Liste der abhängigen Variablen anzugeben, gefolgt von dem Schlüsselwort **by** und der unabhängigen Variablen. Die einzelnen Gruppen ergeben sich aus dem aufgeführten Wertebereich (*min*, *max*) in der unabhängigen Variablen. Fälle mit Werten außerhalb dieses Wertebereichs werden nicht in die Analyse aufgenommen und eventuelle Nachkommastellen werden abgeschnitten.

- Mit Hilfe des Unterbefehls **/matrix=** kann man eine SPSS-Matrix-Datendatei einlesen (**in**) bzw. ausgeben (**out**). Steht in Klammern anstelle eines Dateinamens ein *, so wird die Arbeitsdatei benutzt. Die SPSS-Matrix-Datendatei enthält die Mittelwerte, Standardabweichungen und Gruppenumfänge.
- Mit dem Unterbefehl **/polynomial=** testet man polynomielle Kontraste bis zum angegebenen Grad. Andere spezielle Kontraste sind jeweils mit Hilfe des Unterbefehls **/contrast=** mit der Liste der Kontrastkoeffizienten dahinter zu spezifizieren.
- Für multiple Mittelwertvergleiche nach den verschiedensten Methoden steht der Unterbefehl **/ranges=** zur Verfügung. Liegen ungleiche Gruppengrößen vor, so wird für die Mittelwertvergleiche entweder das harmonische Mittel der jeweiligen beiden Mittelwerte zugrunde gelegt (**/harmonic=none** oder **pair**) oder das harmonische Mittel aller Gruppenmittelwerte (**/harmonic=all**).

Beispiel

Im folgenden Programm wird getestet, ob die Note im Abitur von der Schulbildung des Vaters abhängig ist. Speziell soll noch ein linearer Trend getestet werden.

```
get file='a:\data\frabo.sav'.
oneway abi_note by schule_v
  /polynomial=1.
```

4.3.5 Prozeduren für die mehrfaktorielle Varianzanalyse (UNIANOVA, ANOVA)

Dies ist eine sehr leistungsfähige Prozedur zur Durchführung einer mehrfaktoriellen Varianz- und Kovarianzanalyse für unabhängige Gruppen. Dabei können die zu berücksichtigenden Effekte (das zu analysierende Modell) angegeben werden, außerdem zu testende Kontraste sowie durchzuführende post hoc Mittelwertvergleiche. Neben deskriptiven Statistiken lassen sich auch Profildiagramme anfordern. Die vorhergesagten Werte, Residuen u.a. können als neue Variablen in die Arbeitsdatei übernommen werden.

DIE PROZEDUR "ALLGEMEINES LINEARES MODELL: UNIVARIAT"

Aufruf

1 Menü

Analysieren

Allgemeines lineares Modell ▷
Univariat...

2 Dialogfenster "Univariat"

Abhängige Variable:

[_____]

(die abhängigen Variable aus der Variablenliste auswählen)

Feste Faktoren:

[_____]

(die unabhängigen Variablen mit fixen Effekten aus der Variablenliste auswählen)

Zufallsfaktoren:

[_____]

(die unabhängigen Variablen mit Random Effekten aus der Variablenliste auswählen)

Kovariaten:

[_____]

(für eine Kovarianzanalyse hier die Kovariaten aus der Variablenliste auswählen)

WLS-Gewichtung:

[_____]

[Modell...]

[Kontraste...]

[Diagramme...]

[Post Hoc...]

[Speichern...]
[Optionen...]

3 Dialogfenster "Univariat: Modell"

Modell angeben:

Gesättigtes Modell

Anpassen

Term(e) konstruieren:

[Wechselwirkung

[Haupteffekte]

[Alle 2-Weg]

[Alle 3-Weg]

[Alle 4-Weg]

[Alle 5-Weg]

Quadratsumme: [Typ III

[Typ I]

[Typ II]

[Typ IV]

Konstanten Term in Modell einschließen

(falls kein gesättigtes Modell getestet werden soll, müssen hier die Terme des anzupassenden Modells mit Hilfe der vorgegebenen Liste der Faktoren und Kovariaten angegeben werden)

(hier ist die Methode der Quadratsummenzerlegung anzugeben)

4 Dialogfenster "Univariat: Kontraste"

Faktoren: [_____]

Kontrast: [Keine

[Abweichung]

[Einfach]

[Differenz]

[Helmert]

[Wiederholt]

[Polynomial]

Referenzkategorie:

Letzte

Erste

(den entsprechenden Faktor auswählen)

(hier können zu dem ausgewählten Faktor lineare Kontraste spezifiziert werden)

(für einfache oder Abweichungskontraste)

5 Dialogfenster "Univariat: Profilplots" (zur graphischen Darstellung von Haupteffekten und/oder Interaktionen)

Horizontale Achse:

[_____]

Separate Linien:

[_____]

Separate Diagramme:

[_____]

(aus der Liste der Faktoren jenen auswählen, der auf der x-Achse abgetragen werden soll)

(jede Stufen dieses Faktors soll eine separate Linie bilden)

(für jede Stufen des hier angegebenen Faktors wird ein separates Diagramm gezeichnet)

6 Dialogfenster "Univariat: Post-Hoc-Mehrfachvergleiche für den beobachteten Mittelwert"

Post-Hoc_Tests für:

[_____]

Varianz-Gleichheit angenommen:

LSD

Bonferroni

Sidak

Scheffé

F nach R-E-G-W

Q nach R-E-G-W

S-N-K

Tukey

Tukey-B

Duncan

GT2 nach Hochberg

Gabriel

Waller-Duncan

Typ I/Typ II Fehler-Quotient: [_100_]

(die entsprechenden Faktoren auswählen)

(geringste signifikante Differenz)

(Student-Newman-Keuls)

(Tukey ehrlich signifikante Differenz)

(Duncans Test für multiple Mittelwertsvergleiche)

Dunnett
 Kontrollkategorie: [Letzte [☐]]
 [Erste]

Test:
 Zweiseitig
 < Kontrolle
 > Kontrolle

Keine Varianz-Gleichheit angenommen:

Tamhane-T2
 Dunnett-T3
 Games-Howell
 Dunnett-C

7 Dialogfenster "Univariat: Speichern"

Vorhergesagte Werte:

Nicht standardisiert
 Gewichtet
 Standardfehler

Residuen:

Nicht standardisiert
 Gewichtet
 Standardisiert
 Studentisiert
 Ausgeschlossen

Diagnose:

Cook-Distanz
 Hebelwerte

In neuer Datei speichern:

Koeffizientenstatistik [Datei...]

8 Dialogfenster "Univariat: Optionen"

Geschätzte Randmittel:

Mittelwerte anzeigen für:

[_____]

Haupteffekte vergleichen

Anpassung des Konfidenzintervalls:

[LSD (kein) [☐]]

[Bonferroni]

[Sidak]

(aus der Liste der Effekte jene auswählen, für die eine Mittelwertstabelle ausgegeben werden soll)

Anzeigen:

Deskriptive Statistik
 Schätzer der Effektgröße
 Beobachtete Schärfe
 Parameterschätzer
 Matrix Kontrastkoeffizienten
 Homogenitätstests
 Diagramm: Streubreite vs. mittleres Niveau
 Residuen-Diagramm
 Fehlende Anpassung
 Allgemeine schätzbare Funktion

Signifikanzniveau: [__,05__]

DIE UNIANOVA-PROZEDUR

Funktion

- Neuere leistungsstarke Prozedur zur Varianz- und Kovarianzanalyse für faktorielle Designs

Syntax

unianova

```

dependent_var [by factorlist [with covariate]]
[/random=factor [factor ...]]
[/regwt=var]
[/method=sstype({1 | 2 | 3** | 4})]
[/intercept={include** | exclude}]
[/missing={include | exclude**}]
[/criteria=[eps({1e-8** | a}) [alpha({0.05** | a})]]
[/print=[descriptive] [homogeneity] [parameter] [etasq] [gef] [lof] [opower] [test(lmatrix)]]
[/plot=[spreadlevel] [residuals] [profile(factor factor*factor factor*factor*factor ...)]
[/test=effect vs {linear_combination [df(df)] | value df(df)}]
[/lmatrix={"label"} effect list effect list ...; ... | {"label"} effect list effect list ... | {"label"} all list; all ... | {"label"} all list }
[/kmatrix={number | number; ...}]
[/contrast(factorname)={deviation(refcat)** | simple(refcat) | difference | helmert | repeated | polynomial({1,2,3,... | metric}) | special(matrix)}]
[/posthoc=effect [effect ...] ([snk] [tukey] [btukey] [duncan] [scheffe] [dunnett(refcat)] [dunnett1(refcat)] [dunnett2(refcat)] [bonferroni] [lsd] [sidak] [gt2] [gabriel] [fregw] [qregw] [t2] [t3] [gh] [c] [waller({100** | kratio})] [vs effect]
[/emmeans=[tables({overall | factor | factor*factor...}) [compare adj([lsd(none)] [bonferroni] [sidak])]
[/save=[tempvar(name)] [tempvar(name)]...]
[/outfile={covb(filename) | corb(filename)}] [effect(filename)] [design(filename)]
[/design={intercept ... | effect effect...}

```

Bedeutung der Spezifikationen

- Bei **criteria**:

eps (a)	Toleranzkriterium
alpha (a)	Irrtumswahrscheinlichkeit für Schätzung der Teststärke und für Konfidenzintervall
- Bei **print**:

descriptives	deskriptive Statistiken
homogeneity	Varianzhomogenitätstests
parameter	Parameterschätzungen
etasq	Effektstärke (Eta-Quadrat)
gef	Tabelle der schätzbaren Funktionen
lof	Anpassungsfehler
opower	beobachtete Teststärke
test	Matrix der Kontrastkoeffizienten
- Bei **plot**:

spreadlevel	Streubreite vs. mittleres Niveau plotten
residuals	Residuendiagramm
profile	Liniendiagramme
- Bei **posthoc**:

snk	Student-Newman-Keuls
tukey	Tukey's HSD
btukey	Tukey's b
duncan	Duncan
scheffe	Scheffé
dunnett(refcat)	Dunnett zweiseitig (Referenzstufe)
dunnett1(refcat)	Dunnett einseitig: Mittelwerte kleiner? (Referenzstufe)
dunnett2(refcat)	Dunnett einseitig: Mittelwerte größer? (Referenzstufe)
dunnett	Dunnett (zweisei)
bonferroni	Bonferroni
lsd	Least significant difference
sidak	Sidak t Test
gt2	Hochberg's GT2
gabriel	Gabriel's multipler Parrvergleichstest
fregw	Verfahren von Ryan-Einot-Gabriel-Welsch, basierend auf einem F Test

qregw	Ryan-Einot-Gabriel-Welsch, basierend auf dem Studentized Range Test
t2	Tamhane's T2
t3	Dunnett's T3
gh	Games und Howell
c	Dunnett's C
waller	t-Test nach Waller-Duncan

- Bei **outfile:**
 - covb** Kovarianzmatrix der Parameterschätzungen
 - corb** Korrelationsmatrix der Parameterschätzungen
 - effect** Statistiken der Varianzanalysetabelle
 - design** Designmatrix

Hinweise

- Als erstes ist die abhängige Variable anzugeben, gefolgt von **by** und der Liste der Faktoren. Falls eine Kovarianzanalyse gerechnet werden soll, folgt danach hinter dem Schlüsselwort **with** die Liste der Kovariaten.
- Faktoren mit Zufallseffekten sind im Unterbefehl **/random=** anzugeben.
- Der Unterbefehl **/method=** legt die Methode der Quadratsummenzerlegung fest (relevant bei nicht-orthogonalen Versuchsplänen).
- Mit **/regwgt=** spezifiziert man die Gewichtsvariable für ein WLS-Modell.
- Der Unterbefehl **/intercept=** gibt an, ob das Modell eine additive Konstante enthalten soll.
- Der Unterbefehl **/test=** erlaubt dem Benutzer, für einen Effekte anzugeben, gegen welchen Fehlerterm er getestet werden soll.
- Mit **/lmatrix=** kann der Benutzer eine Koeffizientenmatrix nach dem ALM spezifizieren.
- Der Unterbefehl **/kmatrix=** erlaubt es dem Benutzer, seine Hypothesen in Form einer Kontrastergebnismatrix näher zu spezifizieren. Per Voreinstellung ist diese Matrix die Nullmatrix.
- Mit Hilfe des Unterbefehls **/contrast=** kann man den Typ der Kontrastmatrix für jeden Faktor spezifizieren.
- Der Unterbefehl **/posthoc=** gestattet die Durchführung von Post-Hoc-Mittelwertsvergleichen.
- Mit dem Unterbefehl **/emmeans=** spezifiziert man die Ausgabe von Mittelwertstabellen.
- Mit **/save=** kann man eine oder mehrere berechnete temporäre Variablen in die Arbeitsdatei übernehmen. Dabei kann man die Variablen auch umbenennen (neuer Name in Klammern dahinter). Folgende Variablen können verwendet werden:
 - **pred** unstandardisierte vorhergesagte Werte
 - **wpred** gewichtete unstandardisierte vorhergesagte Werte
 - **resid** unstandardisierte Residuen
 - **wresid** gewichtete unstandardisierte Residuen
 - **dresid** unstandardisierte angepaßte Residuen ("deleted residuals")
 - **zresid** standardisierte Residuen
 - **sresid** studentisierte Residuen
 - **sepred** Standardfehler der vorhergesagten Werte
 - **cook** Cook Distanzen
 - **lever** unzentrierte Leverage-Werte
- Der Unterbefehl **/design=** erlaubt es, das Design zu spezifizieren. Fehlt dieser Unterbefehl, so wird ein gesättigtes Modell angepaßt (vollständiges faktorielles Design).
- Beim Unterbefehl **/outfile=** erfolgt die Ausgabe auf eine externe SPSS-Datei oder - bei Angabe von "*" - in die Arbeitsdatei.

