

Zur Verwendung dieses Buches

1	Zielsetzung des Buches	2
2	Daten und Skalen	4
3	Einteilung multivariater Analysemethoden	7
3.1	Strukturen-prüfende Verfahren	8
3.2	Strukturen-entdeckende Verfahren	12
3.3	Zusammenfassende Betrachtung	15
4	Zur Verwendung von SPSS	15
4.1	Die Daten	16
4.1.1	Der Daten-Editor	18
4.1.2	Erstellung einer neuen Datendatei	19
4.2	Einfache Statistiken und Grafiken	24
4.3	Die Kommandosprache	31
4.3.1	Aufbau einer Syntaxdatei	31
4.3.2	Syntax der Kommandos	32
4.3.3	Kommandos zur Datendefinition	34
4.3.4	Prozedurkommandos	35
4.3.5	Hilfskommandos	35
4.3.6	Erstellen, Öffnen und Speichern einer Syntaxdatei	36
4.3.7	Ausführen der Syntaxdatei	40
4.4	Die Systeme von SPSS	41
5	Literaturhinweise	43

1 Zielsetzung des Buches

Multivariate Analysemethoden sind heute eines der Fundamente der empirischen Forschung in den Realwissenschaften. Die Methoden sind immer noch in stürmischer Entwicklung. Es werden ständig neue methodische Varianten entwickelt, neue Anwendungsbereiche erschlossen und neue oder verbesserte Computer-Programme, ohne die eine praktische Anwendung der Verfahren nicht möglich ist, entwickelt.

Mancher Interessierte aber empfindet Zugangsbarrieren zur Anwendung der Methoden, die aus

- Vorbehalten gegenüber mathematischen Darstellungen,
- einer gewissen Scheu vor dem Einsatz des Computers und
- mangelnder Kenntnis der Methoden und ihrer Anwendungsmöglichkeiten

resultieren. Es ist eine Kluft zwischen interessierten Fachleuten und Methodenexperten festzustellen, die bisher nicht genügend durch das Angebot der Fachliteratur überbrückt wird.

Die Autoren dieses Buches haben sich deshalb das Ziel gesetzt, zur Überwindung dieser Kluft beizutragen. Daraus ist ein Text entstanden, der folgende Charakteristika besonders herausstellt:

1. Es ist größte Sorgfalt darauf verwendet worden, die Methoden *allgemeinverständlich* darzustellen. Der Zugang zum Verständnis durch den mathematisch ungeschulten Leser hat in allen Kapiteln Vorrang gegenüber dem methodischen Detail. Dennoch wird der rechnerische Gehalt der Methoden in den wesentlichen Grundzügen erklärt, damit sich der Leser, der sich in die Methoden einarbeitet, eine Vorstellung von der Funktionsweise, den Möglichkeiten und Grenzen der Methoden verschaffen kann.
2. Das Verständnis wird erleichtert durch die ausführliche Darstellung von *Beispielen*, die es erlauben, die Vorgehensweise der Methoden leicht nachzuvollziehen und zu verstehen.
3. Darüber hinaus wurde - soweit die Methoden das zulassen - ein Beispiel durchgehend für mehrere Methoden benutzt, um das Einarbeiten zu erleichtern und um die Ergebnisse der Methoden vergleichen zu können. Die Rohdaten der Beispiele können über den Bestellschein am Ende des Buches oder über die Internetadresse www.multivariate.de angefordert werden.

Die Beispiele sind dem Marketing-Bereich entnommen. Die Darstellung ist jedoch so gehalten, daß jeder Leser die Fragestellung versteht und auf seine spezifischen Anwendungsprobleme in anderen Bereichen übertragen kann.

4. Der Umfang des zu verarbeitenden Datenmaterials ist in aller Regel so groß, daß die Rechenprozeduren der einzelnen Verfahren mit vertretbarem Aufwand nur computergestützt durchgeführt werden können. Deshalb erstreckt sich die Darstellung der Methoden sowohl auf die Grundkonzepte der Methoden als auch auf die *Nutzung geeigneter Computer-Programme* als Arbeitshilfe. Es existiert heute eine Reihe von Programmpaketen, die die Anwend-

ung multivariater Analysemethoden nicht nur dem Computer-Spezialisten erlauben. Insbesondere bedingt durch die zunehmende Verbreitung und Leistungsfähigkeit des PCs sowie die komfortablere Gestaltung von Benutzeroberflächen wird auch die Nutzung der Programme zunehmend erleichtert. Damit wird der Fachmann für das Sachproblem unabhängig vom Computer-Spezialisten.

Das Programmpaket bzw. Programmsystem, mit dem die meisten Beispiele durchgerechnet werden, ist *SPSS* (ursprünglich: *Statistical Package for the Social Sciences*, jetzt: *Statistical Product and Service Solutions*). Als Programmsystem wird dabei eine Sammlung von Programmen mit einer gemeinsamen Benutzeroberfläche bezeichnet. SPSS hat sehr weite Verbreitung gefunden, besonders im Hochschulbereich, aber auch in der Praxis. Es ist unter vielen Betriebssystemen auf Großrechnern, Workstations und PC verfügbar.

5. Das vorliegende Buch hat den Charakter eines *Arbeitsbuches*. Die Darstellungen sind so gewählt, daß der Leser in jedem Fall alle Schritte der Lösungsfindung nachvollziehen kann. Alle Ausgangsdaten, die den Beispielen zugrunde liegen, sind abgedruckt und können für die umfangreicheren Fallbeispiele über www.multivariate.de bestellt werden. Die Syntaxkommandos für die Computer-Programme werden im einzelnen aufgeführt, so daß der Leser durch eigenes Probieren sehr schnell erkennen kann, wie leicht letztlich der Zugang zur Anwendung der Methoden unter Einsatz des Computers ist, wobei er seine eigenen Ergebnisse gegen die im vorliegenden Buch ausgewiesenen kontrollieren kann.
6. Die Ergebnisse der computergestützten Rechnungen in den einzelnen Methoden werden jeweils anhand der betreffenden *Programmausdrucke* erläutert und kommentiert. Dadurch kann der Leser, der sich in die Handhabung der Methoden einarbeitet, schnell in den eigenen Ergebnissen eine Orientierung finden.
7. Besonderes Gewicht wurde auf die *inhaltliche Interpretation* der Ergebnisse der einzelnen Verfahren gelegt. Wir haben es uns dabei zur Aufgabe gemacht, die *Ansatzpunkte für Ergebnismanipulationen* in den Verfahren offenzulegen und die Gestaltungsspielräume aufzuzeigen, damit der Anwender der Methoden objektive und subjektive Bestimmungsfaktoren der Ergebnisse unterscheiden kann. Dies macht u.a. erforderlich, daß methodische Details offengelegt werden. Dabei wird auch deutlich, daß dem Anwender der Methoden eine Verantwortung für seine Interpretation der Ergebnisse zukommt.

Faßt man die genannten Merkmale des Buches zusammen, dann ergibt sich ein Konzept, das geeignet ist, sowohl dem Anfänger, der sich in die Handhabung der Methoden einarbeitet, als auch demjenigen, der mit den Ergebnissen dieser Methoden arbeiten muß, die erforderliche Hilfe zu geben. Die Konzeption läßt es dabei zu, daß *jede dargestellte Methode für sich verständlich* ist. Der Leser ist also an keine Reihenfolge der Kapitel gebunden.

Im folgenden wird ein knapper Überblick über die Verfahren der multivariaten Analysetechnik gegeben. Da sich die einzelnen Verfahren vor allem danach unterscheiden lassen, welche Anforderungen sie an das Datenmaterial stellen, seien

hierzu einige Bemerkungen vorausgeschickt, die für Anfänger gedacht und deshalb betont knapp gehalten sind.¹

2 Daten und Skalen

Das "Rohmaterial" für multivariate Analysen sind die (vorhandenen oder noch zu erhebenden) *Daten*. Die Qualität von Daten wird u.a. bestimmt durch die Art und Weise der *Messung*. Daten sind nämlich das Ergebnis von Meßvorgängen. Messen bedeutet, daß Eigenschaften von Objekten nach bestimmten Regeln in Zahlen ausgedrückt werden.

Im wesentlichen bestimmt die jeweils betrachtete Art einer Eigenschaft, wie gut man ihre Ausprägung messen, d.h. wie gut man sie in Zahlen ausdrücken kann. So wird z.B. die Körpergröße eines Menschen sehr leicht in Zahlen auszudrücken sein, seine Intelligenz, seine Motivation oder sein Gesundheitszustand dagegen sehr schwierig.

Die "Meßlatte", auf der die Ausprägungen einer Eigenschaft abgetragen werden, heißt *Skala*. Je nachdem, in welcher Art und Weise eine Eigenschaft eines Objektes in Zahlen ausgedrückt (gemessen) werden kann, unterscheidet man Skalen mit unterschiedlichem *Skalenniveau*:

1. Nominalskala
2. Ordinalskala
3. Intervallskala
4. Ratioskala.

Das Skalenniveau bedingt sowohl den *Informationsgehalt der Daten* wie auch die *Anwendbarkeit von Rechenoperationen*. Nachfolgend sollen die Skalentypen und ihre Eigenschaften kurz umrissen werden.

Die *Nominalskala* stellt die primitivste Grundlage des Messens dar. Beispiele für Nominalskalen sind

- Geschlecht (männlich - weiblich)
- Religion (katholisch - evangelisch - andere)
- Farbe (rot - gelb - grün - blau ...)
- Werbemedium (Fernsehen - Zeitungen - Plakattafeln).

Nominalskalen stellen also Klassifizierungen qualitativer Eigenschaftsausprägungen dar. Zwecks leichter Verarbeitung mit Computern werden die Ausprägungen von Eigenschaften häufig durch Zahlen ausgedrückt. So lassen sich z.B. die Farben einer Verpackung wie folgt kodieren:

¹ Vgl. z.B. Bley Müller, J./Gehlert, G./Gülicher, H., 2002, Kapitel 1.5 oder Mayntz, R./Holm, K./Hübner, P., 1978, Kap. 2.

rot = 1
 gelb = 2
 grün = 3

Die Zahlen hätten auch in anderer Weise zugeordnet werden können, solange diese Zuordnung eineindeutig ist, d.h. solange durch eine Zahl genau eine Farbe definiert ist. Mit derartigen Zahlen sind daher keine arithmetischen Operationen (wie Addition, Subtraktion, Multiplikation oder Division) erlaubt. Vielmehr lassen sich lediglich durch Zählen der Merkmalsausprägungen (bzw. der sie repräsentierenden Zahlen) Häufigkeiten ermitteln.

Eine *Ordinalskala* stellt das nächsthöhere Meßniveau dar. Die Ordinalskala erlaubt die Aufstellung einer Rangordnung mit Hilfe von Rangwerten (d.h. ordinalen Zahlen). Beispiele: Produkt A wird Produkt B vorgezogen, Herr M. ist tüchtiger als Herr N. Die Untersuchungsobjekte können immer nur in eine Rangordnung gebracht werden. Die Rangwerte 1., 2., 3. etc. sagen nichts über die Abstände zwischen den Objekten aus. Aus der Ordinalskala kann also nicht abgelesen werden, um wieviel das Produkt A besser eingeschätzt wird als das Produkt B. Daher dürfen auch ordinale Daten, ebenso wie nominale Daten, nicht arithmetischen Operationen unterzogen werden. Zulässige statistische Maße sind neben Häufigkeiten z.B. der Median oder Quantile.

Das wiederum nächsthöhere Meßniveau stellt die *Intervallskala* dar. Diese weist gleichgroße Skalenabschnitte aus. Ein typisches Beispiel ist die Celsius-Skala zur Temperaturmessung, bei der der Abstand zwischen Gefrierpunkt und Siedepunkt des Wassers in hundert gleichgroße Abschnitte eingeteilt wird. Bei intervallskalierten Daten besitzen auch die Differenzen zwischen den Daten Informationsgehalt (z.B. großer oder kleiner Temperaturunterschied), was bei nominalen oder ordinalen Daten nicht der Fall ist.

Oftmals werden - auch in dem vorliegenden Buch - Skalen benutzt, von denen man lediglich annimmt, sie seien intervallskaliert. Dies ist z.B. der Fall bei Ratingskalen: Eine Auskunftsperson ordnet einer Eigenschaft eines Objektes einen Zahlenwert auf einer Skala von 1 bis 7 (oder einer kürzeren oder längeren Skala) zu. Solange die Annahme gleicher Skalenabstände unbestätigt ist, handelt es sich allerdings strenggenommen um eine Ordinalskala.

Intervallskalierte Daten erlauben die arithmetischen Operationen der Addition und Subtraktion. Zulässige statistische Maße sind, zusätzlich zu den oben genannten, z.B. der Mittelwert (arithmetisches Mittel) und die Standardabweichung, nicht aber die Summe.

Die *Ratio- (oder Verhältnis)skala* stellt das höchste Meßniveau dar. Sie unterscheidet sich von der Intervallskala dadurch, daß zusätzlich ein natürlicher Nullpunkt existiert, der sich für das betreffende Merkmal im Sinne von "nicht vorhanden" interpretieren läßt. Das ist z.B. bei der Celsius-Skala oder der Kalenderzeit nicht der Fall, dagegen aber bei den meisten physikalischen Merkmalen (z.B. Länge, Gewicht, Geschwindigkeit) wie auch bei den meisten ökonomischen Merkmalen (z.B. Einkommen, Kosten, Preis). Bei verhältnisskalierten Daten besitzen nicht nur die Differenz, sondern, infolge der Fixierung des Nullpunktes, auch der Quotient bzw. das Verhältnis (Ratio) der Daten Informationsgehalt (daher der Name).

Ratioskalierte Daten erlauben die Anwendung aller arithmetischen Operationen wie auch die Anwendung aller obigen statistischen Maße. Zusätzlich sind z.B. die Anwendung des geometrischen Mittels oder des Variationskoeffizienten erlaubt.

Nominalskala und Ordinalskala bezeichnet man als nichtmetrische oder auch kategoriale Skalen, Intervallskala und Ratioskala als metrische Skalen.

In Abbildung 1 sind noch einmal die vier Skalenniveaus mit ihren Merkmalen zusammengestellt.

