4. Kapitel

Regressions- und kanonische Analyse

Werner Schubö, Klaus Haagen und *Walter Oberhofer*

Im Rahmen des allgemeinen linearen Modells ist das Ziel der Regressions- und kanonischen Analyse, Zusammenhänge zwischen quantitativen Variablen zu beschreiben und inferenzstatistisch zu deuten. Die Regressionsanalyse beschäftigt sich mit dem Zusammenhang zwischen einer einfachen Variablen und einer Gruppe von Variablen. Die kanonische Analyse beschreibt hingegen den Zusammenhang zwischen zwei Gruppen von Variablen und kann als multiva-

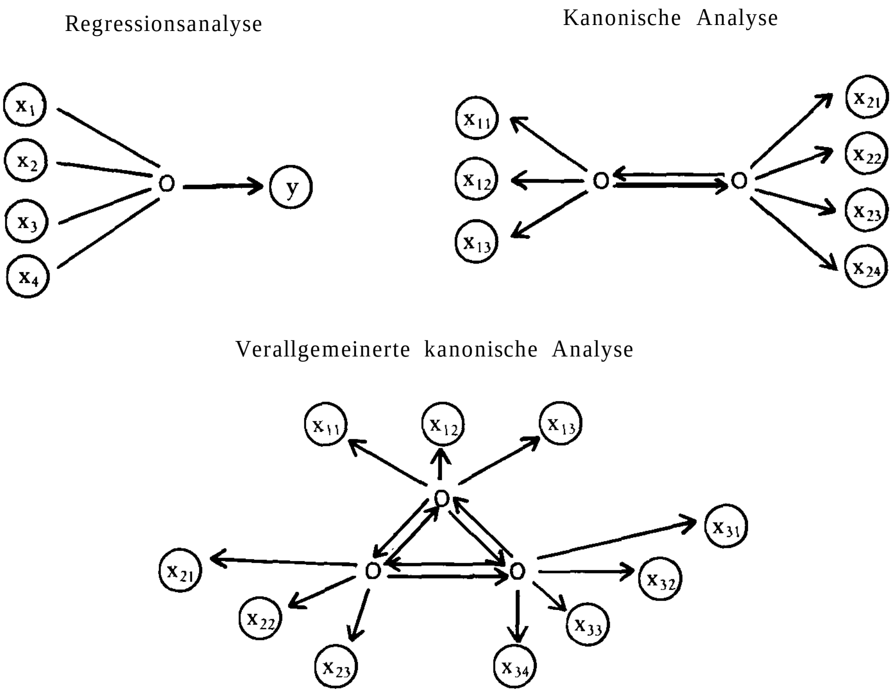

Abb. 1: Die Struktur der Beziehungen bei Regressionsanalyse, kanonischer Analyse und verallgemeinerter kanonischer Analyse.

riate Verallgemeinerung der Regressionsanalyse aufgefaßt werden. Die kanonische Analyse selbst wiederum wird zur kanonischen Analyse für mehrere Variablengruppen verallgemeinert (s. Abb. 1).

Entsprechend der unterschiedlichen Komplexität einzelner Fragestellungen variieren auch die Anforderungen an die Einsatzbereitschaft und die mathematisch-statistischen Vorkenntnisse des Lesers beträchtlich. Da in diesem Beitrag eine systematische Übersicht über das behandelte Gebiet gegeben werden soll, lassen sich schwerer lesbare Passagen nicht vermeiden. Es werden hauptsächlich die Voraussetzungen und Ergebnisse der statistischen Ansätze besprochen; was die mathematischen Ableitungen betrifft, begnügen wir uns mit Literaturhinweisen.

Im Kap. 1 (Regressionsanalyse) wird wohl der Abschnitt 1.8 (Modelle mit Fehlern in den Prädiktoren) am schwierigsten sein. Die Abschnitte 1.10 (Suppression und Kollinearität), 1.11 (Schrittweise Regression) und 1.12 (Teststärke) enthalten nicht nur in der Darstellung, sondern auch im Inhalt bisher nicht beachtete und teils auch nicht bekannte Gesichtspunkte zur Anlage und Interpretation von Regressionsanalysen; sie sind deswegen bewußt besonders leicht lesbar abgefaßt.

Das Kap. 2 (Kanonische Analyse) wird vor allem im Abschnitt 2.6 (Kanonische Analyse für mehrere Variablengruppen) recht komplex. Als Strategie für mathematisch weniger bewanderte Leser sei vorgeschlagen, bei der ersten Lektüre alle nicht unmittelbar verständlichen Formeln zu ignorieren und nur die eingestreuten inhaltlichen Betrachtungen zur Kenntnis zu nehmen. Wie auch in anderen Gebieten der Statistik ist bei diesen Kapiteln nicht zuletzt wegen des mathematischen Aufwandes ein intensives Studium erforderlich, um die hier dargestellten Methoden auch in der Praxis sinnvoll einsetzen zu können.

Zur Notation sei angemerkt, daß Zufallsvariablen und Realisationen von Zufallsvariablen gleich bezeichnet werden, wobei aus der sprachlichen Formulierung des zugehörigen Textes hervorgeht, was gemeint ist. Vektoren werden durch kleine fette Buchstaben, Matrizen durch große fette Buchstaben symbolisiert, während skalare Größen (Merkmale, Zufallsgrößen und Konstanten) im normalen, mageren Druck erscheinen. Tritt eine Matrix oder ein Vektor als Index auf, so unterbleibt der Fettdruck.

1. *Regressionsanalyse*

Zunächst werde die Gliederung des Kapitels erläutert. Ausgangspunkt ist die beschreibende lineare Regression, die unmittelbar für den multiplen Fall mit mehreren Prädiktoren formuliert wird. In Abschnitt 1.2 wird ein allgemeines Modell für nichtstochastische Prädiktoren eingeführt, das den Rahmen für die

Kapitel 1.3 bis 1.6 und 1.12 absteckt. In Kap. 1.3 und 1.4 werden beste Schätzungen für unbekannte Größen des Modells und beste Prognosen für zukünftige Werte im Kriterium angegeben. Dies umfaßt sowohl Punkt- als auch Bereichsschätzungen. In Kap. 1.5 stellen wir statistische Tests zur Prüfung von Hypothesen über das Modell vor. Darauf folgt ein Abschnitt, der die Ridge-Regression als eine Möglichkeit beschreibt, im Falle angenäherter Kollinearität der Prädiktoren die Regressionskoeffizienten des Modells zuverlässiger zu schätzen. Während Kap. 1.2 ein Modell für nichtstochastische Prädiktoren enthält, wird in Kap. 1.7 das entsprechende Modell für Prädiktoren formuliert, die selbst Zufallsvariablen sind. Mit fehlerhaft gemessenen Prädiktoren beschäftigt sich Kap. 1.8. Als eine Anwendung des allgemeinen Ansatzes aus Kap. 1.2 enthält Kap. 1.9 Hinweise auf die regressionsanalytische Behandlung von Zeitreihen. Hauptsächlich praktische Probleme bei der Planung und Interpretation von regressionsanalytischen Studien stehen bei den drei letzten Abschnitten im Vordergrund.

Die hier nicht oder nicht ausführlich behandelten Themen sollen wenigstens durch Literaturhinweise ersetzt werden. Dazu gehören:

Interpretation einer Computerprogramm-Ausgabe (Nie et al., 1975, p. 358; Gaensslen & Schubö, 1976, p. 319; Schuchart-Ficher et al., 1980, p. 49),

Dummy-Variablen als Prädiktoren zur regressionsanalytischen Berechnung einer Varianzanalyse (Cohen & Cohen, 1975, p. 291-424; Gaensslen & Schubö, 1976, p. 147; Sievers, 1977),

Behandlung von Repeated Measurement Designs (Cohen & Cohen, 1975, p. 413; Edwards, 1979, p. 117),

Regression mit qualitativen und quantitativen Variablen zur Skalierung (Young et al., 1976),

Regression auf Polynome und trigonometrische Polynome (Cohen & Cohen, 1975, p.213; Graybill, 1976, p. 302),

Nichtlineare Regression (Marquardt, 1963; Bard, 1974; Milliken & Graybill, 1970),

Monotone Regression (Carroll & Chang, 1964; Kruskal, 1971; de Leeuw, 1977; Sulz, 1980, p. 124),

Missing Data-Behandlung (Timm, 1970; Lösel & Wüstendörfer, 1974; Cohen & Cohen, 1975; Frane, 1976),

Inverse Regression und Kalibrierung (Krutchkoff, 1967; Hoadley, 1970),

Stückweise Regression (McGee & Carleton, 1970; Wainer, 1971),

Konfidenzbereiche für Regressionskoeffizienten von nicht-ellipsoidischer Gestalt (Bowden, 1970),

Eingeschränkter Wertebereich als Nebenbedingung an die Regressionskoeffizienten (Lovell & Prescott, 1970),

Regressionsmodelle für die Sterbetafel-Analyse (Cox, 1972), Kontingenztafelanalyse im linearen Modell nach Grizzle, Starner und Koch und nach Goodman (Küchler, 1979, p. 154-255),

Kovarianzstrukturanalyse und partial least squares (Jöreskog, 1973, 1977, 1978; McDonald, 1974; Lohmöller, 1979),

Pfadanalyse (Seibel & Nygreen, 1972; Vetter, 1972; Hummell & Ziegler, 1976) und

Anwendung der Regressionsanalyse zur Lösung des Interpretationsproblems in der mehrdimensionalen Skalierung und Faktorenanalyse (Gigerenzer, 1981, p. 338).

1.1 Beschreibende lineare Regression

Zunächst soll nur rein beschreibend die lineare Regression in einer Gesamtheit von Beobachtungen dargestellt werden. Es ist dabei nicht beabsichtigt, die Regressionsbeziehung über die vorliegende Gesamtheit hinaus einer Grundgesamtheit oder einem Modell zuzuordnen. Die hier dargestellte deskriptive Regression versteht sich nur als eine lineare Approximation einer Gesamtheit von Beobachtungen.

Wir gehen davon aus, daß von einer Variablen y und von weiteren K Variablen $x_1, x_2, x_3, \ldots, x_K$ jeweils N Messungen vorliegen. Die Variable y nennen wir das *Kriterium* oder auch abhängige *Variable, Regressand* oder *endogene Variable.* Die Variablen $x_1, x_2, x_3, \ldots, x_K$ werden als *Prädiktoren, unabhängige* Variablen, *Regressoren* oder *exogene Variablen* bezeichnet.

In einem Beispiel aus der Wirtschaftspsychologie könnte etwa als Kriterium y die Produktivität oder auch die Dauer des Arbeitsverhältnisses betrachtet werden, deren Zusammenhang mit den Prädiktoren x_k: Arbeitsbedingungen, Zufriedenheit mit dem Vorgesetzten, Bezahlung und Vielfältigkeit der Tätigkeitsbereiche untersucht werden könnte. In einem Therapieexperiment zur Reduktion des Rauchens könnten als Kriterium y der tägliche Zigarettenkonsum und als Prädiktoren x_k die Variablen Arbeits-/Freizeittag, Ausmaß des eingesetzten Treatments und schließlich die fortlaufende Nummer des Tages dienen. Im letztgenannten Beispiel sind die Ergebniseinheiten oder Merkmalsträger die verschiedenen Tage, an denen eine Person beobachtet wurde, während im ersten Beispiel die Variablen jeweils an verschiedenen Beschäftigten in Form einer Querschnittsuntersuchung erhoben werden.

Bei der linearen Regression ist man bestrebt, den Zusammenhang zwischen dem Kriterium y und den Prädiktoren $x_1, x_2, \ldots, x_K$ durch

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_K x_K$$

zu beschreiben. Wenn diese Beziehung zutreffen würde, könnte man mit ihrer Hilfe bei gegebenen Werten der Prädiktoren und bei Kenntnis der für die Gesamtheit aller Beobachtungen konstanten Koeffizienten $\beta_0, \beta_1, \ldots, \beta_K$ eindeutig die Werte des Kriteriums reproduzieren. In der Empirie ist es jedoch im allgemeinen nicht möglich, die Werte des Kriteriums durch eine Funktion so einfacher Bauart zu reproduzieren. Als Ausweg bieten sich daher zwei Möglichkeiten an. Entweder man sucht eine kompliziertere Funktion, die eine perfekte Reproduktion ermöglicht, oder man betrachtet die sich aus $\beta_0 + \beta_1 x_1 + \ldots + \beta_K x_K$ ergebenden Werte als Approximation für das Kriterium y. Wir folgen hier dem zweiten Weg und schreiben für diese Approximation von y

(1) $$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_K x_K.$$

Verwenden wir also (1) als Approximation für das Kriterium, so ergibt sich als *Approximationsfehler (Residuum)*

$$u = y - \hat{y}.$$

Damit läßt sich der Zusammenhang zwischen dem Kriterium y, der Approximation $\hat{y}$ und dem Approximationsfehler u auch schreiben als

(2) $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_K x_K + u.$$

Die Koeffizienten $\beta_0, \beta_1, \ldots, \beta$ sollen nun so bestimmt werden, daß die Approximation von y durch $\hat{y}$ möglichst gut ist, oder, was dasselbe besagt, daß der Approximationsfehler u möglichst klein ist. Diese Aufgabe, deren Lösung im Aufsuchen des Satzes von Koeffizienten besteht, wird in der numerischen Mathematik als Ausgleichung bezeichnet. Eine Voraussetzung zur Lösung ist, daß genauer angegeben wird, was mit „möglichst kleinen Fehlern“ gemeint sein soll.

Man könnte etwa fordern, daß der größte Fehler u in der betrachteten Gesamtheit dem Betrage nach so klein wie möglich sein soll. Dies führt zum Problem der Ausgleichung nach Tschebyscheff, die z.B. in Stiefel (1961, p. 47) auf einfache Weise beschrieben ist. Die Lösung kann durch die Simplex-Methode, einen Algorithmus aus der linearen Programmierung, erfolgen. Dieser Ansatz hat den Nachteil, daß spätestens dann erhebliche Schwierigkeiten auftauchen, wenn inferenzstatistisch von Stichproben auf Grundgesamtheiten geschlossen werden soll.

Die bislang am weitesten verbreitete und in ihren Folgen am besten untersuchte Forderung ist, die Summe der quadrierten Fehler als Maß der Güte einer Approximation heranzuziehen. Diese Forderung wird *Prinzip der kleinsten Quadrate* oder *Minimum-Quadrat-Prinzip (MQ-Prinzip, LS-Prinzip)* genannt und soll auch hier im Vordergrund stehen. Einerseits ist sie rechentechnisch leichter zu handhaben, und andererseits ist sie für inferenzstatistische Schlüsse besser geeignet.

Zur Formulierung des Minimum-Quadrat-Prinzips soll zunächst eine übersichtliche Schreibweise in Form von Vektoren und Matrizen eingeführt werden. Beispielsweise schreiben wir für die Werte der Variablen y bzw. u bzw. x_k in einer (stets als endlich angenommenen) Gesamtheit vom Umfang N

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \text{ bzw. } \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix} \text{ bzw. } \mathbf{x}_k = \begin{bmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ x_{Nk} \end{bmatrix} .$$

Der Satz von Koeffizienten $\beta_0, \beta_1, \ldots, \beta_K$ bildet den Spaltenvektor

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{bmatrix} .$$

Fassen wir die Vektoren x_k als Spalten einer Matrix auf, so erhalten wir die Matrix X der Prädiktorwerte

$$\mathbf{X} = (\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_K) = \begin{bmatrix} x_{10} & x_{11} & \ldots & x_{1K} \\ x_{20} & x_{21} & \ldots & x_{2K} \\ \vdots & \vdots & & \vdots \\ x_{N0} & x_{N1} & \ldots & x_{NK} \end{bmatrix} ,$$

wobei die Variable x_0 als Multiplikator zu $ß_0$ zugunsten einer einheitlichen Schreibweise eingeführt wurde und stets den Wert 1 hat.

Mit dieser Schreibweise wird aus dem linearen Ansatz (2)

(3) $$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}.$$

Aus (1) wird

(4) $$\hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta}.$$

Die Werte der Restvariablen u ergeben sich dementsprechend als

(5) $$\mathbf{u} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}.$$

Das Prinzip der kleinsten Quadrate läßt sich nun als

(6) $$S(\boldsymbol{\beta}) = \mathbf{u}'\mathbf{u} = \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \overset{!}{=} \text{Minimum}$$

formulieren.

$S(\boldsymbol{\beta})$ ist die Summe der quadrierten Werte der Restvariablen. Die Summe hängt von ß ab, was durch die Angabe von ß in Klammern hervorgehoben wird.

Natürlich hängt die Summe auch von den Werten y_n für n = 1, 2,. . .,N und x_{nk} für k = 1, 2,. . .,K und n = 1, 2,. . ., N in der Gesamtheit ab; diese Werte betrachten wir für die Minimierungsaufgabe jedoch als fest, weshalb ihr Einfluß auf die Summe S nicht dokumentiert werden soll.

Es läßt sich zeigen, daß notwendige und hinreichende Bedingungen für die Werte des Vektors ß mit minimaler Fehlerquadratsumme S(ß) die sog. Normalgleichungen

(7) $$\mathbf{X'X\beta = X'y}$$

sind. Daß diese Normalgleichungen für (6) notwendige Bedingungen sind, wird in beinahe allen einschlägigen Lehrbüchern (z.B. Johnston, 1963, p. 11, 109; Draper & Smith, 1966, p. 10, 59; Gaensslen & Schubö, 1976, p. 299; Haagen & Pertler, 1976, p. 70; Hummell & Ziegler, 1976, p.40, 46; Moosbrugger, 1978, p. 51, 59; Edwards, 1979, p. 13, 40; Küchler, 1979, p. 77) gezeigt. Eine vollständige Darstellung erfordert aber auch den Nachweis, daß jede Lösung von (7) auch die Minimum-Quadrat-Forderung (6) erfüllt (siehe z.B. Schönfeld, 1969, p. 30; Wonnacott & Wonnacott, 1970, p. 13; Theil, 1971, p. 34; Rao, 1973, p. 180; Schneeweiß, 1974, p. 43; Graybill, 1976, p. 174, 343; Schach & Schäfer, 1978, p. 12).

Für die Normalgleichungen (7) gibt es stets wenigstens einen Lösungsvektor, dessen Komponenten dann Regressionskoeffizienten heißen. Wenn die Matrix X'X nichtsingulär (invertierbar, von vollem Rang) ist, gibt es genau eine Lösung

(8) $$\mathbf{\beta = X^+y}$$

mit

$$\mathbf{X^+ = (X'X)^{-1}X'}.$$

Dies ist für empirische Daten gewöhnlich der Fall.