Beispiel

Im folgenden Programm geht es um die Analyse der Deutschpunkte in Abhängigkeit vom Geschlecht und vom Bundesland.

```
get file='a:\data\frabo.sav'.
unianova pkte_deu by geschl land.
```

DIE ANOVA-PROZEDUR

Funktion

- Ältere Prozedur zur Varianzanalyse für faktorielle Designs (nur über die Syntax aufrufbar)

Syntax

```
anova
  [variables=] varlist by varlist(min,max) [by...] [with varlist] [variables=...]
  [covariates={first**|with|after}]
  [maxorders={all**|n|none}]
  [method={unique**|experimental|hierarchical}]
  [statistics={mca|reg|mean|all|none}]
  [missing=[{exclude**|include}] ]
  [format={labels**|nolabels}]
```

Bedeutung der Spezifikationen

- Bei statistics:

mca	Klassifikationsanalyse
reg	unstandardisierte Regressionskoeffizienten
mean	Tabelle der Mittelwerte und Gruppenumfänge
- Bei method:

unique	Regressionsansatz
experimental	klassischer experimenteller Ansatz
hierarchical	hierarchischer Ansatz

Hinweise

- Als erstes ist die Liste der abhängigen Variablen anzugeben, gefolgt von **by** und der Liste der unabhängigen Variablen. Die einzelnen Stufen der Faktoren ergeben sich jeweils aus dem aufgeführten Wertebereich (*min*, *max*). Fälle mit Werten außerhalb dieses Wertebereichs werden nicht in die Analyse aufgenommen und eventuelle Nachkommastellen werden abgeschnitten.
- Soll eine Kovarianzanalyse gerechnet werden, so muß nach den Faktoren das Schlüsselwort **with** folgen und dahinter die Liste der Kovariaten. Mit Hilfe des Unterbefehls **/covariates=** kann man festlegen, ob bei der Quadratsummenzerlegung die Kovariaten vor oder nach den Haupteffekten (**first** bzw. **after**) oder gleichzeitig mit ihnen (**with**) verarbeitet werden sollen.
- Mit **/maxorders=** kann man die Ordnung der zu testenden Interaktionen einschränken.
- Im Unterbefehl **/method=** wird die zu wählende Methode für die Quadratsummenzerlegung angegeben:
 - Beim Regressionsansatz werden aus allen Effekten alle anderen jeweils auspartialisiert.
 - Beim klassischen experimentellen Ansatz werden zuerst die Kovariaten, dann die Haupteffekte, dann die 2-fachen Interaktionen, usw. berücksichtigt.
 - Beim hierarchischen Ansatz werden zuerst die Haupteffekte und Kovariaten berücksichtigt. Dabei werden jeweils aus einem Effekt die bisher berücksichtigten Effekte auspartialisiert. Die Reihenfolge der Effekte ergibt sich aus der Reihenfolge der Faktoren in der Faktorenliste hinter **by** und aus der Spezifikation bei **/covariates=**. Nach den Haupteffekten und Kovariaten werden dann die Interaktionen berücksichtigt.

Beispiel

Im folgenden Programm wird eine zweifaktorielle Varianzanalyse durchgeführt. Die Faktoren sind Geschlecht und PC-Besitz, die abhängige Variablen die Punkte in Deutsch.

```
get file='a:\data\frabo.sav'.
anova pkte_deu by geschl(1,2) pc(0,1).
```


5.1.1 Balkendiagramme

Diese Prozedur erlaubt die Erstellung von einfachen, gruppierten oder gestapelten Balkendiagrammen. Dabei können Gruppendaten für verschiedene Werte einer unabhängigen Variable, Gruppendaten für verschiedene abhängige Variablen (z.B. deren Mittelwerte) oder auch individuelle Daten in einer oder mehreren abhängigen Variablen graphisch dargestellt werden.

DIE GRAPHIKPROZEDUR "BALKEN"

Aufruf

1 Menü

Grafiken

Balken...

2 Auswahlbox "Balkendiagramme": *(Typ des Diagramms auswählen)*

Einfach

Gruppirt

Gestapelt

Daten im Diagramm:

Auswertung über Kategorien einer Variablen

Auswertung über verschiedene Variablen

Werte einzelner Fälle

3a Dialogfenster "Einfaches Balkendiagramm definieren: Auswertung über Kategorien einer Variablen"

Bedeutung der Balken: *(die y-Achse definieren)*

Anzahl der Fälle

% der Fälle

Kumul. Anzahl der Fälle

Kum. % der Fälle

Andere Auswertungsfunktion:

Variable:

[_____]

(die abhängige Variable auswählen sowie die Funktion, die auf sie angewandt werden soll)

[Auswertungsfunktion...]

Kategorienachse:

(die Variable für die x-Achse auswählen)

[_____]

Vorlage:

Diagrammeinstellungen verwenden aus:

[Datei...]

(Graphikdatei auswählen, die als Muster für die Graphikeinstellungen dienen soll)

[Titel...]

[Optionen...]

3b Dialogfenster "Einfaches Balkendiagramm definieren: Auswertung über verschiedene Variablen"

Bedeutung der Balken: *(die Achsen definieren)*

[_____]

(die Variablen auswählen und die Funktion, die jeweils dargestellt werden soll; auf der x-Achse werden die angegebenen Variablen abgetragen, auf der y-Achse deren Funktionswerte)

[Auswertungsfunktion...]

Vorlage:

Diagrammeinstellungen verwenden aus:

[Datei...]

(Graphikdatei auswählen, die als Muster für die Graphikeinstellungen dienen soll)

[Titel...]

[Optionen...]

3c Dialogfenster "Einfaches Balkendiagramm definieren: Werte einzelner Fälle"

Bedeutung der Balken: *(die y-Achse definieren)*

[_____]

(die Variable auswählen, deren Werte dargestellt werden sollen)

Kategorienbeschriftungen:

- Fallnummer
 Variable [_____]

(Beschriftung der x-Achse: entweder die Fallnummer oder Wert in einer anzugebenden Variablen)

Vorlage:

- Diagrammeinstellungen verwenden aus:

[Datei...]

(Graphikdatei auswählen, die als Muster für die

[Titel...]

Graphikeinstellungen dienen soll)

3d Dialogfenster "Gruppiertes Balkendiagramm definieren: Auswertung über Kategorien einer Variable"

Bedeutung der Balken: (die y-Achse definieren)

- Anzahl der Fälle
 Kumul. Anzahl der Fälle
 Andere Auswertungsfunktion:

% der Fälle

Kum. % der Fälle

Variable:

[_____]

(die abhängige Variable auswählen sowie die Funktion, die auf sie angewandt werden soll)

[Auswertungsfunktion...]

Kategorienachse:

[_____]

(die unabhängige Variable angeben, deren Werte die einzelnen Balkengruppen auf der x-Achse bilden sollen)

Gruppen definieren durch:

[_____]

(die unabh. Variable angeben, deren Werte die einzelnen Balken innerhalb einer Balkengruppe bilden sollen)

Vorlage:

- Diagrammeinstellungen verwenden aus:

[Datei...]

(Graphikdatei auswählen, die als Muster für die

[Titel...]

Graphikeinstellungen dienen soll

[Optionen...]

3e Dialogfenster "Gruppiertes Balkendiagramm definieren: Auswertung über verschiedene Variablen"

Bedeutung der Balken: (die y-Achse definieren)

[_____]

[Auswertungsfunktion...]

(die abhängigen Variablen auswählen und jeweils die Funktion, die auf sie angewandt werden soll; die angegebenen Variablen bilden die verschiedenen Balken in einer Balkengruppe)

Kategorienachse:

[_____]

(die unabhängige Variable angeben, deren Werte die einzelnen Balkengruppen auf der x-Achse bilden sollen)

Vorlage:

- Diagrammeinstellungen verwenden aus:

[Datei...]

(Graphikdatei auswählen, die als Muster für die

[Titel...]

Graphikeinstellungen dienen soll)

[Optionen...]

3f Dialogfenster "Gruppiertes Balkendiagramm definieren: Werte einzelner Fälle"

Bedeutung der Balken: (die y-Achse definieren)

[_____]

(die abhängigen Variablen angeben, deren Werte dargestellt werden sollen; sie bilden die verschiedenen Balken innerhalb einer Balkengruppe; eine einzelne Balkengruppe entspricht einem einzelnen Fall)

Kategorienbeschriftungen

- Fallnummer
 Variable [_____]

(Beschriftung der x-Achse: entweder die Fallnummer oder Wert in einer anzugebenden Variable)

Vorlage:

- Diagrammeinstellungen verwenden aus:

[Datei...]

(Graphikdatei auswählen, die als Muster für die

[Titel...]

Graphikeinstellungen dienen soll)

3g Dialogfenster "Gestapeltes Balkendiagramm definieren: Auswertung über Kategorien einer Variable"

Bedeutung der Balken: (die y-Achse definieren)

- Anzahl der Fälle
 Kumul. Anzahl der Fälle
 Andere Auswertungsfunktion:

% der Fälle

Kum. % der Fälle

Variable:

(die abhängige Variable auswählen sowie die Funktion,

[_____]
[Auswertungsfunktion...] *die auf sie angewandt werden soll)*

Kategorienachse: [_____] *(die unabhängige Variable angeben, deren Werte die einzelnen Balkenstapel auf der x-Achse bilden sollen)*

Stapel definieren durch: [_____] *(die unabhängige Variable angeben, deren Werte die einzelnen Teilbalken innerhalb eines Balkenstapels darstellen sollen)*

Vorlage:
 Diagrammeinstellungen verwenden aus:
 [Datei...]
 [Titel...]
 [Optionen...]

3h Dialogfenster "Gestapeltes Balkendiagramm definieren: Auswertung über verschiedene Variablen"

Bedeutung der Balken: *(die y-Achse definieren)*
 [_____] *(die abhängigen Variablen auswählen und jeweils die Funktion, die auf sie angewandt werden soll; die angegebenen Variablen bilden die verschiedenen Teilbalken eines Balkenstapels)*
 [Auswertungsfunktion...]

Kategorienachse: [_____] *(die unabhängige Variable angeben, deren Werte die einzelnen Balkenstapel auf der x-Achse bilden sollen)*

Vorlage:
 Diagrammeinstellungen verwenden aus:
 [Datei...] *(Graphikdatei auswählen, die als Muster für die Graphikeinstellungen dienen soll)*
 [Titel...]
 [Optionen...]

3i Dialogfenster "Gestapeltes Balkendiagramm definieren: Werte einzelner Fälle"

Bedeutung der Balken: *(die y-Achse definieren)*
 [_____] *(die abhängigen Variablen angeben, deren Werte dargestellt werden sollen; sie bilden die verschiedenen Teilbalken innerhalb eines Balkenstapels; ein einzelner Balkenstapel entspricht einem einzelnen Fall)*

Kategorienbeschriftungen
 Fallnummer *(Beschriftung der x-Achse: entweder die Fallnummer oder Wert in einer anzugebenden Variable)*
 Variable [_____]

Vorlage:
 Diagrammeinstellungen verwenden aus:
 [Datei...] *(Graphikdatei auswählen, die als Muster für die Graphikeinstellungen dienen soll)*
 [Titel...]

4 Dialogfenster "Titel"

Titel
 Zeile 1: [_____]

 Zeile 2: [_____]

 Untertitel: [_____]

 Fußnote
 Zeile 1: [_____]

 Zeile 2: [_____]

5 Dialogfenster "Optionen"

Missing Werte:
 Listenweiser Fallausschluß
 Fälle Variable für Variable ausschließen
 Fehlende Werte als Kategorie anzeigen
 Grafik mit Fallbeschriftungen anzeigen

5.1.2 Liniendiagramme

Diese Prozedur erlaubt die Erstellung von einfachen, mehrfachen und verbundenen Liniendiagrammen. Einfache Liniendiagramme bestehen aus einem einzelnen Linienzug und mehrfache Liniendiagramme aus mehreren übereinandergezeichneten Linienzügen. Bei verbundenen Liniendiagrammen werden die Daten anstatt durch Linienzüge durch verschiedene Symbole dargestellt, die durch senkrechte Linien verbunden sind. Durch diese sollen Unterschiede (z.B. zwischen Gruppen) besonders augenfällig aufgezeigt werden. Dabei können Gruppendaten für verschiedene Werte einer unabhängigen Variable, Gruppendaten für verschiedene abhängige Variablen (z.B. deren Mittelwerte) oder auch individuelle Daten in einer oder mehreren abhängigen Variablen graphisch dargestellt werden.

DIE GRAPHIKPROZEDUR "LINIEN"

Aufruf

1 Menü

Grafiken
Linie...

2 Auswahlbox "Liniendiagramme" (Typ des Diagramms auswählen):

- Einfach
 Mehrfach
 Verbundlinie

Daten im Diagramm:

- Auswertung über Kategorien einer Variablen
 Auswertung über verschiedene Variablen
 Werte einzelner Fälle

3a Dialogfenster "Einfaches Liniendiagramm definieren: Auswertung über Kategorien einer Variablen"

Linie entspricht: *(die y-Achse definieren)*

- Anzahl der Fälle
 Kumul. Anzahl der Fälle
 Andere Auswertungsfunktion:

% der Fälle
 Kum. % der Fälle
(die abhängige Variable auswählen sowie die Funktion, die auf sie angewandt werden soll)

Variable:
 [_____]
 [Auswertungsfunktion...]

Kategorienachse:

[_____]

(die Variable für die x-Achse auswählen)

Vorlage:

Diagrammeinstellungen verwenden aus:

[Datei...]

(Graphikdatei auswählen, die als Muster für die Graphikeinstellungen dienen soll)

[Titel...]

[Optionen...]

3b Dialogfenster "Einfaches Liniendiagramm definieren: Auswertung über verschiedene Variablen"

Linie entspricht: *(die Achsen definieren)*

[_____]

[Auswertungsfunktion...]