Abbildung 1: Skalenniveau

Skala		Merkmale	Mögliche rechnerische Handhabung
nicht-metrische Skalen	NOMINAL-SKALA	Klassifizierung qualitativer Eigenschaftsausprägungen	Bildung von Häufigkeiten
	ORDINAL-SKALA	Rangwert mit Ordinalzahlen	Median, Quantile
metrische Skalen	INTERVALL-SKALA	Skala mit gleichgroßen Abschnitten ohne natürlichen Nullpunkt	Subtraktion, Mittelwert
	RATIO-SKALA	Skala mit gleichgroßen Abschnitten und natürlichem Nullpunkt	Summe, Division, Multiplikation

Zusammenfassend läßt sich sagen: Je höher das Skalenniveau ist, desto größer ist auch der Informationsgehalt der betreffenden Daten und desto mehr Rechenoperationen und statistische Maße lassen sich auf die Daten anwenden.

Es ist generell möglich, Daten von einem höheren Skalenniveau auf ein niedrigeres Skalenniveau zu transformieren, nicht aber umgekehrt. Dies kann sinnvoll sein, um die Übersichtlichkeit der Daten zu erhöhen oder um ihre Analyse zu vereinfachen. So werden z.B. häufig Einkommensklassen oder Preisklassen gebildet. Dabei kann es sich um eine Transformation der ursprünglich ratio-skalierten Daten auf eine Intervall-, Ordinal- oder Nominal-Skala handeln. Mit der Transformation auf ein niedrigeres Skalenniveau ist natürlich immer auch ein Informationsverlust verbunden.

3 Einteilung multivariater Analysemethoden

In diesem Buch werden die nachfolgenden Verfahren behandelt:

Kapitel 1:	Regressionsanalyse
Kapitel 2:	Varianzanalyse
Kapitel 3:	Diskriminanzanalyse
Kapitel 4:	Kontingenzanalyse
Kapitel 5:	Faktorenanalyse
Kapitel 6:	Strukturgleichungsmodelle
Kapitel 7:	Logistische Regressionsanalyse
Kapitel 8:	Clusteranalyse
Kapitel 9:	Conjoint Measurement
Kapitel 10:	Multidimensionale Skalierung
Kapitel 11:	Korrespondenzanalyse
Kapitel 12:	Neuronale Netze

Im folgenden nehmen wir eine Einordnung dieser multivariaten Analysemethoden vor dem Hintergrund des Anwendungsbezuges vor. Dabei sei jedoch betont, daß eine *überschneidungsfreie Zuordnung* der Verfahren zu praktischen Fragestellungen nicht immer möglich ist, da sich die Zielsetzungen der Verfahren z.T. überlagern. Versucht man jedoch eine Einordnung der Verfahren nach anwendungsbezogenen Fragestellungen, so bietet sich eine Einteilung in primär *strukturen-entdeckende Verfahren* und primär *strukturen-prüfende Verfahren* an. Diese beiden Kriterien werden in diesem Zusammenhang wie folgt verstanden:

1. *Strukturen-prüfende Verfahren* sind solche multivariaten Verfahren, deren primäres Ziel in der *Überprüfung von Zusammenhängen* zwischen Variablen liegt. Der Anwender besitzt eine auf sachlogischen oder theoretischen Überlegungen basierende Vorstellung über die Zusammenhänge zwischen Variablen und möchte diese mit Hilfe multivariater Verfahren überprüfen.
Verfahren, die diesem Bereich der multivariaten Datenanalyse zugeordnet werden können, sind die Regressionsanalyse, die Varianzanalyse, die Diskriminanzanalyse, die Kontingenzanalyse sowie die Logistische Regression, Strukturgleichungsmodelle und das Conjoint Measurement zur Analyse von Präferenzstrukturen.
2. *Strukturen-entdeckende Verfahren* sind solche multivariaten Verfahren, deren Ziel in der *Entdeckung von Zusammenhängen* zwischen Variablen oder zwischen Objekten liegt. Der Anwender besitzt zu Beginn der Analyse noch keine Vorstellungen darüber, welche Beziehungszusammenhänge in einem Datensatz existieren.
Verfahren, die primär eingesetzt werden, um mögliche Beziehungszusammenhänge aufzudecken, sind die Faktorenanalyse, die Clusteranalyse,

die Multidimensionale Skalierung, die Korrespondenzanalyse und die Neuronalen Netze.

3.1 Strukturen-prüfende Verfahren

Die strukturen-prüfenden Verfahren werden primär zur Durchführung von *Kausalanalysen* eingesetzt, um herauszufinden, ob und wie stark sich z.B. das Wetter, die Bodenbeschaffenheit sowie unterschiedliche Düngemittel und -mengen auf den Ernteertrag auswirken oder wie stark die Nachfrage eines Produktes von dessen Qualität, dem Preis, der Werbung und dem Einkommen der Konsumenten abhängt.

Voraussetzung für die Anwendung der entsprechenden Verfahren ist, daß der Anwender *a priori (vorab)* eine sachlogisch möglichst gut fundierte Vorstellung über den Kausalzusammenhang zwischen den Variablen entwickelt hat, d.h. er weiß bereits oder vermutet, welche der Variablen auf andere Variablen einwirken. Zur Überprüfung seiner (theoretischen) Vorstellungen werden die von ihm betrachteten Variablen i.d.R. in *abhängige* und *unabhängige* Variablen eingeteilt und dann mit Hilfe von multivariaten Analysemethoden an den empirisch erhobenen Daten überprüft. Nach dem Skalenniveau der Variablen lassen sich die grundlegenden strukturen-prüfenden Verfahren gemäß Abbildung 2 charakterisieren.

Abbildung 2: Grundlegende strukturen-prüfende Verfahren

		UNABHÄNGIGE VARIABLE	
		metrisches Skalenniveau	nominales Skalenniveau
ABHÄNGIGE VARIABLE	metrisches Skalenniveau	Regressionsanalyse	<ul style="list-style-type: none"> • Varianzanalyse, • Regression mit Dummies
	nominales Skalenniveau	<ul style="list-style-type: none"> • Diskriminanzanalyse, • Logistische Regression 	Kontingenzanalyse

Regressionsanalyse

Die Regressionsanalyse ist ein außerordentlich vielseitiges und flexibles Analyseverfahren, das sowohl für die *Beschreibung* und *Erklärung von Zusammenhängen* als auch für die *Durchführung von Prognosen* große Bedeutung besitzt. Sie ist damit sicherlich das wichtigste und am häufigsten angewendete multivariate Analyseverfahren. Insbesondere kommt sie in Fällen zur Anwendung, wenn Wirkungsbeziehungen zwischen einer abhängigen und einer oder mehreren unabhängigen Variablen untersucht werden sollen. Mit Hilfe der Regressionsanalyse können derartige Beziehungen quantifiziert und damit weitgehend exakt beschrieben werden. Außerdem lassen sich mit ihrer Hilfe Hypothesen über Wirkungsbeziehungen prüfen und auch Prognosen erstellen.

Ein Beispiel bildet die Frage, ob und wie die Absatzmenge eines Produktes vom Preis, den Werbeausgaben, der Zahl der Verkaufsstätten und dem Volkseinkommen abhängt. Sind diese Zusammenhänge mit Hilfe der Regressionsanalyse quantifiziert und empirisch bestätigt worden, so lassen sich Prognosen (What-if-Analysen) erstellen, die beantworten, wie sich die Absatzmenge verändern wird, wenn z.B. der Preis oder die Werbeausgaben oder auch beide Variablen zusammen verändert werden.

Die Regressionsanalyse ist prinzipiell anwendbar, wenn sowohl die abhängige als auch die unabhängigen Variablen metrisches Skalenniveau besitzen. Dies ist der klassische Fall. Durch Anwendung der sog. *Dummy-Variablen-Technik* lassen sich aber auch qualitative (nominal skalierte) Variable in die Regressionsanalyse einbeziehen und deren Anwendungsbereich somit ausweiten. Dummy-Variable sind binäre Variable, die nur die Werte 0 oder 1 annehmen. Stellen wir uns vor, es sollen die Einflüsse verschiedener Produkteigenschaften auf das Kaufverhalten von Konsumenten untersucht werden. Die Dummy-Variable q_1 würde dann in allen Fällen, bei denen das Produkt eine rote Verpackung hat, den Wert 1 annehmen, und wenn dies nicht der Fall ist, den Wert 0.

$$q_1 = \begin{cases} 1 & \text{falls Farbe} = \text{rot} \\ 0 & \text{sonst} \end{cases}$$

In analoger Weise lassen sich auch eine Dummy-Variable q_2 für die Farbe Gelb und eine Dummy-Variable q_3 für die Farbe Grün definieren. Wenn allerdings nur Verpackungen in den drei Farben Rot, Gelb und Grün vorkommen, so wäre eine der drei Dummies überflüssig. Denn wenn $q_1 = 0$ und $q_2 = 0$ gilt, so muß zwangsläufig $q_3 = 1$ gelten. Die drei Farben lassen sich also eindeutig mittels der zwei Dummies (q_1, q_2) beschreiben: rot = (1, 0), gelb = (0, 1), grün = (0, 0). Generell gilt, daß sich eine nominale Variable mit n Ausprägungen durch $n-1$ Dummy-Variablen ersetzen läßt.

Die Bedeutung von Dummy-Variablen liegt darin, daß sie sich wie metrische Variable behandeln lassen. Somit lassen sich mit ihrer Hilfe auch nominal skalierte Variable in eine Regressionsanalyse einbeziehen. Dies gilt aber generell nur für die unabhängigen Variablen und nicht für die abhängige Variable. Nachteilig ist, daß sich dadurch u.U. die Zahl der Variablen und der damit verbundene Kodierungs-

und Rechenaufwand stark erhöht. Deshalb kann in solchen Fällen die Anwendung einer Varianzanalyse einfacher und übersichtlicher sein.

Varianzanalyse

Werden die unabhängigen Variablen auf nominalem Skalenniveau gemessen und die abhängigen Variablen auf metrischem Skalenniveau, so findet die Varianzanalyse Anwendung. Dieses Verfahren besitzt besondere Bedeutung für die *Analyse von Experimenten*, wobei die nominalen unabhängigen Variablen die experimentellen Einwirkungen repräsentieren. So kann z.B. in einem Experiment untersucht werden, welche Wirkung alternative Verpackungen eines Produktes oder dessen Platzierung im Geschäft auf die Absatzmenge haben.

Diskriminanzanalyse

Ist die abhängige Variable nominal skaliert, und besitzen die unabhängigen Variablen metrisches Skalenniveau, so findet die Diskriminanzanalyse Anwendung. Die Diskriminanzanalyse ist ein Verfahren zur *Analyse von Gruppenunterschieden*. Ein Beispiel bildet die Frage, ob und wie sich die Wähler der verschiedenen Parteien hinsichtlich soziodemografischer und psychografischer Merkmale unterscheiden. Die abhängige nominale Variable identifiziert die Gruppenzugehörigkeit, hier die gewählte Partei, und die unabhängigen Variablen beschreiben die Gruppenelemente, hier die Wähler.

Ein weiteres Anwendungsgebiet der Diskriminanzanalyse bildet die *Klassifizierung von Elementen*. Nachdem für eine gegebene Menge von Elementen die Zusammenhänge zwischen der Gruppenzugehörigkeit der Elemente und ihren Merkmalen analysiert wurden, läßt sich darauf aufbauend eine Prognose der Gruppenzugehörigkeit von neuen Elementen vornehmen. Derartige Anwendungen finden sich z.B. bei der Kreditwürdigkeitsprüfung (Einstufung von Kreditkunden einer Bank in Risikoklassen) oder bei der Personalbeurteilung (Einstufung von Außendienstmitarbeitern nach erwartetem Verkaufserfolg).

Kontingenzanalyse

Eine weitere Methodengruppe, die der Analyse von Beziehungen zwischen ausschließlich nominalen Variablen dient, wird als Kontingenzanalyse bezeichnet. Hier kann es z.B. darum gehen, die Frage nach dem Zusammenhang zwischen Rauchen (Raucher versus Nichtraucher) und Lungenerkrankung (ja, nein) statistisch zu überprüfen. Die Überprüfung erfolgt dabei auf der Basis von in Form einer Kreuztabelle (Kontingenztafel) angeordneten Daten. Mit Hilfe weiterführender Verfahren, wie der sog. Logit-Analyse, läßt sich weiterhin auch die Abhängigkeit einer nominalen Variablen von mehreren nominalen Einflußgrößen untersuchen (vgl. hierzu auch das Verfahren der logistischen Regression).

Logistische Regression

Ganz ähnliche Fragestellungen, wie mit der Diskriminanzanalyse können auch mit dem Verfahren der logistischen Regression untersucht werden. Hier wird die *Wahrscheinlichkeit* der Zugehörigkeit zu einer Gruppe (einer Kategorie der abhängigen Variablen) in Abhängigkeit von einer oder mehrerer unabhängiger Variablen

bestimmt. Dabei können die unabhängigen Variablen sowohl nominales als auch metrisches Skalenniveau aufweisen. Über die Analyse der Gruppenunterschiede hinaus kann z.B. auch das Herzinfarkttrisiko von Patienten in Abhängigkeit von ihrem Alter und ihrem Cholesterin-Spiegel ermittelt werden. Da zur Schätzung der Eintrittswahrscheinlichkeiten der Kategorien der abhängigen Variablen auf die (s-förmige) logistische Funktion zurückgegriffen wird, gehört dieses Verfahren zu den *nicht-linearen Analyseverfahren*.

Strukturgleichungsmodelle

Die bisher betrachteten Analysemethoden gehen davon aus, daß alle Variablen in der Realität beobachtbar und gegebenenfalls auch meßbar sind. Bei vielen theoriegestützten Fragestellungen hat man es aber auch mit nicht beobachtbaren Variablen zu tun, sog. *hypothetischen Konstrukten* oder *latenten Variablen*. Beispiele hierfür sind psychologische Konstrukte wie Einstellung und Motivation oder soziologische Konstrukte wie Kultur und soziale Schicht. In solchen Fällen kann die Analyse von Strukturgleichungen zur Anwendung kommen.