Es gibt jedoch zwei Situationen, in denen die Matrix X'X im „wissenschaftlichen Alltag“ singulär werden kann:

1. Die Anzahl der Beobachtungen N ist zu klein; sie ist nicht größer als die Anzahl der Prädiktoren K.
2. Es besteht ein exakter linearer Zusammenhang eines Prädiktors mit anderen Prädiktoren. Wenn etwa eine gemessene Variable x_k konstant ist, besteht ein linearer Zusammenhang zu x_0. Auch wenn x_k der summative Gesamtwert aus den Prädiktoren x_1, x_2 und x_3 ist, besteht eine derartige lineare Abhängigkeit.

Wenn die Matrix X'X singulär ist, gibt es unendlich viele Lösungen der *Nor*malgleichungen, die alle zur gleichen minimalen Quadratsumme S(ß) führen.

Dies kann sich so auswirken, daß ein oder mehrere Regressionskoeffizienten (Anzahl entspricht dem Rangdefekt) frei wählbar sind; in der Regel kann jedoch die Vielfalt der Lösungen nicht so einfach angegeben werden. Wird die Regressionsanalyse durch ein Computerprogramm ausgeführt, äußert sich die Singularität der Matrix durch einen Abbruch der Berechnungen mit der Fehlermeldung „Division durch Null“ oder „Determinante ist Null“, da in den Programmen keine Vorsorge für eine derartige Datenkonstellation getroffen ist. Im allgemeinen wird man dann einige Prädiktoren ausschließen, um wenigstens für eine abgewandelte Regressionsaufgabe eine Lösung zu erhalten. Später werden wir unter den Stichwörtern Kollinearität und Ridge-Regression noch auf ähnliche Situationen zurückkommen, wo X'X nur „beinahe“ singulär ist; ansonsten soll in Zukunft der in diesem Absatz besprochene Fall ausgeschlossen sein.

Verwendet man im linearen Beschreibungsmodell (3) die Lösung (8) für ß aufgrund der Minimum-Quadrat-Forderung (6), lassen sich eine Reihe von Eigenschaften der Restvariablen u ableiten, ohne daß zusätzliche Annahmen notwendig sind.

So sind die Produktsummen der Residuen mit jeder Prädiktorvariablen Null

(9) $$X'u = 0,$$

woraus sich spezieller ergibt, daß der Mittelwert der Fehlervariablen Null ist, und daß die empirischen Kovarianzen zwischen der Fehlervariablen und allen Prädiktorvariablen verschwinden:

(10) $$\bar{u} = 0$$

(11) $$s_{ux_k} = 0 \text{ für alle } k = 1,\ldots,K.$$

Eine weitere Folge von (9) ist

(12) $$\hat{\mathbf{y}}'\mathbf{u} = 0,$$

weshalb die Kovarianz zwischen der Approximation $\hat{\mathbf{y}}$ und der Fehlervariablen Null ist. Weiterhin gilt die wichtige Zerlegung

(13) $$\mathbf{y}'\mathbf{y} = \hat{\mathbf{y}}'\hat{\mathbf{y}} + \mathbf{u}'\mathbf{u}$$

die auch in Form empirischer Varianzen geschrieben werden kann

(14) $$s^2_y = s^2_{\hat{y}} + s^2_u.$$

Hieraus läßt sich die Interpretation des Bestimmtheitsmaßes

(15) $$R^2 = \frac{s^2_{\hat{y}}}{s^2_y} = 1 - \frac{s^2_u}{s^2_y}$$

als Maß der Anpassungsgüte ablesen; es liegt um so näher bei 1, je kleiner die Fehlerwerte sind, während es sich bei großem Fehleranteil der Zahl Null nähert.

Obwohl im Rahmen der beschreibenden Regression noch eine Reihe wichtiger Themen wie polynomische Regression, Bedeutung einzelner Prädiktoren usw. besprochen werden könnte, sollen zunächst die Populationsmodelle für inferenzstatistische Aufgaben dargestellt werden, um den Zusammenhang der Argumentation nicht durch zu viele Einzelheiten zu verwischen.

1.2 Das allgemeine regressionsanalytische Modell

Bei den meisten Anwendungen der Regression möchte man zu Aussagen kommen, die über die konkrete Stichprobe hinausreichen. In den anfangs erwähnten Beispielen interessiert eher der Zusammenhang zwischen der Produktivität y und den Prädiktoren Arbeitsbedingungen, Zufriedenheit *usw.* bei *zukünftig* Beschäftigten bzw. das Ausmaß des Rauchens bei *zukünftiger* Anwendung von Treatments als die Zusammenhänge in der Stichprobe. Dies entspricht der Situation von Galileo Galilei, der 1604 sein Fallgesetz v = gs mit einem universellen Anspruch auch für zukünftig fallende Massen aufstellte. Die Form des Fallgesetzes sowie die nähere Kennzeichnung der Umstände, unter denen es zutreffen soll, bilden ein Populationsmodell (= Satz von Oberhypothesen) mit dem freien Parameter Fallbeschleunigung g, der dann durch einige Beobachtungen näher eingegrenzt werden kann. Das Populationsmodell sollte natürlich auch durch Beobachtungen begründet und überprüft werden. Obwohl wir vornehmlich Aussagen über Parameter des Modells treffen wollen, werden wir auch einige Möglichkeiten zur Überprüfung des Modells kennenlernen. übrigens erwies sich auch Galileis ursprüngliches Modell als revisionsbedürftig, da es dem später allgemein verwendeten Fallgesetz v = gt unterlegen war. Außer dem in diesem Kapitel einzuführenden regressionsanalytischen Modell wird in Kap. 1.7 noch das korrelationsanalytische Modell vorgestellt.

Für die Formulierung der Populationsmodelle wird auf den Begriff der Zufallsvariablen mit zugehörigen Verteilungsfunktionen (2. B. Haagen & Pertler, 1976, p. 169) zurückgegriffen. Wir wollen nämlich annehmen, daß die Beobachtungen in Stichproben als Realisierungen von Zufallsvariablen aufgefaßt werden können.

Das allgemeine *regressionsanalytische Modell (general linear model, GLS-Modell)* umfaßt folgende Annahmen:

(A1) Es besteht die Beziehung

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u},$$

wobei X eine Nx(K+1)-Matrix von festen, beobachtbaren

Werten, ß ein Vektor von festen, unbekannten Parametern mit K+ 1 Komponenten,
y ein Vektor von N beobachtbaren Zufallsgrößen und
u ein Vektor von N unbeobachtbaren Zufallsgrößen sind.

(A2) Die Matrix X ist vom Rang K + 1

(A3) Der Erwartungswert von u ist der Nullvektor: Eu = 0

(A4) Die Kovarianzmatrix $\mathbf{\Sigma}_{uu}$ zu u existiert und ist nichtsingulär. $\mathbf{\Sigma}_{uu}$ läßt sich als $\mathbf{\Sigma}_{uu} = \sigma^2\mathbf{V}$ mit $\sigma^2 = \frac{1}{N}$ Spur $\mathbf{\Sigma}_{uu}$ darstellen.

(A5) Der Vektor u ist N-dimensional normalverteilt.

Diese Annahmen bedürfen der Erläuterung.

In Al wird die Art der in der Regression beteiligten Größen und ihr Zusammenhang eingeführt. Jede der N Zeilen von X umfaßt die K + 1 Werte der Prädiktorvariablen eines Merkmalsträgers. Ein Merkmalsträger zur Zeile n soll eine Ziehung aus der Grundgesamtheit der Merkmalsträger zu den zugehörigen festen Werten $x_{n0}, x_{n1}, \ldots, x_{nK}$ der Prädiktorvariablen sein. Mit der Grundgesamtheit von Merkmalsträgern ist zugleich eine Wahrscheinlichkeitsverteilung der n-ten Komponente y_n der Kriteriumsvariablen durch die n-te Komponente u, der Fehlervariablen verknüpft. Die Realisierungen von y_n und u_n und damit auch ihre Wahrscheinlichkeitsverteilungen unterscheiden sich nur additiv durch den festen Wert der n-ten Kompenente $(\mathbf{X\beta})_n$.

Eine wichtige Konsequenz der Annahme, daß die Werte von X fest sind, ist, daß sich die Interpretation der Wahrscheinlichkeiten als relative Häufigkeiten für sehr große Anzahlen von Versuchswiederholungen auf wiederholte Ziehungen zu denselben Prädiktorwerten stützt. Das bedeutet allerdings nicht, daß sich Inferenzen nur auf die betrachteten Prädiktorwerte beziehen können; Inferenzen sind stets in dem Bereich von Prädiktorwerten möglich, in dem die Modellannahmen gültig sind (Graybill, 1976, p. 147). Die Annahme fester Werte in X verlangt vor allem, daß die Prädiktoren fehlerfrei gemessen werden. Bei kontinuierlichen Prädiktorvariablen ist diese Forderung natürlich nie erfüllbar, weshalb man sich dort mit der Annahme kleiner relativer Meßfehler zufrieden geben muß; inferenzstatistische Schlüsse bleiben dann nur noch näherungsweise gültig.

Durch die Annahme A2 wird lediglich der unbequeme Fall der Nicht-Identifizierbarkeit des Vektors ß ausgeschlossen. Wäre X nicht vom Rang K + 1, gäbe es beliebig viele verschiedene Vektoren ß, die unter sonst gleichen Umständen zu derselben Verteilung des Zufallsvektors y führen. A2 ist der Forderung gleichwertig, daß X'X nichtsingulär ist. Die Forderung kann immer durch Umorganisation der Prädiktorvariablen und Verkleinerung ihrer Anzahl erfüllt werden, ohne daß der Erklärungswert der Prädiktoren geschmälert würde.

Die Annahme A3 besagt, daß für jede betrachtete Konstellation der Prädiktorwerte der Erwartungswert („long-run-Mittelwert“) der Fehlervariablen Null ist, was damit gleichwertig ist, daß der Erwartungswert der Kriteriumsvariablen

(16) $$\mathrm{E}\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$$

linear durch die Prädiktoren bestimmt ist. Es besteht eine Ähnlichkeit von A3 mit der Eigenschaft (10) der Minimum-Quadrat-Methode, daß der Mittelwert des Approximationsfehlers verschwindet. Eine spezielle Minimum-Quadrat-Forderung benötigt man wegen Annahme A3 nicht.

In Annahme A4 wird zunächst lediglich festgelegt, daß die Varianzen der Fehlerwerte und ihre Kovarianzen zu verschiedenen Prädiktorkombinationen existieren. Damit werden z.B. Autokorrelationen der Fehlerwerte $\mathrm{cov}\,(u_n, u_m) \neq 0$ für $n \neq m$ oder Heterogenität der Fehlervarianzen $\mathrm{var}\, u_n \neq \mathrm{var}\, u_m$ für $n \neq m$ zugelassen. Weil manchmal zwar die absoluten Größen der Varianzen und Kovarianzen unbekannt, aber ihre Verhältnisse untereinander bekannt sind, wird meist die Kovarianzmatrix in die Faktoren σ^2 und V zerlegt. Wir behandeln hier nur den Fall, daß V nichtsingulär ist (sonst s. Theil, 1971, p. 274). Man fordert dann, daß die NxN-Matrix V bekannt sein muß, während σ^2 als einzelner unbekannter Parameter geschätzt wird. Beispielsweise wird im *klassischen regressionsanalytischen Modell (ordinary least squares model, OLS-Modell),* das die meisten Computerprogramme zur Regression voraussetzen, als Matrix V die Einheitsmatrix I vorausgesetzt:

(K4) $$\boldsymbol{\Sigma}_{uu} = \sigma^2\mathbf{I},$$
was Unkorreliertheit der Fehlervariablen und Varianzhomogenität (Homoskedastizität) bedeutet.

Das klassische regressionsanalytische Modell ist also ein Spezialfall des allgemeinen regressionsanalytischen Modells.

Die Annahme A5 schließlich fügt eine Voraussetzung über die Form der Verteilung der Kriteriumsvariablen für feste Prädiktorwerte hinzu. Diese Annahme ist für die Beurteilung der Eigenschaften der noch zu besprechenden Schätzfunktionen nicht erforderlich, sondern wird erst bei der Konstruktion von Konfidenzbereichen und statistischen Tests benötigt. Allerdings wird für Schätzer nach dem *Maximum-Likelihood-Prinzip (ML-Prinzip) (z.B.* Schönfeld, 1969, p. 110; Graybill, 1976, p. 75) bereits bei der Konstruktion des Schätzers der Typ der Wahrscheinlichkeitsverteilung von u vorausgesetzt, wobei wir annehmen, daß die Wahrscheinlichkeitsverteilung parametrisierbar ist.

1.3 Die Schätzung der Parameter im allgemeinen regressionsanalytischen Modell

Wenn man weiß, daß ein allgemeines regressionsanalytisches Modell mit den Annahmen Al bis A4 vorliegt, und wenn eine Stichprobe vom Umfang N mit den Prädiktorwerten X und den Kriteriumswerten y vorliegt, dann besteht die Aufgabe darin, die noch unbekannten Parameter ß und σ^2 einzugrenzen. Die Matrix V aus A4 sei bekannt. Zunächst sollen Punktschätzungen und ihre Eigenschaften angegeben werden.

Der *beste (effiziente, kleinste Varianz) lineare erwartungstreue* (kein Bias) Schätzer für ß, den man ohne a priori-Information über ß oder σ^2 gewinnen kann (Seifert, 1975, p. 44), ist

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}, \tag{17}$$

der *sog. GLS-Schätzer (generalized least squares estimator).* Wenn die Voraussetzung der Normalverteilung A5 hinzugenommen wird, ist dieser Schätzer sogar in der Klasse der erwartungstreuen Schätzer (linear oder nichtlinear) effizient (Theil, 1971, p. 390). Unter dieser Voraussetzung ist ß zugleich auch der Maximum-Likelihood-Schätzer.

Wendet man das Minimum-Quadrat-Prinzip auf die Daten der Stichprobe an, dann erhält man entsprechend Gleichung (8) den sogenannten *OLS-Schätzer (ordinavy least squares estimator):*

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \tag{18}$$

der unter der zusätzlichen Annahme K4 des klassischen regressionsanalytischen Modells dem GLS-Schätzer gleich ist. Auch der OLS-Schätzer ist ein linearer erwartungstreuer Schätzer (Johnston, 1963, p. 188), der aber nur im *klassischen* regressionsanalytischen Modell kleinste Varianz hat. Der Effizienzverlust im *allgemeinen* regressionsanalytischen Modell ist für Punktschätzungen jedoch nicht groß (Wonnacott & Wonnacott, 1970, p. 334), so daß die Verwendung des OLS-Schätzers vertretbar scheint.

übrigens kann auch der GLS-Schätzer $\hat{\boldsymbol{\beta}}$ im allgemeinen regressionsanalytischen Modell nach einem abgewandelten Minimum-Quadrat-Prinzip konstruiert werden. Man erhält ihn, indem man das Minimum von

$$(\mathbf{y}-\hat{\mathbf{y}})'\mathbf{V}^{-1}(\mathbf{y}-\hat{\mathbf{y}}) = \sum_{k=1}^{K}\sum_{l=1}^{K}(y_k-\hat{y}_k)(y_l-\hat{y}_l)(\mathbf{V}^{-1})_{kl}$$

$$\text{mit } \hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

für ß sucht. Deshalb wird der GLS-Schätzer häufig auch als WLS-Schätzer (weighted least squares) bezeichnet. $\hat{\mathbf{y}}$ ist die Schätzung des Erwartungswertes Ey = $\mathbf{X\beta}$.

Ein erwartungstreuer quadratischer (z.B. Schönfeld, 1969, p. 67) Schätzer für σ^2 ist

(19) $$\hat{\sigma}^2 = \frac{1}{N-K-1} (\mathbf{y}-\hat{\mathbf{y}})'\mathbf{V}^{-1}(\mathbf{y}-\hat{\mathbf{y}})$$

oder gleichwertig

$$\hat{\sigma}^2 = \frac{1}{N-K-1} \mathbf{y}'(\mathbf{V}^{-1}-\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})\mathbf{y}.$$

Dieser Schätzer ist unter der zusätzlichen Annahme der Normalverteilung A5 effizient in der Klasse der erwartungstreuen quadratischen Schätzer. Der ML-Schätzer für σ^2 weicht von (19) etwas ab, ist allerdings auch nicht erwartungstreu. Im klassischen regressionsanalytischen Modell entfällt V^{-1} (= Einheitsmatrix), und $\hat{\sigma}^2$ ist die Schätzung der Störvarianz.

Durch eine *Bereichsschätzung* wird ein Bereich um die Punktschätzung angegeben, der mit der Wahrscheinlichkeit *l-a* den zu schätzenden Parameter enthält (Haagen & Seifert, 1979, p. 140). Die Wahrscheinlichkeit *a*, daß der Bereich den Parameter nicht überdeckt, wird Irrtumswahrscheinlichkeit genannt. Der Konfidenzbereich wird aufgrund der Werte des Zufallsvektors y festgelegt und ist insofern von der Realisierung der Zufallsvariablen abhängig. Aus der Bereichsschätzung kann auch ein statistischer Test zur Prüfung einer Hypothese über den Parameter abgeleitet werden.

Zunächst sollen Bereichsschätzungen für die Regressionskoeffizienten ß eingeführt werden. Um das Spektrum der Anwendungen zu vergrößern, wird die Bereichsschätzung allgemeiner für beliebige Linearkombinationen

(20) $\gamma = \mathbf{L\beta}$
mit fester Mx(K+1)−Matrix $\mathbf{L}$
mit $M \leqq K+1$ und Rang $\mathbf{L} = M$

durchgeführt. Die Matrix L wird auch als *Kontrast* bezeichnet. Setzt man die Einheitsmatrix für L ein, ist γ = ß. Man weiß übrigens (Seifert, 1975, p. 57), daß die beste lineare erwartungstreue Schätzung von γ durch

(21) $$\hat{\gamma} = \mathbf{L}\hat{\boldsymbol{\beta}}$$

aus dem GLS-Schätzer für ß berechnet werden kann.