(die Variablen auswählen und die Funktion, die jeweils dargestellt werden soll; auf der x-Achse werden die angegebenen Variablen abgetragen, auf der y-Achse deren Funktionswerte)

Vorlage:

Diagrammeinstellungen verwenden aus:

[Datei...]

(Graphikdatei auswählen, die als Muster für die Graphikeinstellungen dienen soll)

[Titel...]

[Optionen...]

3c Dialogfenster "Einfaches Liniendiagramm definieren: Werte einzelner Fälle"

Linie entspricht: *(die y-Achse definieren)*

[_____] *(die Variable auswählen, deren Werte dargestellt werden sollen)*

Kategorienbeschriftungen:

- Fallnummer *(Beschriftung der x-Achse: entweder die Fallnummer oder Wert in einer anzugebenden Variable)*
 Variable [_____]

Vorlage:

Diagrammeinstellungen verwenden aus:

[Datei...] *(Graphikdatei auswählen, die als Muster für die*

[Titel...] *Graphikeinstellungen dienen soll)*

3d Dialogfenster "Mehrfachliniendiagramm definieren: Auswertung über Kategorien einer Variable"

Linien entsprechen: *(die y-Achse definieren)*

- Anzahl der Fälle % der Fälle
 Kumul. Anzahl der Fälle Kum. % der Fälle
 Andere Auswertungsfunktion:
 Variable: *(die abhängige Variable auswählen sowie die Funktion, die auf sie angewandt werden soll)*
 [_____]
 [Auswertungsfunktion...]

Kategorienachse: *(die unabhängige Variable angeben, deren Werte die x-Achse bilden sollen)*

[_____]

Linien definieren durch: *(die unabhängige Variable angeben, deren Werte die verschiedenen Linien bilden sollen)*

[_____]

Vorlage:

Diagrammeinstellungen verwenden aus:

[Datei...] *(Graphikdatei auswählen, die als Muster für die*

[Titel...] *Graphikeinstellungen dienen soll*

[Optionen...]

3e Dialogfenster "Mehrfaches Liniendiagramm definieren: Auswertung über verschiedene Variablen"

Linien entsprechen: *(die y-Achse definieren)*

[_____] *(die abhängigen Variablen auswählen und jeweils die Funktion, die auf sie angewandt werden soll; sie bilden die einzelnen Linien)*
 [Auswertungsfunktion...]

Kategorienachse: *(die unabhängige Variable angeben, deren Werte die x-Achse bilden sollen)*

[_____]

Vorlage:

Diagrammeinstellungen verwenden aus:

[Datei...] *(Graphikdatei auswählen, die als Muster für die*

[Titel...] *Graphikeinstellungen dienen soll)*

[Optionen...]

3f Dialogfenster "Mehrfaches Liniendiagramm definieren: Werte einzelner Fälle"

Linien entsprechen: *(die y-Achse definieren)*

[_____] *(die abhängigen Variablen angeben, deren Werte dargestellt werden sollen; sie bilden die verschiedenen Linien; auf der x-Achse sind die einzelnen Fälle abgetragen)*

Kategorienbeschriftungen:

- Fallnummer *(Beschriftung der x-Achse: entweder die Fallnummer oder Wert in einer anzugebenden Variable)*
 Variable [_____]

Vorlage:

Diagrammeinstellungen verwenden aus:

[Datei...] *(Graphikdatei auswählen, die als Muster für die*

[Titel...] *Graphikeinstellungen dienen soll)*

3g Dialogfenster "Verbundliniendiagramm definieren: Auswertung über Kategorien einer Variable"

Punkte entsprechen: *(die y-Achse definieren)*

- Anzahl der Fälle
 Kumul. Anzahl der Fälle
 Andere Auswertungsfunktion:

Variable:
 [_____
 [Auswertungsfunktion...]

(die abhängige Variable auswählen sowie die Funktion, die auf sie angewandt werden soll)

Kategorienachse:
 [_____]

(die unabhängige Variable angeben, deren Werte die x-Achse bilden sollen)

Punkte definieren durch:
 [_____]

(die unabhängige Variable angeben, deren Werte die verschiedenen verbundenen Punkte darstellen sollen)

Vorlage:

- Diagrammeinstellungen verwenden aus:
 [Datei...]

[Titel...]

[Optionen...]

3h Dialogfenster "Verbundliniendiagramm definieren: Auswertung über verschiedene Variablen"

Punkte entsprechen: *(die y-Achse definieren)*

[_____
 [Auswertungsfunktion...]

(die abhängigen Variablen auswählen und jeweils die Funktion, die auf sie angewandt werden soll; die angegebenen Variablen bilden die verschiedenen miteinander verbundenen Punkte)

Kategorienachse:
 [_____]

(die unabhängige Variable angeben, deren Werte die x-Achse bilden sollen)

Vorlage:

- Diagrammeinstellungen verwenden aus:
 [Datei...]

(Graphikdatei auswählen, die als Muster für die Graphikeinstellungen dienen soll)

[Titel...]

[Optionen...]

3i Dialogfenster "Verbundliniendiagramm definieren: Werte einzelner Fälle"

Punkte entsprechen: *(die y-Achse definieren)*

[_____]

(die abhängigen Variablen angeben, deren Werte dargestellt werden sollen; sie bilden die verschiedenen verbundenen Punkte innerhalb eines Falles)

Kategorienbeschriftungen

- Fallnummer
 Variable [_____]
(Beschriftung der x-Achse: entweder die Fallnummer oder Wert in einer anzugebenden Variable)

Vorlage:

- Diagrammeinstellungen verwenden aus:
 [Datei...]

(Graphikdatei auswählen, die als Muster für die Graphikeinstellungen dienen soll)

[Titel...]

4 Dialogfenster "Titel"

Titel

Zeile 1: [_____]

Zeile 2: [_____]

Untertitel: [_____]

Fußnote

Zeile 1: [_____]

Zeile 2: [_____]

5 Dialogfenster "Optionen"

Missing Werte:

- Listenweiser Fallausschluß
 Fälle Variable für Variable ausschließen
 Fehlende Werte als Kategorie anzeigen
 Grafik mit Fallbeschriftungen anzeigen

5.1.3 Flächendiagramme

Diese Prozedur erlaubt die Erstellung von einfachen oder gestapelten Flächendiagrammen. Einfache Flächendiagramme entsprechen eigentlich einem Liniendiagramm, bei dem die Fläche zwischen dem Linienzug und der x-Achse ausgefüllt ist. Ein gestapeltes Flächendiagramm besteht im Prinzip aus mehreren einfachen Flächendiagrammen, die übereinander gestapelt sind, wobei die obere Begrenzungslinie einer Fläche jeweils die Basis für die nächste Fläche darstellt. Es können Gruppendaten, z.B. Mittelwerte für die Kategorien einer Variable oder für verschiedene Variablen, oder auch individuelle Daten graphisch dargestellt werden.

DIE GRAPHIKPROZEDUR "FLÄCHEN"

Aufruf

1 Menü

Grafiken
Fläche...

2 Auswahlbox "Flächendiagramme" (Typ des Diagramms auswählen):

- Einfach
 Gestapelt

Daten im Diagramm:

- Auswertung über Kategorien einer Variablen
 Auswertung über verschiedene Variablen
 Werte einzelner Fälle

3a Dialogfenster "Einfaches Flächendiagramm definieren: Auswertung über Kategorien einer Variablen"

Fläche entspricht: (die y-Achse definieren)

- Anzahl der Fälle
 Kumul. Anzahl der Fälle
 Andere Auswertungsfunktion:

Variable:

[_____]

[Auswertungsfunktion...]

% der Fälle
 Kum. % der Fälle
(die abhängige Variable auswählen sowie die Funktion, die auf sie angewandt werden soll)

Kategorienachse:

[_____]

(die Variable für die x-Achse auswählen)

Vorlage:

- Diagrammeinstellungen verwenden aus:

[Datei...]

(Graphikdatei auswählen, die als Muster für die Graphikeinstellungen dienen soll)

[Titel...]

[Optionen...]

3b Dialogfenster "Einfaches Flächendiagramm definieren: Auswertung über verschiedene Variablen"

Fläche entspricht: (die Achsen definieren)

[_____]

[Auswertungsfunktion...]

(die Variablen auswählen und die Funktion, die jeweils dargestellt werden soll; auf der x-Achse werden die angegebenen Variablen abgetragen, auf der y-Achse deren Funktionswerte)

Vorlage:

- Diagrammeinstellungen verwenden aus:

[Datei...]

(Graphikdatei auswählen, die als Muster für die Graphikeinstellungen dienen soll)

[Titel...]

[Optionen...]

3c Dialogfenster "Einfaches Flächendiagramm definieren: Werte einzelner Fälle"

Fläche entspricht: (die y-Achse definieren)

[_____]

(die Variable auswählen, deren Werte dargestellt werden sollen)

Kategorienbeschriftungen:

- Fallnummer *(Beschriftung der x-Achse: entweder die Fallnummer oder Wert in einer anzugebenden Variable)*
 Variable [_____]

Vorlage:

- Diagrammeinstellungen verwenden aus:
 [Datei...] *(Graphikdatei auswählen, die als Muster für die Graphikeinstellungen dienen soll)*
 [Titel...]

3d Dialogfenster "Gestapeltes Flächendiagramm definieren: Auswertung über Kategorien einer Variable"

Flächen entsprechen: *(die y-Achse definieren)*

- Anzahl der Fälle % der Fälle
 Kumul. Anzahl der Fälle Kum. % der Fälle
 Andere Auswertungsfunktion:
 Variable: *(die abhängige Variable auswählen sowie die Funktion, die auf sie angewandt werden soll)*
 [_____]
 [Auswertungsfunktion...]

Kategorienachse: *(die unabhängige Variable angeben, deren Werte die x-Achse bilden sollen)*
 [_____]

Flächen definieren durch: *(die unabhängige Variable angeben, deren Werte die verschiedenen Flächen bilden sollen)*
 [_____]

Vorlage:

- Diagrammeinstellungen verwenden aus:
 [Datei...] *(Graphikdatei auswählen, die als Muster für die Graphikeinstellungen dienen soll)*
 [Titel...]
 [Optionen...]

3e Dialogfenster "Gestapeltes Flächendiagramm definieren: Auswertung über verschiedene Variablen"

Flächen entsprechen: *(die y-Achse definieren)*

- [_____] *(die abhängigen Variablen auswählen und jeweils die Funktion, die auf sie angewandt werden soll; sie bilden die einzelnen Flächen)*
 [Auswertungsfunktion...]

Kategorienachse: *(die unabhängige Variable angeben, deren Werte die x-Achse bilden sollen)*
 [_____]

Vorlage:

- Diagrammeinstellungen verwenden aus:
 [Datei...] *(Graphikdatei auswählen, die als Muster für die Graphikeinstellungen dienen soll)*
 [Titel...]
 [Optionen...]

3f Dialogfenster "Gestapeltes Flächendiagramm definieren: Werte einzelner Fälle"

Linien entsprechen: *(die y-Achse definieren)*

- [_____] *(die abhängigen Variablen angeben, deren Werte dargestellt werden sollen; sie bilden die verschiedenen Flächen; auf der x-Achse sind die einzelnen Fälle abgetragen)*

Kategorienbeschriftungen:

- Fallnummer *(Beschriftung der x-Achse: entweder die Fallnummer oder Wert in einer anzugebenden Variable)*
 Variable [_____]

Vorlage:

- Diagrammeinstellungen verwenden aus:
 [Datei...] *(Graphikdatei auswählen, die als Muster für die Graphikeinstellungen dienen soll)*
 [Titel...]

4 Dialogfenster "Titel"

Titel

Zeile 1: [_____]

Zeile 2: [_____]

Untertitel: [_____]

Fußnote

Zeile 1: [_____]
 Zeile 2: [_____]

5 Dialogfenster "Optionen"

Missing Werte:

- Listenweiser Fallausschluß
- Fälle Variable für Variable ausschließen
- Fehlende Werte als Kategorie anzeigen
- Grafik mit Fallbeschriftungen anzeigen

5.1.4 Kreisdiagramme

Diese Prozedur erlaubt die Erstellung von Kreisdiagrammen (Tortendiagrammen). Dabei können Gruppendaten für verschiedene Werte einer unabhängigen Variable oder für verschiedene abhängige Variablen (z.B. deren Mittelwerte) oder auch individuelle Daten in einer oder mehreren abhängigen Variablen graphisch dargestellt werden.

DIE GRAPHIKPROZEDUR "KREIS"

Aufruf

- 1 Menü
 - Grafiken
 - Kreis...
- 2 Auswahlbox "Kreisdiagramme" (Typ des Diagramms auswählen):

Daten im Diagramm:

 - Auswertung über Kategorien einer Variablen
 - Auswertung über verschiedene Variablen
 - Werte einzelner Fälle
- 3a Dialogfenster "Kreisdiagramm definieren: Auswertung über Kategorien einer Variablen"

Segmente entsprechen: *(den Inhalt der Segmente definieren)*

 - Anzahl der Fälle % der Fälle
 - Kumul. Anzahl der Fälle Kum. % der Fälle
 - Andere Auswertungsfunktion:

Variable: *(die abhängige Variable auswählen sowie die Funktion, die auf sie angewandt werden soll)*
 [_____]
 [Auswertungsfunktion...]

Segmente definieren durch: *(die Segmentvariable auswählen; ihre Werte in der o.g. Auswertungsfunktion definieren die Größe der Segmente)*
 [_____]

Vorlage:

 - Diagrammeinstellungen verwenden aus:
 - [Datei...] *(Graphikdatei auswählen, die als Muster für die Graphikeinstellungen dienen soll)*
 - [Titel...]
 - [Optionen...]
- 3b Dialogfenster "Kreisdiagramm definieren: Auswertung über verschiedene Variablen"

Segmente entsprechen: *(den Inhalt der Segmente definieren)*

 - [_____] *(die Variablen auswählen und die Funktion, die jeweils dargestellt werden soll)*
 - [Auswertungsfunktion...]

Vorlage:

Diagrammeinstellungen verwenden aus:

[Datei...]