Zur Behandlung von Strukturgleichungsmodellen wird in diesem Buch auf das Programmpaket AMOS (Analysis of Moment Structures) zurückgegriffen, das Datenmatrizen aus SPSS analysieren und Ergebnisse mit SPSS austauschen kann.² Mit Hilfe von AMOS lassen sich komplexe Kausalstrukturen überprüfen. Insbesondere können Beziehungen mit mehreren abhängigen Variablen, mehrstufigen Kausalbeziehungen und mit nicht beobachtbaren (latenten) Variablen überprüft werden. Der Benutzer muß, wenn er latente Variable in die Betrachtungen einbeziehen will, zwei Modelle spezifizieren:

- Das *Meßmodell*, das die Beziehungen zwischen den latenten Variablen und geeigneten Indikatoren vorgibt, mittels derer sich die latenten Variablen indirekt messen lassen.
- Das *Strukturmodell*, welches die Kausalbeziehungen zwischen den latenten Variablen vorgibt, die letztlich dann zu überprüfen sind.

Die Variablen des Strukturmodells können alle latent sein, müssen es aber nicht. Ein Beispiel, bei dem nur die unabhängigen Variablen latent sind, wäre die Abhängigkeit der Absatzmenge von der subjektiven Produktqualität und Servicequalität eines Anbieters.

Conjoint Measurement

Bei den bisher aufgezeigten Verfahren wurde nur zwischen metrischem und nominalem Skalenniveau der Variablen unterschieden. Ein Verfahren, bei dem die abhängige Variable häufig auf ordinalem Skalenniveau gemessen wird, ist das Conjoint Measurement. Insbesondere lassen sich mit Hilfe des Conjoint Measurement ordinal gemessene Präferenzen analysieren. Ziel ist es dabei, den *Beitrag einzelner Merkmale* von Produkten oder sonstigen Objekten *zum Gesamtnutzen* dieser Objekte herauszufinden. Einen wichtigen Anwendungsbereich bildet die Gestaltung

² Bis zur 9. Auflage wurde bei der Behandlung von Strukturgleichungsmodellen auf das Programm LISREL (Linear Structural RELationships) zurückgegriffen.

neuer Produkte. Dazu ist es von Wichtigkeit, den Einfluß oder Beitrag alternativer Produktmerkmale, z.B. alternativer Materialien, Formen, Farben oder Preisstufen, auf die Nutzenbeurteilung zu kennen.

Beim Conjoint Measurement muß der Forscher vorab festlegen, welche Merkmale in welchen Ausprägungen berücksichtigt werden sollen. Hierauf basierend wird sodann ein Erhebungsdesign gebildet, im Rahmen dessen Präferenzen, z.B. bei potentiellen Käufern eines neuen Produktes, gemessen werden. Auf Basis dieser Daten erfolgt dann die Analyse zur Ermittlung der Nutzenbeiträge der berücksichtigten Merkmale und ihrer Ausprägungen. Das Conjoint Measurement bildet damit also eine *Kombination aus Erhebungs- und Analyseverfahren*.

3.2 Strukturen-entdeckende Verfahren

Die hier den strukturen-entdeckenden Verfahren zugeordneten Analysemethoden werden primär zur *Entdeckung von Zusammenhängen* zwischen Variablen oder zwischen Objekten eingesetzt. Es erfolgt daher vorab durch den Anwender *keine* Zweiteilung der Variablen in abhängige und unabhängige Variablen, wie es bei den strukturen-prüfenden Verfahren der Fall ist.

Faktorenanalyse

Die Faktorenanalyse findet insbesondere dann Anwendung, wenn im Rahmen einer Erhebung eine Vielzahl von Variablen zu einer bestimmten Fragestellung erhoben wurde, und der Anwender nun an einer Reduktion bzw. *Bündelung der Variablen* interessiert ist. Von Bedeutung ist die Frage, ob sich möglicherweise sehr zahlreiche Merkmale, die zu einem bestimmten Sachverhalt erhoben wurden, auf einige wenige "zentrale Faktoren" zurückführen lassen. Ein einfaches Beispiel hierzu bildet die Verdichtung der zahlreichen technischen Eigenschaften von Kraftfahrzeugen auf wenige Dimensionen, wie Größe, Leistung und Sicherheit.

Einen wichtigen Anwendungsbereich der Faktorenanalyse bilden *Positionierungsanalysen*. Dabei werden die subjektiven Eigenschaftsbeurteilungen von Objekten (z.B. Produktmarken, Unternehmen oder Politiker) mit Hilfe der Faktorenanalyse auf zugrundeliegende Beurteilungsdimensionen verdichtet. Ist eine Verdichtung auf zwei oder drei Dimensionen möglich, so lassen sich die Objekte im Raum dieser Dimensionen grafisch darstellen. Im Unterschied zu anderen Formen der Positionierungsanalyse spricht man hier von faktorieller Positionierung.

Clusteranalyse

Während die Faktorenanalyse eine Verdichtung oder Bündelung von Variablen vornimmt, wird mit der Clusteranalyse eine *Bündelung von Objekten* angestrebt. Das Ziel ist dabei, die Objekte so zu Gruppen (Clustern) zusammenzufassen, daß die Objekte in einer Gruppe möglichst ähnlich und die Gruppen untereinander möglichst unähnlich sind. Beispiele sind die Bildung von Persönlichkeitstypen auf

Basis der psychografischen Merkmale von Personen oder die Bildung von Marktsegmenten auf Basis nachfragerrelevanter Merkmale von Käufern.

Zur Überprüfung der Ergebnisse einer Clusteranalyse kann die Diskriminanzanalyse herangezogen werden. Dabei wird untersucht, inwieweit bestimmte Variablen zur Unterscheidung zwischen den Gruppen, die mittels Clusteranalyse gefunden wurden, beitragen bzw. diese erklären.

Multidimensionale Skalierung

Den Hauptanwendungsbereich der Multidimensionalen Skalierung (MDS) bilden Positionierungsanalysen, d.h. die *Positionierung von Objekten im Wahrnehmungsraum* von Personen. Sie bildet somit eine Alternative zur faktoriellen Positionierung mit Hilfe der Faktorenanalyse.

Im Unterschied zur faktoriellen Positionierung werden bei Anwendung der MDS nicht die subjektiven Beurteilungen von Eigenschaften der untersuchten Objekte erhoben, sondern es werden nur wahrgenommene globale Ähnlichkeiten zwischen den Objekten erfragt. Mittels der MDS werden die diesen Ähnlichkeiten zugrundeliegenden Wahrnehmungsdimensionen abgeleitet. Wie schon bei der faktoriellen Positionierung lassen sich sodann die Objekte im Raum dieser Dimensionen positionieren und grafisch darstellen. Die MDS findet insbesondere dann Anwendung, wenn der Forscher keine oder nur vage Kenntnisse darüber hat, welche Eigenschaften für die subjektive Beurteilung von Objekten (z.B. Produktmarken, Unternehmen oder Politiker) von Relevanz sind.

Zwischen der Multidimensionalen Skalierung und dem Conjoint Measurement besteht sowohl inhaltlich wie auch methodisch eine enge Beziehung, obgleich wir sie hier unterschiedlich zum einen den strukturen-entdeckenden und zum anderen den strukturen-prüfenden Verfahren zugeordnet haben. Beide Verfahren befassen sich mit der Analyse psychischer Sachverhalte und bei beiden Verfahren können auch ordinale Daten analysiert werden, weshalb sie z.T. auch identische Algorithmen verwenden. Ein gewichtiger Unterschied besteht dagegen darin, daß der Forscher bei Anwendung des Conjoint Measurement bestimmte Merkmale auszuwählen hat.

Korrespondenzanalyse

Die Korrespondenzanalyse dient, wie auch die Faktorenanalyse und die Multidimensionale Skalierung (MDS), zur Visualisierung komplexer Daten. Sie wird daher in der Marktforschung ebenfalls zur Durchführung von Positionierungsanalysen verwendet. Insbesondere kann sie als ein Verfahren der multidimensionalen Skalierung von nominal skalierten Variablen charakterisiert werden. Sie ermöglicht es, die Zeilen und Spalten einer zweidimensionalen Kreuztabelle (Kontingenztafel) grafisch in einem gemeinsamen Raum darzustellen.

Beispiel: Gegeben sei eine Häufigkeitstabelle, deren Zeilen Automarken betreffen und in deren Spalten wünschenswerte Merkmale von Autos (z.B. hohe Sicherheit, schönes Design) stehen. Die Zellen der Matrix sollen beinhalten, mit welcher Häufigkeit ein bestimmtes qualitatives Merkmal den verschiedenen Automarken im Rahmen einer Käuferbefragung zugeordnet wurde. Marken und Merkmale lassen sich sodann mit Hilfe der Korrespondenzanalyse in einem gemeinsamen Raum

als Punkte darstellen. Dadurch läßt sich dann erkennen, wie die Automarken relativ zueinander und in bezug auf die Merkmale von den Käufern beurteilt werden. Für die Korrespondenzanalyse spielt es dabei *keine* Rolle (im Unterschied zur Faktorenanalyse), welche Elemente in den Zeilen und welche in den Spalten angeordnet werden.

Ein besonderer Vorteil der Korrespondenzanalyse liegt darin, daß sie kaum Ansprüche an das Skalenniveau der Daten stellt. Die Daten müssen lediglich nicht-negativ sein. Die Korrespondenzanalyse kann daher auch zur Quantifizierung qualitativer Daten verwendet werden. Da sich qualitative Daten leichter erheben lassen als quantitative Daten, kommt diesem Verfahren eine erhebliche praktische Bedeutung zu.

Neuronale Netze

Neuronale Netze werden heute in der Praxis in zunehmendem Maße sowohl ergänzend zu den klassischen multivariaten Methoden eingesetzt, als auch in den Fällen, in den die klassischen Methoden versagen. Anwendungsgebiete sind Klassifikationen von Objekten, Prognosen von Zuständen oder Probleme der Gruppenbildung. Insofern bestehen hinsichtlich der Aufgabenstellungen Ähnlichkeiten zur Diskriminanzanalyse und zur Clusteranalyse. Die Methodik neuronaler Netze lehnt sich an biologische Informationsverarbeitungsprozesse im Gehirn an (daher der Name). Es werden künstliche neuronale Netze gebildet, die in der Lage sind, selbstständig aus Erfahrung zu lernen. Insbesondere vermögen sie, komplexe Muster in vorhandenen Daten (z.B. Finanzdaten, Verkaufsdaten) zu erkennen und eröffnen so eine sehr einfache Form der Datenanalyse. Besonders vorteilhaft lassen sie sich zur Behandlung von schlecht strukturierten Problemstellungen einsetzen.

Innerhalb neuronaler Netze werden künstliche Neuronen (Nervenzellen) als Grundelemente der Informationsverarbeitung in Schichten organisiert, wobei jedes Neuron mit denen der nachgelagerten Schicht verbunden ist. Dadurch lassen sich auch hochgradig nicht-lineare und komplexe Zusammenhänge ohne spezifisches Vorwissen über die etwaige Richtung und das Ausmaß der Wirkungsbeziehungen zwischen einer Vielzahl von Variablen modellieren.

Zum Erlernen von Strukturen wird das Netz zunächst in einer sog. *Trainingsphase* mit beobachteten Daten "gefüttert". Dabei wird unterschieden zwischen Lernprozessen, bei denen die richtigen Ergebnisse bekannt sind und diese durch das Netz reproduziert werden sollen (*überwachtes Lernen*), und solchen, bei denen die richtigen Ergebnisse nicht bekannt sind und lediglich ein konsistentes Verarbeitungsmuster erzeugt werden soll (*unüberwachtes Lernen*). Nach der Trainingsphase ist das Netz konfiguriert und kann für die Analyse neuer Daten eingesetzt werden.

3.3 Zusammenfassende Betrachtung

Die vorgenommene Zweiteilung der multivariaten Verfahren in strukturenprüfende und strukturen-entdeckende Verfahren kann keinen Anspruch auf Allgemeingültigkeit erheben, sondern kennzeichnet nur den vorwiegenden Einsatzbereich der Verfahren. So kann und wird auch die Faktorenanalyse zur Überprüfung von hypothetisch gebildeten Strukturen eingesetzt, und viel zu häufig werden in der empirischen Praxis auch Regressions- und Diskriminanzanalyse im heuristischen Sinne zur Auffindung von Kausalstrukturen verwendet. Diese Vorgehensweise wird nicht zuletzt auch durch die Verfügbarkeit leistungsfähiger Rechner und Programme unterstützt. Der gedankenlose Einsatz von multivariaten Verfahren kann leicht zu einer Quelle von Fehlinterpretationen werden, da ein statistisch signifikanter Zusammenhang keine hinreichende Bedingung für das Vorliegen eines kausal bedingten Zusammenhangs bildet. ("Erst denken, dann rechnen!") Es sei daher generell empfohlen, die strukturenprüfenden Verfahren auch in diesem Sinne, d. h. zur empirischen Überprüfung von theoretisch oder sachlogisch begründeten Hypothesen, einzusetzen. In Abbildung 3 sind die oben skizzierten multivariaten Verfahren noch einmal mit jeweils einem Anwendungsbeispiel zusammengefaßt.

4 Zur Verwendung von SPSS

Zur rechnerischen Durchführung der Analysen, die in diesem Buch behandelt werden, wurde vornehmlich das Programmsystem SPSS verwendet, da dieses in Wissenschaft und Praxis eine besonders große Verbreitung gefunden hat. Der Name 'SPSS' stand ursprünglich als Akronym für *Statistical Package for the Social Sciences*. Der Anwendungsbereich von SPSS reicht allerdings weit über den Bereich der Sozialwissenschaften hinaus und umfaßt auch verschiedene Systeme. Vermutlich deshalb steht heute SPSS für *Statistical Product and Service Solutions*.

In den einzelnen Kapiteln sind jeweils die erforderlichen Kommando-Sequenzen zum Nachvollzug der Analysen wiedergegeben. An dieser Stelle sollen in sehr kurzer Form einige allgemeine Hinweise zur Handhabung von SPSS angeführt werden. Bezüglich näherer Ausführungen muß auf die einschlägige Literatur verwiesen werden.³

³ Vgl. hierzu insbesondere die Handbücher von Norusis, M.J./SPSS Inc., die im Literaturverzeichnis aufgeführt sind, sowie das deutschsprachige Handbuch von Bühl, A./Zöfel, P., 2000.