Als Konfidenzbereich für γ zum Konfidenzgrad $1-\alpha$ wird der Bereich für γ als

$$\text{(22)} \qquad \frac{1}{M\hat{\sigma}^2}\,(\gamma-\hat{\gamma})'(L(X'V^{-1}X)^{-1}L')^{-1}(\gamma-\hat{\gamma}) \leqq F_\alpha\,(M,N-K-1)$$

konstruiert. Der angegebene Bereich ist im allgemeinen das Innere eines M-dimensionalen Ellipsoids mit Mittelpunkt $\hat{\gamma}$, für den Fall nur einer einkomponentigen Linearkombination ergibt sich ein Intervall. Ein wichtiger Spezialfall ist das Konfidenzintervall für einen einzelnen Regressionskoeffizienten β_K im klassischen regressionsanalytischen Modell (V = 1), das man durch den Ansatz L = (0,0 ,..., 0,1,0 ,..., 0) erhält:

$$\text{(23)} \qquad \frac{(\beta_k-b_k)^2}{\hat{\sigma}^2_{b_k}} \leqq F_\alpha(1,\, N-K-1)$$

$$\text{mit } \hat{\sigma}^2_{b_k} = \sigma^2\,[(X'X)^{-1}]_{kk}\ ;$$

die Größe $\sqrt{\hat{\sigma}^2_{b_k}}$ wird bei Computerprogrammen zur Regression unter dem Titel Standardfehler des Regressionskoeffizienten ausgedruckt. Eine Bereichsschätzung für den Parameter σ^2, der im klassischen Regressionsmodell die Varianz der Fehlervariablen u_k ist, kann bei einem Signifikanzniveau von a durch

$$\frac{(N-K-1)\,\hat{\sigma}^2}{\chi^2_{1-\alpha/2}(N-K-1)} \leqq \sigma^2 \leqq \frac{(N-K-1)\,\hat{\sigma}^2}{\chi^2_{\alpha/2}(N-K-1)}$$

angegeben werden (Schneeweiß, 1974, p. 65, 111).

1.4 Prognose im allgemeinen regressionsanalytischen Modell

Im vorangehenden Kapitel hatten wir diskutiert, wie in einem bestimmten Modell (A1 bis A5, K4) aus den Werten einer Stichprobe die unbekannten Parameter der Verteilung geschätzt werden können. Durch die Schätzung erhalten wir empirische Aussagen über den Zusammenhang von Prädiktor- und Kriterienwerten. Wenn Art und Ausmaß des Zusammenhangs für zukünftige Beobachtungen weiter gelten, können wir versuchen, zukünftige Werte des Kriteriums aus denen der Prädiktoren vorherzusagen. In dem am Anfang von Kap. 1.1 geschilderten Beispiel wollen wir also wissen, welchen Wert der Produktivität wir für zukünftige Beschäftigte vermuten, wenn wir ihre Arbeitsbedingungen, die Zufriedenheit mit dem Vorgesetzten, die Bezahlung und die Vielfältigkeit der Tätigkeitsbereiche kennen. Die so beschriebene Fragestellung der Prognose ist im Grunde ebenfalls eine Aufgabe der Schätzung, wobei im Gegensatz zum letzten Kapitel nicht feste Parameter einer Verteilung, sondern die Werte einer Zufallsvariablen geschätzt werden sollen.

Nach dem Ziel der Prognose können passive und aktive Prognose unterschieden werden. Von der passiven Prognose soll bei einer festen, von außen vorgegebenen zukünftigen Konstellation der Prädiktoren abgeschätzt werden, wie die Werte des Kriteriums liegen werden. In der aktiven Prognose werden verschiedene alternative zukünftige Konstellationen der Prädiktoren in ihrer Auswirkung auf das Kriterium untersucht; das betrachtete System wird dann der Prädiktorkonstellation ausgesetzt, die den nach externen Beurteilungsmaßstäben günstigsten Wert des Kriteriums erwarten läßt. Die aktive Prognose ist etwa auf Verlaufsuntersuchungen von klinisch-psychologischen Therapien anwendbar, wo Treatments im Hinblick auf günstigen Symptomverlauf ausgewählt werden sollen. Auch bei der Auswahl von Bewerbern durch Regression wird aktive Prognose verwendet. Nur wenn die Prädiktorkonstellation nicht beeinflußt werden kann, wenn sie etwa ihrerseits prognostiziert werden muß, kann nur passiv prognostiziert werden. Die darzustellenden Prognoseansätze sind sowohl für aktive wie passive Prognose verwendbar.

Zusätzlich zu den Beobachtungen der Zufallsvariablen y_n für $n = 1,2,...,N$ und der festen Variablenwerte x_{nk} für $n = 1,2,\ldots,N$ und $k = 0,1,\ldots,K$ seien M neue Merkmalsträger mit den Werten für die Prädikatorvariablen x_{nk} für $n = N+1, N+2,\ldots,N+M$ und $k = 0,1,\ldots,K$ beobachtet, ohne daß zugleich die Werte der Kriteriumsvariablen bekannt wären. Die Prädiktorwerte fassen wir in der $M\times(K+1)$-Matrix X_z und die zugehörigen Zufallsvariablen y_n im Vektor y_z mit M Komponenten zusammen. Die zusätzlichen Prädiktorwerte seien fest, d.h., daß sie fehlerfrei gemessen und nicht zufällig sind. Für viele Anwendungen genügt es, die Prognose nur für einen weiteren Merkmalsträger $(M = 1)$ durchzuführen. Im allgemeinen regressionsanalytischen Modell können jedoch auch Zeitreihen mit voneinander abhängigen Kriteriumsvariablen (z.B. Korrelation von y_n mit y_{n+1}) behandelt werden, weshalb die Prognose einer Serie von unbekannten Werten wünschenswert ist.

Für die Prognose muß gelten, daß das allgemeine regressionsanalytische Modell auch unter Hinzunahme von X_z und y_z noch gilt. Weil in den Annahmen A1 bis A5 diese zusätzlichen Variablen nicht berücksichtigt waren, müssen wir sie in Form eines *Prognosemodells* erweitern:

(P1) Es besteht die Beziehung

$$\mathbf{y}_z = \mathbf{X}_z\boldsymbol{\beta}+\mathbf{u}_z$$

mit demselben Vektor ß wie in A1.

Dies bedeutet mit $\mathbf{y}^* = \begin{pmatrix} \mathbf{y} \\ \mathbf{y}_z \end{pmatrix}$, $\mathbf{X}^* = \begin{pmatrix} \mathbf{X} \\ \mathbf{X}_z \end{pmatrix}$ und $\mathbf{u}^* = \begin{pmatrix} \mathbf{u} \\ \mathbf{u}_z \end{pmatrix}$ insgesamt

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta}+\mathbf{u}^*$$

(P2) Die Matrix X_z ist vom Rang M, $M \leqq K+1$

(P3) Der Erwartungswert von u_z ist der Nullvektor: $\mathbf{E}\mathbf{u}_z = 0$

(P4) Die Kovarianzmatrix $\mathbf{\Sigma}_{u^*u^*}$ zu u* existiert und ist nichtsingulär.

$$\text{Zerlegung: } \mathbf{\Sigma}_{u^*u^*} = \sigma^2\mathbf{V}^* = \sigma^2 \begin{pmatrix} \mathbf{V} & \mathbf{W} \\ \mathbf{W}' & \mathbf{V}_z \end{pmatrix}$$

(P5) Der Vektor u_z ist M-dimensional normalverteilt.

Zuerst wollen wir uns der Prognose von

(24) $$\boldsymbol{\mu}_z = \mathbf{E}\mathbf{y}_z,$$

also der *Prognose des Erwartungswertes* des Kriteriums für die zukünftigen Beobachtungen zuwenden. Die beste lineare erwartungstreue *Punktschätzung* dafür ist

(25) $$\hat{\boldsymbol{\mu}}_z = \mathbf{X}_z\hat{\boldsymbol{\beta}}$$

mit $\hat{\boldsymbol{\beta}}$ aus Gleichung (17) (s. Schönfeld, 1969, p. 77), wobei nur die Annahmen Al bis A4, Pl und P3 erforderlich sind. Sind alle Annahmen Al bis A5 und Pl bis P5 erfüllt, ist eine *Bereichsschätzung* für $\boldsymbol{\mu}_z$ durch

(26) $$\frac{1}{M\hat{\sigma}^2} (\boldsymbol{\mu}_z-\hat{\boldsymbol{\mu}}_z)'\mathbf{X}_z\ (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'_z(\boldsymbol{\mu}_z-\hat{\boldsymbol{\mu}}_z) \leqq F_\alpha(M,\ N-K-1)$$

mit der Irrtumswahrscheinlichkeit *a* gegeben (Seifert, 1975, p. 87).

Bestehen Zusammenhänge zwischen den Fehlern u bei der ursprünglichen Stichprobe und den Fehlern u_z bei den hinzugekommenen Erhebungseinheiten, d.h. ist die Matrix W aus Annahme P4 keine Nullmatrix, dann kann man aus den geschätzten Fehlern $\hat{\mathbf{u}}$ sogar die Fehler u_z schätzen, so daß $y_z = X\ ß + u_z$ insgesamt genauer geschätzt werden kann, als dies wie in Gleichung (25) über eine alleinige Schätzung von ß möglich ist. Insofern geht die *Prognose der Zufallsgröße* y_z über die Prognose des Erwartungswertes $\boldsymbol{\mu}_z$ hinaus. Sind die Annahmen Al bis A4 und Pl, P3 und P4 erfüllt, dann ist die beste (im Sinne des mittleren quadratischen Prognosefehlers) erwartungswerterhaltende (d.h. der Erwartungswert der Schätzung $\hat{\mathbf{y}}_z$ ist gleich dem Erwartungswert von y_z) lineare *Punktschätzung* für y_z

(27) $$\hat{\mathbf{y}}_z = \mathbf{X}_z\hat{\boldsymbol{\beta}} + \mathbf{W}'\mathbf{V}^{-1}\ (\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})\ .$$

Für diese Punktschätzung muß außer der Matrix V, die wir grundsätzlich als bekannt vorausgesetzt haben, auch noch die Matrix W aus P4 bekannt sein. Wenn wie im klassischen regressionsanalytischen Modell W = 0 gilt, fällt die Punktschätzung für y_z mit der für $\boldsymbol{\mu}_z$ zusammen.

Als Bereichsschätzung für y_z wünscht man sich eigentlich einen aufgrund der Realisation von y konstruierten Bereich, der dann mit der Wahrscheinlichkeit 1 - α den Zufallsvektor y_z enthält. Das bedeutet im „long run", daß $(1 - \alpha)\cdot 100\%$ der zu prognostizierenden Werte y_z in dem aufgrund der Realisierung von y konstruierten Bereich enthalten sind. Aufgrund rechnerischer Schwierigkeiten ist die Konstruktion solcher Bereiche im allgemeinen (wie auch im klassischen) regressionsanalytischen Modell noch nicht gelungen. Ein nicht voll befriedigender Lösungsansatz besteht in der Konstruktion von Toleranzintervallen und ist in Schneeweiß (1974, p. 81) skizziert. Wir wollen uns mit einer anderen Lösung begnügen; ihre Spezialisierung auf die klassische Regression mit nur einer Prädiktorvariablen führt zu den Vertrauensintervallen für Werte des Kriteriums, die in der elementaren Literatur (z.B. Hays, 1963, p. 523; Gaensslen & Schubö, 1976, p. 56; Bortz, 1977, p. 231) angegeben sind.

Wir stellen uns vor, daß einer Realisation von y auch nur eine Realisation von y_z zugeordnet wird. Die Bereichsschätzung wird so konstruiert, daß die Wahrscheinlichkeit dafür 1 - α ist, daß die Zufallsgröße y_z in dem durch y bestimmten Bereich liegt. Man stellt sich für die long run frequency-Interpretation von 1 - α also die wiederholte Realisation von y und y_z vor. Wenn die Annahmen Al bis A5 und Pl bis P5 erfüllt sind, ist

$$(28)\qquad \frac{1}{M\hat{\sigma}^2}(\mathbf{y}_z-\hat{\mathbf{y}}_z)'\,(\mathbf{V}_z+\tilde{\mathbf{X}}_z\,(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\tilde{\mathbf{X}}_z'-\mathbf{W}'\mathbf{V}^{-1}\mathbf{W})^{-1}\,(\mathbf{y}_z-\hat{\mathbf{y}}_z) \leqq$$
$$\leqq F_\alpha(M, N-K-1)$$
$$\text{mit } \tilde{\mathbf{X}}_z = \mathbf{X}_z-\mathbf{W}'\mathbf{V}^{-1}\mathbf{X}$$

eine Bereichsschätzung für y_z (s. Theil, 1971, p. 288) zur Irrtumswahrscheinlichkeit α. Zur Berechnung dieser Bereichsschätzung ist die Kenntnis der ganzen Matrix V* aus P4 erforderlich, also die Kenntnis von V, W und V_z.

Im klassischen regressionsanalytischen Modell (Unabhängigkeit aller Fehlervariablen aus u*) wird V = I, W = 0 und V_z = I vorausgesetzt, wodurch sich (28) zur Bereichsschätzung

$$(29)\qquad \frac{1}{M\hat{\sigma}^2}(\mathbf{y}_z-\hat{\mathbf{y}}_z)'\,(\mathbf{I}+\mathbf{X}_z\,(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}_z')^{-1}\,(\mathbf{y}_z-\hat{\mathbf{y}}_z) \leqq F_\alpha(M, N-K-1)$$

vereinfacht. Bei weiterer Vereinfachung auf eine Prädiktorvariable x (K = 1) und eine zusätzliche Erhebungseinheit (M = 1) erhält man die bekannte einfachste Bereichsschätzung für die skalare Zufallsvariable y_z

$$(30)\qquad \frac{(y_z - \hat{y}_z)^2}{\hat{\sigma}^2\left(1 + \dfrac{x_z^2 - 2x_z\bar{x} + \overline{x^2}}{N\,(\overline{x^2} - \bar{x}^2)}\right)} \leqq F_\alpha(1,N-2).$$

1.5 Statistische Tests im klassischen regressionsanalytischen Modell

In den beiden letzten Kapiteln haben wir im Rahmen eines Modells über den funktionalen Zusammenhang versucht, Parameter oder von diesen abhängige Zufallsgrößen zu schätzen. Insofern war die Aufgabe die „Quantifizierung einer Theorie“. Jetzt wollen wir uns der Frage zuwenden, ob unsere Annahmen im Widerspruch zu den Beobachtungen stehen. Die neue Aufgabe ist also die „empirische Überprüfung einer Theorie“ (Seifert, 1975, p. 130). Dabei ist es häufig unmöglich, aus den Beobachtungen abzulesen, welche spezielle Annahme nicht zutrifft, weil der Widerspruch in der Regel nur innerhalb eines ganzen Satzes von Annahmen (Hypothesen und Oberhypothesen) hergestellt werden kann. Man muß dann nach dem Ausschlußverfahren der Reihe nach alle Annahmen bis auf eine von ihnen als Quelle des Widerspruchs ausschließen.

Eine Übersicht über die Theorie der statistischen Tests wird in diesem Handbuchband von Wendt gegeben. Doch wollen wir hier die Nomenklatur der statistischen Tests (s. auch Haagen & Seifert, 1979, Teil III) präzisieren. Unter einem *Signifikanztest* verstehen wir die Konstruktion eines kritischen Bereichs derart, daß für alle Parameter aus der Nullhypothese die Wahrscheinlichkeit dafür a ist, daß die Prüfgröße in den kritischen Bereich fällt. Die Entscheidungsregel ist: Fällt die Realisation der Prüfgröße in den kritischen Bereich, verwerfen wir die Nullhypothese; sonst enthalten wir uns des Urteils. Im Neyman-Pearson-Test (NP-Test) wird der kritische Bereich so konstruiert, daß die Wahrscheinlichkeit dafür, daß die Prüfgröße für alle Parameter aus der Nullhypothese einen Wert in diesem Bereich annimmt, a ist, und möglichst so, daß sie für alle Parameter aus der Alternativhypothese maximal ist; gelingt dies, nennt man den Test einen gleichmäßig mächtigsten Test (UMP-Test). Bei sog. zweiseitigen Alternativhypothesen gibt es keinen gleichmäßig mächtigsten Test; wenn man vom kritischen Bereich jedoch fordert, daß die Wahrscheinlichkeit dafür, daß die Prüfgröße für die Parameter der Alternativhypothese einen Wert in diesem Bereich annimmt, stets mindestens a ist, existiert auch dann ein gleichmäßig mächtigster Test, den man einen gleichmäßig mächtigsten unverfälschten Test (UMPU-Test) nennt. Beim *Likelihood-Quotienten-Test* (LQ-Test) wird als Prüfgröße der Likelihood-Quotient verwendet, der Quotient der maximalen Likelihood aus der Nullhypothese und der maximalen Likelihood aus der Alternativhypothese. Als kritischer Bereich wird ein Intervall zwischen 0 und dem kritischen LQ gewählt. Sowohl beim NP-Test als auch beim LQ-Test entscheidet man sich für die Nullhypothese, wenn die Prüfgröße nicht in den kritischen Bereich fallt.

Aus Bereichsschätzungen zu einem evtl. vektoriellen Parameter $\boldsymbol{\Theta}$ kann immer zumindest ein Signifikanztest konstruiert werden, wenn die Nullhypothese im zu schätzenden Parameter einfach ist. Hat die Bereichsschätzung zum Signifikanzniveau α die Gestalt $f(\boldsymbol{\Theta},\mathbf{y}) \leqq g_\alpha$ und spezifiziert die Nullhypothese den Parameter $\boldsymbol{\Theta} = \boldsymbol{\Theta}^{(0)}$, dann weiß man, daß $f(\boldsymbol{\Theta}^{(0)},\mathbf{y}) \leqq g_\alpha$ mit Wahrscheinlichkeit $1-\alpha$ gilt, wenn die Nullhypothese zutrifft. Als Prüfgröße kann daher $f(\boldsymbol{\Theta}^{(0)},\mathbf{y})$ verwendet werden; ist ihre Realisation $f(\boldsymbol{\Theta}^{(0)},\mathbf{y}) > g$, entscheidet man sich gegen $\boldsymbol{\Theta}^{(0)}$ und damit gegen die Nullhypothese. Da die Tests für das allgemeine regressionsanalytische Modell durch die beschriebenen Bereichsschätzungen bereits explizit gegeben sind, werden hier die entsprechenden Tests für das *klassische regressionsanalytische Modell* mit Normalverteilungsannahme A5 angegeben.