(Graphikdatei auswählen, die als Muster für die

[Titel...]

Graphikeinstellungen dienen soll)

[Optionen...]

3c Dialogfenster "Kreisdiagramm definieren: Werte einzelner Fälle"

Segmente entsprechen: *(den Inhalt der Segmente definieren)*

[_____] *(die Variable auswählen, deren Werte dargestellt werden sollen)*

Segmentbeschriftung:

Fallnummer

(Beschriftung der Segmente: entweder die Fallnummer

Variable [_____]

oder Wert in einer anzugebenden Variable)

Vorlage:

Diagrammeinstellungen verwenden aus:

[Datei...]

(Graphikdatei auswählen, die als Muster für die

[Titel...]

Graphikeinstellungen dienen soll)

4 Dialogfenster "Titel"

Titel

Zeile 1: [_____]

Zeile 2: [_____]

Untertitel: [_____]

Fußnote

Zeile 1: [_____]

Zeile 2: [_____]

5 Dialogfenster "Optionen"

Missing Werte:

Listenweiser Fallausschluß

Fälle Variable für Variable ausschließen

Fehlende Werte als Kategorie anzeigen

Grafik mit Fallbeschriftungen anzeigen

5.1.5 Boxplots

Diese Prozedur erlaubt die Erstellung von einfachen oder gruppierten Boxplots. Die Boxen können repräsentieren: Gruppendaten für verschiedene Werte einer unabhängigen Variablen oder Gruppendaten für verschiedene abhängige Variablen.

BOXPLOTS	
<i>Aufruf</i>	
1	Menü Grafiken Boxplot...
2	Auswahlbox "Boxplots": (<i>Typ des Diagramms auswählen</i>) <input checked="" type="radio"/> Einfach <input type="radio"/> Gruppirt Daten im Diagramm: <input checked="" type="radio"/> Auswertung über Kategorien einer Variablen <input type="radio"/> Auswertung über verschiedene Variablen
3a	Dialogfenster "Einfachen Boxplot definieren: Auswertung über Kategorien einer Variablen" Variable: <i>(die abhängige Variable auswählen)</i> [.....] Kategorienachse: <i>(die Variable für die x-Achse auswählen)</i> [.....] Fallbeschriftung: <i>(die Variable für die Fallbeschriftung auswählen)</i> [.....] [Optionen...]
3b	Dialogfenster "Einfachen Boxplot definieren: Auswertung über verschiedene Variablen" Box entspricht: <i>(die Variablen auswählen, für die die Plots erstellt werden sollen; sie werden auf der x-Achse abgetragen)</i> [.....] Fallbeschriftung: <i>(die Variable für die Fallbeschriftung auswählen)</i> [.....] [Optionen...]
3c	Dialogfenster "Gruppierten Boxplot definieren: Auswertung über Kategorien einer Variable" Variable: <i>(die abhängige Variable auswählen)</i> [.....] Kategorienachse: <i>(die Variable für die x-Achse auswählen)</i> [.....] Gruppen definieren durch: <i>(die unabhängige Variable auswählen, deren Werte die verschiedenen Boxplots bilden sollen)</i> [.....] Fallbeschriftung: <i>(die Variable für die Fallbeschriftung auswählen)</i> [.....] [Optionen...]
3d	Dialogfenster "Gruppierten Boxplot definieren: Auswertung über verschiedene Variablen" Box entspricht: <i>(die Variablen auswählen, für die die Plots erstellt werden sollen; sie werden auf der x-Achse abgetragen)</i> [.....] Kategorienachse: <i>(die Variable für die x-Achse auswählen)</i> [.....] Fallbeschriftung: <i>(die Variable für die Fallbeschriftung auswählen)</i> [.....] [Optionen...]

- 4 Dialogfenster "Optionen"
Missing Werte:
- Listenweiser Fallausschluß
 - Fälle Variable für Variable ausschließen
 - Fehlende Werte als Kategorie anzeigen
 - Grafik mit Fallbeschriftungen anzeigen

5.1.6 Fehlerbalkendiagramme

Diese Prozedur erlaubt die Erstellung von einfachen oder gruppierten Fehlerbalkenplots. Die einzelnen Fehlerbalken können sich auf eine abhängige Variable für verschiedene Teilgruppen beziehen oder auf verschiedene abhängige Variablen.

DIE GRAPHIKPROZEDUR "FEHLERBALKEN"

Aufruf

- 1 Menü
Grafiken
Fehlerbalken...
- 2 Auswahlbox "Fehlerbalken": *(Typ des Diagramms auswählen)*
 - Einfach
 - Gruppirt

Daten im Diagramm:

 - Auswertung über Kategorien einer Variablen
 - Auswertung über verschiedene Variablen
- 3a Dialogfenster "Einfaches Fehlerbalkendiagramm definieren: Auswertung über Kategorien einer Variablen"

Variable: *(die abhängige Variable auswählen)*
[_____]

Kategorienachse: *(die Variable für die x-Achse auswählen)*
[_____]

Bedeutung der Balken:

[Konfidenzintervall für Mittelwert	[<input checked="" type="radio"/>]
[Standardfehler des Mittelwerts]
[Standardabweichung]

Niveau: [_95_] % Multiplikator [_2_] *(die Prozentangabe wird für das KI verwendet, der Multiplikator für die anderen beiden Alternativen)*

Vorlage:

 - Diagrammeinstellungen verwenden aus:
[Datei...] *(Graphikdatei auswählen, die als Muster für die Graphikeinstellungen dienen soll)*
 - [Titel...]
 - [Optionen...]
- 3b Dialogfenster "Einfaches Fehlerbalkendiagramm definieren: Auswertung über verschiedene Variablen"

Fehlerbalken: *(die Variablen auswählen, für die die Plots erstellt sollen; sie werden auf der x-Achse abgetragen)*
[_____]

Bedeutung der Balken:

[Konfidenzintervall für Mittelwert]
 [Standardfehler des Mittelwerts]
 [Standardabweichung]

Niveau: [_95_] % Multiplikator [_2_] *(die Prozentangabe wird für das KI verwendet, der Multiplikator für die anderen beiden Alternativen)*

Vorlage:

Diagrammeinstellungen verwenden aus:
 [Datei...]

(Graphikdatei auswählen, die als Muster für die Graphikeinstellungen dienen soll)

[Titel...]

[Optionen...]

3c Dialogfenster "Gruppiertes Fehlerbalkendiagramm definieren: Auswertung über Kategorien einer Variablen"

Variablen:

(die abhängige Variable auswählen)

[_____]

Kategorienachse:

(die Variable für die x-Achse auswählen)

[_____]

Gruppen definieren durch:

(die unabhängige Variable auswählen, deren Werte die verschiedenen Plots bilden sollen)

[_____]

Bedeutung der Balken:

[Konfidenzintervall für Mittelwert]
 [Standardfehler des Mittelwerts]
 [Standardabweichung]

Niveau: [_95_] % Multiplikator [_2_] *(die Prozentangabe wird für das KI verwendet, der Multiplikator für die anderen beiden Alternativen)*

Vorlage:

Diagrammeinstellungen verwenden aus:
 [Datei...]

(Graphikdatei auswählen, die als Muster für die Graphikeinstellungen dienen soll)

[Titel...]

[Optionen...]

3d Dialogfenster "Gruppiertes Fehlerbalkendiagramm definieren: Auswertung über verschiedene Variablen"

Variablen:

(die abhängigen Variablen auswählen)

[_____]

Kategorienachse:

(die Variable für die x-Achse auswählen)

[_____]

Bedeutung der Balken:

[Konfidenzintervall für Mittelwert]
 [Standardfehler des Mittelwerts]
 [Standardabweichung]

Niveau: [_95_] % Multiplikator [_2_] *(die Prozentangabe wird für das KI verwendet, der Multiplikator für die anderen beiden Alternativen)*

Vorlage:

Diagrammeinstellungen verwenden aus:
 [Datei...]

(Graphikdatei auswählen, die als Muster für die Graphikeinstellungen dienen soll)

[Titel...]

[Optionen...]

4 Dialogfenster "Titel"

Titel

Zeile 1: [_____]

Zeile 2: [_____]

Untertitel: [_____]

Fußnote

Zeile 1: [_____]

Zeile 2: [_____]

- 5 Dialogfenster "Optionen"
- Missing Werte:
- Listenweiser Fallausschluß
 - Fälle Variable für Variable ausschließen
 - Fehlende Werte als Kategorie anzeigen
 - Grafik mit Fallbeschriftungen anzeigen

5.1.7 Histogramme

Diese Prozedur erlaubt die Erstellung von Histogrammen, auf Wunsch auch mit überlagerter Normalverteilungskurve

DIE GRAPHIKPROZEDUR "HISTOGRAMM"

Aufruf

- 1 Menü
 - Grafiken
 - Histogramm...
- 2 Dialogfenster "Histogramm"

Variable: _____ *(die abhängige Variable auswählen)*

Vorlage:

 - Diagrammeinstellungen verwenden aus: _____ *(Graphikdatei auswählen, die als Muster für die Graphikeinstellungen dienen soll)*
 - Normalverteilungskurve anzeigen
 - [Titel...]
- 4 Dialogfenster "Titel"

Titel

 - Zeile 1: _____
 - Zeile 2: _____

Untertitel: _____

Fußnote

 - Zeile 1: _____
 - Zeile 2: _____

5.1.8 Verteilungsplots

Zwei Graphikprozeduren erlauben die Erstellung von Verteilungsplots für eine oder mehrere Variable. Alle gebräuchlichen Testverteilungen stehen zur Verfügung. Die Variablen können vorher auch logarithmiert, standardisiert, differenziert oder saisonal bereinigt werden. Zu jedem Plot wird zudem jeweils noch ein trendbereinigter Plot ausgegeben.

DIE GRAPHIKPROZEDUR "P-P"	
<i>Aufruf</i>	
1	Menü Grafiken P-P...
2	Dialogfenster "P-P-Diagramme"
	Variablen: _____ <i>(die abhängigen Variablen auswählen, für die die Plots erstellt werden sollen)</i>
	Testverteilung:
	[Normal <input checked="" type="radio"/>]
	[Beta]
	[Chi-Quadrat]
	[Exponentiell]
	[Gamma]
	[Halb-Normalverteilung]
	[Laplace]
	[Logistisch]
	[Log-Normal]
	[Pareto]
	[Student-T]
	[Weibull]
	[Gleichverteilung]
	df: [_1_]
	Verteilungsparameter:
	<input checked="" type="checkbox"/> Aus den Daten schätzen
	Lage: [_0_] <i>(die Bezeichnung der Verteilungsparameter ist abhängig von der gewählten Testverteilung!)</i>
	Skala: [_1_] <i>(bei Bedarf die gewünschten Transformation der Variablen angeben)</i>
	Transformieren:
	<input type="checkbox"/> Natürlicher Logarithmus
	<input type="checkbox"/> Werte standardisieren
	<input type="checkbox"/> Differenz: [_1_]
	<input type="checkbox"/> Saisonale Differenz: [_1_]
	Formel für Anteilsschätzungen
	<input checked="" type="radio"/> Blom <input type="radio"/> Rankit <input type="radio"/> Tukey <input type="radio"/> Van der Waerden
	Zugewiesener Rang bei Rangbindungen
	<input checked="" type="radio"/> Mittelwert <input type="radio"/> Maximum <input type="radio"/> Minimum <input type="radio"/> Bindungen willkürlich lösen

DIE GRAPHIKPROZEDUR "Q-Q"

Aufruf

1 Menü

Grafiken

Q-Q

2 Dialogfenster "Q-Q-Diagramme"

Variablen:

[_____]

(die abhängigen Variablen auswählen, für die die Plots erstellt werden sollen)

Testverteilung:

[Normal]

[Beta]

[Chi-Quadrat]

[Exponentiell]

[Gamma]

[Halb-Normalverteilung]

[Laplace]

[Logistisch]

[Log-Normal]

[Pareto]

[Student-T]

[Weibull]

[Gleichverteilung]

df: [_1_]

Verteilungsparameter:

Aus den Daten schätzen

Lage: [_0_]

Skala: [_1_]

(die Bezeichnung der Verteilungsparameter ist abhängig von der gewählten Testverteilung!)

Transformieren:

Natürlicher Logarithmus

Werte standardisieren

Differenz: [_1_]

Saisonale Differenz: [_1_]

(bei Bedarf die gewünschten Transformation der Variablen angeben)

Formel für Anteilsschätzungen

Blom Rankit Tukey Van der Waerden

Zugewiesener Rang bei Rangbindungen

Mittelwert Maximum Minimum Bindungen willkürlich lösen

5.1.9 Streudiagramme

Diese Prozedur dient zur Erstellung von einfachen Streudiagrammen (Scatterplots), von überlagerten Streudiagrammen, von Streudiagramm-Matrizen und von 3D-Streudiagrammen.

Einfache Streudiagramme stellen die gemeinsame Verteilung zweier Variablen graphisch dar, und zwar für die Gesamtdaten oder gemeinsam für Untergruppen von Fällen. Bei überlagerten Streudiagrammen werden die Streudiagramme mehrerer Variablenpaare übereinander gezeichnet. Streudiagramm-Matrizen fassen ähnlich einer Korrelationsmatrix alle paarweisen Streudiagramme mehrerer Variablen in einer quadratischen Matrix zusammen (jede Zelle dieser Matrix besteht aus einem einfachen Streudiagramm). 3D-Streudiagramme schließlich dienen der Veranschaulichung des Zusammenhangs zwischen drei Variablen.

DIE GRAPHIKPROZEDUR "STREUDIAGRAMM"

Aufruf

1 Menü

Grafiken

Streudiagramm...