Abbildung 3: Synopsis der multivariaten Analyseverfahren

Verfahren	Beispiel
Regressionsanalyse	Abhängigkeit der Absatzmenge eines Produktes von Preis, Werbeausgaben und Einkommen.
Varianzanalyse	Wirkung alternativer Verpackungsgestaltungen auf die Absatzmenge eines Produktes.
Diskriminanzanalyse	Unterscheidung der Wähler der verschiedenen Parteien hinsichtlich soziodemografischer und psychografischer Merkmale.
Kontingenzanalyse	Zusammenhang zwischen Rauchen und Lungenerkrankung.
Faktorenanalyse	Verdichtung einer Vielzahl von Eigenschaftsbeurteilungen auf zugrundeliegende Beurteilungsdimensionen.
Strukturgleichungsmodelle	Abhängigkeit der Käufertreue von der subjektiven Produktqualität und Servicequalität eines Anbieters.
Logistische Regression	Ermittlung des Herzinfarkttrisikos von Patienten in Abhängigkeit ihres Alters und ihres Cholesterinspiegels.
Clusteranalyse	Bildung von Persönlichkeitstypen auf Basis der psychografischen Merkmale von Personen.
Conjoint Measurement	Ableitung der Nutzenbeiträge alternativer Materialien, Formen u. Farben von Produkten zur Gesamtpräferenz.
Multidimensionale Skalierung	Positionierung von konkurrierenden Produktmarken im Wahrnehmungsraum der Konsumenten.
Korrespondenzanalyse	Darstellung von Produktmarken und Produktmerkmalen in einem gemeinsamen Raum.
Neuronale Netze	Untersuchung von Aktienkursen und möglichen Einflußfaktoren zwecks Prognose von Kursentwicklungen.

4.1 Die Daten

Die Datenanalyse mit SPSS setzt voraus, daß die Daten in Form einer *Matrix* angeordnet werden (vgl. Abbildung 4). SPSS erwartet, daß die *Spalten der Matrix* sich auf *Variablen* (variables), z.B. Eigenschaften, Merkmale, Dimensionen, beziehen.

Die *Zeilen der Matrix* bilden *Beobachtungen bzw. Fälle (cases)*, die sich auf unterschiedliche Personen, Objekte oder Zeitpunkte beziehen können. Ein kleines Beispiel zeigt Abbildung 5.

Abbildung 4: Datenmatrix

Fälle k	Variablen				
	1	2	3	J
1	x_{11}	x_{21}	x_{31}	x_{J1}
2	x_{12}	x_{22}	x_{32}	x_{J2}
.	.				.
.	.				.
.	.		Werte x_{jk}		.
.	.				.
.	.				.
K	x_{1K}	x_{2K}	x_{3K}	x_{JK}

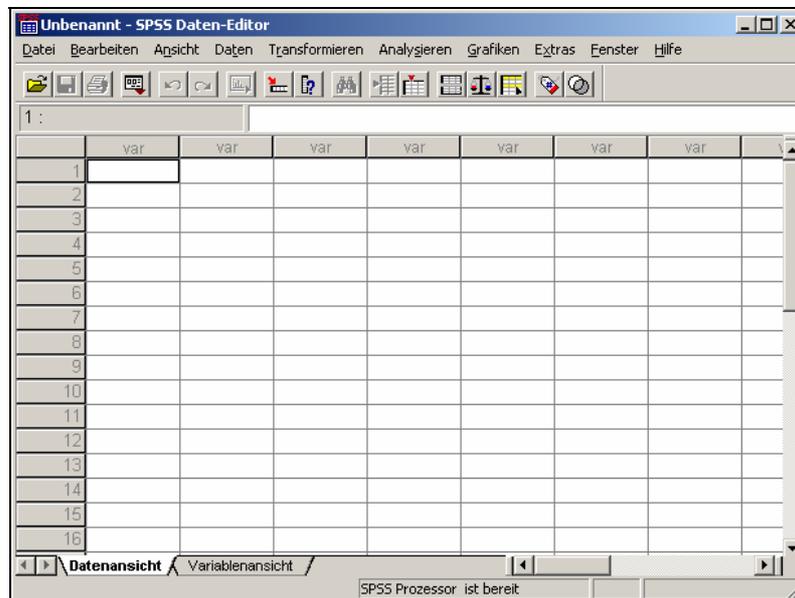
Abbildung 5: Beispiel einer Datenmatrix

Person	Geschlecht	Größe [cm]	Gewicht [kg]
1	1	178	68
2	0	166	50
3	1	183	75
4	0	168	52
5	1	195	100
6	1	175	73

4.1.1 Der Daten-Editor

Der Daten-Editor dient der Eingabe der zu analysierenden Daten in SPSS. Neben der Erstellung neuer Datensätze können hier aber auch bereits bestehende Datensätze modifiziert werden. Abbildung 6 zeigt zunächst den Aufbau des Daten-Editors. Er besteht ähnlich einem Spreadsheet aus Zeilen und Spalten. Die einzelnen Zeilen entsprechen dabei den Beobachtungen bzw. Fällen (z.B. Personen, Marken) und die Spalten den Variablen (Merkmalen). In die einzelnen Felder sind für jeden Fall die jeweiligen Meßwerte der entsprechenden Variablen einzugeben. Die Größe des rechteckigen Daten-Tableaus wird folglich durch die Anzahl der Fälle und Variablen bestimmt. So liegen für das Beispiel aus Abbildung 5 für sechs Personen bezüglich der drei Variablen Geschlecht, Größe und Gewicht Meßwerte vor, die in den Daten-Editor eingegeben werden können. Neben dem Eingabefeld enthält der Daten-Editor auch eine Menüleiste mit den Optionen "Datei", "Bearbeiten", "Ansicht", "Daten", "Transformieren", "Analysieren", "Grafiken", "Extras", "Fenster", "Hilfe", "Datei", "Bearbeiten", "Ansicht" etc. Auf deren Anwendung bzw. Nutzung wird innerhalb der einzelnen Analyseverfahren näher eingegangen.

Abbildung 6: Der Daten-Editor



4.1.2 Erstellung einer neuen Datendatei

4.1.2.1 Variablen definieren

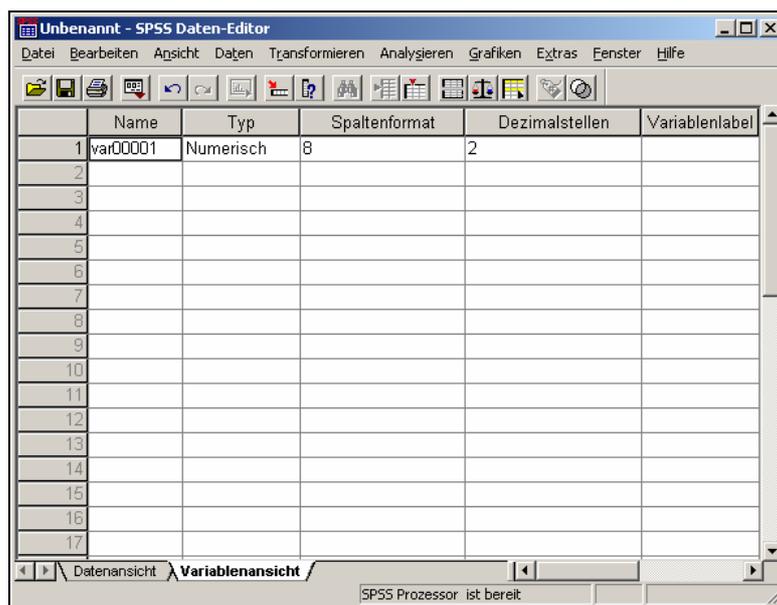
Bevor mit der Eingabe der zu analysierenden Daten in den Daten-Editor begonnen werden kann, ist es in einem ersten Schritt erforderlich, die relevanten Variablen (z.B. Geschlecht, Größe, Gewicht) zu definieren. Der Eintrag "var" in den jeweiligen Spaltenköpfen zeigt zunächst an, daß für die entsprechende Spalte noch keine Variable definiert wurde. Folgende Eigenschaften der Variablen können im Rahmen der Variablendefinition festgelegt werden: Variablenname, Variablentyp, Variablen- und Wertelabels, fehlende Werte, Spaltenformat und Meßniveau.

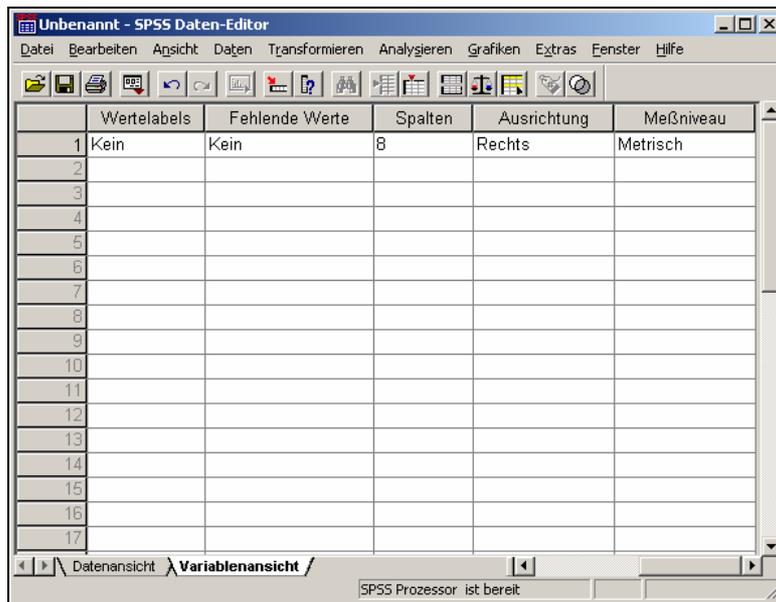
Um eine Variable zu definieren, stehen verschiedene Herangehensweisen alternativ zur Verfügung:

- Aufruf der Option "Ansicht/ Variablen" aus der Menüleiste (oben),
- Doppelklick auf einen mit "var" betitelten Spaltenkopf bzw. auf den entsprechenden Spaltenkopf bei Änderung einer bereits definierten Variable,
- Klick auf die Registerkarte „Variablenansicht“ (links unten).

Durch die alternativen Vorgehensweisen wird das in Abbildung 7 dargestellte Tableau „Variablenansicht“ im SPSS-Daten-Editor geöffnet.

Abbildung 7: Die Variablenansicht





Zunächst werden in der Variablenansicht für eine Variable im Dateneditor die Voreinstellungen angezeigt, die sodann vom Benutzer verändert werden können. So kann im Eingabefeld "Variablenname" der Variablen ein Name zugewiesen werden. Hierbei sind jedoch einige Beschränkungen zu berücksichtigen, wie z.B.:

- Der Name muß mit einem Buchstaben beginnen.
- Der Name darf nicht länger als acht Zeichen sein.
- Der Name kann aus Buchstaben und Ziffern sowie einigen Sonderzeichen (., \$, @, #) gebildet werden.

Aufgrund der Beschränkung bei der Festlegung des Variablennamens ist es in SPSS möglich, jeder Variable noch ein sog. Label, d.h. eine nähere Beschreibung, die maximal 120 Zeichen umfassen kann, zuzuordnen. Diese kann in das Feld "Variablenlabels" eingegeben werden (vgl. Abbildung 7). Über das Feld "Wertelabels" ist es auch möglich, den einzelnen Werten einer Variablen Beschreibungen zuzuordnen. Dies ist insbesondere bei der numerisch kodierten Eingabe von nominalen Variablen nützlich, damit später nachvollzogen werden kann, wie die Kodierung erfolgte (z.B. Geschlecht: 1 = männlich, 2 = weiblich). Hierzu dient das Dialogfenster "Wertelabels definieren", das sich bei Anklicken des Feldes "Wertelabels" öffnet (Abbildung 8).

Abbildung 8: Dialogfenster "Wertelabels definieren:"

Zusätzlich läßt sich in der Variablenansicht (vgl. Abbildung 7) auch das "Meßniveau" der Variable (metrisch, ordinal und nominal) spezifizieren. Voreingestellt (default) ist das Skalenniveau "metrisch". So wäre zum Beispiel für die Variable Geschlecht der Variablenname "geschlec" möglich und als Meßniveau wäre "nominal" zu definieren. Die Variablen Größe ("groesse") und Gewicht ("gewicht") wurden dahingegen auf dem metrischen Skalenniveau gemessen.

Die zu analysierenden Daten weisen häufig sehr unterschiedliche *Variablentypen* auf. So können neben einfachen numerischen Werten z.B. auch Datums- und Währungsformate oder auch Stringformate⁴ vorliegen. Klickt man auf die Schaltfläche "Typ...", wird das Dialogfenster "Variablentyp definieren:" (vgl. Abbildung 9) geöffnet. Hier wird es dem Nutzer ermöglicht, zwischen verschiedenen Variablentypen zu wählen. Je nach gewähltem Typ können für die Variable zusätzlich unterschiedliche Spezifikationen vorgenommen werden. So kann z.B. im Rahmen der Definition eines numerischen Variablentyps (Voreinstellung) zum einen die Anzahl der Zeichen (einschließlich Nachkommastellen und Dezimaltrennzeichen) angegeben werden, die die Werte der Variablen umfassen dürfen (Breite, maximal 40 Zeichen). Zum anderen ist es möglich, die Anzahl der Dezimalstellen (maximal 16) festzulegen. Ähnliche Einstellungen sind auch innerhalb der anderen Variablentypen möglich. Für die drei Variablen Geschlecht, Größe und Gewicht kann die Voreinstellung numerisch beibehalten werden. Lediglich die Anzahl der Dezimalstellen ließe sich hier auf Null herabsetzen (vgl. Daten in Abbildung 5).