Zur Überprüfung von *Hypothesen über die Regressionskoeffizienten* verwenden wir wieder die Linearkombination $\boldsymbol{\gamma} = \mathbf{L}\boldsymbol{\beta}$ aus Gleichung (20) mit der Mx(K+l)-Kontrastmatrix L. Die Nullhypothese sei $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{(0)}$, die Alternativhypothese $\boldsymbol{\gamma} \neq \boldsymbol{\gamma}^{(0)}$. Die Nullhypothese wird verworfen, wenn

$$(31) \qquad \frac{1}{M\hat{\sigma}^2}\,(\boldsymbol{\gamma}^{(0)} - \hat{\boldsymbol{\gamma}})'\,(\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}')^{-1}\,(\boldsymbol{\gamma}^{(0)} - \hat{\boldsymbol{\gamma}}) > F_\alpha(M,N-K-1)$$

$$\text{mit } \hat{\boldsymbol{\gamma}} = \mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

$$\text{und } \hat{\sigma}^2 = \frac{1}{N-K-1}\,\mathbf{y}'\,(\mathbf{I} - \mathbf{X}\,(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}$$

ist. Dies ist auch gleichwertig dem LQ-Test und für M = 1 (nur eine skalare Linearkombination) ein gleichmäßig mächtigster unverfälschter NP-Test, wenn in der Entscheidungsregel die Entscheidung H_0 bei „$\leqq$" in Gleichung (31) einbezogen wird. Für M > 1 existiert kein gleichmäßig mächtigster unverfälschter NP-Test (Kendall & Stuart, 1976, p. 238). Wenn als Kontrastmatrix L die Einheitsmatrix verwendet wird, kommt man zum Test der Regressionskoeffizienten. Bei der speziellen Nullhypothese $\boldsymbol{\beta}^{(0)} = 0$ erhält man mit dem aus Gleichung (15) übernommenen Bestimmtheitsmaß (Quadrat der empirischen multiplen Korrelation)

$$(32) \qquad R^2 = \frac{s^2_{\hat{y}}}{s^2_y} = \frac{\hat{\mathbf{y}}'\,\hat{\mathbf{y}} - N\bar{y}^2}{\mathbf{y}'\mathbf{y} - N\bar{y}^2} = \frac{\mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - N\bar{y}^2}{\mathbf{y}'\mathbf{y} - N\bar{y}^2}$$

die Entscheidungsregel, daß die Nullhypothese verworfen wird, wenn

$$(33) \qquad \frac{(N-K-1)\left(R^2 + \dfrac{\bar{y}^2}{s_y^2}\right)}{(K+1)\,(1-R^2)} > F_\alpha(K+1,N-K-1).$$

Dieser Test wird in der Regel nur sinnvoll sein, wenn sowohl die Prädiktorvariablen als auch die Kriteriumsvariable auf Verhältnisskalen (z.B. Gigerenzer, 1981, p. 116) gemessen sind, weil nur dann die Komponente β_0 des Vektors (der Achsenabschnitt der Regressionsfläche) in allen zulässigen Skalierungen der Variablen zugleich den Wert Null hat. Beschränkt man die Nullhypothese auf die „echten" Regressionskoeffizienten $\beta_1^{(0)} = \beta_2^{(0)} = \ldots = \beta_K^{(0)} = 0$ und läßt das Absolutglied $\beta_0^{(0)}$ (wie stets auch σ^2) offen, erhält man aus Gleichung (31) mit einer geeigneten Matrix L die bekannte Entscheidungsregel zum Verwerfen der Nullhypothese:

$$(34) \qquad \frac{(N-K-1)\,R^2}{K\,(1-R^2)} > F_\alpha(K,N-K-1)$$

(Cooley & Lohnes, 1971, p. 51; Kerlinger & Pednazur, 1973, p. 63; Cohen & Cohen, 1975, p. 104; Gaensslen & Schubö, 1976, p. 108; Bortz, 1977, p. 592; Moosbrugger, 1978, p. 70; Küchler, 1979, p. 127). Mit diesem Test kann also über die Frage entschieden werden, ob wenigstens eine der Prädiktorvariablen $x_1, \ldots, x_K$ einen von Null verschiedenen Regressionskoeffizienten hat, ob also insgesamt ein Zusammenhang zwischen Kriterium und Prädiktoren besteht.

Bei der Prüfung der Frage nach der *Relevanz einzelner Prädiktorvariablen* x_k ist der schwerwiegendere Fehler, wenn eine bedeutsame Variable vergessen wird, als wenn eine Variable mit $\beta_k = 0$ überflüssigerweise aufgenommen wird. Denn das Vergessen einer bedeutsamen Variablen ist ein Spezifikationsfehler in Al, der zur Folge haben kann, daß die Schätzung der Regressionskoeffizienten zu den übrigen Variablen nicht erwartungstreu ist, sowie daß sonst vermeidbare Heteroskedastizität, Autokorrelation oder Strukturbruch auftreten. Daher wäre der Signifikanztest der Nullhypothese $\beta_k^{(0)} \neq 0$ angemessen; ein derartiger Signifikanztest bereitet jedoch erhebliche theoretische Schwierigkeiten und kann höchstens für feste Werte, z.B. $\beta_k^{(0)} = -0.5$, durchgeführt werden. Man begnügt sich in der Regel mit der Prüfung der Nullhypothese $\beta_k^{(0)} = 0$, wodurch nur das Signifikanzniveau α für den weniger gravierenden Fehler kontrolliert wird. Entsprechend der Bereichsschätzung (23) kommt man zur Entscheidungsregel, daß die Nullhypothese abgelehnt wird, wenn

$$(35) \qquad \frac{b_k^2}{\hat{\sigma}_{b_k}^2} > F_\alpha(1,N-K-1)$$

ist. Auch dieser Test kann durch entsprechende Korrelationskoeffizienten ausgedrückt werden. Dazu muß aber noch die empirische semipartielle Korrelation (vgl. Gaensslen & Schubö, 1976, p. 87)

$$(36) \qquad r_{y\,(k-12\ldots\not{k}\ldots K)} = b_k \sqrt{\frac{1-R^2}{\hat{\sigma}_{b_k}^2(N-K-1)}}$$

eingeführt werden. Man erhält dann dieselbe Entscheidungsregel wie bei (35) mit

$$(37) \qquad \frac{(N-K-1)\,r^2_{y\,(k-12\ldots\not{k}\ldots K)}}{1-R^2} > F_\alpha(1,N-K-1).$$

Weil hier nur ein Regressionskoeffizient geprüft wird, ist der Test als NP-Test ein gleichmäßig mächtigster unverfälschter Test. Einen gleichmäßig mächtigsten Test für die einseitige Alternativhypothese $\beta_k > 0$ erhält man durch

$$\frac{b_k}{\sqrt{\hat{\sigma}^2_{b_k}}} > \sqrt{F_{2\alpha}(1, N-K-1)} = t_\alpha(N-K-1).$$

Als Beispiel für die Anwendung einer Kontrastmatrix soll der Test auf Strukturbruch vorgestellt werden. Unter Strukturbruch versteht man eine Verletzung des klassischen regressionsanalytischen Modells derart, daß entweder die Regressionskoeffizienten ß oder die Fehlervarianz σ^2 zu den ersten N_1 Beobachtungen einen anderen Wert haben als die zu den zweiten $N_2 = N - N_1$ Beobachtungen. Das Problem und dessen Behandlung werden unter dem Stichwort „*differentielle Prognostizierbarkeit*" angesprochen (z.B. Moosbrugger, 1980b). Wenn der Strukturbruch erkannt ist, müssen entweder in den beiden Gruppen getrennte Modelle aufgestellt werden oder man muß das Regressionsmodell um sog. Interaktionen erweitern. Für den Test auf Strukturbruch der Fehlervarianz werden wir das erste Vorgehen wählen, das zweite Vorgehen wählen wir für den Test auf Strukturbruch in den Regressionskoeffizienten. Hier soll zunächst der *Strukturbruch in den Regressionskoeffizienten* getestet werden. Für die Darstellung beschränken wir uns auf die Einfachregression K = 1 mit ß' = (β_0, β_1). β_0 ist der Koeffizient zur Konstante $x_{0n} = 1$, β_1 der zu $x_{1n} = x_n$.

Die Regressionskoeffizienten in der 1. Beobachtungsgruppe seien $\beta_0^{(1)}$, $\beta_1^{(1)}$, die in der zweiten Beobachtungsgruppe $\beta_0^{(2)}$, $\beta_1^{(2)}$. Die Fehlervarianz sei in beiden Gruppen einheitlich σ^2. Um den Test (31) anwenden zu können, muß das Modell auf 4 Regressionskoeffizienten, d.h. auch auf 4 Prädiktoren erweitert werden:

(38) $$y_n = \beta_0^{(1)}\,\delta_n^{(1)} + \beta_1^{(1)}\,\delta_n^{(1)}x_n + \beta_0^{(2)}\,\delta_n^{(2)} + \beta_1^{(2)}\,\delta_n^{(2)}x_n + u_n$$

mit den „Dummyvariablen"

(39) $$\delta_n^{(i)} = \begin{cases} 1 \text{ falls Beobachtung n zur Gruppe i gehört} \\ 0 \text{ sonst.} \end{cases}$$

Dadurch, daß nun insgesamt 4 Prädiktoren vorliegen, von denen für jede Gruppe jeweils zwei konstant den Wert Null haben, liegen in einer Regressionsgleichung zwei verschiedene Regressionsansätze vor. Obwohl wegen des schwerwiegenderen Fehlers, den Strukturbruch zu übersehen, eigentlich die gegensätzliche Hypothesenstellung angemessen wäre, wählt man wegen der erwähnten Schwierigkeiten als Nullhypothese

(40) $$\beta_0^{(1)} = \beta_0^{(2)} \text{ und } \beta_1^{(1)} = \beta_1^{(2)}$$

oder

$$\gamma = \begin{pmatrix} \beta_0^{(1)} - \beta_0^{(2)} \\ \beta_1^{(1)} - \beta_1^{(2)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Die Kontrastmatrix L ist demnach

(41) $$\mathbf{L} = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}$$

womit die Prüfung nach (31) erfolgen kann. Praktisch wird dies nur möglich sein, wenn ein Computerprogramm zum Linearen Modell vorliegt. Für Computerprogramme zur Regressionsanalyse ist der Test für den zusätzlichen Beitrag von Variablengruppen nach Kap. 1.11 für den folgenden gleichwertigen Modellansatz günstiger

(42) $$Y_n = \beta_0 + \beta_1 x_n + \beta_2 (\delta_n^{(1)} - \delta_n^{(2)}) + \beta_3 (\delta_n^{(1)} - \delta_n^{(2)}) x_n + u_n,$$

wobei der Zusammenhang zu den Regressionskoeffizienten aus dem Ansatz (38) durch

(43) $$\begin{aligned} \beta_0 &= (\beta_0^{(1)} + \beta_0^{(2)})/2 \\ \beta_1 &= (\beta_1^{(1)} + \beta_1^{(2)})/2 \\ \beta_2 &= (\beta_0^{(1)} - \beta_0^{(2)})/2 \\ \beta_3 &= (\beta_1^{(1)} - \beta_1^{(2)})/2 \end{aligned}$$

gegeben ist. Die Nullhypothese (40) lautet in dieser Parametrisierung jetzt einfacher

(44) $$\beta_2 = \beta_3 = 0$$

und kann als Test direkt zu den Regressionskoeffizienten praktisch leichter durchgeführt werden. Eine derartige Reparametrisierung ist bei jeder Kontrastmatrix L möglich, wobei lediglich die Prädiktorvariablen entsprechend transformiert werden müssen.

Der Test auf *Strukturbruch der Fehlervarianz* prüft, ob die Fehlervarianzen $\sigma^2_{(1)}$ und $\sigma^2_{(2)}$ zu den beiden Beobachtungsgruppen gleich sind. Insofern handelt es sich um eine einfachste Form der Prüfung auf *Varianzheterogenität* (Heteroskedastizität); differenziertere, kompliziertere und auch nur asymptotisch gültige Tests hierzu sind z.B. in Ramsey (1969), Kmenta (1971, p. 269) und Gartside (1972) beschrieben. Wie angekündigt stellen wir zwei getrennte Regressionsansätze in den beiden Beobachtungsgruppen auf; dadurch wird auch zugelassen, daß zugleich Strukturbruch in den Regressionskoeffizienten besteht. Aufgrund der beiden Teilstichproben vom Umfang N_1 und N_2 bildet man Schätzungen der Fehlervarianzen $\hat\sigma^2_{(1)}$ und $\hat\sigma^2_{(2)}$. Für die einseitige Fragestellung mit der Alternativhypothese $\sigma^2_{(1)} < \sigma^2_{(2)}$ erhält man mit der Entscheidungsregel

(45) $$\frac{\hat\sigma^2_{(1)}}{\hat\sigma^2_{(2)}} < F_{1-\alpha}(N_1-K-1,\ N_2-K-1) = \frac{1}{F_\alpha(N_2-K-1,\ N_1-K-1)}$$

für die Alternativhypothese einen gleichmäßig mächtigsten NP-Test zum Niveau a. Bei ungerichteter Alternativhypothese $\sigma^2_{(1)} \neq \sigma^2_{(2)}$ ist wegen der Asymmetrie der F-Verteilung folgender zweiseitiger Test mit jeweils gleichen Irrtumswahrscheinlichkeiten ein nur näherungsweise gleichmäßig mächtigster unverfälschter NP-Test: Verwirf die Nullhypothese zugunsten der Alternativhypothese, wenn

(46)
$$\frac{\hat\sigma^2_{(1)}}{\hat\sigma^2_{(2)}} < F_{1-\alpha/2}\,(N_1-K-1, N_2-K-1) \quad \text{oder}$$
$$\frac{\hat\sigma^2_{(1)}}{\hat\sigma^2_{(2)}} > F_{\alpha/2}\,(N_1-K-1,\ N_2-K-1).$$

Besonders schwierig zu interpretieren sind *Prognosetests,* weil durch sie die Nullhypothese „der ganze Satz von Annahmen A1 bis A5 und P1 bis P5 trifft zu" geprüft wird. Wird durch den Test die Nullhypothese verworfen, muß durch das Vorwissen eingegrenzt werden, welche der Annahmen speziell fragwürdig ist, ob etwa ein Strukturbruch bei der Erhebung der zusätzlichen Beobachtungseinheiten vorliegt, ob ein Modellfehler in Form eines vergessenen Prädiktors vorliegt oder die Normverteilungsannahme verletzt ist (oder im allgemeinen regressionsanalytischen Modell die Matrix V falsch spezifiziert wurde). Dennoch hat der Prognosetest für die Beurteilung einer Regressionsbeziehung als ganzes seine unbestrittene Bedeutung; er ist für *Kreuzvalidierungsstudien* der Test der Wahl. Weil die Anwendung der Bereichsschätzung (28) für den Prognosetest im allgemeinen regressionsanalytischen Modell allenfalls rechnerische Schwierigkeiten mit sich bringt, wird auch hier wie überall in diesem Kapitel 1.5 nur die Spezialisierung auf das klassische regressionsanalytische Modell dargestellt. Gegenüber der Anwendung der Bereichsschätzung (29) als Signifikanztest kann ein etwas vereinfachter Test dadurch erreicht werden, daß die Prädiktoren in der Matrix X „ortho-standardisiert" werden; dies kann z.B. durch eine Hauptkomponentenanalyse mit Bildung von standardisierten Hauptkomponentenwerten geschehen, wobei die Konstante $x_0 = 1$ nicht transformiert werden muß. Für derartige Prädiktoren gilt

(47)
$$X'X = NI,$$

womit der Prognosetest durch die Entscheidungsregel zur Verwerfung der Nullhypothese

(48)
$$\frac{1}{M\hat\sigma^2}\,(\mathbf{y}_z - \hat{\mathbf{y}}_z)'\,(\mathbf{I}+\frac{1}{N}\,\mathbf{X}_z\mathbf{X}'_z)^{-1}\,(\mathbf{y}_z - \hat{\mathbf{y}}_z) > F_\alpha\,(M, N-K-1)$$

gegeben ist. Dabei ist $\hat\sigma^2$ die Schätzung der Fehlervarianz aus der ursprünglichen Stichprobe und M die Anzahl der Beobachtungseinheiten in der zusätzlichen Stichprobe. Die Schwierigkeit bei dieser Entscheidungsregel ist, daß sie mit den üblichen Computerprogramm nicht auszuwerten ist. Doch kann man für die linke Seite der Ungleichung (48) eine obere und untere Schranke angeben, deren Komponenten z.B. mit dem Statistik-Programmsystem SPSS (Nie et al., 1975; Beutel et al., 1980) bestimmt werden können. Wenn (47) nach wie vor erfüllt ist, gilt mit

$$h_{ob} = \sum_{n=N+1}^{N+M} (y_n - \hat{y}_n)^2 \quad \text{und}$$

(49)

$$h_{unt} = \sum_{n=N+1}^{N+M} (y_n - \hat{y}_n)^2 - \frac{1}{N} \sum_{k=0}^{K} \Big(\sum_{n=N+1}^{N+M} (y_n - \hat{y}_n) x_{nk}\Big)^2$$

die Ungleichung

$$h_{unt} \leqq (\mathbf{y}_z - \hat{\mathbf{y}}_z)' \Big(\mathbf{I} + \frac{1}{N}\, \mathbf{X}_z \mathbf{X}'_z\Big)^{-1} (\mathbf{y}_z - \hat{\mathbf{y}}_z) \leqq h_{ob}.$$

Mit diesen Größen kommen wir zur Entscheidungsregel:
Für

$$\frac{h_{unt}}{M\hat{\sigma}^2} > F_\alpha(M, N-K-1)\text{: Entscheidung für } H_1$$

(50)

$$\frac{h_{ob}}{M\hat{\sigma}^2} < F_\alpha(M, N-K-1)\text{: Entscheidung für } H_0$$

sonst: keine Entscheidung oder Rückgriff auf (48).