2 Auswahlbox "Streudiagramm" (Typ des Diagramms auswählen):

- Einfach
- Überlagert
- Matrix
- 3D

3a Dialogfenster "Einfaches Streudiagramm"

Y-Achse:

[_____]

(die Variable für die y-Achse auswählen)

X-Achse:

[_____]

(die Variable für die x-Achse auswählen)

Markierungen festlegen durch:

[_____]

(falls die Streudiagramme für verschiedene Teilgruppen übereinander geplottet werden sollen, hier die Gruppierungsvariable angeben)

Fallbeschriftung:

[_____]

(optional; eine Variable angeben, deren Werte zur Kennzeichnung individueller Datenwerte, z.B. für Ausreißer benutzt werden sollen)

Vorlage:

Diagrammeinstellungen verwenden aus:

[Datei...]

(Graphikdatei auswählen, die als Muster für die Graphikeinstellungen dienen soll)

[Titel...]

[Optionen...]

3b Dialogfenster "Überlagertes Streudiagramm"

Y-X Paare:

[_____]

(hier sind die Variablenpaare anzugeben, deren Streudiagramme zusammen in einem gemeinsamen Koordinatensystem dargestellt werden sollen: die beiden Variablen werden jeweils nacheinander angeklickt und über die Pfeilschaltfläche in die Liste der Y-X Paare übernommen)

Fallbeschriftung:

[_____]

(optional; eine Variable angeben, deren Werte zur Kennzeichnung individueller Datenwerte, z.B. für Ausreißer benutzt werden sollen)

Vorlage:

Diagrammeinstellungen verwenden aus:

[Datei...]

(Graphikdatei auswählen, die als Muster für die Graphikeinstellungen dienen soll)

[Titel...]

[Optionen...]

3c Dialogfenster "Streudiagramm-Matrix"

Matrixvariablen:

[_____]

(die gewünschten Variablen auswählen)

Markierungen festlegen durch:

[_____]

*(falls die Streudiagramme für verschiedene Teilgruppen
übereinander geplottet werden sollen, hier die
Gruppierungsvariable angeben)*

Fallbeschriftung:

[_____]

*(optional; eine Variable angeben, deren Werte zur
Kennzeichnung individueller Datenwerte, z.B. für
Ausreißer benutzt werden sollen)*

Vorlage:

 Diagrammeinstellungen verwenden aus:

[Datei...]

*(Graphikdatei auswählen, die als Muster für die
Graphikeinstellungen dienen soll)*

[Titel...]

[Optionen...]

3d Dialogfenster "3D-Streudiagramm"

X-Achse:

[_____]

(die Variable für die x-Achse auswählen)

Y-Achse:

[_____]

(die Variable für die y-Achse auswählen)

Z-Achse:

[_____]

(die Variable für die z-Achse auswählen)

Markierungen festlegen durch:

[_____]

*(falls die Streudiagramme für verschiedene Teilgruppen
übereinander geplottet werden sollen, hier die
Gruppierungsvariable angeben)*

Fallbeschriftung:

[_____]

*(optional; eine Variable angeben, deren Werte zur
Kennzeichnung individueller Datenwerte, z.B. für
Ausreißer benutzt werden sollen)*

Vorlage:

 Diagrammeinstellungen verwenden aus:

[Datei...]

*(Graphikdatei auswählen, die als Muster für die
Graphikeinstellungen dienen soll)*

[Titel...]

[Optionen...]

4 Dialogfenster "Titel"

Titel

Zeile 1: [_____]

Zeile 2: [_____]

Untertitel: [_____]

Fußnote

Zeile 1: [_____]

Zeile 2: [_____]

5 Dialogfenster "Optionen"

Missing Werte:

 Listenweiser Fallausschluß Fälle Variable für Variable ausschließen Fehlende Werte als Kategorie anzeigen Grafik mit Fallbeschriftungen anzeigen

5.1.10 Die GRAPH-Prozedur

DIE GRAPH-PROZEDUR

Funktion

- Prozedur zur Erstellung von folgenden Typen von Standardgraphiken: Balkendiagramme, Liniendiagramme, Flächendiagramme, Kreisdiagramme. Histogramme und Streuungsdiagramme. Bei den meisten Diagrammtypen sind noch verschiedene Untertypen realisierbar.

Syntax (allgemein)

graph

```
[title='zeile1' ['zeile2']]
[subtitle='zeile']
[footnote='zeile1' ['zeile2']]
unterbefehl für den diagrammtyp
[/template=file]
[/missing={listwise**variable} [{noreport**report}] [{exclude**include}]].
```

o Unterbefehl für Balkendiagramme

```
/bar[{(simple)|(grouped)|(stacked)|(range)}] = ...
```

Balkendiagramme über Kategorien einer Variable:

```
... = [countfby] var [by var]
```

Balkendiagramme über verschiedene Variablen:

```
... = {sumf(varlist) | sumf(var) sumf(var) ...}[by var [by var]]
```

Balkendiagramme über Werte einzelner Fälle:

```
... = value(varlist) [by var]
```

o Unterbefehl für Linien- bzw. Flächendiagramme

```
/line[{(simple)|(multiple)|(area)|(difference)|(drop)}] = ...
```

Liniendiagramme über Kategorien einer Variable:

```
... = [countfby] var [by var]
```

Liniendiagramme über verschiedene Variablen:

```
... = {sumf(varlist) | sumf(var) sumf(var) ...}[by var]
```

Liniendiagramme über Werte einzelner Fälle:

```
... = value(varlist) [by var]
```

o Unterbefehl für Kreisdiagramme

```
/pie = ...
```

Kreisdiagramme über Kategorien einer Variable:

```
... = [countfby] var
```

Kreisdiagramme über verschiedene Variablen:

```
... = {sumf(varlist) | sumf(var) sumf(var) ...}
```

Kreisdiagramme über Werte einzelner Fälle:

```
... = value(var) [by var]
```

o Unterbefehl für Fehlerbalkendiagramme

```
/errorbar [{ (ci [{2n}]) | sterror[{2n}] | stddev[{2n}] }] = { varlist by var | var by var by var }
```

o Unterbefehl für Histogramme

`/histogram[(normal)]=var`

o Unterbefehl für Scatterdiagramme

`/scatterplot[{{(bivariate)|(overlay)|(matrix)|(xyz)}}] = ...`

Zweidimensionale Scatterdiagramme:

`... = var with var [by var] [by var(name)]`

Overlay-Scatterdiagramme:

`... = varlist with varlist[(pair)]`

Scatterplot-Matrix:

`... = varliste [by var] [by var(name)]`

Dreidimensionale Scatterdiagramme:

`... = var with var with var`

Bedeutung der Spezifikationen

- Bei **bar**:

simple	einfaches Balkendiagramm
grouped	gruppiertes Balkendiagramm
stacked	gestapeltes Balkendiagramm
range	Bereichsbalkendiagramm
- Bei **line**:

simple	einfaches Liniendiagramm
multiple	mehrfaches Liniendiagramm
drop	Verbundliniendiagramm
area	Flächendiagramm
difference	Differenzliniendiagramm
- Bei **histogram**:

normal	Histogramm mit überlagerter Normalverteilungskurve
---------------	--
- Bei **scatterplot**:

bivariate	zweidimensionales Scatterdiagramm
overlay	Overlay-Scatterdiagramm, d.h. mehrere zweidimensionale Plots in einem gemeinsamen Koordinatensystem
matrix	Scatterplot-Matrix
xyz	dreidimensionales Scatterdiagramm
- Bei den Unterbefehlen für die einzelnen Diagrammtypen:

<i>countf</i>	Zählfunktion für die Anzahl der Fälle; folgende Funktionen stehen zur Verfügung:
count	absolute Anzahl (Voreinstellung)
pct	Prozent
cufreq	kumulative absolute Anzahl
cupct	kumulative Prozent
<i>sumf</i>	Funktion für Variablenstatistiken; folgende Funktionen stehen zur Verfügung:
n	Anzahl valider Werte
sum	Summe
cusum	kumulierte Summe
mean	arithmetischer Mittelwert
median	Median
gmedian	Gruppenmedian
mode	Modalwert
stddev	Standardabweichung
variance	Varianz
minimum	Minimum
maximum	Maximum
ptile(x)	x-ter Perzentilwert
plt(x)	Prozent der Werte kleiner als x

pgt (x)	Prozent der Werte größer als x
nlt (x)	Anzahl der Werte kleiner x
ngt (x)	Anzahl der Werte größer x
pin ($x1, x2$)	Prozent der Werte zwischen $x1$ und $x2$
nin ($x1, x2$)	Anzahl der Werte zwischen $x1$ und $x2$

- Bei **missing**:
 - listwise** fallweiser Ausschluß fehlender Werte
 - pairwise** variablenweiser Ausschluß fehlender Werte
 - report** fehlende Werte werden berichtet und geplottet
 - noreport** Unterdrückung von Kategorien mit Missing-Werten
 - exclude** Ausschluß von Missing-Werten
 - include** Einbeziehung von Missing-Werten

Hinweise

- Fehlt bei den Diagrammen über Werte einzelner Fälle die BY-Variable, so wird als x-Variable die Fallnummer benutzt.
- Fehlt bei Balkendiagrammen die Angabe des Untertyps, so tritt (in Abhängigkeit von der Anzahl der **by**-Angaben) als Voreinstellung entweder **simple** oder **grouped** in Kraft. Analoges gilt für die Voreinstellungen **simple** oder **multiple** bei Liniendiagrammen.
- Bei den Fehlerbalken werden in Abhängigkeit von der Variablenspezifikation einfache oder gruppierte Diagramme erstellt.
- Bei Differenzliniendiagrammen muß die letzte BY-Variable dichotom sein.
- Zweidimensionale Scatterdiagramme (bzw. Scatterdiagramm-Matrizen) für Untergruppen werden durch Angabe einer Kontrollvariable (Gruppierungsvariable) hinter **by**= angefordert. Steht hinter dem Namen der Kontrollvariable in Klammern das Schlüsselwort **name**, so dient diese Variable lediglich als Labelvariable.
- Bei Overlay-Scatterdiagrammen werden - getrennt durch das Schlüsselwort **with** - zwei Variablenlisten angegeben. Normalerweise werden dann alle Variablenpaare aus den beiden Variablenlisten gebildet und deren Verteilung in einem gemeinsamen Koordinatensystem dargestellt. Durch Angabe von (**pair**) kann man die Variablenpaare aber einschränken: dann wird lediglich die erste Variable mit der ersten, die zweite mit der zweiten usw. kombiniert.
- Mit dem Unterbefehl **/template**= kann man den Namen einer SPSS-Diagrammvorlage (*.sct) angeben, die als Muster für das zu erstellende Diagramm dienen soll. Man erstellt eine Diagrammvorlage mit Hilfe des SPSS-Diagrammeditors, indem man ein fertig bearbeitetes Diagramm als Diagrammvorlage für eine spätere Verwendung abspeichert. Dadurch läßt sich die manuelle Nachbearbeitung von Diagrammen ähnlichen Typs in der Zukunft umgehen.
- Die GRAPH-Prozedur erlaubt auch die Erstellung von Bereichsbalken- (Hoch-Tief-) und von Pareto-Diagrammen; diese werden aber hier nicht weiter behandelt.
- Boxplots werden mit Hilfe der Prozedur EXAMINE erstellt und Normalverteilungsplots mit Hilfe der Prozedur PLOT.

Beispiel

Im folgenden Programm wird ein mehrfaches Liniendiagramm für das mittlere Körpergewicht in Abhängigkeit vom Alter und vom Geschlecht erstellt.

```
get file='a:\data\beispiel.sav'.
graph /line(multiple)=mean(gewicht) by alter by geschl.
```

5.1.11 Die PLOT-Prozedur

DIE PLOT-PROZEDUR

Funktion

- Prozedur zur Erstellung von normalen und trendbereinigten Verteilungsdiagrammen vom Typ Q-Q und P-P.

Syntax

pplot

```
[variables=] varlist
[/distribution={normal[(a,b)]** | exponential(a) | weibull(a,b) | pareto(a,b) | lnormal(a,b) | beta(a,b) |
gamma(a,b) | logistic(a,b) | laplace(a,b) | uniform(a,b) | hnormal(a) | chi(df) | student(df)}]
[/fraction={blom** | rankit | tukey | vw}]
[/ties={mean** | low | high | condense}]
[/nonstandardize** | standardize}]
[/type={q-q** | p-p}]
[/plot={both** | normal | detrended}]
[/diff={1 | n}]
[/sdiff={1 | n}]
[/period=n]
[/nolog** | ln]
[/apply[='modelname']].
```

Bedeutung der Spezifikationen

- **distribution:** Testverteilung

normal	Normalverteilung
exponential	Exponentialverteilung
lnormal	Lognormalverteilung
uniform	Gleichverteilung
hnormal	Halb-Normalverteilung (<i>a</i> ist der Streuungsparameter)
chi	Chiquadratverteilung
student	t-Verteilung
- **fraction:** Formel für Schätzung der Anteilswerte

blom	Blom
rankit	Rankit
tukey	Tukey
vw	van der Waerden
- **ties:** zugewiesener Rang bei Rangbindungen

mean	Mittelwert
low	Maximum
high	Minimum
condense	Bindungen willkürlich lösen
- **standardize:** Werte vorher standardisieren
- **type:** Typ des Verteilungsdiagramms
- **plot:** Art des Diagramms

normal	normales Verteilungsdiagramm
detrended	trendbereinigtes Verteilungsdiagramm
- **diff:** Werte vorher differenzieren

<i>n</i>	Differenzierungsgrad
----------	----------------------
- **sdiff:** Werte vorher saisonal differenzieren

<i>n</i>	Differenzierungsgrad
----------	----------------------
- **period:** Periodenlänge für die Berechnung der saisonalen Differenzen (für den Unterbefehl /sdiff=)
- **apply:** Verwendung von Spezifikationen, die in einem vorherigen PLOT-Befehl definiert wurden

<i>modelname</i>	Name des Modells, auf das man sich beziehen möchte; fehlt er, so werden die Spezifikationen des letzten PLOT-Befehls benutzt
------------------	--

Hinweis

- Sind bei dem Unterbefehl /**distribution**= die Parameter der Testverteilung nicht angegeben, so werden sie aus den Daten geschätzt.