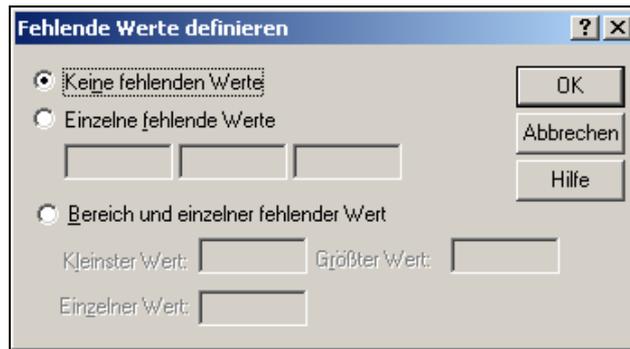
⁴ Stringvariablen arbeiten mit Zeichenketten. Gültige Werte umfassen Buchstaben, Ziffern und Sonderzeichen.

Abbildung 9: Dialogfenster "Variablentyp definieren:"

Ein Problem, das bei der praktischen Anwendung statistischer Methoden häufig auftritt, bilden *fehlende Werte* (*missing values*). Hierbei handelt es sich um Variablenwerte, die von den Befragten entweder außerhalb des zulässigen Beantwortungsintervalls vergeben oder überhaupt nicht eingetragen wurden. So kann zum Beispiel eine "0" für das Gewicht einer Person bedeuten, daß der Wert nicht bekannt ist. Um eine Fehlinterpretation zu vermeiden, kann dies dem SPSS-Programm angezeigt werden. Zunächst kann zwischen zwei Arten von fehlenden Werten unterschieden werden. Werden die Felder im Daten-Editor, für die keine Angaben vorliegen leer gelassen bzw. entspricht der Eintrag nicht dem Variablenformat, erzeugt SPSS automatisch fehlende Werte. Diese werden als *systemdefinierte fehlende Werte* bezeichnet. Für den Nutzer werden diese Werte automatisch durch ein Komma in dem entsprechenden Feld kenntlich gemacht.⁵ Es ist jedoch durch den Nutzer auch möglich, fehlende Werte selbst zu definieren. Zur Festlegung dieser *benutzerdefinierten fehlenden Werte* wird über die Schaltfläche "Fehlende Werte..." das Dialogfenster "Fehlende Werte definieren:" (vgl. Abbildung 10) aufgerufen. Für jede Variable stehen hier drei Optionen zur Festsetzung der fehlenden Werte zur Verfügung:

- keine fehlenden Werte (keine benutzerdefinierten fehlenden Werte),
- einzelne fehlende Werte (Eingabe von bis zu drei einzelnen Werten möglich, die als fehlende Werte behandelt werden sollen),
- Bereich und einzelner Wert (Eingabe eines Wertebereiches für fehlende Werte und eines einzelnen Wertes außerhalb dieses Bereiches, nur für numerische Variablen verfügbar).

⁵ Allerdings gilt dies nicht für String-Variablen, da diese auch einen leeren Eintrag enthalten können.

Abbildung 10: Dialogfenster "Fehlende Werte definieren:"

Die so definierten fehlenden Werte unterliegen im Rahmen der einzelnen Analyseverfahren automatisch einer speziellen Handhabung oder werden von vielen Berechnungen ausgeschlossen. Da in unserem Beispiel sämtliche Variablenwerte vorliegen, kann die Voreinstellung "Keine fehlenden Werte" beibehalten werden.⁶

Schließlich ist es über die Schaltfläche "Spalten" für jede Variable möglich, die Spaltenbreite und über die Schaltfläche "Ausrichtung" die Textausrichtung (Links, Rechts, Mitte) festzulegen.

Über die Schaltfläche "OK" werden die vorgenommenen Einstellungen bezüglich der einzelnen Variablen aktiviert.

4.1.2.2 Dateneingabe

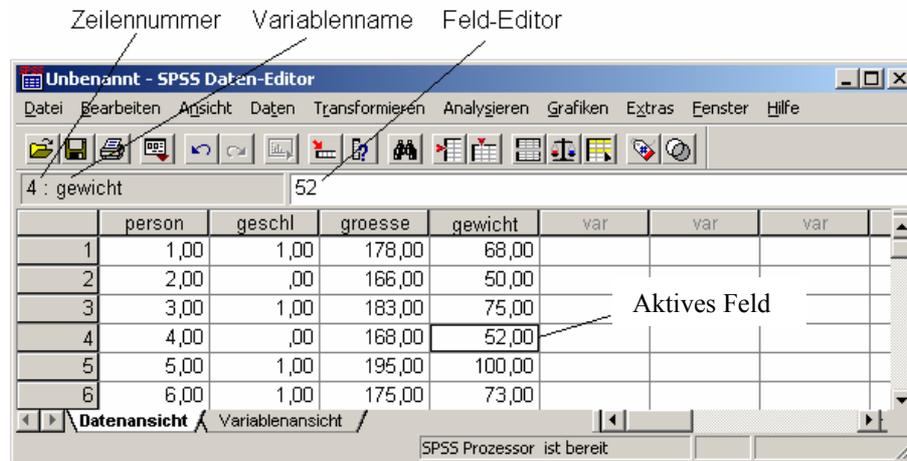
Nachdem die Variablen definiert wurden, können die Daten direkt in den Daten-Editor eingegeben werden. Dabei kann man sowohl fall- als auch variablenweise vorgehen. Das jeweils aktive Feld, in das ein Wert eingegeben werden kann, ist durch eine starke Umrandung hervorgehoben. Die eingegebenen Daten werden allerdings zunächst in die Bearbeitungszeile geschrieben, die sich über den einzelnen Spalten befindet. Weist das aktive Feld bereits einen Eintrag auf, wird in der Bearbeitungszeile auch die entsprechende Zeilennummer und der Variablenname ausgewiesen (vgl. Abbildung 11). Bei der Eingabe der Daten ist jedoch zu beachten, daß nur Werte entsprechend des definierten Variablentyps eingegeben werden können. Das heißt, daß beispielsweise beim Variablentyp "numerisch" keine Buchstaben eingegeben werden können. Die Zulässigkeit überprüft SPSS bereits während der Eingabe, indem unzulässige Zeichen gar nicht erst aufgenommen werden.

Nachdem die neuen Daten in den Daten-Editor eingegeben oder eine bereits bestehende Datei geändert wurde, muß die Datei vor dem Schließen bzw. dem Beenden von SPSS gespeichert werden. Hierzu ist aus dem Menü der Befehl

⁶ Mit dem SPSS-Modul Missing Value Analysis können Muster von fehlenden Daten beschrieben werden. Ebenfalls können Mittelwerte und andere statistische Größen geschätzt sowie Werte für fehlende Beobachtungen ersetzt werden.

"Datei, Speichern unter..." auszuwählen. Es wird die Dialogbox "Daten speichern unter" geöffnet, über die die Datei unter Angabe eines Dateinamens gespeichert werden kann. Die für Datendateien erforderliche Erweiterung .sav wird von SPSS automatisch vorgegeben.

Abbildung 11: Aufbau des Daten-Editors



4.2 Einfache Statistiken und Grafiken

Wurden die Daten in den Daten-Editor eingegeben, ist es in der Regel sinnvoll, nicht sofort mit umfangreichen näheren Analysen zu beginnen, sondern zunächst die Daten selbst etwas ausführlicher zu betrachten. Somit erlangt man zum einen einen ersten Eindruck von den Daten selbst und kann zum zweiten mögliche Hypothesen über den Zusammenhang zwischen einzelnen Variablen aufstellen. SPSS bietet hier die Möglichkeit, die Daten z.B. durch entsprechende Kennzahlen (Mittelwert, Standardabweichung, Spannweite etc.) zu beschreiben oder ihre Verteilung zu überprüfen (z.B. Darstellung der Verteilung in Form eines Histogrammes, Berechnung von Kurtosis und Schiefe). Diese einfachen Analysen sind insbesondere auch für die Aufdeckung etwaiger Eingabefehler hilfreich. Mittels eines Streudiagrammes ist es beispielsweise aber auch möglich, zwei Variablen gegenüberzustellen, um so eine erste Vermutung über deren Zusammenhang zu erhalten. Im folgenden soll auf einige dieser einfachen Analysen eingegangen werden.

Abbildung 12: Daten-Editor mit Auswahl der Option "Analysieren/Deskriptive Statistiken/Häufigkeiten"

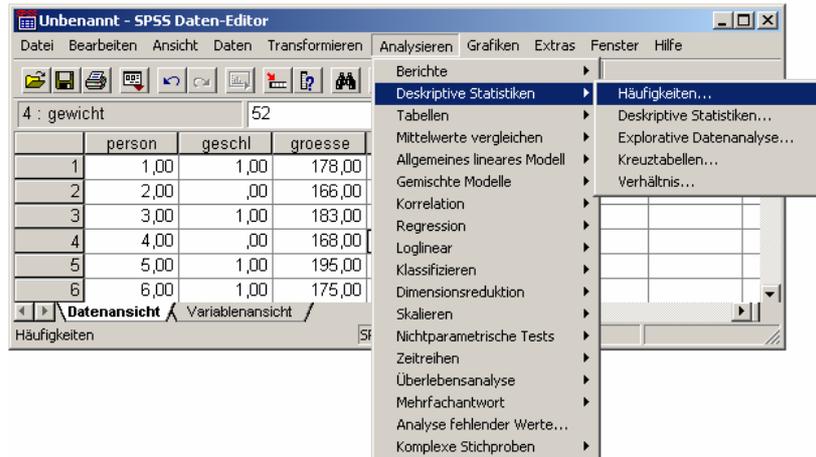
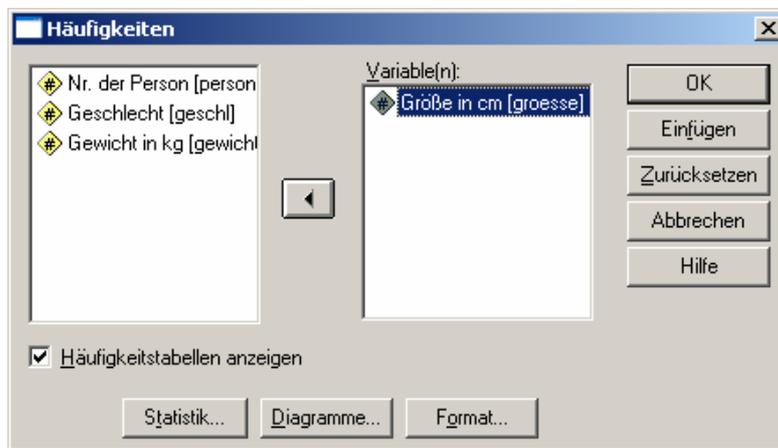


Abbildung 13: Dialogfenster "Häufigkeiten"

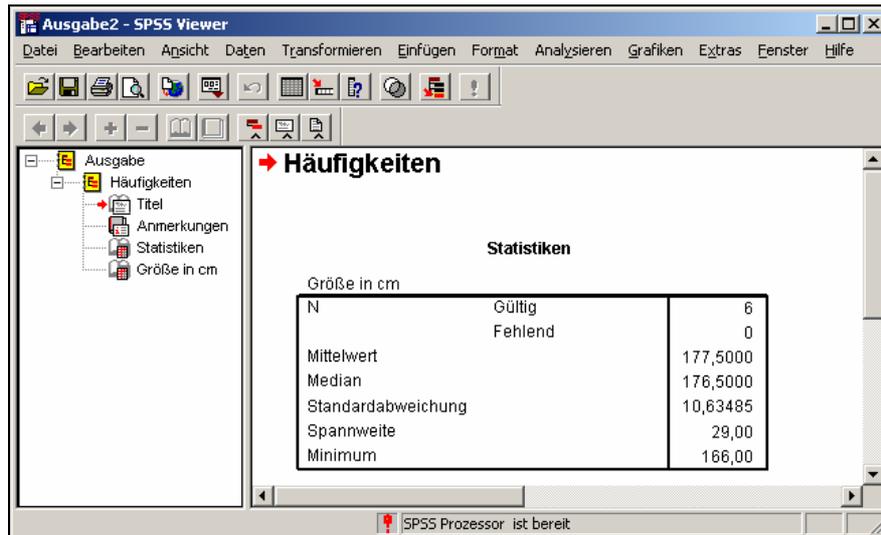


Unter dem Menüpunkt "Analysieren/ Deskriptive Statistiken/ Häufigkeiten" (vgl. Abbildung 12) ist es möglich, ein Dialogfenster aufzurufen, daß die Optionen bietet, zum einen verschiedene statistische Kennzahlen zu berechnen und zum anderen die Häufigkeitsverteilung tabellarisch und grafisch darzustellen.⁷ Um diese

⁷ Die statistischen Kennzahlen lassen sich aber auch unter den Menüoptionen "Deskriptive Statistiken" und "Explorative Datenanalyse" berechnen und unter dem Menüpunkt "Explorative Datenanalyse" ist es ebenso möglich, zur grafischen Veranschaulichung der Häufigkeitsverteilung das Histogramm zu wählen.

Auswertungen zu berechnen bzw. anzuzeigen, sind in dem Dialogfenster "Häufigkeiten" (vgl. Abbildung 13) zunächst aus der linken Quellvariablenliste die relevanten Variablen auszuwählen und über den Variablen-Selektionsschalter (kleine Pfeil-Schaltfläche) in die nebenstehende Wahlvariablenliste zu übertragen. Abbildung 13 verdeutlicht dies am Beispiel der Variable "Größe". Im folgenden Schritt können dann über die entsprechenden Schaltflächen "Statistik..." und "Diagramme..." weitere Dialogfenster aufgerufen werden, die es ermöglichen, die erforderlichen statistischen Kennzahlen bzw. grafischen Darstellungen für die selektierten Variablen optional auszuwählen. Nach Auswahl der gewünschten Optionen ist auf "OK" zu klicken.