Der *Durbin- Watson-Test* auf Verletzung der Annahme des klassischen regressionsanalytischen Modells, daß die Fehlervariablen unkorreliert sind, wird in Kapitel 1.9 besprochen. Die *Normalverteilungsannahme* verbleibt schließlich als eine schwer überprüfbare Annahme. Im Modell-Ansatz verlangt A5, daß die Fehler u_n multivariat normalverteilt sind. Dafür genügt es nicht, daß jede Fehlervariable u_n für sich normalverteilt ist, wenn nicht zusätzlich über K4 hinaus die Unabhängigkeit dieser Fehlervariablen angenommen werden kann. Es ist wegen des zentralen Grenzwertansatzes richtig, daß die Fehler dann mit guter Näherung als multivariat normalverteilt betrachtet werden können, wenn sie additiv auf viele, unabhängige Einflußfaktoren zurückgehen. Doch niemand kann letztlich ausschließen, daß der Fehler fast ausschließlich auf eine im Modell nicht berücksichtigte, nicht normal verteilte Variable zurückgeht. Draper & Smith (1966, p. 87) diskutieren einige Möglichkeiten der grafischen Beurteilung der Normalverteilungsannahme, die sich auf die empirische Verteilung der N geschätzten Residuen stützen. Eine einfache Methode zur Beurteilung der univariaten Normalität, die vor allem auf Ausreißer in der empirischen Verteilung der Residuen empfindlich ist, besteht darin, nachzuprüfen, ob im Bereich

$$[-\sqrt{\hat{\sigma}^2}\, t_{\alpha/2}(N-K-1),\ +\sqrt{\hat{\sigma}^2}\, t_{\alpha/2}(N-K-1)]$$

auch wirklich der Anteil 1 - α der Residuen liegt. In Tests zur Normalverteilungsannahme führen Lienert (1967, p. 172) und Koerts & Abrahamse (1969, Kap. 7) ein. Über die Robustheit der Bereichsschätzungen und Signifikanztests gegenüber Abweichungen von der Normalverteilung findet man in Malinvaud (1966, p. 297) Hinweise.

1.6 Ridge-Regression

Wir betrachten das klassische Regressionsmodell $y = \mathbf{X\beta} + u$, bei dem die Annahmen A1, A2, A3 und K4 erfüllt sind, so daß die Fehler u_k unkorreliert sind und gleiche Varianz haben. Es sollen nun genauer die Eigenschaften des Minimum-Quadrat-Schätzers

$$b = (X'X)^{-1}X'y$$

untersucht werden. Im Zusammenhang mit Gl. (18) war berichtet worden, daß b der beste lineare erwartungstreue Schätzer ist. Diese Aussage des Satzes von Gauß-Markov bedeutet, daß für alle Linearkombinationen $l'b$ mit festem, beliebigem Gewichtevektor l die Varianz kleiner ist als für die mit irgendeinem anderen linearen erwartungstreuen Schätzer ß gebildete Linearkombination $\mathbf{l'\beta}$:

$$\mathrm{var}\,(\mathbf{l'b}) \leqq \mathrm{var}\,(\mathbf{l'\check{\beta}})$$

für beliebige l und alle linearen erwartungstreuen Schätzer $\check{\boldsymbol{\beta}}$. Setzt man für l den Einsvektor j ein, erhält man als Varianz

$$\text{(51)} \qquad \mathrm{var}\,(\mathbf{j'b}) = \mathrm{E}\,(\mathbf{b} - \boldsymbol{\beta})'(\mathbf{b} - \boldsymbol{\beta}) = \mathrm{E} \sum_{k=0}^{K} (b_k - \beta_k)^2$$

den Erwartungswert der Summe der quadrierten Schätzfehler, den man auch den mittleren quadratischen Fehler (mean square error) nennt. Dieser mittlere quadratische Fehler ist für den Minimum-Quadrat-Schätzer b also minimal unter den linearen und erwartungstreuen Schätzern. Dieser Fehler wird sehr groß, wenn die Prädiktoren angenähert kollinear sind, wie sich aus den folgenden Überlegungen ergibt.

Für die Kovarianzmatrix von b ergibt sich

$$\text{(52)} \qquad \boldsymbol{\Sigma}_{bb} = \sigma^2\,(\mathbf{X'X})^{-1}.$$

Nach Definition der Kovarianzmatrix gilt

(53) $$E\,\mathbf{bb}' = \boldsymbol{\beta\beta}' + \boldsymbol{\Sigma}_{bb}$$

und damit gemäß (52)

(54) $$E\,\mathbf{bb}' = \boldsymbol{\beta\beta}' + \sigma^2\,(\mathbf{X}'\mathbf{X})^{-1}.$$

Daraus ergibt sich weiter

(55) $$E \sum_k b_k^2 = \text{Spur}\,(E\,\mathbf{bb}') = \sum_k \beta_k^2 + \sigma^2\,\text{Spur}\,(\mathbf{X}'\mathbf{X})^{-1}.$$

Bezeichnen wir mit $\lambda_1, \ldots, \lambda_K$ die Eigenwerte von X'X, dann ergibt sich aus (55)

(56) $$E \sum_k b_k^2 = \sum_k \beta_k^2 + \sigma^2 \sum_k (1/\lambda_k).$$

Aus (56) ist ersichtlich, daß der Erwartungswert der Summe der Quadrate der b_k um so mehr von der Summe der Quadrate der β_k abweicht, je kleinere Eigenwerte von X'X auftreten. Wenn aber die Prädikatoren fast kollinear sind, so ist mindestens ein Eigenwert von X'X relativ klein. Dementsprechend wird auch der mittlere quadratische Fehler

(57) $$E \sum_k (b_k - \beta_k)^2 = \sigma^2 \sum_k (1/\lambda_k)$$

relativ groß, so daß sich im Fall angenäherter Kollinearität sehr ungenaue Schätzungen ergeben. Die Ungenauigkeit der Schätzungen beruht letztlich auf mangelnder Information und könnte nur durch zusätzliche a priori Information kompensiert werden.

Um nun zu verhindern, daß die erwartete Quadratsumme der Schätzungen zu groß wird (siehe (56)), werden die β_k nach der Minimum-Quadrat-Methode durch einen Vektor ß unter der folgenden Nebenbedingung

(58) $$\sum_k \tilde{\beta}_k^2 = c$$

geschätzt. Mit Hilfe des Lagrange-Ansatzes ergibt sich das Minimierungsproblem

(59) $$(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) + h\,\tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{\beta}} \stackrel{!}{=} \text{Minimum},$$

wobei h der Lagrange-Multiplikator ist. Daraus resultieren die Normalgleichungen

(60) $$\mathbf{X}'\mathbf{X}\tilde{\boldsymbol{\beta}} + h\tilde{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$$

bzw. die für $h \geqq 0$ stets existierende Lösung

(61) $$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X} + h\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}.$$

Man könnte nun wegen (58) h durch c ausdrücken und als Funktion von c ersetzen. Da dieser Ersetzungsprozeß nur implizit möglich ist, parametrisieren wir ß durch h, d.h. für jedes h wird vermöge (61) ein Schätzer festgelegt. Speziell resultiert für h = 0 der gewöhnliche Minimum-Quadrat-Schätzer. Der Schätzer ß heißt Ridge-Regression-Schätzer (R.R.-Schätzer).

Es gibt noch eine Verallgemeinerung, die sogenannte generalisierte Ridge-Regression-Schätzung (G.R.R.-Schätzer), bei der das Vielfache der Einheitsmatrix in (61) durch eine Diagonalmatrix H ersetzt wird

(62) $$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X} + \mathbf{H})^{-1}\mathbf{X}'\mathbf{y}.$$

Im folgenden wird nur die R.R.-Schätzung betrachtet. Zuerst wird gezeigt, daß der R.R.-Schätzer verzerrt ist. Es gilt nämlich

(63) $$\mathrm{E}\,\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X} + \mathrm{h}\mathbf{I})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta}) = \boldsymbol{\beta} - \mathrm{h}(\mathbf{X}'\mathbf{X} + \mathrm{h}\mathbf{I})^{-1}\boldsymbol{\beta}.$$

Also ergibt sich die Verzerrung

(64) $$\mathrm{E}(\tilde{\boldsymbol{\beta}}) - \boldsymbol{\beta} = -\,\mathrm{h}(\mathbf{X}'\mathbf{X} + \mathrm{h}\mathbf{I})^{-1}\boldsymbol{\beta}.$$

die von ß abhängt.

Der Vorteil des R.R.-Schätzers ist, daß es ein h gibt, so daß der mittlere quadratische Fehler $\mathrm{E}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta})'(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta})$ kleiner ist als der des gewöhnlichen Minimum-Quadrat-Schätzers $\mathrm{E}(\mathbf{b}-\boldsymbol{\beta})'(\mathbf{b}-\boldsymbol{\beta})$, wie Hoerl & Kennard (1970) gezeigt haben. Von Theobald (1974) wurde gezeigt, daß für

$$< h < h_{max} = \frac{2\sigma^2}{\boldsymbol{\beta}'\boldsymbol{\beta}}$$ und alle positiv definiten Matrizen W gilt

(65) $$\mathrm{E}\,(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{W}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) < \mathrm{E}\,(\mathbf{b} - \boldsymbol{\beta})'\mathbf{W}(\mathbf{b} - \boldsymbol{\beta}).$$

Wird speziell W = I gewählt, so ergibt sich eine Aussage, die der von Hoerl & Kennard (1970) entspricht. Da aber h_{max} von den unbekannten β_k abhängt, hat die Aussage (65) nur geringe praktische Bedeutung.

Es gibt auch Versuche, h_{max}, und damit h, in Abhängigkeit von y und X zu schätzen. Dann werden h_{max} und h stochastisch und der entsprechende R.R.-Schätzer nichtlinear. Die theoretischen Eigenschaften dieses Schätzers sind unbekannt.

Um sich einen Überblick über das Verhalten des R.R.-Schätzers in Abhängigkeit von h zu verschaffen, werden die β_k als Funktionen von h grafisch dargestellt. Diese Darstellung wird als Ridge Trace bezeichnet.

Für die praktische Anwendung des R.R.-Schätzers ist es von Bedeutung, daß die Minimum-Quadrat-Schätzung in folgendem Sinne instabil ist: Bei angenähert kollinearen Prädiktoren ändern sich die b_k aufgrund geringerer Änderungen der Daten in erheblichem Maße. Für die R.R.-Schätzung verringert sich diese Instabilität; obwohl (und weil) sie nicht erwartungstreu ist, hat sie einen kleineren mittleren quadratischen Fehler.

Als ergänzende Lektüre können Hoerl & Kennard (1970), Theobald (1974), Vinod (1978) und Jahn (1979) empfohlen werden.

1.7 Klassisches korrelationsanalytisches Modell und multiple Korrelation

Bei der Diskussion des regressionsanalytischen Modells wurde auf die besondere Bedeutung der Annahme hingewiesen, daß die Werte X der Prädiktoren feste Werte sind. Dies bedeutet für die Interpretation der Wahrscheinlichkeit als relative Häufigkeit im „long run", daß die Replikation der Beobachtungen zu immer denselben Prädiktorwerten erfolgen muß. Mit Replikation der Beobachtungen ist allerdings nicht die Wiederholung der Beobachtungen zum Aufbau einer Stichprobe, sondern die Wiederholung der Untersuchung an einer neuen Stichprobe gemeint. Die Annahme fester Prädiktoren ist für viele psychologische Fragestellungen unrealistisch. In diesem Kapitel soll gezeigt werden, unter welchen veränderten Annahmen mit etwas abweichender Interpretation die bisher berichteten Ergebnisse beibehalten werden können, wenn stochastische Prädiktoren vorliegen, also die Prädiktoren selbst Zufallsvariablen sind. Als nicht leicht verständliche, aber umfassende Lektüre zu diesem Thema können Schönfeld (1971, p.1) oder Graybill (1976, p. 373) empfohlen werden. Eine erste Einführung in das korrelationsanalytische Modell bieten Gaensslen & Schubö (1976, p. 47).

Das *klassische korrelationsanalytische Modell (general linear regression model, unabhängige stochastische lineare Regression)* umfaßt folgende Annahmen:

(S1) Es besteht die Beziehung
$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u},$$
wobei X eine Nx(K+ l)-Matrix von beobachtbaren Zufallsgrößen, ß ein Vektor von festen, unbekannten Parametern mit K + l Komponenten, y ein Vektor von N beobachtbaren Zufallsgrößen und u ein Vektor von N unbeobachtbaren Zufallsgrößen ist. X und u (und damit y) besitzen eine gemeinsame Wahrscheinlichkeitsverteilung.

(S2) Der Rang der Zufallsmatrix X ist mit Wahrscheinlichkeit 1 gleich der Spaltenzahl K + 1.

(S3) $E_u(\mathbf{u}|\mathbf{X}) = 0$ für alle Realisationen $\mathbf{X}$.

(S4) $\boldsymbol{\Sigma}(\mathbf{u},\mathbf{u}|\mathbf{X}) = E_u(\mathbf{u}\,\mathbf{u}'|\mathbf{X}) = \sigma^2\mathbf{I}$ für alle Realisationen $\mathbf{X}$.

(S5) Für jede Realisation X ist die bedingte Verteilung von u N-dimensional normal.

(S6) Es existieren EX'X, $E(X'X)^{-1}$ und $E(X'X)^{-1}X'$.
EX'X ist nichtsingulär.

(S7) Die Rand-Dichtefunktion $f_x(X)$ ist für alle Realisationen X positiv.

Die Annahmen Sl bis S5 entsprechen weitgehend den früheren Annahmen Al bis A5. Um in nicht zu große Schwierigkeiten bei der Darstellung zu gelangen, wurde für S4 auf die Möglichkeit einer beliebigen Kovarianzmatrix verzichtet. Ansonsten wurde dem Umstand Rechnung getragen, daß nunmehr Zufallsvariablen auch für X vorliegen, wodurch auch die zusätzlichen Annahmen S6 und S7 erforderlich werden. Weiterhin wird vorausgesetzt, daß die Prädiktoren fehlerfrei gemessen sind.

Wichtig ist die Unterscheidung zwischen dem bedingten und unbedingten Erwartungswert; der bedingte Erwartungswert $E_u(\mathbf{u}|\mathbf{X})$ bedeutet den Erwartungswert von u unter der Bedingung, daß X einen festen Wert annimmt, der unbedingte Erwartungswert Ex,,(u) = Eu ist der Erwartungswert in der Randverteilung von u. Unter den Annahmen S3 und S7 gilt jedoch auch

(66) $$E_{X,u}(\mathbf{u})\ E_X(E_u(\mathbf{u}\,|\,\mathbf{X})) = E_X(0) = 0,$$

so daß der unbedingte Erwartungswert wie der bedingte verschwindet. Ebenso ist wegen S4 und S7

(67) $$\mathbf{\Sigma}_{uu} = \sigma^2\mathbf{I}.$$

Durch Bildung des bedingten Erwartungswerts für die Regressionsbeziehung in Sl erhält man für jede Realisierung X

(68) $$E_u(\mathbf{y}\,|\,\mathbf{X}) = \mathbf{X\beta}.$$

Insofern entspricht $E_u(y \mid X)$ einerseits E_y aus Gleichung (16) und andererseits der Approximation $\hat{y}$ aus Gleichung (4) der beschreibenden linearen Regression (vgl. Steyer, 1979; Moosbrugger, 1980a, 1980b). Als Konsequenzen der Annahmen des stochastischen Modells (ohne die Normalverteilungsannahme S5) gilt: Die Fehler u und die Prädiktoren X sind unkorreliert, können aber stochastisch abhängig sein:

(69) $$E_{X,u}(\mathbf{X}'\mathbf{u}) = 0.$$

Unter den Annahmen Sl bis S7 behalten alle Ergebnisse (bis auf eine noch zu erwähnende Ausnahme) Gültigkeit, die wir im klassischen regressionsanalytischen Modell (A1 bis A3, K4, A5) dargestellt hatten, alle Schätzer bleiben erwartungstreu, die Bereichsschätzungen sind nach wie vor gültig. Dies ist der Fall, obwohl für die Verteilungsform der Prädiktoren keine Aussage gemacht wurde. So ist z.B.

(70) $$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

nicht notwendig normalverteilt, wie dies im klassischen regressionsanalytischen Modell der Fall ist. Der Schätzer (70) ist übrigens in den Zufallsvariablen nichtlinear. Daraus resultiert die angekündigte Ausnahme: Er kann daher auch kein bester linearer erwartungstreuer Schätzer sein; er ist jedoch asymptotisch effizient (Schönfeld, 1971, p. 17).

Im korrelationsanalytischen Modell ist es auch natürlich, ein Maß für die Stärke des Zusammenhangs

(71) $$\varrho^2 = \varrho^2_{y(01\ldots K)} = \frac{\operatorname{Spur} \boldsymbol{\Sigma}_{yy} - N\sigma^2}{\operatorname{Spur} \boldsymbol{\Sigma}_{yy}}$$

$$= \frac{\boldsymbol{\beta}'[E_X(\mathbf{X}'\mathbf{X}) - \frac{1}{N} E_X(\mathbf{X}')\,\mathbf{J}\,E_X(\mathbf{X})]\,\boldsymbol{\beta}}{\boldsymbol{\beta}'[E_X(\mathbf{X}'\mathbf{X}) - \frac{1}{N} E_X(\mathbf{X}')\,\mathbf{J}\,E_X(\mathbf{X})]\,\boldsymbol{\beta} + N\sigma^2}$$

mit der Einsmatrix J, d.h. $(J)k_l = 1$,

in Form des multiplen Bestimmungsmaßes aufzustellen. Die Wurzel ϱ aus dem Bestimmungsmaß heißt *multiple Korrelation*. Da durch die Konstante x_0 kein Zusammenhang vermittelt wird, läßt man sie häufig in der Notation der multiplen Korrelation weg:

(72) $$\varrho = \varrho_{y(012\ldots K)} = \varrho_{y(12\ldots K)}.$$

Eine alternative Darstellungsform der multiplen Korrelation für den Spezialfall, daß Kriterium und (echte) Prädiktoren insgesamt (K + 1)-dimensional normalverteilt sind, wird in Graybill (1976, p. 113, 418) beschrieben. Wählt man für die Prädiktoren eine Punktverteilung, was dem klassischen regressionsanalytischen Modell entspricht, ist die Definition (71) nach wie vor anwendbar, wobei die Symbole E_X entfallen. In diesem Spezialfall wird besonders deutlich, daß der multiple Korrelationskoeffizient von den gewählten Prädiktorwerten abhängt; allgemein hängt er von den Charakteristika der Verteilung der Prädiktoren ab, wie natürlich auch von ß und σ^2.