Beispiel

Im folgenden Programm wird für die Variable Körpergewicht ein normales und ein trendbereinigtes QQ-Normalverteilungsdiagramm erstellt.

```
get file='a:\data\beispiel.sav'.  
pplot variables=gewicht.
```

5.2 Die Bearbeitung von Standard-Graphiken

Beim Aufruf einer SPSS-Graphikprozedur hat der Benutzer auf das äußere Erscheinungsbild und die Ausgestaltung des Diagramms wenig Einfluß. Eine Vielzahl an Gestaltungsmöglichkeiten ergibt sich erst durch eine Nachbearbeitung des Diagramms. Dies kann allerdings nur interaktiv mit Hilfe des SPSS-Diagramm-Editors geschehen (Diagramm im Viewer selektieren und Menü "Datei|Objekt: SPSS-Diagramm|Öffnen" wählen oder Doppelklick auf das Diagramm).

Bei der Bearbeitung eines Diagramms geht man objektorientiert vor: die einzelnen Bestandteile des Diagramms (z.B. Achsen, Linien, usw.) stellen Objekte dar, die selektiert und deren Attribute verändert werden können. Das Diagramm kann ausgedruckt oder als Datei im internen oder in einem Standardformat abgespeichert werden.

GRAPHIKEN BEARBEITEN

Menüs

Galerie	<i>Ändern des Diagrammtyps bzw. Erstellen eines neuen Diagramms mit einem anderen Typ bei gleicher Variablenspezifikation</i>
Balken...	<i>ein Balkendiagramm erstellen</i>
Linien...	<i>ein Liniendiagramm erstellen</i>
Flächen...	<i>ein Flächendiagramm erstellen</i>
Gemischt...	<i>eine Mischung aus Balken-, Linien-, bzw. Flächendiagramm, erstellen</i>
Kreis...	<i>ein Kreisdiagramm erstellen</i>
Hoch-Tief...	<i>ein Hoch-Tief-Diagramm erstellen</i>
Scatterplot...	<i>einen Scatterplot erstellen</i>
Histogramm...	<i>ein Histogramm erstellen</i>
Diagramme	<i>Ändern von globalen Attributen des Diagramms</i>
Optionen...	<i>spezielle Optionen, abhängig von der jeweiligen Graphikprozedur</i>
Achse...	<i>Attribute der x-, y- und z-Achse</i>
Balkenabstand...	<i>Angaben über die Größenverhältnisse bei Balken</i>
Titel...	<i>Titel des Diagramms</i>
Fußnote...	<i>Definition von Fußnoten</i>
Legende...	<i>Angabe zur Legende</i>
Anmerkung...	<i>Hinzufügen von Anmerkungen in das Diagramm</i>
Bezugslinie...	<i>Zeichnen von Bezugslinien (Referenzlinien)</i>
Äußerer Rahmen	<i>Zeichnen eines äußeren Rahmens um das gesamte Diagramm</i>
Innerer Rahmen	<i>Zeichnen eines Rahmens um das Innere des Diagramms</i>
Aktualisieren	<i>Aktualisieren des Diagramms</i>
Datenreihen	<i>Selektion von Datenwerten, Zuordnung der unabhängigen Variablen</i>
Anzeigen...	<i>z.B. um bestimmte Daten auszublenden, Definition der Achsen zu ändern</i>
Transponieren	<i>Variablen für Kategorienachse und Gruppierung vertauschen</i>
Grafikattribute	<i>Ändern der Attribute ausgewählter Objekte:</i>
Füllmuster...	<i>Füllmuster von Flächen</i>
Farbe...	<i>Farbe von Objekten</i>
Markierung...	<i>Typ der Markierung bei Liniendiagrammen</i>
Linienart...	<i>Linienmuster</i>
Balkenart...	<i>Art der Balken: normal, schattiert oder 3D</i>
Labels...	<i>Labels in Balken: keine, Standard oder umrahmt</i>
Interpolation...	<i>Art der Interpolation bei Liniendiagrammen</i>
Text...	<i>Textattribute, z.B. Schriftart und -größe</i>
3D-Rotation...	<i>Rotieren von 3D-Diagrammen: 6 Drehrichtungen</i>
Achsen vertauschen	<i>Vertauschung von x- und y-Achse</i>
Kreissegment absetzen	<i>bei Kreisdiagrammen: ausgewähltes Segment herausziehen</i>

Unterbrechung bei Missing Wert

Linienzug bei Missing Werten unterbrechen?

Grafik drehen

Drehen eines 3D-Diagramms

Dialogfenster für Diagramme/Optionen

Dialogfenster "Optionen für Balkendiagramme" (nur für gruppierte oder gestapelte Balken)

Skala auf 100 % setzen

Balkentyp

Gruppirt Gestapelt

Dialogfenster "Optionen für Liniendiagramme" (nur für mehrfache und Verbundliniendiagramme)

Optionen für Linie:

Verbindungslinie zwischen Markierungen (nur bei Verbundliniendiagrammen)

Projektion [Kategorie...] (nicht bei Verbundliniendiagrammen)

Dialogfenster "Optionen für Flächendiagramme" (nur für gestapelte Flächendiagramme)

Skala auf 100 % setzen

Dialogfenster "Optionen für Kreisdiagramm"

Erstes Segment bei [_12_] Uhr

Zusammenfassung aller Segmente kleiner als [_5_]

Beschriftung:

Text Werte Prozente

[Text bearbeiten...] [Format...]

Dialogfenster "Kreisdiagramm: Beschriftungen bearbeiten"

Segmentbeschriftung:

Label [_____] (hier kann das Label geändert werden)

Beschriftung für Restsegment [_Andere_____]

Dialogfenster "Kreisdiagramm: Format für Beschriftungen"

Position [Außen, links-/rechtsbündig

[Außen]

[Innen]

[Beste Anpassung]

[Zahlen innen, Text außen]

Verbindungslinie zu äußeren Beschriftungen

Linie mit Pfeil

Rahmen um:

Äußere Beschriftungen Innere Beschriftungen

Werte:

1000er Trennung Dezimalstellen [_2_]

Dialogfenster "Optionen für Boxplots"

Anzeigen:

Ausreißer Extremwerte

Fallbeschriftungen: [Ein

[Aus]

[Wie vorgegeben]

Herkunft der Labels:

ID-Variable

Fallnummer

Häufigkeiten für Kategorien anzeigen

Dialogfenster "Optionen für Fehlerbalken"

Häufigkeiten für Kategorien anzeigen

Dialogfenster "Optionen für Histogramme"

Anzeigen:

- Normalverteilungskurve
 Statistik in der Legende
 Fallhäufigkeitsgewichtung verwenden *(nur wenn Gewichtung der Fälle eingeschaltet ist)*

Dialogfenster "Optionen für Streudiagramme" (auch für Normalverteilungsplots)

Optionen für Anzeige:

- Untergruppen anzeigen
Fallbeschriftungen: [Ein [0]]
[Aus]
[Wie vorgegeben]

Herkunft der Labels:

- ID-Variable
 Fallnummer

Sonnenblumen

- Sonnenblumen anzeigen [Optionen für Sonnenblumen...]

Kurvenanpassung

- Gesamt Untergruppen [Anpassungsoptionen...]

Linie(n) für Y-Mittelwert

- Gesamt Untergruppen Projektionslinien anzeigen
 Fallhäufigkeitsgewichtung verwenden *(nur wenn Gewichtung der Fälle eingeschaltet ist)*

Dialogfenster "Optionen für Streudiagramm: Sonnenblumen"

Ein Blatt entspricht:

- Anzahl Fälle [_1_] Automatisch

Auflösung:

- Grob
 Fein

Position:

- Zentriert
 Mittelwert

Dialogfenster "Optionen für Streudiagramm: Anpassungslinie"

Anpassungsmethode

- Lineare Regression
 Quadratische Regression
 Kubische Regression
 Lowess
% der Punkte für die Anpassung [_50_]
Anzahl der Iterationen [_3_]

Linien für Regressionsschätzer

- Mittelwert Einzel
Konfidenzintervall [_95_] %

Regressionsoptionen

- Konstante in Gleichung einschließen
 R-Quadrat in Legende anzeigen

Dialogfenster "3D-Streudiagramm: Optionen" Untergruppen anzeigen

- Fallbeschriftungen: [Ein [0]]
[Aus]
[Wie vorgegeben]

Herkunft der Labels:

- ID-Variable
 Fallnummer

- Fallhäufigkeitsgewichtung verwenden *(nur wenn Gewichtung der Fälle eingeschaltet ist)*

- Projektionslinien: [Keine [0]]
[Parallel]

[Zentroid]
 [Ursprung]

Rahmen:

Vollständiger Rahmen (12 Ränder) Halber Rahmen (9 Ränder) Kein Rahmen

Sonstige Dialogfenster

Dialogfenster "Skalenachse"

Achsenlinie anzeigen

Achsentitel: [Absolute Werte _____]

Ausrichtung des Titels [Links/unten
 [Rechts/oben]
 [Mitte]

Skala:

Linear

Log

Bereich Minimum Maximum
 Angezeigt: [_____] [_____]

1. Unterteilung:

Inkrement: [_____]

Teilstriche Gitter

2. Unterteilung:

Inkrement: [_____]

Teilstriche Gitter

Basislinie für Balken bei [_0_] *(nur bei Balkendiagrammen)*

Abgeleitete Achse anzeigen

[Abgeleitete Achse...]

Labels anzeigen

[Beschriftungen...]

Dialogfenster "Skalenachse: Beschriftungen"

Dezimalstellen [_0_]

Führendes Zeichen [____]

Abschlußzeichen [____]

1000er-Trennzeichen

Skalierungsfaktor [_1_]

Orientierung: [Horizontal
 [Vertikal]
 [Diagonal]
 [Versetzt]
 [Automatisch]

Dialogfenster "Skalenachse: Abgeleitete Achse "

Definition: Skalenachse: Abgeleitete Achse:

Verhältnis: [_1_] Einheit(en) gleich [_1_] Einheit(en)

Übereinstimmung: [_0_] Wert gleich [_0_] Wert

Titel

Text: [_____]

Ausrichtung: [oben
 [unten]
 [Mitte]

Inkmente

1. Unterteilung: [__10__] Teilstriche

2. Unterteilung: [__10__] Teilstriche

Achsenlinie anzeigen

Beschriftungen:

Anzeigen

Dezimalstellen [_0_]

Führendes Zeichen [____]

Abschlußzeichen [____]

1000er-Trennzeichen
Skalierungsfaktor [_1_]

Dialogfenster "Kategorienachse"

Achsenlinie anzeigen
Achsentitel: [_____]
Ausrichtung des Titels [Links/unten []
[Rechts/oben]
[Mitte]

Achsenmarkierungen
 Teilstriche Gitterlinien
 Labels anzeigen
[Beschriftungen...]

Dialogfenster "Kategorienachse: Beschriftungen"

Anzeigen:
 Alle Beschriftungen
 Jede [_3_] Beschriftung
 Teilstriche für ausgelassene Beschriftungen

Beschriftungstext *(defaultmäßig werden die Wertelabels bzw. die Datenwerte verwendet; sie können hier aber geändert werden)*
Label: [_____]

Orientierung: [Automatisch []
[Horizontal]
[Vertikal]
[Diagonal]
[Versetzt]

Dialogfenster "Balkenabstand"

Balkenrand: [_5_] % der Rahmenbreite *(Raum ganz links und ganz rechts, zusammen)*
Balkenabstand [_0_] % der Balkenbreite *(Raum zwischen den Balken)*
Gruppenabstand [_0_] % der Gruppenbreite *(bei gruppierten Balken)*

Dialogfenster "Titel"

Titel 1: [_____]
Titel 2: [_____]
Ausrichtung des Titels: [Linksbündig []
[Rechtsbündig]
[Mitte]

Untertitel: [_____]
Ausrichtung des Untertitels: [Linksbündig []
[Rechtsbündig]
[Mitte]

Dialogfenster "Fußnote(n)"

Fußnote 1: [_____]
Fußnote 2: [_____]
Position der Fußnote: [Linksbündig []
[Rechtsbündig]
[Mitte]

Dialogfenster "Legende"

Legende anzeigen
Legendentitel: [_____]
Ausrichtung: [Linksbündig []
[Rechtsbündig]
[Mitte]

Beschriftungen:
[_____]
Ausgewählte Beschriftung *(die ausgewählte Beschriftung werden hier*

Zeile 1: [_____]
 Zeile 2: [_____]

*angezeigt; sie kann hier bearbeitet und über
 [Ändern] übernommen werden)*

Dialogfenster "Anmerkung"

Anmerkung(en):

[_____]

Anmerkung:

Text: [_____]

*(Anmerkungen können über entsprechende
 Befehlsschaltflächen hinzugefügt, geändert
 oder entfernt werden)*

Textrahmen anzeigen

Ausrichtung: [Linksbündig]
 [Rechtsbündig]
 [Mitte]

Position Skalenachse: [_____] (y-Achse)

Pos. Kategorienachse: [_____] (x-Achse)

Dialogfenster "Bezugslinien für Skalenachse" und "Bezugslinien für Kategorienachse"

Position der Linie(n):: [_____]

*(für die ausgewählte Achse können Bezugslinien
 hinzugefügt, geändert oder entfernt werden)*

Ausblenden

5.3 Interaktive Graphiken

Neben den Standardgraphiken werden seit der Version 8 von SPSS die sog. interaktiven Graphiken angeboten. Das Wort "interaktiv" bedeutet hier nicht etwa, daß diese Graphiken nur im interaktiven Modus benutzbar sind, sondern daß die Erzeugung von Diagrammen damit hochgradig interaktiv und die Bedienung intuitiv ist¹². So wird beispielsweise bei der Auswahl der Variablen von der Technik des "Drap and Drop" Gebrauch gemacht und wie bei der Bearbeitung der Standardgraphiken ist die Arbeitsweise objektorientiert.