Abbildung 14: Ausgabedatei

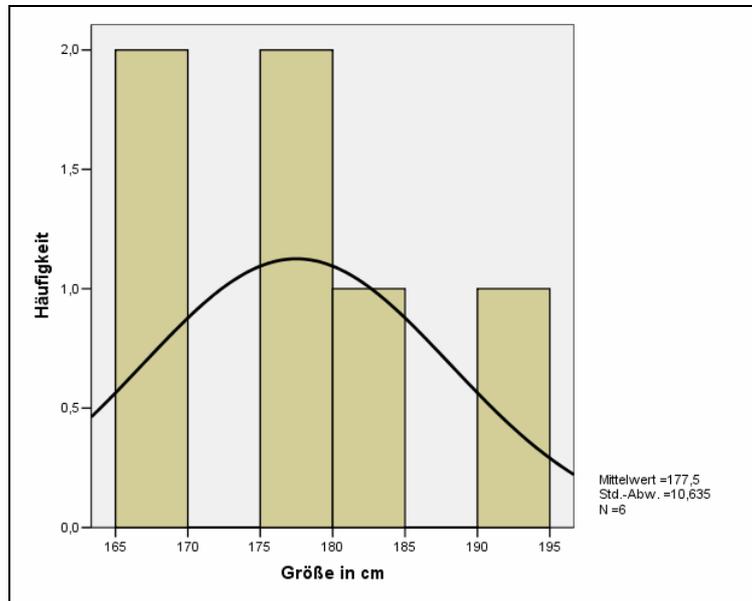


Die Ergebnisse dieser Analysen werden von SPSS automatisch in eine gesonderte Ausgabedatei (Viewer) geschrieben, die man bei Bedarf ausdrucken kann. Wie Abbildung 14 verdeutlicht, unterteilt sich diese Ausgabedatei in zwei Fenster. Das linke Fenster enthält einen Überblick über die Inhalte des Outputs und im rechten werden statistische Tabellen, Diagramme und sonstige Texte (z.B. auch Fehlermeldungen) ausgewiesen.

Abbildung 14 enthält bereits für die Variable "Größe" einige statistische Kennzahlen (Mittelwert, Median, Standardabweichung, Spannweite, Minimum), die nach der dargestellten Vorgehensweise optional ausgewählt wurden. Wie erwähnt, können neben diesen Statistiken aber auch Diagramme ausgegeben werden, wie zum Beispiel ein Histogramm für die Variable "Größe" (vgl. Abbildung 15). Diese Darstellung verdeutlicht, daß zwei Personen eine Größe im Bereich von 165 bis 174 cm aufweisen, drei Personen im Bereich von 175 bis 184 cm liegen und eine

Person zwischen 195 und 204 cm groß ist. Sämtliche Tabellen, Grafiken etc. lassen sich in der Ausgabedatei auch weiter bearbeiten. Die Ausgabedatei selbst kann unter der Erweiterung .spo abgespeichert werden.

Abbildung 15: Histogramm mit Normalverteilungskurve



Um für zwei (oder drei) metrische Variablen die gemeinsame Verteilung darstellen und somit auch einen ersten Einblick in deren möglichen Zusammenhang zu erhalten, bietet es sich an, diese Variablen in einem Streudiagramm abzubilden. Hierzu ist aus dem Menü der Befehl "Grafik/Streudiagramm..." auszuwählen, wodurch das Dialogfenster "Streu-/Punktdiagramm" (vgl. Abbildung 16) geöffnet wird.⁸

⁸ Dieser Befehl läßt sich sowohl im Daten-Editor als auch in der Ausgabedatei (Viewer) aufrufen.

Abbildung 16: Dialogfenster "Streu-/Punktdiagramm"

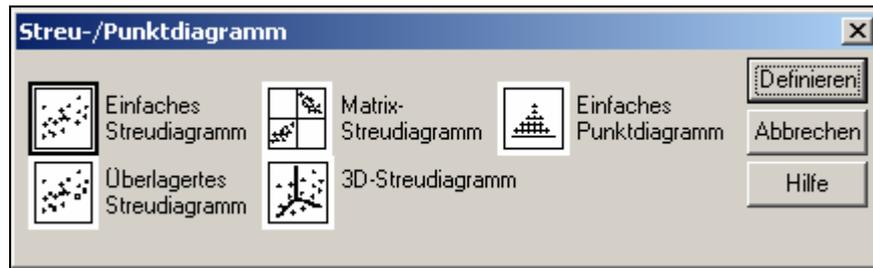
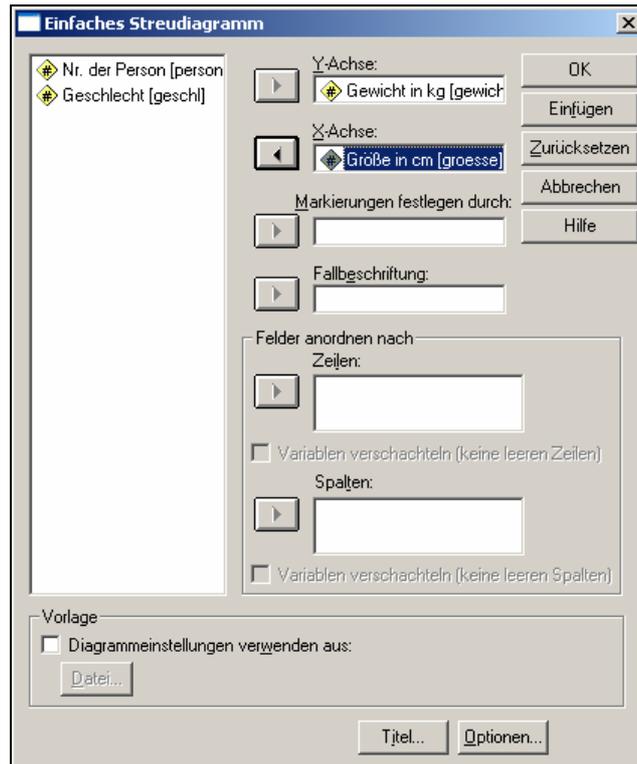


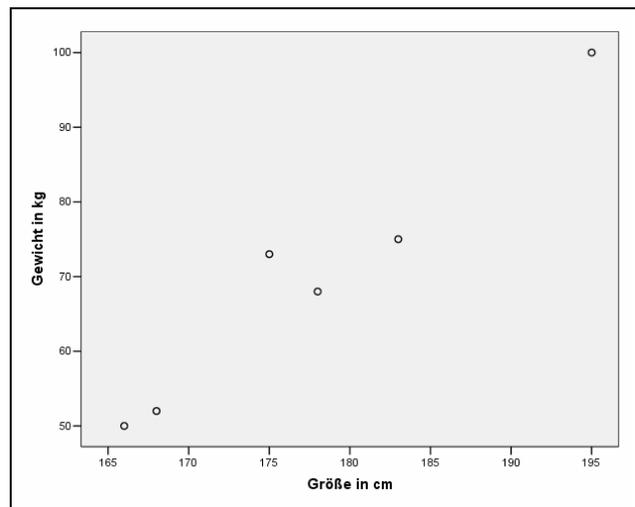
Abbildung 17: Dialogfenster "Einfaches Streudiagramm"



Um die gemeinsame Verteilung zweier Variablen darzustellen, ist das einfache Streudiagramm zu definieren. Hierzu sind aus der Variablenliste des Dialogfensters "Einfaches Streudiagramm" (vgl. Abbildung 17) die entsprechenden Variablen (hier z.B. Gewicht und Größe) auszuwählen und der Y- bzw. X-Achse zuzuordnen.

Im Ergebnis liefert SPSS in der Ausgabedatei ein Streudiagramm entsprechend Abbildung 18. Hier wird jedes Wertepaar durch ein Quadrat angezeigt.⁹ Wie das Streudiagramm verdeutlicht, besteht zwischen den Variablen Gewicht und Größe scheinbar ein positiver Zusammenhang. Das heißt, daß mit zunehmender Größe auch das Gewicht zunimmt. Gestützt wird dieser vermutete Zusammenhang auch durch den Korrelationskoeffizienten, der sich durch SPSS ebenfalls leicht berechnen läßt (Menü: "Analysieren/ Korrelation..."). Dieser liegt in diesem Fall bei 0,975. Dieser Zusammenhang läßt sich noch deutlicher erkennen, wenn in die Grafik eine Regressionsgerade (siehe zur Regression ausführlich Kapitel 1) eingefügt wird. Dabei ist wie folgt vorzugehen: Durch einen Doppelklick auf das Streudiagramm wird ein neues Fenster geöffnet, der Diagramm-Editor. In diesem Editor ist es möglich, das Diagramm weiter zu bearbeiten.

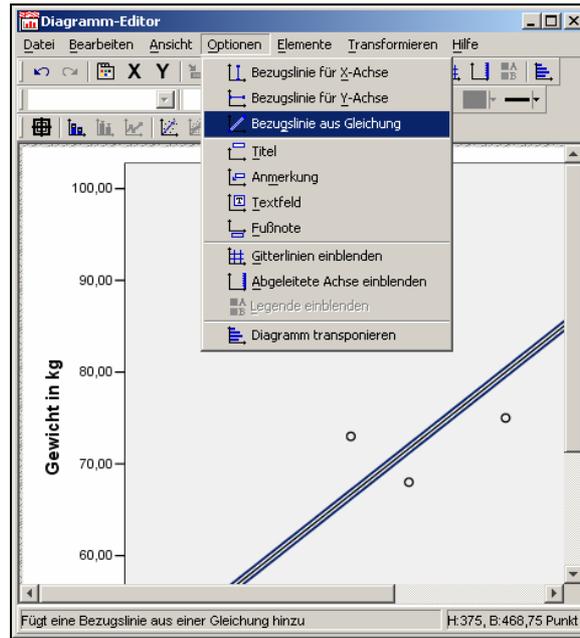
Abbildung 18: Einfaches Streudiagramm für die Variablen Größe und Gewicht



Zum Einfügen der Regressionsgeraden ist aus dem Menü der Befehl "Diagramme/ Optionen..." der Punkt "Bezugslinie aus Gleichung" zu wählen. Die dadurch in die Grafik eingefügte Regressionsgerade (vgl. Abbildung 20) bestätigt die Vermutung aus dem einfachen Scatterplot.

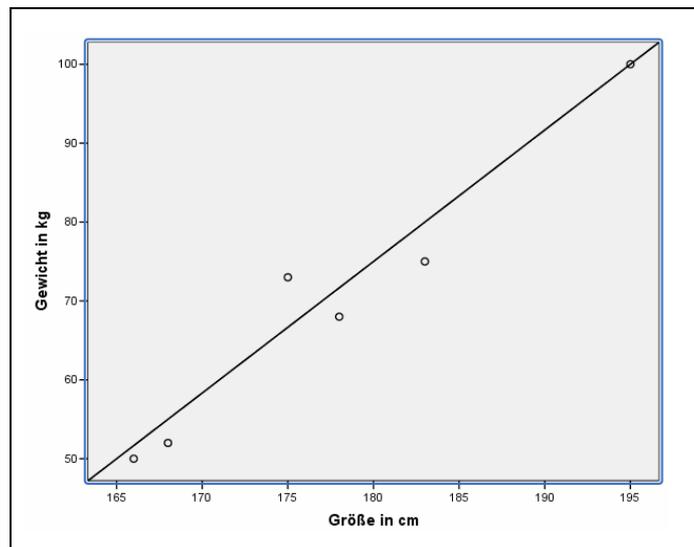
⁹ Zusätzlich wäre es möglich, diese Markierungen zum Beispiel durch die Variable Geschlecht festzulegen (vgl. Abbildung 17). Im Output erscheint die Markierung dann je nach Ausprägung des Geschlechtes in einer anderen Farbe, so daß die einzelnen Wertepaare zugeordnet werden können.

Abbildung 19: Dialogfenster "Optionen für Streudiagramme"



Die einzelnen Wertepaare weisen nur sehr geringe Abweichungen von der Geraden auf.

Abbildung 20: Einfaches Streudiagramm mit linearer Regressionskurve



4.3 Die Kommandosprache

Das Programmsystem SPSS existiert in unterschiedlichen Versionen für PC und Großrechner. Allen Versionen liegt eine gemeinsame Kommandosprache zugrunde. Auf diese wird auch von der grafischen Benutzeroberfläche von SPSS für Windows zugegriffen, d.h. wenn der Benutzer über die Dialogfelder des Programmes Befehle auswählt, werden diese automatisch in die Kommandosprache übersetzt und in eine Syntaxdatei geschrieben. Es handelt sich dabei um eine einfache Textdatei, die gelesen und bearbeitet werden kann. Alternativ kann man aber auch direkt eine Syntaxdatei erstellen und damit den Programmablauf steuern.

Wenngleich sich mit *SPSS für Windows* auch ohne Kenntnis der Kommandosprache arbeiten läßt, so ist es doch vorteilhaft, einige Grundkenntnisse hierüber zu haben. Zum einen sind einige Funktionen von *SPSS für Windows* nur über die Kommandosprache zugänglich und zum anderen ist es bei komplexeren Problemen von Vorteil, mit Syntaxdateien zu arbeiten. Die Erstellung einer Syntaxdatei wird bei Verwendung der Windows-Version dem Anwender sehr erleichtert, indem ihm die beim Dialogbetrieb intern erzeugte Kommandosequenz über ein Dialogfenster zugänglich gemacht wird. Dort kann er sie wie einen Text weiterbearbeiten und sodann erneut starten. Bei Bedarf kann er sie in einer Datei abspeichern, auf die sich später wieder zugreifen läßt. Hierauf wird aber später noch etwas ausführlicher eingegangen.

4.3.1 Aufbau einer Syntaxdatei

Abbildung 21 zeigt ein Beispiel für eine Syntaxdatei. Neben den Syntaxkommandos enthält diese Datei auch den Datensatz aus Abbildung 5.

Die Syntaxdatei gliedert sich in zwei Teile:

- Datendefinition
- Prozedur (Datenanalyse).

Die Datendefinition beschreibt die Daten und kann auch, wie in der Syntaxdatei in Abbildung 21, die Daten selbst enthalten. Bei größeren Datensätzen kann es dagegen vorteilhaft sein, diese in einer separaten Datei abzulegen. In diesem Fall ist in der Datendefinition der Name der betreffenden Datendatei anzugeben.

Der Prozedurteil weist SPSS an, welche Analysen mit den Daten vorzunehmen sind. Das Kommando DESCRIPTIVES im Beispiel weist SPSS an, für die drei Variablen Geschlecht, Größe und Gewicht einfache Statistiken wie den arithmetischen Mittelwert und die Standardabweichung zu berechnen. Es können beliebig viele Prozedurkommandos folgen. Mittels FREQUENCIES werden die Häufigkeiten der Geschlechter ausgezählt.