Als *Punktschätzung* für ϱ^2 kann entsprechend Gleichung (71)

(73) $$R^2 = \frac{\mathbf{b}'\left[\mathbf{X}'\mathbf{X} - \frac{1}{N}\mathbf{X}'\mathbf{J}\mathbf{X}\right]\mathbf{b}}{\mathbf{b}'\left[\mathbf{X}'\mathbf{X} - \frac{1}{N}\mathbf{X}'\mathbf{J}\mathbf{X}\right]\mathbf{b} + (N-K-1)\hat{\sigma}^2}$$

$$= \frac{\mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{y} - \frac{1}{N}\,\mathbf{y}'\mathbf{J}\mathbf{y}}{\mathbf{y}'\mathbf{y} - \frac{1}{N}\,\mathbf{y}'\mathbf{J}\mathbf{y}}$$

verwendet werden; dies ist genau gleich der schon in Gleichung (32) als empirische multiple Korrelation eingeführten Größe R^2. Mit der Matrix der empirischen Kovarianzen der Prädiktoren

(74) $$\mathbf{S}_{XX} = \frac{1}{N}(\mathbf{X}'\mathbf{X} - \frac{1}{N}\ \mathbf{X}'\mathbf{J}\mathbf{X}) = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & s_1^2 & s_{12} & \dots & s_{1K} \\ 0 & s_{21} & s_2^2 & \dots & s_{2K} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & s_{K1} & s_{K2} & \dots & s_K^2 \end{bmatrix},$$

dem Vektor der empirischen Kovarianzen von Prädiktoren und Kriterium

(75) $$\mathbf{s}_{Xy} = \frac{1}{N}(\mathbf{X}'\mathbf{y} - \frac{1}{N}\ \mathbf{X}'\mathbf{J}\mathbf{y}) = \begin{bmatrix} 0 \\ s_{1y} \\ s_{2y} \\ \vdots \\ s_{Ky} \end{bmatrix}$$

und der Varianz des Kriteriums

(76) $$s_{yy} = \frac{1}{N}\ (\mathbf{y}'\mathbf{y} - \frac{1}{N}\ \mathbf{y}'\mathbf{J}\mathbf{y}) = \overline{y^2} - \bar{y}^2 = s_y^2$$

erhält man

(77) $$R^2 = \frac{\mathbf{b}'\ \mathbf{S}_{XX}\mathbf{b}}{s_{yy}} = \frac{\mathbf{s}'_{Xy}\ \mathbf{b}}{s_{yy}} = \frac{\mathbf{s}'_{Xy}\ \mathbf{S}_{XX}^{-1}\ \mathbf{s}_{Xy}}{s_{yy}}\ .$$

Wenn die Konstante x_0 aus den Matrizen und Vektoren ausgeschlossen wird, ändert sich am Ergebnis nichts, d.h. daß in (74) die erste Zeile und erste Spalte, in (75) die erste Komponente entfallen kann. Wenn für alle Variablen y und x_k eine multivariate Normalverteilung vorliegt, ist R^2 der Maximum-Likelihood-Schätzer. R^2 überschätzt im Mittel die Modell-Korrelation und ist demgemäß nicht erwartungstreu (siehe Olkin & Pratt, 1958). Wenn man die Schrumpfungskorrektur

(78) $$\tilde{R}^2 = 1 - (1 - R^2)\frac{N-1}{N-K-1}$$

anwendet, ist der Bias dieser modifizierten Schätzung geringer. Für den Fall der einfachen Korrelation ist

(79) $$\tilde{r} = r + \frac{r\ (1 - r^2)}{2\ (N - 4)}$$

eine gute Näherung für den unter der Voraussetzung der Binomalverteilung konstruierbaren (Olkin & Pratt, 1958; Graybill, 1976, p. 396) besten unverzerrten Schätzer von ϱ.

Eine *Bereichsschätzung* für die multiple Korrelation ϱ im Fall der Multinormalverteilung ist in Graybill (1976, p. 419, 667) in tabellarischer Form angegeben. Für den Fall der einfachen Korrelation (K = 1) nützt man die Tatsache aus, daß für die bivariate Normalverteilung

(80) $$z = \frac{\operatorname{arctanh}(r) - \operatorname{arctanh}(\varrho)}{\sqrt{N-3}}$$

näherungsweise standardnormalverteilt ist, um Bereichsschätzungen zu bilden. Ein näherungsweiser Konfidenzbereich zum Konfidenzgrad 1 - α ist daher durch

(81) $$\tanh\left(\operatorname{arctanh}\left(r - \frac{N_{\alpha/2}}{\sqrt{N-3}}\right)\right) \leqq \varrho \leqq \tanh\left(\operatorname{arctanh}\left(r + \frac{N_{\alpha/2}}{\sqrt{N-3}}\right)\right)$$

gegeben.

Aus den genannten Bereichsschätzungen lassen sich wiederum *Signifikanztests* konstruieren. Die Nullhypothese $\varrho^{(0)} = 0$ ist der Nullhypothese $\beta_1^{(0)} = \beta_2^{(0)} = \ldots = \beta_K^{(0)} = 0$ gleichwertig. Deswegen eignet sich zur Prüfung dieser Nullhypothese ebenso die Entscheidungsregel (34), die auch im korrelationsanalytischen Modell anwendbar ist.

Da viele Begriffsbildungen und Ergebnisse für das klassische regressionsanalytische Modell und das klassische korrelationsanalytische Modell gleichermaßen gelten, soll im folgenden vom *klassischen Modell* oder vom *Regressionsmodell* (manche Autoren sprechen vom Allgemeinen linearen Modell) die Rede sein, wenn es sich um gemeinsame Eigenschaften der beiden Ansätze handelt.

1.8 Modelle mit Fehlern in den Prädiktoren

Üblicherweise geht man in der Regressionsanalyse davon aus, daß die beobachtbaren Variablen exakt, d.h. ohne Fehler gemessen werden können. In der Regel sind die Variablen aber fehlerbehaftet; die Ursache liegt in zahlreichen Fehlerquellen, z.B. in der Erhebungstechnik (falsche Abgrenzungen, falsche Antworten), am Stichprobenfehler, an der Aufbereitung der Daten usw. Im folgenden werden Modelle mit fehlerbehafteten Variablen betrachtet. Als Literatur hierzu können Klein (1953, p. 282-305), Malinvaud (1966, p. 326--363), Cochran (1970), Schönfeld (1971, p. 100-124), Schneeweiß (1974, p. 216-232), Stroud (1974), Namboodiri et al. (1975), Wainer & Thyssen (1976), Aigner & Goldberger (1977), Maddala (1977, p. 292-319), Rock et al. (1977) und Dhrymes (1978, p. 242-268) herangezogen werden.

Es müssen nun für das Kriterium und jeden Prädiktor jeweils zwei Werte unterschieden werden: der richtige bzw. der tatsächlich beobachtete und damit fehlerbehaftete. Um ersteren vom zweiten Wert in der Notation zu unterscheiden, wird er jeweils mit einem Stern versehen.

Aus schreibtechnischen Gründen wird nun für das Kriterium nicht mehr y, sondern s_0 und für die Prädiktoren einschließlich der bisherigen Konstante $x_0 = 1$ in abgeänderter Numerierung x_k geschrieben ($1 \leqq k \leqq K$). Es bezeichnet also x^*_{0n} bzw. x_{0n} das tatsächliche bzw. beobachtete Kriterium zur Erhebungseinheit n und x^*_{kn} bzw. X_{kn} den tatsächlichen bzw. beobachteten k-ten Prädiktor zur Erhebungseinheit n.

Der Fehler in der Variablen X_{kn} ($0 \leqq k \leqq K$, $1 \leqq n \leqq N$) wird mit v_{kn} bezeichnet. Dieser Fehler kann additiv oder prozentual sein. Es wird ein additiver Zusammenhang unterstellt:

$$x_{kn} = x^*_{kn} + v_{kn} \quad (0 \leqq k \leqq K,\ 1 \leqq n \leqq N).$$

Die Regressionsbeziehung lautet:

$$x^*_{0n} = \sum_{k=1}^{K} \beta_k x^*_{kn} + \varepsilon_n \quad (1 \leqq n \leqq N),$$

wobei ε_n den Fehler in der Gleichung erfaßt.

In kompakter Matrizenform schreiben sich obige Beziehungen

(82) $$\mathbf{X} = \mathbf{X}^* + \mathbf{V}, \quad \mathbf{x}_0 = \mathbf{x}_0^* + \mathbf{v}_0$$

(83) $$\mathbf{x}_0^* = \mathbf{X}^* \boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

wobei X bzw. X* bzw. V jeweils NxK-Matrizen darstellen, ß einen Spaltenvektor mit K Komponenten und x_0^* bzw. x_0 bzw. $\boldsymbol{\varepsilon}$ Spaltenvektoren mit je N Komponenten. Wird x_0^* in (83) gemäß (82) ersetzt, so ergibt sich

(84) $$\mathbf{x}_0 = \mathbf{X}^* \boldsymbol{\beta} + \boldsymbol{\varepsilon} + \mathbf{v}_0.$$

Aus (84) resultiert, daß der Effekt des Fehlers in der Gleichung, nämlich $\boldsymbol{\varepsilon}$, derselbe ist wie der des Fehlers in der Variable x_0, nämlich v_0. Wir setzen daher ohne Beschränkung der Allgemeinheit ε zu Null.

Wird weiter X* in (84) gemäß (82) ersetzt, so ergibt sich die entsprechende Beziehung in den beobachtbaren Variablen

(85) $$\mathbf{x}_0 = \mathbf{X}\boldsymbol{\beta} + (\mathbf{v}_0 - \mathbf{V}\boldsymbol{\beta}).$$

Der Fehler in der Gleichung (85) wird zusammengefaßt zu

$$\mathbf{u} = \mathbf{v}_0 - \mathbf{V}\boldsymbol{\beta}.$$

Damit resultiert die übliche Regressionsbeziehung

(86) $$\mathbf{x}_0 = \mathbf{X\beta}+\mathbf{u}.$$

Folgende Annahmen die Fehler betreffend werden getroffen:

(F1) Die Fehlervariable v_{kn} ist unabhängig von v_{lm} für $n \neq m$ und $0 \leq k$, $1 \leq K$.

(F2) $(v_{0n}, v_{1n}, \ldots, v_{Kn})$ besitzt eine von n unabhängige Verteilung und ω_{kl} bezeichnet die Kovarianz von v_{kn} und v_{ln} bzw. $\mathbf{\Omega}$ die entsprechende Kovarianzmatrix. Mit $\mathbf{\Omega}_{11}$ bezeichnen wir die Matrix, die aus $\mathbf{\Omega}$ durch Weglassen der ersten Zeile und Spalte entsteht, und mit $\boldsymbol{\omega}_1$ die erste Spalte von $\mathbf{\Omega}$ aber ohne ω_{00}.

(F3) Die Fehlervariablen v_{kn} sind alle unabhängig von den x^*_{ln} für $0 \leq k$, $l \leq K$, $1 \leqq n \leqq N$.

(F4) $Ev_{kn} = 0$ für $0 \leqq k \leqq K$, $1 \leqq n \leqq N$.

Es wird zugelassen, daß einige der Variablen fehlerfrei beobachtbar sind, d.h. daß einige ω_{kk} verschwinden können. Nur ω_{00} sei ungleich Null, da in v_0 auch der Fehler in der Gleichung miterfaßt wird.

Nun müssen noch die tatsächlichen Prädiktoren spezifiziert werden. Es gibt zwei Möglichkeiten:

1. Die stochastische Spezifikation, nach der alle x^*_{kn} $(1 \leqq k \leqq K,\ 1 \leqq n \leqq N)$ stochastische Variable sind (und damit auch x^*_{0n});
2. Die deterministische Spezifikation, nach der alle x^*_{kn} fest vorgegeben sind. In diesem Falle ist wegen (83) (man beachte $\boldsymbol{\varepsilon}=0$) auch x^*_{0n} fest vorgegeben.

Bei stochastischer Spezifikation soll noch gelten

(F5) $$\lim_N \frac{1}{N} E(\mathbf{X}^{*\prime}\mathbf{X}^*) = \operatorname*{plim}_N \frac{1}{N}\mathbf{X}^{*\prime}\mathbf{X}^*,$$
und die Grenzwertmatrix M sei nichtsingulär.

Bei deterministischer Spezifikation soll gelten

(F5') $$\lim_N \frac{1}{N}\mathbf{X}^{*\prime}\mathbf{X}^*$$
existiert und die Grenzwertmatrix sei nichtsingulär.

Zuerst wird unabhängig von der Spezifikation der Minimum-Quadrat-Schätzer für (86) betrachtet. Nach Definition von u und gemäß (86) ergibt sich

$$E\mathbf{X}'\mathbf{u} = E(\mathbf{X}^*-\mathbf{V})'(\mathbf{v}_0-\mathbf{V\beta}) = -E\mathbf{V}'\mathbf{v}_0+(E\mathbf{V}'\mathbf{v})\boldsymbol{\beta}.$$

Daraus ergibt sich (auch für den Fall, daß V und v_0 unkorreliert sind) eine Korrelation der beobachtbaren Prädiktoren mit der Fehlervariablen u. Daher ist der Minimum-Quadrat-Schätzer

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}$$

verzerrt.

Gemäß (83) gilt aber (beachte $\boldsymbol{\varepsilon}=0$)

$$\mathbf{X}^{*\prime}\mathbf{X}^{*}\boldsymbol{\beta} = \mathbf{X}^{*\prime}\mathbf{x}_0^{*}.$$

Darin ersetzen wir X^* und x_0^* gemäß (82). Es ergibt sich also

$$(\mathbf{X}-\mathbf{V})'(\mathbf{X}-\mathbf{V})\boldsymbol{\beta} = (\mathbf{X}-\mathbf{V})'(\mathbf{x}_0-\mathbf{v}_0)$$

bzw. durch Ausmultiplizieren,

$$(\mathbf{X}'\mathbf{X}-\mathbf{V}'\mathbf{X}-\mathbf{X}'\mathbf{V}+\mathbf{V}'\mathbf{V})\boldsymbol{\beta} = \mathbf{X}'\mathbf{x}_0-\mathbf{V}'\mathbf{x}_0-\mathbf{X}'\mathbf{v}_0+\mathbf{V}'\mathbf{v}_0. \tag{87}$$

Wegen der getroffenen Annahmen ergibt sich asymptotisch

$$\text{plim } \frac{1}{N}\mathbf{V}'\mathbf{X} = \text{plim } \frac{1}{N}\mathbf{X}'\mathbf{V} = 0$$

$$\text{plim } \frac{1}{N}\mathbf{V}'\mathbf{x}_0 = \text{plim } \frac{1}{N}\mathbf{X}'\mathbf{v}_0 = 0$$

$$\text{plim } \frac{1}{N}\mathbf{V}'\mathbf{V} = \boldsymbol{\Omega}_{11}$$

und

$$\text{plim } \frac{1}{N}\mathbf{V}'\mathbf{v}_0 = \boldsymbol{\omega}_1.$$

Daher ist das folgende Gleichungssystem asymptotisch äquivalent zu (87)

$$\left(\frac{1}{N}\mathbf{X}'\mathbf{X} + \boldsymbol{\Omega}_{11}\right)\tilde{\boldsymbol{\beta}} = \frac{1}{N}\mathbf{X}'\mathbf{x}_0 + \boldsymbol{\omega}_1 \tag{88}$$

Wenn also $\boldsymbol{\Omega}_{11}$ und $\boldsymbol{\omega}_1$ bekannt sind, so ergibt sich gemäß (88) aus den beobachtbaren Variablen der Schätzer ß, der nach obiger Herleitung konsistent ist.

Dieser Schätzer setzt die Kenntnis von $\boldsymbol{\Omega}_{11}$ und $\boldsymbol{\omega}_1$ voraus. In vielen Fällen kann $\boldsymbol{\omega}_1$ gleich Null gesetzt werden und $\boldsymbol{\Omega}_{11}$ als diagonal angenommen werden, d.h. die Fehler in den verschiedenen Variablen sind untereinander unkorreliert. Trotzdem ist über die Beobachtungen hinaus eine gewisse a priori Kenntnis über die Varianzen der Fehler in den Variablen nötig, um die Parameter konsistent zu schätzen. Hier liegt ein Identifikationsproblem vor. Dieses wird im Rahmen der stochastischen Spezifikation und bei Vorliegen von Normalverteilungen noch genauer diskutiert.

Die Struktur der Verzerrung des Minimum-Quadrat-Schätzers kann für den Fall der Einfachregression $x^*_{0n} = \beta_1+\beta_2 x^*_{2n}$ genauer studiert werden. Das Absolutglied $x_{1n} = 1$ kann natürlich ohne Fehler beobachtet werden. Weiter sei $\omega_{01} = 0$.

Damit ergibt sich

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{x}_0$$

bzw.

$$\operatorname{plim} \hat{\boldsymbol{\beta}} = \left(\operatorname{plim} \frac{1}{N}\mathbf{X}^{*\prime}\mathbf{X}^* + \begin{pmatrix} 0 & 0 \\ 0 & \omega_{22} \end{pmatrix}\right)^{-1} \left(\operatorname{plim}\frac{1}{N} \mathbf{X}^{*\prime}\mathbf{X}^*\boldsymbol{\beta}\right)$$

und

$$\operatorname{plim} \hat{\beta}_2 = \beta_2 \frac{1}{1 + \frac{\omega_{22}}{s}}$$

wobei $s = \operatorname{plim} \frac{1}{N} \sum_n (x^*_{2n}-\bar{x}^*_2)^2$, $\bar{x}^*_2 = \frac{1}{N} \sum_n x^*_{2n}$.

Offensichtlich wird β_2 immer unterschätzt und zwar um so mehr, je größer das Verhältnis der Varianz von v_{2n} zu x^*_{2n} ist. Wenn wir z.B. $N = 10$ annehmen und $x^*_{2n} = c_1$ für $1 \leqq n \leqq 5$ bzw. $x^*_{2n} = c_2 \neq c_1$ für $6 \leqq n \leqq 10$, so ist es relativ klein. Grafisch ergibt sich die Darstellung in Abb. 2.

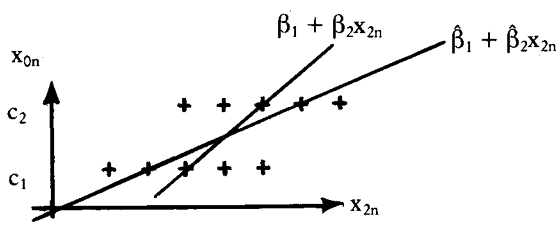

Abb. 2: Verzerrung der Regressionskoeffizienten, wenn der Prädiktor mit Meßfehler gemessen wurde.