Die interaktiven Graphiken umfassen in wesentlichen alle Diagrammtypen, die man von den Standardgraphiken her kennt. Dabei hat der Benutzer aber von vornherein die Kontrolle über alle Attribute für die einzelnen Teilobjekte einer Graphik, so daß eine Nachbearbeitung weitgehend entfallen kann.

5.3.1 Die IGRAPH-Prozedur

DIE IGRAPH-PROZEDUR

Funktion

- Prozedur zur Erstellung von interaktiven Graphiken

Syntax

```
igraph
  /y=var
    [type={scale [min=value] [max=value] | categorical}]
    [title='string']
  /x1=var
    [type={scale [min=value] [max=value] | categorical}]
    [title='string']
  /x2=[var]
    [type={scale [min=value] [max=value] | categorical}]
    [title='string']
  /ylength=value
  /x1length=value
  /x2length=value
  /catorder=var
    ({count | occurrence | label | value}
    [{ascending | descending}]
    [{showempty | omitempty}))
  /color=var
    [type={scale [min=value] [max=value] | categorical}]
    [legend={on | off}]
    [title='string']
    [{cluster | stack}]
  /style=var
    [legend={on | off}]
    [title='string']
    [{cluster | stack}]
  /size=var
    [type={scale [min=value] [max=value] | categorical}]
    [legend={on | off}]
    [title='string']
  /refline=var value
    [label={on | off}]
    [spike={on | off}]
  /normalize
  /cluster=var
  /summaryvar=var
```

¹² Aus diesem Grunde wird im folgenden lediglich der IGRAPH-Befehl erläutert und nicht die interaktive Benutzung dieser Graphiken.

```

/panel=varlist
/pointlabel=var [{none | all}]
/coordinate={horizontal | vertical | three}
/effect={ none | three }
/title='string'
/subtitle='string'
/caption='string'
/viewname='string'
/chartlook='filename'
/format=[spike [color={on | off}] [style={on | off}] ]
/spike={x1 | x2 | y | corner | origin | floor | centroid [total] [meffect]}

```

Unterbefehl für den Diagrammtyp:

```

/bar [(summaryFunction)]
  [label {inside|outside} [val] [n]]
  [shape={rectangle|pyramid|obelisk}]
  [barbase={square|round}]
  [baseline={auto | value}]
/line [(summaryFunction)]
  style={dotline|line|dot|none}
  [dropline={on|off}]
  [label=[val] [n] [pct]]
  [linelabel=[cat] [n] [pct]]
  [interpolate={straight|step|cstep|rstep|jump|cjump|rjump|spline|lagrange3|lagrange5}]
  [break={missing|none}]
/area [(summaryFunction)]
  [key={on|off}]
  [pointlabel=[val] [n] [pct]]
  [arealabel=[n] [pct]]
  [interpolate={straight|step|cstep|rstep}]
  [baseline={auto|value}]
  [style={missing|none}]
/pie [(summaryFunction)]
  [start=value]
  [{cw|ccw}]
  [slice={inside|outside|textin|numin} [label] [pct] [val] [n]]
  [cluster={uright|right|uleft|left} [label] [pct] [val] [n]]
/box
  [outliers={on|off}]
  [extreme={on|off}]
  [median={on|off}]
  [label=[n]]
  [boxbase={square|round}]
  [whisker={t|fancy|line}]
  [capwidth(value)]
/errorbar=
  [{ci(value)|sd(value)|se(value)}]
  [label=[val] [n]]
  [direction={both|up|down|sign}]
  [capwidth(value)]
  [capstyle={none|t|fancy}]
  [symbol={on|off}]
  [baseline=value]
/histogram=
  [cum]
  [shape=(histogram)]
  [x1interval={auto|num=n|width=value}]
  [x2interval={auto|num=n|width=value}]
  [x1start=value]
  [x2start=value]
  [curve={on|off}]

```

```

[surface={on|off}]
/scatter
[coincident={none|jitter[(value)]}]
/fitline=
[method={
  none|
  regression linear|
  origin linear|
  mean|
  llr [(normal|epanechnikov|uniform)]
  [bandwidth={fast|constrained}]
  [x1multiplier=value] [x2multiplier=value]
}]
[interval[(value)]=[mean] [individual]]
[line=[total]|meffect]

```

Bedeutung der Spezifikationen

- **y, x1, x2**: Variablen für die einzelnen Koordinatenachsen (die Achse x1 verläuft von links nach rechts, die Achse x2 von vorn nach hinten, die Achse y von unten nach oben); als Variablen können neben benutzerdefinierten Variablen auch folgende 'eingebaute' Variablen verwendet werden: **\$count** (Anzahl), **\$pct** (Prozent), **\$case** (Fallnummer)

type	Variablentyp: metrisch (evtl. eingeschränkt auf einen Wertebereich) oder kategorial
title	Achsenbeschriftung
- **ylength, x1length, x2length**: Achsenlängen in Inch
- **catorder**: Reihenfolge für die Anordnung der Kategorien auf der jeweiligen Achse

count	Kategorienhäufigkeit
occurrence	Reihenfolge des ersten Auftretens der Kategorien in der Arbeitsdatei
label	Value Labels der Kategorien (falls keine definiert sind: Kategorienwert)
value	Kategorienwert
showempty	leere Kategorien ins Diagramm aufnehmen
omitempty	leere Kategorien weglassen
- **color, style, size**: Legendenvariablen, deren Werte Farbe, Stil bzw. Größe der im Diagramm zu verwendenden Symbole definieren; auf diese Weise kann man in einem Diagramm die Daten für verschiedene Untergruppen darstellen (z.B. erzeugt man damit gruppierte oder gestapelte Balkendiagramme, mehrfache Liniendiagramme, überlagerte Streudiagramme od. ähnl.)

<i>var</i>	Variable, die als Legendenvariable für das betreffende Attribut verwendet werden soll
type	Typ der Legendenvariable (metrisch - evtl. eingeschränkt auf einen Wertebereich - oder kategorial)
title	Legendentitel
cluster/stack	Erzeugung von gruppierten/gestapelten Diagrammen
- **refline**: Ausgabe einer Referenzlinie

<i>var</i>	Variable, für die eine Referenzlinie erstellt werden soll
<i>value</i>	Variablenwert
label	Label für die Referenzlinie
spike	Projektionslinien von den Datenpunkten auf die Referenzlinie
- **normalize**: 100%-Skalierung bei gestapelten Balkendiagrammen
- **cluster**: Erstellung eines gruppierten Kreisdiagramms

<i>var</i>	Variable, die dafür verwendet werden soll (ihr Typ muß kategorial sein)
------------	---
- **summaryvar**: abhängige Variable, die für die Erstellung eines Kreisdiagramms verwendet werden soll

<i>var</i>	eine der internen Variablen \$count oder \$pct oder eine benutzerdefinierte Variable; auf diese wird die im /pie -Unterbefehl angegebene Summaryfunktion angewandt (dort sind erlaubt: sum , sumav , sumsq , nlt(x) , nle(x) , neq(x) , ngt(x) und nge(x))
------------	--
- **panel**: Erstellung eines Paneldiagramms; dabei werden für alle Kategorie der Panelvariable (bzw. bei mehreren Panelvariablen für alle Kategorienkombination), d.h. für alle Untergruppen, einzelne Diagramme erstellt; all diese Einzeldiagramm werden dann in Form einer Matrix angeordnet

<i>varlist</i>	Variablen, die als Panelvariablen verwendet werden sollen
----------------	---
- **pointlabel**: Fallbeschriftung für Boxplots und Scatterplots

<i>var</i>	Variable, die dafür verwendet werden soll
none	anfangs keine Labels ausgeben (Labels werden bei Bedarf später interaktiv hinzugefügt)
all	anfangs alle Labels ausgeben
● coordinate:	Orientierung des Koordinatensystems
horizontal	Orientierung der y-Achse waagrecht
vertical	Orientierung der y-Achse senkrecht
three	3D-Diagramm (die Orientierung kann nicht beeinflusst werden)
● effect:	3D-Effekt für 2D-Diagramme
three	Diagramm mit 3D-Effekt
none	kein 3D-Effekt
● title, subtitle, caption:	Zeilen, die als Überschrift oder unterhalb des Diagramms ausgegeben werden sollen
<i>string</i>	beliebiger Text (max. 255 Zeichen lang); für Zeilenvorschub kann man \n verwenden
● viewname:	Angabe eines Diagrammnamens für den Ausgabenavigator
<i>string</i>	beliebiger Text (max. 255 Zeichen lang)
● chartlook:	Angabe einer Datei, die als Vorlage für die kosmetische Ausgestaltung des Diagramms verwendet werden soll; damit lassen sich auch Attribute beeinflussen, die über den IGRAPH-Befehl nicht erreichbar sind
<i>filename</i>	Name der Datei (CHARTLOOK-Dateien haben die Extension *.CLO)
● format:	Attribute der Projektionslinien (bei Diagrammen mit COLOR bzw. STYLE-Variablen)
spike	für Projektionslinien die Attribute der COLOR/STYLE-Variablen verwenden
color	Farbattribute ein- bzw. ausschalten
style	Musterattribute ein- bzw. ausschalten
● spike:	Ausgabe von Projektionslinien der individuellen Datenpunkte auf ...
x1	die x1-Achse
x2	die x2-Achse
y	die y-Achse
corner	den Punkt der Koordinatenminima
origin	den Koordinatenursprung
floor	die Basisachse bzw. -ebene
centroid	total: das Zentroid der Gesamtgruppe meffect: das Zentroid der Untergruppe (siehe Unterbefehle /color, ...)
● Bei den Unterbefehlen für den Diagrammtyp:	
<i>SummaryFunction</i>	Funktion, die auf die abhängige Variable angewandt werden soll; folgende Funktionen stehen zur Verfügung:
first	Variablenwert beim ersten in der jeweiligen Kategorie auftretenden Fall
last	Variablenwert beim letzten in der jeweiligen Kategorie auftretenden Fall
minimum	Minimum
maximum	Maximum
mean	arithmetischer Mittelwert
median	Median
mode	Modalwert
stddev	Standardabweichung
variance	Varianz
skew(x)	Schiefe
kurtosis(x)	Exzeß (Kurtosis)
sesmean(x)	Standardfehler des arithmetischen Mittelwerts
seskew(x)	Standardfehler der Schiefe
sekurt(x)	Standardfehler des Exzesses (Kurtosis)
nlt(x)	Anzahl der Werte kleiner <i>x</i>
nle(x)	Anzahl der Werte kleiner oder gleich <i>x</i>
neq(x)	Anzahl der Werte gleich <i>x</i>
ngt(x)	Anzahl der Werte größer <i>x</i>
nge(x)	Anzahl der Werte größer oder gleich <i>x</i>
nin(x1,x2)	Anzahl der Werte zwischen <i>x1</i> und <i>x2</i>
plt(x)	Prozent der Werte kleiner <i>x</i>
ple(x)	Prozent der Werte kleiner oder gleich <i>x</i>
pgt(x)	Prozent der Werte größer <i>x</i>
pge(x)	Prozent der Werte größer oder gleich <i>x</i>
pin(x1,x2)	Prozent der Werte zwischen <i>x1</i> und <i>x2</i>

	ptile(x)	x-ter Perzentilwert
	sum	Summe
	sumav	Summe der absoluten Werte
	sumsq	Quadratsumme
key		Erläuterungen zum Diagramm anzeigen (on) bzw. nicht anzeigen (off)
● /bar:	Erstellung eines Balkendiagramms	
label		Balkenbeschriftung mit: y-Wert (val) und/oder die Anzahl der Werte (n); Anordnung der Beschriftung: innerhalb (inside) bzw. außerhalb (outside) der Balken
shape		Balkenform: Rechteck (rectangle), Pyramide (pyramid) oder Rechteck mit Dach (obelisk)
barbase		Balkengrundfläche bei 3D-Effekt: quadratisch (square) bzw. rund (round)
baseline		Balkengrundlinie: automatisch (auto) bzw. benutzerdefiniert beim angegebenen Wert
● /line:	Erstellung eines Liniendiagramms	
style		Muster: Linienzug (line), Linienzug mit Punkten (dotline), nur Punkte (dot) oder nichts (none)
dropline		Verbundlinien (bei mehrfachen Liniendiagrammen)
label		Punktebeschriftung mit: y-Wert (val), Anzahl der Werte (n) und/oder Prozent der Werte (pct oder percent)
linelabel		Linienbeschriftung mit: Anzahl der Werte (n), Prozent der Werte (pct) und /oder Kategorienwert (cat); eine Linienbeschriftung ist nur bei mehrfachen Liniendiagrammen sinnvoll
interpolate		Art der Interpolation zwischen den Punkten: gerade (straight), Stufe links (lstep), Stufe in der Mitte (cstep), Stufe rechts (rstep), Sprung links (ljump), Sprung in der Mitte (cjump), Sprung rechts (rjump), Spline (spline), Lagrange 3. Ordnung (lagrange3), Lagrange 5. Ordnung (lagrange5)
break		Unterbrechung des Linienzugs bei fehlenden Kategorien
● /area:	Erstellung eines Flächendiagramms	
key		= {on off}
pointlabel		Punktebeschriftung mit: y-Wert (val), Anzahl der Werte (n) und/oder Prozent der Werte (pct)
arealabel		Flächenbeschriftung mit: Anzahl der Werte (n) und/oder Prozent der Werte (pct); eine Flächenbeschriftung ist nur bei gestapelten Flächendiagrammen sinnvoll
interpolate		Interpolationsmethode zwischen den Punkten: gerade (straight), Stufe links (lstep), Stufe in der Mitte (cstep), Stufe rechts (rstep)
baseline		Flächengrundlinie: automatisch (auto) bzw. benutzerdefiniert mit angegebenem Wert (der Defaultwert ist 0)
break		Unterbrechung des Linienzugs bei fehlenden Kategorien
● /pie:	Erstellung eines Kreisdiagramms	
start		Winkel in Grad, bei dem das kleinste Segment endet (0 = 3 Uhr, 90 = 12 Uhr, usw.)
cw/ccw		Anordnung der Segmente im bzw. gegen den Uhrzeigersinn
slice		Segmentbeschriftung mit: Kategorie (label), Wert (val), Prozent der Werte (pct), Anzahl der Werte (n); Anordnung der Beschriftung: innerhalb (inside), außerhalb (outside), Text innerhalb und Zahlen außerhalb (textin), Zahlen innerhalb und Text außerhalb (numin)
cluster		Gruppenbeschriftung bei einem gruppierten Kreisdiagramm mit: Kategorie (label), Wert (val), Prozent der Werte (pct), Anzahl der Werte (n); Anordnung der Beschriftung: oben rechts (uright), unten rechts (lright), oben links (uleft), unten links (lleft)
● /box:	Erstellung eines Boxplots	
outliers		Ausgabe von Ausreißern (d.h. Werten zwischen 1.5 und 3 Interquartilabständen)
extreme		Ausgabe von Extremwerten (d.h. Werten ab 3 Interquartilabständen)
median		Ausgabe der Medianlinie
label		Boxbeschriftung: Häufigkeiten (n)
boxbase		Grundfläche bei 3D-Effekt: quadratisch (square) bzw. rund (round)
whisker		Aussehen der Whisker-Enden: gerade - ohne Box-Enden (line), T-Form (t) oder Box-Enden mit Serifen (fancy)
capwidth		Länge der Whisker-Enden in Prozent der Boxbreiten (Voreinstellung: 45)
● /errorbar:	Erstellung eines Fehlerbalkendiagramms	
ci		Konfidenzintervall mit angegebenem Niveau (in Prozent)