Abbildung 21: Beispiel einer Syntaxdatei für SPSS

```

† MVA: Einführung

* Datendefinition

DATA LIST FREE
  /Person Geschl Groesse Gewicht.
VARIABLE LABELS Person "Nr. der Person"
                 /Geschl "Geschlecht"
                 /Groesse "Groesse in cm"
                 /Gewicht "Gewicht in kg".
VALUE LABELS Geschl 0 "weiblich"
              1 "maennlich".

BEGIN DATA
  1 1 178 68
  2 0 166 50
  3 1 183 75
  4 0 168 52
  5 1 195 100
  6 1 175 73
END DATA.

* Prozeduren

DESCRIPTIVES VARIABLES = Geschl Groesse Gewicht.

FREQUENCIES VARIABLES = Geschl
  /HISTOGRAM.

```

4.3.2 Syntax der Kommandos

Die Kommandos entsprechen den Sätzen einer Sprache. Sie sind nach einfachen syntaktischen Regeln aufgebaut.

Ein *Kommando* besteht aus einem

- *Schlüsselwort* (keyword), das gleichzeitig auch den Namen des Kommandos bildet (z.B. TITLE, DATA LIST oder DESCRIPTIVES) und
- *Spezifikationen*, die zusätzliche Informationen enthalten.

Spezifikationen können folgende Elemente enthalten:

- Schlüsselwörter, z.B. FREE oder VARIABLES,
- Namen, z.B. Person oder Geschl,
- Zahlen, z.B. Daten oder Parameter,
- sonstige Zeichenketten (Strings), die durch Hochkommata oder Anführungszeichen eingeschlossen sein müssen, z.B. Titel oder Labels.

Beispiel: DATA LIST-Kommando

Kommando	Spezifikation
DATA LIST	FREE / Person Geschl Groesse Gewicht.

Schlüsselwörter sind hier DATA LIST und FREE.

Spezifikationen bilden hier die Formatangabe FREE und die Variablenliste mit den Namen der Variablen. Mehrere Spezifikationen sind durch Schrägstrich (/) zu trennen.

Zur Unterscheidung von Namen und Strings werden hier Schlüsselwörter mit Großbuchstaben geschrieben. SPSS unterscheidet dagegen nicht zwischen Klein- und Großbuchstaben.

Ein Kommando kann auch *Unterkommandos* enthalten, die ebenso aufgebaut sind. Wie alle Kommandos beginnen auch Unterkommandos mit einem Schlüsselwort, das gleichzeitig dessen Namen bildet. Kommandos wie Unterkommandos können Spezifikationen enthalten, müssen es aber nicht. Z.B. ist HISTOGRAM ein Unterkommando des Kommandos FREQUENCIES. Es erzeugt eine Darstellung der Häufigkeitsverteilung, die durch FREQUENCIES ermittelt wird. Mehrere Unterkommandos sind durch Schrägstrich (/) zu trennen. Falls das Unterkommando Spezifikationen umfaßt, so sind diese durch das Gleichheitszeichen (=) vom Kommando-Schlüsselwort zu trennen (z.B. VARIABLES = Geschl).

Ein Kommando kann beliebig viele Zeilen umfassen. Es muß aber immer in einer neuen Zeile begonnen und durch einen Punkt (.) abgeschlossen werden. Alternativ kann auch eine Leerzeile angehängt werden. Leerzeichen innerhalb eines Kommandos werden vom Programm überlesen.

Neben den Kommandos kann eine Syntaxdatei auch Kommentarzeilen enthalten, die durch einen Stern (*) einzuleiten sind. Sie dienen der besseren Lesbarkeit der Syntaxdatei. Ein Kommentar kann auch mehrere Zeilen umfassen, wobei Fortsetzungszeilen ebenfalls durch einen Stern einzuleiten oder um wenigstens eine Spalte einzurücken sind.

Die *SPSS-Kommandos* lassen sich grob in drei Gruppen einteilen:

- Kommandos zur Datendefinition (z.B. DATA LIST, VALUE LABELS),
- Prozedurkommandos (z.B. DESCRIPTIVES, REGRESSION),
- Hilfskommandos (z.B. TITLE).

4.3.3 Kommandos zur Datendefinition

Durch das Kommando DATA LIST wird dem SPSS-Programm mitgeteilt, wo die Eingabedaten stehen und wie sie formatiert sind. Falls die Eingabedaten nicht, wie hier im Beispiel, in der Syntaxdatei stehen, könnte hier der Name der Datendatei angegeben werden.

Der Parameter FREE besagt, daß die Eingabedaten formatfrei (freefield) zu lesen sind. Erforderlich ist hierfür, daß die Zahlen durch Leerzeichen (blanks) oder Kommata voneinander getrennt stehen. Wenn den Variablen feste Spalten zugewiesen werden sollen, ist der Parameter FIXED zu verwenden. In diesem Fall ist kein Trennzeichen zwischen den Variablenwerten erforderlich.

Mittels der folgenden Liste von Variablennamen wird angezeigt, wieviele Variablen der Datensatz enthält. Ein Variablenname darf maximal 8 Zeichen umfassen, von denen das erste Zeichen ein Buchstabe sein muß. Falls das Datenformat FIXED spezifiziert wurde, muß hinter jedem Namen angegeben werden, welche Spalten die betreffende Variable belegt.

Mit dem Kommando VALUE LABELS können den Werten einer Variablen Beschreibungen zugeordnet werden, um so den Ausdruck besser lesbar zu machen. Die Labels sollten nicht mehr als 20 Zeichen umfassen und müssen durch Hochkommata oder Anführungsstriche eingeschlossen sein.

Ein ähnliches Kommando ist VARIABLE LABELS, mit dem den Variablen bei Bedarf erweiterte Bezeichnungen oder Beschreibungen (bis zu 120 Zeichen) zugeordnet werden können.

Die Kommandos BEGIN DATA und END DATA zeigen Beginn und Ende der Daten an. Sie müssen unmittelbar vor der ersten und nach der letzten Datenzeile stehen. Die Daten lassen sich auch als eine Spezifikation von BEGIN DATA auffassen.

Ein Problem, das bei der praktischen Anwendung statistischer Methoden häufig auftaucht, bilden *fehlende Werte*. So bedeutet im Beispiel die "0" für das Gewicht von Person 4, daß der Wert nicht bekannt ist. Um eine Fehlinterpretation zu vermeiden, kann dies dem Programm durch das folgende Kommando angezeigt werden:

```
MISSING VALUE Gewicht (0).
```

Der fehlende Wert, für den hier die "0" steht, wird dann bei den Durchführungen von Rechenoperationen gesondert behandelt.

Neben derartigen *vom Benutzer spezifizierten fehlenden Werten* (User-Missing Values) setzt SPSS auch *automatisch fehlende Werte* (System-Missing Values) ein, wenn im Datensatz anstelle einer Zahl ein Leerfeld oder eine sonstige Zeichenfolge steht. Automatisch fehlende Werte werden bei der Ausgabe durch einen Punkt (.) gekennzeichnet. Generell aber ist es von Vorteil, wenn der Benutzer fehlende Werte durch das MISSING VALUE-Kommando spezifiziert.

4.3.4 Prozedurkommandos

Prozedurkommandos sind im Sprachgebrauch von SPSS alle Kommandos, die "etwas mit den Daten machen", z.B. sie einlesen, verarbeiten oder ausgeben. Die Kommandos zur Datendefinition (oder auch Transformationen) werden erst dann wirksam, wenn ein Prozedurkommando das Einlesen der Daten auslöst. Der Großteil der Prozedurkommandos betrifft die statistischen Prozeduren von SPSS. Eine Ausnahme ist z.B. das Kommando LIST, mit dem sich die Daten in das Ausgabeprotokoll schreiben lassen.

Durch Prozedurkommandos wird SPSS mitgeteilt, welche statistischen Analysen mit den zuvor definierten Daten durchgeführt werden sollen. So lassen sich z.B. mit dem Kommando DESCRIPTIVES einfache Statistiken wie Mittelwert und Standardabweichung berechnen oder mit dem Kommando REGRESSION eine multiple Regressionsanalyse durchführen. Weitere Kommandos zur Durchführung multivariater Analysen sind z.B. ANOVA, DISCRIMINANT, FACTOR oder CLUSTER. Sie werden im Zusammenhang mit der Darstellung der Verfahren in den jeweiligen Kapiteln dieses Buches erläutert.

Eine Syntaxdatei kann beliebig viele Prozedurkommandos enthalten. Die Prozedurkommandos sind z.T. sehr komplex und können eine große Zahl von Unterkommandos (subcommands) umfassen.

Viele Kommandos wie auch Unterkommandos besitzen hinsichtlich ihrer möglichen Spezifikationen *Voreinstellungen (defaults)*, die zur Anwendung kommen, wenn durch den Benutzer keine Spezifikation erfolgt. Die Voreinstellungen von Unterkommandos treten z.T. auch in Kraft, wenn das Unterkommando selbst nicht angegeben wird. So wurde hier bei den Prozeduren DESCRIPTIVES und FREQUENCIES jeweils auf Angabe des Unterkommandos STATISTICS verzichtet, mit Hilfe dessen sich steuern läßt, welche statistischen Maße berechnet und ausgegeben werden sollen.

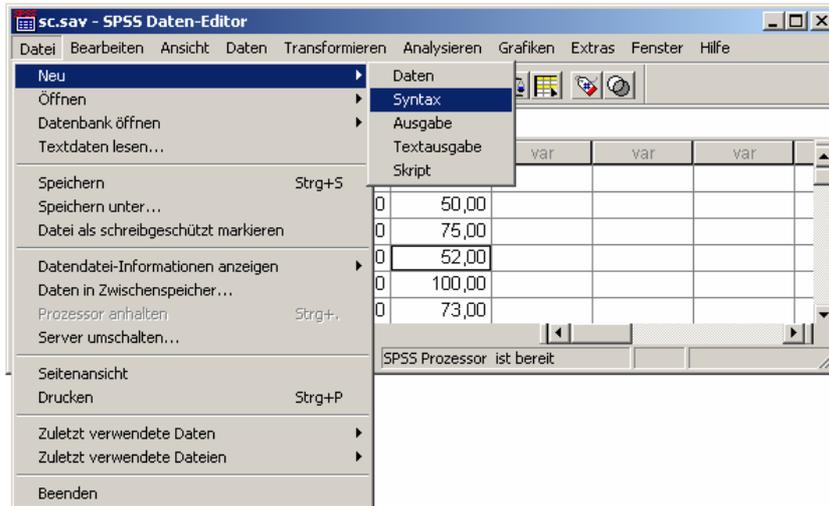
4.3.5 Hilfskommandos

SPSS kennt eine Vielzahl weiterer Kommandos, die weder die Datendefinition noch die Datenanalyse betreffen und die hier der Einfachheit halber als Hilfskommandos bezeichnet werden. Hierunter fallen z.B. die Kommandos TITLE und SUBTITLE, mit denen sich Seitenüberschriften spezifizieren lassen. Weitere Hilfskommandos, die SPSS anbietet, dienen z.B. zur Steuerung der Ausgabe oder zur Selektion, Gewichtung, Sortierung und Transformation von Daten.

4.3.6 Erstellen, Öffnen und Speichern einer Syntaxdatei

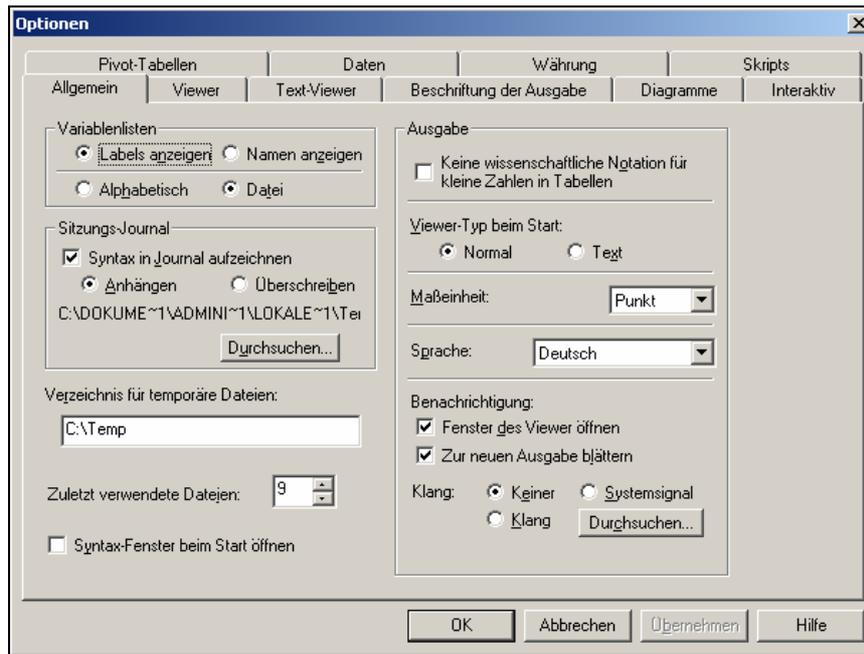
Um eine neue Syntaxdatei zu erstellen, stehen zwei alternative Vorgehensweisen zur Verfügung. Zum einen kann eine neue leere Syntaxdatei nach dem Start von SPSS geöffnet werden. Hierzu ist aus dem Menüpunkt "Datei/ Neu" die Option "Syntax" zu wählen (vgl. Abbildung 22).