Es seien nun alle Fehlervariablen v_{kn} und alle x^*_{kn} für $0 \leqq k \leqq K$, $1 \leqq n \leqq N$ normalverteilt. Mit μ_k wird der Erwartungswert von x^*_{kn} bezeichnet $(0 \leqq k \leqq K)$. Es gilt dann wegen (83)

$$(89) \qquad \mu_0 = \sum_{k=1}^{K} \mu_k \beta_k$$

Mit $\boldsymbol{\Sigma} = (\sigma_{kl})$ wird die (K+l)x(K+l)-Kovarianz-Matrix von $(x_{0n},\ldots,x_{kn})$ bezeichnet und mit $\boldsymbol{\Sigma}^* = (\sigma^*_{kl})$ die KxK-Kovarianz-Matrix von $(x^*_{1n},\ldots,x^*_{kn})$.

Es gilt dann

$$\sigma_{kl} = \sigma^*_{kl}+\omega_{kl} \text{ für } 1 \leqq k, l \leqq K$$

$$\sigma_{00} = \sum_{k,l=1}^{K} \beta_k\sigma^*_{kl}\beta_l+\omega_{00}$$

$$\sigma_{0l} = \sum_{k=1}^{K} \beta_k\sigma^*_{kl}+\omega_{0l} \text{ für } 1 \leqq l \leqq K.$$

Mit diesen Bezeichnungen ergibt sich für die Log-Likelihood-Funktion (bis auf einen unwesentlichen konstanten Term)

$$-\frac{N}{2}\ln(|\det \boldsymbol{\Sigma}|)-\frac{1}{2}\sum_{n=1}^{N}\sum_{k,l=0}^{K}(x_{kn}-\mu_k)\sigma_{kl}^{(-1)}(x_{ln}-\mu_l).$$

Aus den beobachtbaren Variablen können wir bestenfalls die Matrix $\boldsymbol{\Sigma}$ und $\mu_0,\ldots,\mu_K$ erhalten. Die eigentlich interessierenden Parameter sind aber $\mu_1,\ldots,\mu_K$, $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}^*$ und $\boldsymbol{\Omega}$.

Offensichtlich sind letztere viel zahlreicher und damit nicht alle identifizierbar. Durch geeignete a priori Information z.B., daß $\boldsymbol{\Omega}$ diagonal ist und gewisse Hauptdiagonalelemente Null sind, kann Identifizierbarkeit erreicht werden.

Für den Fall der deterministischen Spezifikation kann eine andere Schätzfunktion hergeleitet werden. Die ersten $K_1 \leqq K$ Variablen seien fehlerbehaftet und die restlichen exakt beobachtbar. Diese K_1 ersten Variablen seien als X^*_1 und V_1 definiert. Damit gilt nun anstatt (82) und (83)

(82') $$\mathbf{X}_1 = \mathbf{X}^*_1+\mathbf{V}_1$$

und

(83') $$\mathbf{x}^*_0 = \mathbf{X}^*\boldsymbol{\beta}$$

Sei $\boldsymbol{\Omega}_1$ die nichtsinguläre Kovarianz-Matrix der Zeilen von V_1, dann wird folgende gewichtete Minimum-Quadrat-Methode angewandt

$$\text{Spur } (\mathbf{X}_1-\mathbf{X}^*_1)\boldsymbol{\Omega}_1^{-1}(\mathbf{X}_1-\mathbf{X}^*_1) \overset{!}{=} \text{Minimum}$$

unter der Nebenbedingung (83'). Dabei werden die Größen x^*_{kn} als Parameter angesehen, die „mitgeschätzt" werden. Der resultierende Schätzer ist konsistent.

1.9 Zeitreihenanalyse im allgemeinen regressionsanalytischen Modell

In neuerer Zeit gewinnt in der Psychologie die auch quantitative Untersuchung von zeitlichen Prozessen größere Bedeutung, mit der die Untersuchung von Zuständen ergänzt werden kann. Dies hat wohl nicht zuletzt seinen Grund darin, daß wichtige Teile des quantitativ-methodischen Instrumentariums für die speziellen Bedürfnisse der Veränderungsbeschreibung erst jüngst entwickelt (Yule, 1927; Hannan, 1955; Hannan, 1960; Box und Tiao, 1965; Kendall & Stuart, 1966, p. 403-430; Malinvaud, 1966, p. 373-393, 425-453; Box et al., 1967; Box & Jenkins, 1970; Box & Pierce, 1970, Box & Tiao, 1975; Glass et al., 1975; Anderson, 1976; Kendall, 1976) oder in die psychologische Literatur (Holtzmann, 1963; Huber, 1967; Huber, 1973; Dahme, 1975; Dahme, 1976; Gudat & Revenstorff, 1976; Hersen & Barlow, 1976; Petermann, 1977a, 1977b, 1978; Revenstorff, 1979; Revenstorff & Keeser, 1979; McCleary & Hay, 1980; Keeser, 1981) aufgenommen wurden.

Das klassische regressionsanalytische Modell

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

mit K4, wozu die Unkorreliertheit der Komponenten u_n und u_m für $n \neq m$, $1 \leqq n, m \leq N$, gehört, ist für Zeitreihenanalysen deswegen nicht geeignet, weil in der Regel der nicht durch X erklärbare Anteil des Kriteriums, die Fehlervariable u, durch nicht erfaßte, zeitlich träge Größen beeinflußt wird. Bei einem Experiment zur Rauchertherapie wird zwar einerseits das Kriterium y, die Anzahl gerauchter Zigaretten, wesentlich durch die in den Prädiktorvariablen x_k dargestellten Interventionen bestimmt sein, andererseits werden sich zeitlich kurzfristig und längerfristig wirksame Einflußgrößen bemerkbar machen, deren Wirkung auf das Kriterium durch die Fehlervariable u zusammengefaßt wird. Beispielsweise wird ein zum Zeitpunkt t_n auftretender Partnerkonflikt eines Klienten sich so auswirken, daß für die Anzahl gerauchter Zigaretten die Fehlerwerte u_n, u_{n+1},. . . mit sich abschwächender Tendenz systematisch erhöht sind. Insofern wird das Kriterium nun von drei Varianzquellen bestimmt, den Prädiktoren, den Fehlern und der seriellen Abhängigkeit der Fehler. Unter bestimmten Umständen, die an einem Beispiel noch näher erläutert werden, können derartige Zusammenhänge durch den Ansatz einer von der Einheitsmatrix I verschiedenen Matrix V im allgemeinen regressionsanalytischen Modell berücksichtigt werden.

Leider ist es so, daß in praktischen Fällen nicht einmal die Matrix V als Indikator der Struktur der Kovarianzmatrix der Zufallsvariablen u_n

$$\mathbf{\Sigma}_{uu} = \sigma^2\mathbf{V} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \dots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} \dots & \sigma_{2N} \\ \vdots & \vdots & \\ \sigma_{N1} & \sigma_{N2} \dots & \sigma_{NN} \end{bmatrix} \tag{90}$$

bekannt ist. Die Matrix $\mathbf{\Sigma}_{uu}$ muß daher geschätzt werden, wobei es ausreicht, die wechselseitigen Verhältnisse der Kovarianzen zu schätzen, wie sie in V erfaßt sind. Leider hat man zum Schätzen der (N(N+ 1)/2)- 1 freien Parameter in V nur die N Werte y_n zur Verfügung, so daß eine konsistente Schätzung der Matrix V so nicht möglich ist. Die Zahl der freien Parameter in $\mathbf{\Sigma}_{uu}$ muß also dadurch verringert werden, daß ein angemessenes stochastisches Modell mit erheblich weniger freien Parametern angesetzt wird, das Beziehungen zwischen den Kovarianzen σ_{nm} definiert.

Ist V geschätzt, kann in einer zweiten Phase die Schätzung der Regressionskoeffizienten und σ^2 nach den in Kap. 1.3 beschriebenen Methoden erfolgen. Leider sind die dort angegebenen Eigenschaften der Schätzer nicht für die hier beschriebene *zweiphasige* Schätzung (two step estimator) gültig. Unter welchen Bedingungen andere, nur asymptotische Schätzeigenschaften nachgewiesen werden können, beschreibt etwa Schönfeld (1969, p. 207). Auch bei Vorliegen der erforderlichen Normalverteilungsannahme sind die in Kap. 1.3 berichteten Bereichsschätzungen und die daraus abgeleiteten Signifikanztests nur asymptotisch, für endliche Zeitreihen also nur näherungsweise gültig. Natürlich ist dabei die Näherung erheblich besser, als wenn die Abhängigkeit der Fehler ignoriert und fälschlich das klassische regressionsanalytische Modell angenommen wird (Keeser, 1981, p. 281). Wie sehr das Signifikanzniveau

Tabelle 1: Verzerrung der Irrtumswahrscheinlichkeit bei autokorrelierten Daten; relative Häufigkeiten von signifikanten Ergebnissen auf dem 5 %-Niveau von jeweils 1000 Stichproben von Zufallszahlen beim t-Test und Mann-Whitney-Test für verschiedene Verteilungsformen und verschiedene Werte der Autokorrelation (nach Box, 1974).

		Verteilungsform		
ϱ	Test	Normal	Rechteck	χ^2 (schief)
0.0	t-Test	.054	.056	.047
0.0	M-W-Test	.045	.043	.043
-.4	t-Test	.003	.005	.001
-.4	M-W-Test	.001	.005	.002
+.4	t-Test	.105	.125	.114
+.4	M-W-Test	.096	.110	.101

einfacher t-Tests (der t-Test ist der Test (37)) oder von Mann-Whitney-Tests verfälscht wird, wenn die Fehler autoregressiv korreliert sind (siehe unten), zeigen die Ergebnisse einer Simulationsstudie von Box (1974), die in Tab. 1 wiedergegeben ist.

Für die Anwendung des zweiphasigen Schätzers ist ein nichttriviales Problem das Aufstellen eines angemessenen und zugleich sparsamen stochastischen Modells für die Fehlervariablen, das in der Praxis anhand der vorliegenden Daten erfolgen muß. Eine Klasse von möglichen Modellen wird unter der Bezeichnung ARIMA-Modell (autoregressive-integrated moving average model) zusammengefaßt. Die Eigenschaften des Modells, die Methoden zur Identifikation des speziell für die Daten angemessenen Modells und die Schätzung der Parameter des Modells stellen z.B. McCleary & Hay (1980) und Keeser (1981) ausführlich dar. Computerprogramme zur Bearbeitung dieser Aufgabe sind in den neueren Versionen der Programme BMDP2T (Dixon & Brown, 1981), SPSS-Prozedur BOX-JENKINS (Beutel & Schubö, 1982) und in der Unterprogrammbibliothek IMSL (IMSL, 1979) enthalten.

Als einfachstes und am besten dokumentiertes (z.B. Johnston, 1963, p. 177; Schönfeld, 1969, p. 211; Schönfeld, 1971, p. 31; Schneeweiß, 1974, p. 180; Seifert, 1975, p. 109) Beispiel sollen die Ergebnisse beim autoregressiven Schema *erster Ordnung* (AR(l) oder ARIMA (l,0,0)) vorgestellt werden, woran das Prinzip der Vorgehensweise sichtbar wird. Dies ist der einfachste Ansatz, der jedoch leider für die meisten Anwendungen z.B. in der Psychotherapieforschung nicht ausreicht.

Als stochastisches Modell für die Fehlervariable wird angenommen:

(Z1) $u_n = \varrho u_{n-1} + e_n$ für $n = 2,3,\ldots,N$ mit

(Z2) $|\varrho| < 1$, $Eu_1 = 0$,

wobei für die Zufallsvariablen e_n gilt

(Z3) $E\, e_n = 0$ für $n = 1,2,\ldots,N$

(Z4) $\boldsymbol{\Sigma}_{ee} = \sigma_e^2\, I$

(Z5) $\operatorname{cov}(u_{n-1}, e_n) = 0$ für $n = 2,3,\ldots,N$

Es besteht also ein linearer Zusammenhang zwischen den Fehlervariablen zu unmittelbar aufeinanderfolgenden Zeitpunkten t_{n-1} und t_n, der selbst wieder von unregelmäßigen Zufallsschocks überlagert wird. Für die Beziehung aufeinanderfolgender Fehlervariablen wird demnach ein klassisches korrelationsanalytisches Modell angesetzt. Unter den Bedingungen Zl bis Z5 folgt y_n einem stationären Prozeß. Es gilt

(91) $$u_n = \varrho^{n-1} u_1 + \varrho^{n-2} e_2 + \ldots + \varrho\, e_{n-1} + e_n \text{ für } n = 2,3,\ldots,N$$

woraus sich wegen $|\varrho| < 1$ ablesen läßt, daß die Zufallsschocks für spätere

Zeitpunkte einen immer schwächer werdenden Einfluß haben. Die Kovarianzmatrix der Fehlervariablen hat die Form

(92)
$$\boldsymbol{\Sigma}_{uu} = \sigma^2 \mathbf{V} \text{ mit}$$
$$\sigma^2 = \sigma_u^2 = \frac{\sigma_e^2}{1-\varrho^2} \text{ und}$$
$$\mathbf{V} = \begin{bmatrix} 1 & \varrho & \varrho^2 & \dots \varrho^{N-1} \\ \varrho & 1 & \varrho & \dots \varrho^{N-2} \\ \vdots & \vdots & & \vdots \\ \varrho^{N-1} & \varrho^{N-2} & \varrho^{N-3} & \dots 1 \end{bmatrix}$$

Die für die Schätzung der Regressionskoeffizienten benötigte Inverse ist

(93)
$$\mathbf{V}^{-1} = \frac{1}{1-\varrho^2} \begin{bmatrix} 1 & -\varrho & 0 \dots & 0 & 0 & 0 \\ -\varrho & 1+\varrho^2 & -\varrho \dots & 0 & 0 & 0 \\ 0 & -\varrho & 1+\varrho^2 \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 \dots & -\varrho & 1+\varrho^2 & -\varrho \\ 0 & 0 & 0 \dots & 0 & -\varrho & 1 \end{bmatrix}$$

Kennt man die Autokorrelation ϱ, dann lassen sich damit in der zweiten Phase innerhalb des allgemeinen regressionsanalytischen Modells alle Schätzungen, Prognosen und Tests exakt durchführen, vorausgesetzt die Normalverteilungsannahme A5 ist erfüllt.

Im Falle des autoregressiven Schemas erster Ordnung (Wonnacott & Wonnacott, 1970, p. 328) wie im allgemeinen ARIMA-Modell (Keeser, 1981, p. 207) läßt sich die sog. Aitken-Reduktion auf das klassische regressionsanalytische Modell auf besonders einfache Weise durch Differenzbildung durchführen. Man zerlegt V^{-1} in zwei Faktoren

(94)
$$\mathbf{V}^{-1} = \mathbf{D}'\mathbf{D} \text{ mit}$$
$$\mathbf{D} = \frac{1}{\sqrt{1-\varrho^2}} \begin{bmatrix} \sqrt{1-\varrho^2} & 0 & 0 \dots & 0 & 0 \\ -\varrho & 1 & 0 \dots & 0 & 0 \\ 0 & -\varrho & 1 \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 \dots & 1 & 0 \\ 0 & 0 & 0 \dots & -\varrho & 1 \end{bmatrix}$$

und erhält in

(95)
$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \mathbf{u}^*$$

mit $\mathbf{y}^* = \mathbf{D}\mathbf{y}$, $\mathbf{X}^* = \mathbf{D}\mathbf{X}$, $\mathbf{u}^* = \mathbf{D}\mathbf{u}$

ein klassisches regressionsanalytisches Modell mit unkorrelierten Fehlervariablen $(\mathrm{Du})_n$. Daß es sich bei D um eine Matrix handelt, die Differenzen aufeinanderfolgender Messungen bildet, sei an

(96) $$\mathbf{Dy} = \begin{bmatrix} \sqrt{1-\varrho^2} & y_1 \\ y_2 - & \varrho y_1 \\ y_3 - & \varrho y_2 \\ \vdots & \\ y_N - & \varrho y_{N-1} \end{bmatrix}$$

demonstriert; wie man sieht, muß die erste Messung y_1 etwas abweichend behandelt werden. Es genügt für die Transformation in das klassische Modell übrigens nicht, nur die Kriteriumswerte zu transformieren, auch die Prädiktorwerte müssen, wie aus (95) ersichtlich, mittransformiert werden.

Bisher wurde davon ausgegangen, daß die Autokorrelation ϱ bekannt ist. Im allgemeinen muß man sie jedoch schätzen (Phase 1). Dazu bildet man die OLS-Schätzungen (klassisches Regressionsmodell) für die Residuen

(97) $$\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

und schätzt ϱ durch

(98) $$\hat{\varrho} = \frac{\sum_{n=2}^{N} \hat{u}_{n-1}\, \hat{u}_n}{\sum_{n=1}^{N-1} \hat{u}_n}$$

Dieser Schätzer ist nicht erwartungstreu. Die Schätzung setzt man anstelle des Parameters ϱ in den Gleichungen (92) bis (96) für die zweite Schätzphase ein, die sonst wie in dem Fall erfolgt, in dem ϱ a priori bekannt ist.

Bestehen Zweifel, ob überhaupt Autokorrelation der Residuen vorliegt, kann die Entscheidung darüber auch durch statistische Tests herbeigeführt werden. Pauschal kann etwa mit einem Runs-Test auf die geschätzten Residuen (97) entschieden werden, ob die Vorzeichen der Residuen tatsächlich regellos verteilt sind. Für die spezifische Prüfung der Nullhypothese, daß die Autokorrelation $\varrho = 0$ ist, müssen zwei Varianten unterschieden werden, die beide die Normalverteilungsannahme A5 erfordern.

Für die erste Variante müssen die beiden Residuen selbst bekannt sein, d.h. die Regressionskoeffizienten ß müssen bekannt sein. Dann ist die Verteilung des *von Neumann-Verhältnisses*

$$Q(y) = \frac{\frac{1}{N-1} \sum_{n=2}^{N} (u_n - u_{n-1})^2}{\frac{1}{N} \sum_{n=1}^{N} u_n^2} \tag{99}$$

unter der Nullhypothese bekannt (Tabellen in Hart & von Neumann, 1942). Q ist symmetrisch um $2N/(N-1)$ mit Varianz var $Q \approx 4/N$ verteilt.