sd	Mittelwert +/- <i>value</i> *Standardabweichung
se	Mittelwert +/- <i>value</i> *Standardfehler des Mittelwerts
label	Balkenbeschriftung mit dem Mittelwert (val) und/oder mit der Anzahl (n)
direction	Richtung der Fehlerbalken: nach oben und unten (both), nur nach oben (up), nur nach unten (down), nach oben oder unten, je nachdem der Mittelwert oberhalb oder unterhalb des Baselinewerts liegt (sign)
capstyle	Form der Balkenenden: gerade - ohne Querbalken (none), T-Form (t), T-Form mit Serifen (fancy)
capwidth	Länge der Balkenenden in Prozent der Balkenabstände (Voreinstellung: 45)
symbol	Ausgabe einer Markierung für den Mittelwert
baseline	Lage der Baseline (der Defaultwert ist 0)

- **/histogram:** Erstellung eines Histogramms
 - cum** kumulatives Histogramm
 - shape** Form des Diagramms (zur Zeit nur **histogram** möglich)
 - x1interval/x2interval** Intervallbreite auf x1- bzw. x2-Achse: automatisch (**auto**), angegebene Anzahl von Intervallen (**num**), angegebene Intervallbreite (**width**)
 - x1start/x2start** Abstand vor dem ersten Intervalls in Prozent
 - curve** Histogramm mit Normalverteilungskurve
- **/scatter:** Erstellung eines zwei- oder dreidimensionalen Streudiagramms
 - coincident** Anordnung der Marker für Punkte mit identischen Werten: überlagert (**none**) oder es wird ein kleiner zufälliger Betrag (Werte zwischen 0 und 10) hinzuaddiert, so daß die mehrfachen Punkte sichtbar sind (**jitter**)
- **/fitline:** Anpassung einer Regressionslinie (oder -fläche) für ein Streudiagramm
 - method** Anpassungsmethode:
 - none:** keine Linie
 - regression linear:** lineare Regression
 - origin linear:** lineare Regression ohne additive Konstante
 - mean:** Gerade bzw. Ebene durch den Mittelwert von y
 - llr:** lokale lineare Regression
 - Kern: Normalverteilung (**normal**), Epanechnikov (**epanechnikov**), Gleichverteilung (**uniform**)
 - Bandbreite (**bandwidth**): ohne Restriktionen (**fast**) oder konstante Bandbreite verwenden (**constrained**)
 - interval[*value*]** Ausgabe eines Vertrauensbereichs für die Mittelwerte (**mean**) bzw. die Einzelwerte (**individual**) mit angegebener Konfidenz (Voreinstellung: 95)
 - line** Anpassungslinien für die Gesamtgruppe (**total**) und/oder für Untergruppen (**meffect**)

Hinweise

- Sämtliche Unterbefehle sind optional (in der Syntaxdarstellung wurden alle entsprechenden Optionsklammern weggelassen!). Es sollte zumindest eine abhängige Variable (und meist auch eine unabhängige Variable) sowie ein Diagrammtyp spezifiziert werden, sonst wird ein leeres Diagramm ausgegeben.
- Bei den einzelnen Unterbefehlen kann man anstatt "*var*" auch die ausführlichere Schreibweise "**var**(*var*)" verwenden.
- Die Segmente bei Kreisdiagrammen werden durch die **/color-** bzw. **/style-**Variable definiert. Geplottete Kreisdiagramme kann man mit Hilfe des **/x1-**, **/x2-** und/oder **/y-**Unterbefehls erzeugen: dann werden die Kreisdiagramme für die entsprechenden Untergruppen in einem ein-, zwei- bzw. dreidimensionalen Koordinatensystem angeordnet.

Beispiel

Im folgenden Programm wird die Häufigkeitsverteilung der Variable Alter in einem Balkendiagramm dargestellt.

```
get file='a:\data\beispiel.sav'.
igraph /y=count$ /x1=alter /bar.
```

6 Literaturhinweise

- Brosius, F. (1998): *SPSS 8.0. Professionelle Statistik unter Windows*. Bonn: MITP-Verlag.
- Brosius, G. & Brosius, F. (1995): *SPSS. Base System und Professional Statistics*. Bonn/ Albany: Thompson Publishing.
- Bühl, A., & Zöfel, P. (2000). *SPSS Version 10. Einführung in die moderne Datenanalyse unter Windows*. München: Addison-Wesley.
- Diehl, J.M., & Staufenbiel, T. (2001). *Statistik mit SPSS Version 10.0*. Eschborn: Verlag Dietmar Klotz..
- Kähler, W.-M. (1994): *SPSS für Windows. Datenanalyse unter Windows*. Braunschweig/Wiesbaden: Vieweg.
- Kinnear, P.R., & Gray, C.D. (2000). *SPSS for Windows Made Simple*. Hove East Sussex: Psychology Press.
- Norusis, M.J. (2000). *SPSS 10 Guide to Data Analysis*. Upper Saddle River, NJ: Prentice Hall.
- Schubö, W., Uehlinger, H.-M., Perleth, Ch., Schröger, E., & Sierwald, W. (1991). *SPSS. Handbuch der Programmversionen 4.0 und SPSS-X 3.0*. Stuttgart: Gustav Fischer Verlag.
- SPSS Inc. (1999). *SPSS Base 10.0 User's Guide*. Chicago: SPSS Inc.
- SPSS Inc. (1999). *SPSS 10.0 Syntax Reference Guide*. Chicago: SPSS Inc.
- Voss, W. (2000) *Praktische Statistik mit SPSS*. München: Hanser Verlag.

Index

\$CASENUM.....	36	RECODE.....	61
\$DATE.....	36	SAMPLE.....	73
\$JDATE.....	36	SAVE.....	41
\$LENGTH.....	36	SAVE TRANSLATE.....	42
\$SYSMIS.....	36	SELECT IF.....	73
\$WIDTH.....	36	SORT CASES.....	69
abs().....	17	SPLIT FILE.....	70
Aggregieren.....	95	SPSS-Programm abbrechen.....	25
any().....	18	SPSS-Programmdatei einfügen.....	26
Arbeitsdatei.....	9	STRING.....	37
Arbeitsvariablen.....	36	SUBTITLE.....	57
arithmetische Ausdrücke.....	15	TEMPORARY.....	67
arithmetische Funktionen.....	17	TITLE.....	57
arsin().....	17	USE.....	75
artan().....	17	VALUE LABELS.....	29
Attribute von Variablen.....	27	VARIABLE ALIGNMENT.....	31
Aufteilen der Fälle.....	70	VARIABLE LABELS.....	29
Ausgabefenster.....	10	VARIABLE LEVEL.....	32
Ausgabeformat.....	31	VARIABLE WIDTH.....	31
Auswählen von Fällen.....	71	VECTOR.....	38
Automatisch umkodieren.....	62	WEIGHT.....	68
Bedingte Blöcke.....	78	WRITE.....	54
Bedingte Transformationen.....	66	WRITE FORMATS.....	32
Befehle		Befehlsprotokoll.....	19
ADD FILES.....	50	Befehlssyntax.....	21
AUTORECODE.....	62	Berechnen.....	63
BEGIN DATA.....	45	Bivariate Korrelationen.....	100
BREAK.....	82	cdf.normal.....	17
COMMENT.....	24	cdfnorm.....	17
COMPUTE.....	63	cfvar().....	17
COUNT.....	65	concat().....	18
CREATE.....	48	cos().....	17
DATA LIST.....	44	ctime.days().....	18
DO IF.....	78	date.dmy().....	18
DO REPEAT.....	79	Dateien zusammenfügen.....	49
ERASE.....	53	Datenansicht.....	10
EXECUTE.....	77	Datendatei.....	9
FILTER.....	74	Dateneditor.....	12, 39
FINISH.....	25	Datenfenster.....	10
FLIP.....	76	definierende Variablenlisten.....	14
FORMATS.....	32	Deskriptive Statistiken.....	87
GET.....	41	Diagrammeditor.....	12
GET TRANSLATE.....	42	Diagrammfenster.....	10
IF.....	66	Dialogfelder.....	11
INCLUDE.....	26	Einfaktorielle Varianzanalyse.....	120
INPUT PROGRAM.....	46	Einfügemechanismus.....	19
LIST.....	57	Erzeugung von Zufallszahlen.....	17
LOOP.....	80	exp().....	17
MATCH FILES.....	51	Explorative Datenanalyse.....	89
MISSING VALUES.....	30	Fälle hinzufügen.....	49
N OF CASES.....	74	Fälle zusammenfassen.....	92
NEW FILE.....	52	fehlende Werte.....	30
NUMERIC.....	37	Fenster.....	10
PRINT.....	55	Funktionen.....	17
PRINT EJECT.....	55	Funktionen für Datum und Uhrzeit.....	18
PRINT FORMATS.....	32	Funktionen für fehlende Werte.....	17
PRINT SPACE.....	56	Gewichten von Fällen.....	68

Häufigkeiten	84	IGRAPH	158
Hinzufügen		MEANS	119
Fälle	49	NONPAR CORR	102
Variablen	49	NPART TESTS	112
idf.normal()	17	ONEWAY	121
index()	18	PARTIAL CORR	103
Inline-Daten	9, 19, 45	PLOT	150
Interaktiver Modus	13	REGRESSION	107
Journaldatei	19	SUMMARIZE	93
Kommentare	24	TTEST	116
Kontrollstrukturen	20, 77	UNIANOVA	124
Kopfzeilen	57	range()	18
Korrelation	100	reflexive Variablenlisten	14
Kreuztabellen	97	Regression	105
kurze Strings	28	round()	18
lag()	18	rnd()	17
lange Strings	28	rpad()	18
length()	18	rtrim()	18
ln()	17	rv.normal()	17
log10()	17	rv.uniform()	17
logische Ausdrücke	16	Schleifen	79
logische Funktionen	18	sd()	17
logische Werte	15	Sequenz	20
Log-Tabelle	19	sin()	17
Löschen der Arbeitsdatei	52	Skripteditor	12
lower()	18	Skriptfenster	11
lpad()	18	sonstige Funktionen	18
ltrim()	18	Sortieren von Fällen	69
max()	17, 18	Spaltenformat	31
mean()	17	SPSS-Datenformate	33
Mehrfaktorielle Varianzanalyse	122	SPSS-Programm (Aufbau und Arbeitsweise)	20
Menüs	11	sqrt()	17
Meßniveau	32	statistische Funktionen	17
Metasprache	21	Statusleiste	11
min()	17, 18	string()	18
missing()	15	Stringfunktionen	18
missing()	17	substr()	18
Missing-Werte	30	sum()	17
Mittelwerte	118	Symbolleisten	11
mod()	17	Syntax	21
Nichtparametrische Tests	110	Metasprache	21
normal()	17	SPSS-Befehle	21
number()	18	Syntaxeditor	12
numerierte Variablen	14	Syntaxfenster	10
numerische Intervalle	14	sysmis	15, 17
numerische Werte	14	sysmis()	17
offene Transformationen ausführen	77	systemdefiniert fehlende Werte	10
Partielle Korrelationen	103	System-Missing	30
probit()	17	Systemvariablen	36
Programmdatei	9	Temporäre Transformationen	67
Programmgesteuerter Modus	13	Textausgabe-Editor	12
Programmierhilfen	19	Textviewer	12
Programmierung	19	Transponieren	76
Prozeduren		trunc()	17
AGGREGATE	96	t-Test bei einer Stichprobe	114
ANOVA	127	t-Test bei gepaarten Stichproben	115
CORRELATIONS	101	t-Test bei unabhängigen Stichproben	115
CROSSTABS	98	t-Tests	114
DESCRIPTIVES	88	Umkodieren	59
EXAMINE	90	uniform()	17
FREQUENCIES	85	upcase()	18
GRAPH	147	value()	18

var()	17	Verteilungsfunktionen.....	17
Variablen hinzufügen.....	49	Viewer.....	12
Variablenansicht.....	10	Wahrheitstabellen.....	16
Variablenlabels	14, 28	Werte	14
Variablenlisten.....	14	Wertelabels	15, 29
Variablennamen	14, 28	Wertelisten.....	15
<i>Variablentyp</i>	28	xdate.tday().....	18
Vektoren	37	yrmoda()	18
Vergleichsausdrücke	15	Zählen.....	64