Abbildung 22: Erstellung einer neuen Syntaxdatei nach dem Start von SPSS



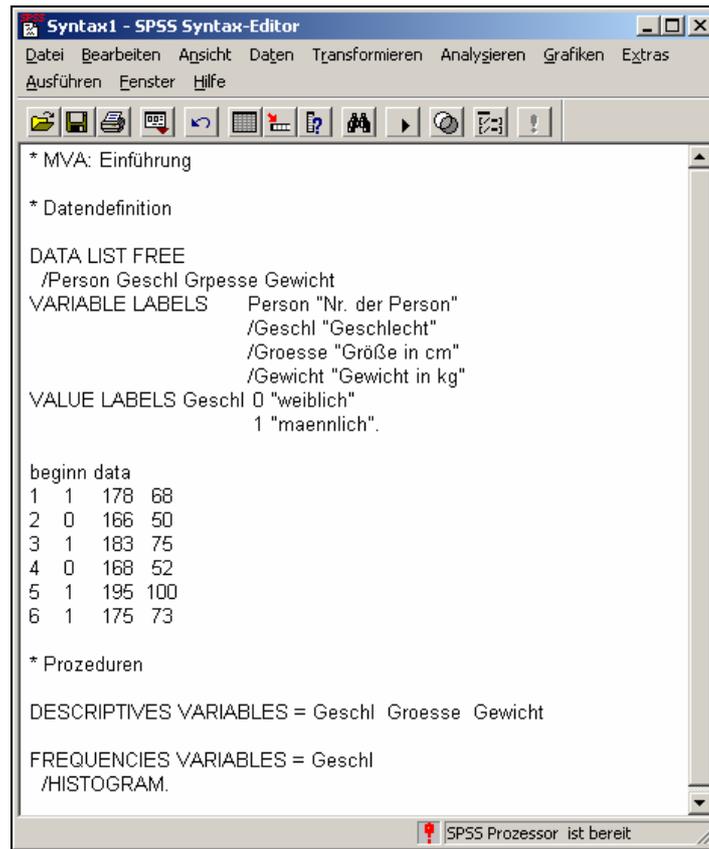
Andererseits ist es möglich festzulegen, daß bei jedem Programmstart von SPSS automatisch eine neue Syntaxdatei geöffnet wird. Hierzu ist zunächst aus dem Menüpunkt "Bearbeiten" der Befehl "Optionen" aufzurufen. Aus dem nunmehr geöffneten Dialogfenster "Optionen" (vgl. Abbildung 23) ist im weiteren die Registerkarte "Allgemein" auszuwählen. Durch Aktivierung der Option "Syntax-Fenster beim Start öffnen" wird bei jedem Start von SPSS automatisch eine neue Syntaxdatei erstellt.

Abbildung 23: Dialogfenster "Optionen/Allgemein"



Neben der Erstellung neuer Syntaxdateien können natürlich auch bereits bestehende während einer SPSS-Sitzung geöffnet werden. Über den Menüpunkt "Datei/ Öffnen" wird hierzu das Dialogfenster "Datei öffnen" aufgerufen. In diesem kann dann die zu öffnende Syntaxdatei ausgewählt werden, wobei zu beachten ist, daß die Syntaxdateien standardmäßig mit der Extension ".sps" versehen sind. Der Inhalt der Syntaxdatei erscheint dann im Syntax-Editor (vgl. Abbildung 24). Beim Speichern einer Syntaxdatei (über den Menüpunkt "Datei/Speichern unter..." bzw. "Datei/ Speichern") wird die Extension .sps automatisch vergeben. Hier ist lediglich der bei einer neuen Syntaxdatei von SPSS automatisch gebildete Dateiname sinnvollerweise zu ändern bzw. bei Bedarf auch der Dateiname einer bestehenden Datei zu variieren.

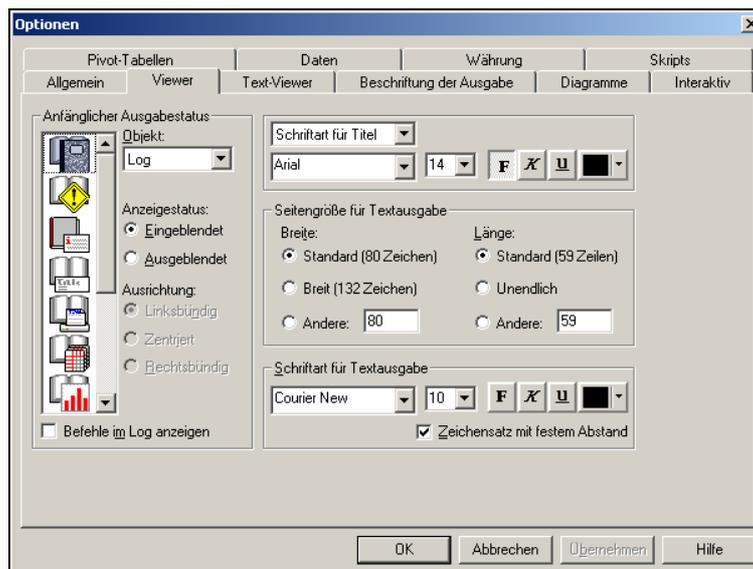
Abbildung 24: Syntax-Editor von SPSS



Neben der direkten Erstellung einer Syntaxdatei, d.h. der manuellen Eingabe der Kommandos durch den Nutzer, stehen auch die folgenden Methoden zur Verfügung, um automatisch eine Syntaxdatei zu erzeugen: Zum einen ist es möglich, die Syntax über die Dialogfenster der jeweils aktuellen Analyse in den Syntax-Editor einzufügen. Hierzu ist in dem jeweiligen Dialogfenster die Schaltfläche "Einfügen" zu aktivieren (vgl. Abbildung 25). Die Syntax wird dann automatisch in den geöffneten Syntax-Editor geschrieben, bzw. es wird automatisch ein neuer Syntax-Editor geöffnet, in den die jeweiligen der Analyse zugrundeliegenden Kommandos eingefügt werden.

Abbildung 25: Übertragung der Syntaxkommandos aus dem Dialogfenster

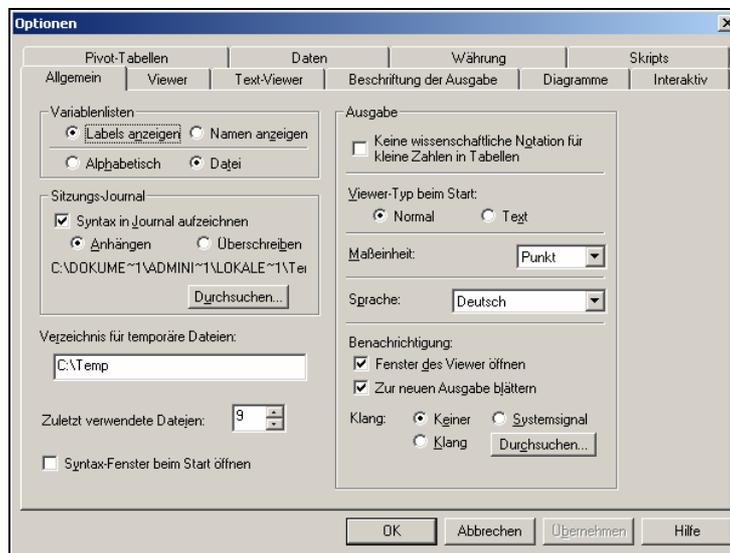
Eine zweite Möglichkeit besteht darin, die Syntax aus dem SPSS-Log der Ausgabedatei manuell in den Syntax-Editor zu kopieren. Damit sämtliche Befehle zunächst in den Log der Ausgabedatei geschrieben werden, ist es vor der Durchführung von Analysen erforderlich, die Option "Befehle im Log anzeigen" auszuwählen. Diese Option ist in der Registerkarte "Viewer" des Dialogfensters "Bearbeiten/Optionen" zu finden (vgl. Abbildung 26).

Abbildung 26: Dialogfenster "Optionen/Viewer"

Werden dann Analysen über die Dialogfenster durchgeführt, werden die entsprechenden SPSS-Kommandos automatisch zusammen mit dem Output im Ausgabefenster angezeigt. Von hier können dann die Kommandos in eine Syntaxdatei manuell kopiert werden.

Letztlich ist es möglich, sämtliche Kommandos in einer Journaldatei zu speichern. Hierbei handelt es sich ebenfalls um eine Textdatei, die auch bearbeitet werden kann. Sie besitzt zwar die Extension .jnl, kann aber als Syntaxdatei (.sps) gespeichert werden, die dann wiederholt zu Datenanalysen verwendet werden kann. Damit das Sitzungsjournal erstellt wird, ist die Option "Befehlssyntax in Journaldatei aufzeichnen" in der Karte "Allgemein" des Dialogfensters "Bearbeiten/ Optionen" zu aktivieren (vgl. Abbildung 27). Per Voreinstellung wird dieses Journal im Verzeichnis C:\Temp\spss.jnl gespeichert. Diese Einstellung kann aber auch variiert werden, d.h. es kann ein anderes Verzeichnis angegeben werden. Hierzu ist über die Schaltfläche "Durchsuchen" ein entsprechender Pfad zu wählen. Je nach Einstellung wird die Journaldatei bei jeder SPSS-Sitzung erweitert ("Anhängen") oder überschrieben ("Überschreiben").

Abbildung 27: Dialogfenster "Optionen/Allgemein"



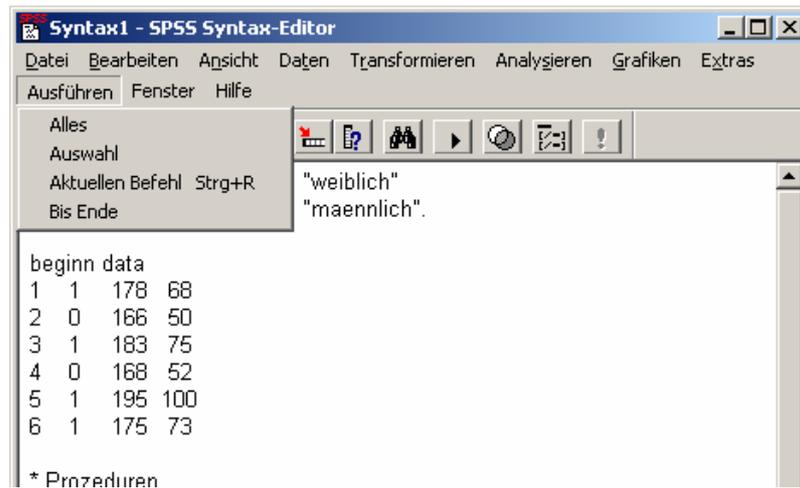
4.3.7 Ausführen der Syntaxdatei

Um eine Syntaxdatei zur Ausführung zu bringen, muß zunächst entsprechend der bereits dargestellten Vorgehensweise nach dem Programmaufruf von SPSS die Syntaxdatei geöffnet werden. Es lassen sich sodann entweder sämtliche Befehle der Datei oder einzelne, unmittelbar aufeinander folgende Befehle ausführen.

Hierzu ist aus dem Menü des Syntax-Editors der Befehl "Ausführen" zu wählen, wobei dieser wie folgt spezifiziert werden kann (vgl. Abbildung 28):

- Alles (Alle Kommandos der Syntaxdatei werden ausgeführt.)
- Auswahl (Nur die markierten Kommandos werden ausgeführt.)
- Aktuellen Befehl (Es werden alle Kommandos ausgeführt, wo sich der Cursor befindet.)
- Bis Ende (Alle Kommandos zwischen der aktuellen Cursorposition und dem Ende der Syntaxdatei werden ausgeführt.)

Abbildung 28: Auswahl der Option "Ausführen" im Syntax-Editor



4.4 Die Systeme von SPSS

Die PC-Versionen von SPSS umfassen jeweils eine Reihe von Modulen oder Systemen, die separat gekauft werden können und für die jeweils getrennte Handbücher existieren. Von den Systemen von *SPSS für Windows* sind für die hier behandelten Verfahren die folgenden relevant:

- Base System
- Advanced Models
- Regression Models
- Conjoint
- AMOS
- Clementine

Mit Ausnahme von AMOS und Clementine laufen alle Systeme unter einer gemeinsamen Benutzeroberfläche. Die folgende Aufstellung in Abbildung 29 zeigt, welche SPSS-Prozeduren für die hier behandelten Methoden benötigt werden und in welchen SPSS-Systemen diese zu finden sind.

Abbildung 29: Synopse der behandelten Methoden und der entsprechenden SPSS-Prozeduren

Methode	SPSS-Prozeduren	in SPSS-System
Regressionsanalyse	REGRESSION	Base
Varianzanalyse	UNIANOVA GLM	Base
Diskriminanzanalyse	DISCRIMINANT	Base
Kontingenzanalyse	CROSSTABS LOGLINEAR HILOGLINEAR	Base Advanced Models Advanced Models
Faktorenanalyse	FACTOR	Base
Strukturgleichungsmodelle		Amos*
Logistische Regression	LOGISTIC REGRESSION NOMREG	Regression Models
Clusteranalyse	CLUSTER QUICK CLUSTER	Base
Conjoint-Analyse	CONJOINT ORTHOPLAN PLANCARDS	Conjoint
Multidimensionale Skalierung	ALSCAL	Base
Korrespondenzanalyse	CORRESPONDENCE	Categories
Neuronale Netze		Clementine*

* Eigenständiges Programm

5 Literaturhinweise

- Bley Müller, J./Gehlert, G./Gülicher, H. (2002): Statistik für Wirtschaftswissenschaftler, 13. Aufl., München.
- Bühl, A./Zöfel, P. (2000): SPSS: Methoden für die Markt- und Meinungsforschung, München.
- Bühl, A./Zöfel, P. (2005): SPSS 12: Einführung in die moderne Datenanalyse unter Windows, 9. Aufl., München.
- Janssen, J./Laatz, W. (2003): Statistische Datenanalyse mit SPSS für Windows, 4. Aufl., Berlin/Heidelberg/New York.
- Mayntz, R./Holm, K./Hübner, P. (1978): Einführung in die Methoden der empirischen Soziologie, 5. Aufl., Opladen.
- Norusis, M. J./SPSS Inc. (2002): SPSS 11.0 guide to data analysis, Chicago.
- SPSS Inc. (1997): SPSS Conjoint 8.0, Chicago.
- SPSS Inc. (1999): SPSS Base 9.0 User's Guide Package, Chicago.
- SPSS Inc. (1999): SPSS Base 9.0 Applications Guide, Chicago.
- SPSS Inc. (2001): SPSS Base 10 Syntax Reference Guide, Chicago.
- SPSS Inc. (2001): SPSS Advanced Models 11.5, Chicago.
- SPSS Inc. (2001): SPSS Regression Models 11.0, Chicago.
- Wittenberg, R. (1998): Grundlagen computerunterstützte Datenanalyse, 2. Aufl., Stuttgart.
- Wittenberg, R./Cramer, H. (2000): Datenanalyse mit SPSS für Windows: mit zahlreichen Tabellen, 2. Aufl., Stuttgart.