Für den realistischeren Fall, daß nur die Schätzungen der Residuen nach (97) vorliegen, gibt es die zweite Variante, den *Durbin-Watson-Test* (Durbin & Watson, 1950, 1951, 1971), der auf der Prüfgröße

$$d = \frac{\sum_{n=2}^{N} (\hat{u}_n - \hat{u}_{n-1})^2}{\sum_{n=1}^{N} \hat{u}_n^2} \tag{100}$$

beruht. Es handelt sich dabei um einen „bounds-Test“, bei dem untere und obere Schranken für die Durbin-Watson-Statistik tabellarisch erfaßt sind (z.B. Koerts & Abrahamse, 1969). Die Tabellen sind in der Regel für die einseitige Alternativhypothese $\varrho > 0$ ausgelegt. Unterschreitet d die untere Schranke $d_{L,\alpha}$, wird die Nullhypothese zugunsten der Alternativhypothese abgelehnt. Wird die obere Schranke $d_{U,\alpha}$ überschritten, entscheidet man sich für die Nullhypothese. Liegt d zwischen $d_{L,\alpha}$ und $d_{U,\alpha}$, enthält man sich des Urteils (s. Abb. 3).

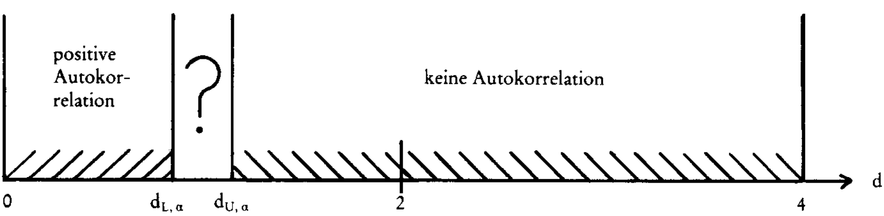

Abb. 3: Durbin-Watson-Test auf positive Autokorrelation.

Die beiden besprochenen Tests sind für die Diagnostik des autoregressiven Prozesses erster Ordnung konstruiert, der in Anwendungsfällen der Psychologie eher selten vorzufinden ist. Für allgemeinere serielle Abhängigkeiten haben Box & Pierce (1970) einen Modellanpassungstest entwickelt, den sie Portemanteau-Test nennen. Zur Entscheidung, ob die Residuen einem periodischen Muster folgen, empfehlen Jenkins & Watts (1968) den Periodogramm-Test.

Wenn auch die Prädiktoren selbst Zufallsvariablen mit seriellen Abhängigkeiten sind, müssen Transfermodelle (distributes lags, dynamische Regression) herangezogen werden (s. Box & Jenkins, 1970; McCleary & Hay, 1980).

1.10 Suppression und Kollinearität

Die Interpretation von Regressionsanalysen ist deswegen häufig nicht einfach, weil die Regressionskoeffizienten stark von den Interkorrelationen der Prädiktoren abhängen. Die Schwierigkeiten bei der Interpretation einer Regressionsanalyse werden an einem Beispiel aufgezeigt, dessen Bestimmungsgrößen wir später variieren. Dabei werden die Begriffe Wichtigkeit einzelner Prädiktoren, Suppression und Kollinearität eingeführt und diskutiert.

Cohen & Cohen (1975, p. 73) stützen sich bei der Diskussion der Zerlegung der Varianz des Kriteriums, die durch eine Regressionsanalyse gegeben ist, auf ein Beispiel, das wir zur Illustration übernehmen wollen. Es wird eine Stichprobe von Mitgliedern des Lehrkörpers von Universitäten untersucht. Als Kriteriumsvariable y dient der akademische Rang einer Person, der, übertragen auf deutsche Verhältnisse, dadurch skaliert wird, daß von der Besoldungsgruppe Hl bis H4 der Buchstabe entfernt wird; Mitgliedern des akademischen Mittelbaus könnte wahlweise der Wert 1 oder 0 zugewiesen werden. Der Zusammenhang zwischen diesem Kriterium und den beiden Prädiktoren x_1 = Anzahl der Publikationen und x_2 = akademisches Alter = Anzahl der Jahre seit der Promotion soll untersucht werden. Wenn die Personen der Stichprobe zufällig aus der Grundgesamtheit entnommen werden, liegt ein klassisches korrelationsanalytisches Modell vor; wegen der Zufallsreihenfolge in der Stichprobe sind serielle Korrelationen ausgeschlossen. Falls für die Argumentation erforderlich, wollen wir für die Residuen Normalverteilung annehmen,

Tabelle 2: Die Ergebnisse zum Beispiel von Cohen & Cohen (1975, p. 74). Der Stichprobenumfang ist N = 15.

Variable	Akad. Rang y	Konstante x_0	Anz. Publik. x_1	Akad. Alter x_2
Mittelwert	1.8	1.0	7.6	9.6
Standardabw. s	.748	.0	4.96	7.25
einf. Korr. mit y	1.0	–	.463	.612
einf. Korr. mit x_1	.463	–	1.0	.683
Regressionskoeff. b_k	–	1.15	.013	.057
Regressionskoeff. b^*_k für stand. Variablen	–	.0	.084	.555

obwohl sie wegen der Diskretheit des Kriteriums und der Beschränkung des Wertebereichs sicher höchstens in schlechter Näherung gegeben sein kann. Die meisten Interpretationen benötigen jedoch die Normalverteilungsannahme nicht. Außerdem sollen zwischen allen drei Variablen nur lineare Zusammenhänge bestehen. Die wichtigsten Ergebnisse der empirischen Regression mit den Schätzungen der Regressionskoeffizienten im klassischen Modell sind in Tab. 2 enthalten.

In dieser Tabelle sind neben den Regressionskoeffizienten aus

(101) $$\hat{y}_n = b_0+b_1x_{n1}+b_2x_{n2} \qquad (1 \leqq n \leqq N)$$

auch die Regressionskoeffizienten für die standardisierten Werte

(102) $$z_{yn} = \frac{y_n-\bar{y}}{s_y}, \; z_{nk} = \frac{x_{nk}-\bar{x}_k}{s_k} \qquad (1 \leqq n \leqq N, 1 \leqq k \leqq K)$$

enthalten. Die Regressionskoeffizienten b^*_k für standardisierte Werte (in der Literatur werden sie häufig auch Beta-Gewichte genannt) sind deshalb angenehmer für die Interpretation, weil sie nicht von der Skalierung, d.h. von den Maßeinheiten der Variablen abhängen. Sie sind die Bestimmungsgrößen zur OLS-Schätzung der standardisierten Kriteriumswerte

(103) $$\hat{z}_{yn} = b_1^*z_{n1}+b_2^*z_{n2} \qquad (1 \leqq n \leqq N),$$

wobei die Variable x_0 nicht mehr berücksichtigt werden muß.

In Abb. 4 werden die Anteile gemeinsamer Varianz durch das Überlappen von Kreisscheiben dargestellt und mit den entsprechenden einfachen semipartiellen (Cohen & Cohen, 1975, p. 81; Gaensslen & Schubö, 1976, p. 88) oder multiplen Korrelationen etikettiert. Die quadrierten Korrelationskoeffizienten sind zugleich die erklärten Varianzen, wenn es sich um standardisierte Variable handelt.

Richtig wiedergegeben wird in Abb. 4, daß für Fläche a verschiedene äquivalente Berechnungsmöglichkeiten bestehen

(104) $$\begin{aligned} a &= R^2_{y(12)}-r^2_{y(1-2)}-r^2_{y(2-1)} \\ &= r^2_{y1}-r^2_{y(1-2)} \\ &= r^2_{y2}-r^2_{y(2-1)} \\ &= .2106. \end{aligned}$$

Doch ist die Abbildung insofern gefährlich, als sie nicht berücksichtigt, daß dieser Betrag bei anderen Daten durchaus negativ sein könnte (s. Tab. 3). Auch können in einer derart vereinfachten Abbildung häufig die Beziehungen der Prädiktoren untereinander nicht angemessen dargestellt werden.

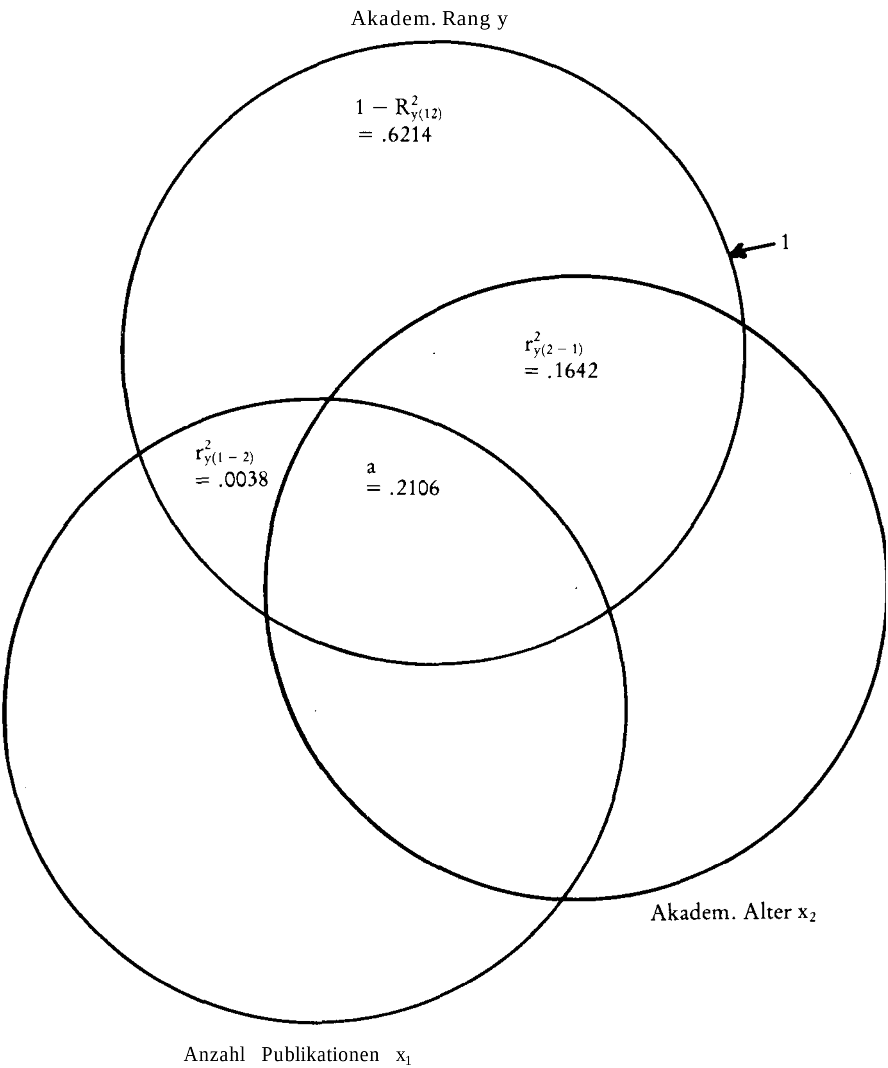

Abb. 4: Die Varianzzerlegung bei der Regression mit zwei Prädiktoren im Beispiel von Cohen & Cohen (1975, p. 74).

Wie läßt sich dieses Ergebnis nun deskriptiv interpretieren? Einerseits besteht ein Zusammenhang zwischen der Anzahl der Publikationen und dem akademischen Rang, der wechselseitig 21% der Varianz erklärt; über eine kausale gerichtete Abhängigkeit läßt sich bei Querschnittsuntersuchungen nie etwas aus den Daten, sondern allenfalls aufgrund einer Theorie über den Gegenstandsbereich aussagen. Andererseits besteht ein noch stärkerer Zusammenhang (erklärter Varianzanteil 37%) zwischen dem akademischen Alter und dem akademischen Rang. Die Betrachtung des gemeinsamen Zusammenhangs der Variablen Anzahl der Publikationen und akademisches Alter mit dem akademischen Rang wirft gegenüber der isolierten Betrachtung der Prädiktoren ein neues Licht vor allem auf die Rolle der Anzahl der Publikationen. Der durch sie zusätzlich erklärte Anteil der Kriteriumsvarianz beträgt nämlich nur $r^2_{y(1-2)}$ = .0038 und ist demnach verschwindend gering. In der Begriffsbildung von Bortz (1977, p. 596) ist die Anzahl der Publikationen demnach eine redundante Prädiktorvariable. Das bedeutet, daß für die Schätzung des akademischen Rangs die Hinzunahme der Anzahl der Publikationen zum akademischen Rang keine Verbesserung erbringt. Dies zeigt sich übrigens nicht so deutlich an den Regressionskoeffizienten b^*_k, wo z_{x1} mit 0.084 gewichtet ist. Durch den Zusammenhang zwischen x_1 und x_2 entsteht eine schwache Kollinearität, die dazu führt, daß aus beiden Prädiktoren gleichermaßen schätzbare Anteile des Kriteriums auf die Prädiktoren aufgeteilt geschätzt werden.

Um zu demonstrieren, wie sehr die Ergebnisse von der Korrelation der Prädiktoren unter sonst gleichen Umständen abhängen, werden einige für die Interpretation einer Regressionsanalyse wichtige Größen in Abhängigkeit von der Prädiktorenkorrelation dargestellt. Die Regressionskoeffizienten für standardisierte Variablen können im hier behandelten Fall zweier Prädiktoren nach

$$(105) \qquad b_1^* = \frac{r_{y1} - r_{y2}r_{12}}{1 - r_{12}^2}, \quad b_2^* = \frac{r_{y2} - r_{y1}r_{12}}{1 - r_{12}^2},$$

die quadrierten semipartiellen Korrelationen, die den zusätzlichen Varianzbeitrag der Variablen angeben, nach

$$(106) \qquad r_{y(1-2)} = \frac{r_{y1} - r_{y2}r_{12}}{\sqrt{1 - r_{12}^2}}, \; r_{y(2-1)} = \frac{r_{y2} - r_{y1}r_{12}}{\sqrt{1 - r_{12}^2}}$$

und der insgesamt erklärte Varianzanteil nach

$$(107) \qquad R^2_{y(12)} = \frac{(r_{y1} - r_{y2})^2}{1 - r_{12}^2} + 2\,\frac{r_{y1}r_{y2}}{1 + r_{12}}$$

berechnet werden. Aus der Bedingung, daß $R^2_{y(12)}$ den Wert 1 nicht übersteigen darf, ergibt sich der zulässige Wertebereich für r_{12}

$$(108) \qquad r_{y1}r_{y2} - \sqrt{(1 - r_{y1}^2)(1 - r_{y2}^2)} \leqq r_{12} \leqq r_{y1}r_{y2} + \sqrt{(1 - r_{y1}^2)(1 - r_{y2}^2)}.$$

Für die Grenzen dieses Intervalls wird der insgesamt erklärte Varianzanteil stets 1. Der minimal erklärte Varianzanteil wird, wenn o.B.d.A. $|r_{y1}| \leqq |r_{y2}|$,

(109) $$r_{12} = \frac{r_{y1}}{r_{y2}}$$

angenommen und sein Betrag ist dann

$$R^2_{y(12)} = r^2_{y2} \,.$$

Die konkreten Ergebnisse für die im Beispiel vorliegende Prädiktorkorrelation und einige weitere Korrelationswerte sind in Tab. 3 wiedergegeben. Es überrascht, wie stark die Ergebnisse voneinander abweichen, obwohl die Stärke des Zusammenhangs jedes einzelnen Prädiktors zum Kriterium stets dieselbe ist.

Leider sind die Ergebnisse nur für zwei Fälle leicht zu interpretieren: Zum ersten im Fall der minimal erklärten Varianz, r_{12} = 0.463/0.612 = .757, bei dem die Schätzung des Kriteriums ausschließlich durch den damit am stärksten korrelierenden Prädiktor erfolgt, und zum zweiten im Fall r_{12}=0, wo die Prädiktoren unkorreliert sind. In allen anderen Fällen ist es prinzipiell unmöglich, befriedigende Maßzahlen für den alleinigen Beitrag einer Prädiktorvariablen innerhalb der gemeinsamen Schätzung herauszunehmen, so sehr der Praktiker sich Hinweise auf die Wichtigkeit oder Bedeutung einzelner Prädiktoren wünscht. Dennoch werden als Indizes der *Wichtigkeit einzelner Prädiktoren* die einfachen Korrelationen mit dem Kriterium, die Regressionskoeffizienten für standardisierte Variablen, die quadrierten semipartiellen Korrelationen oder die Produkte von einfacher Korrelation und Regressionskoeffizient für standardisierte Variablen herangezogen. Alle diese Größen haben ihre sinnvolle Bedeutung, doch sind sie alle nur in den beiden genannten Fällen der leichten Interpretierbarkeit als Indizes für den eigenständigen Beitrag eines Prädiktors uneingeschränkt verwendbar. Die spezifischen Vor- und Nachteile dieser Größen werden am ausführlichsten von Darlington (1968) diskutiert.

Recht kontrovers wird in der Literatur auch die Frage diskutiert, was unter *Suppressorvariablen* verstanden werden soll, und wie sie noch näher charakterisiert und gegliedert werden können. Gemeinsam ist die Forderung, daß eine Suppressorvariable eine Variable sein soll, die im Kontext der übrigen Prädiktoren die erklärte Varianz in einer Weise erhöht, die man zunächst (aufgrund der einfachen Korrelationen) nicht vermutet hätte. Bortz (1977, p. 597) definiert als Suppressorvariablen solche, die den Vorhersagebeitrag einer (oder mehrerer) anderen Variablen erhöht, indem sie irrelevante Varianzen in der (den) anderen Prädiktorvariablen unterdrückt. Bortz operationalisiert dies in Anlehnung an Conger & Jackson (1972) und Conger (1974) für den Fall zweier Prädiktoren dadurch, daß er $b^*_1 r_{y1} > r^2_{y1}$, fordert, wenn x_2 eine Suppressorvariable sein soll. Darlington (1968) polt alle Prädiktoren so um, daß ihre einfache Korrelation mit dem Kriterium nicht negativ ist, und nennt die Prädiktoren

Tabelle 3: Die Ergebnisse zum Beispiel aus Tab. 2 für verschiedene Werte der Prädiktoreninterkorrelation bei festen Korrelationen mit dem Kriterium r_{y1} = .463 und r_{y2} = .612

Prädiktoren-interkorrelation r_{12}	Regressions-koeffizienten b_1^*	b_2^*	Zusätzlicher Varianzbeitrag $r^2_{y(1-2)}$	$r^2_{y(2-1)}$	Bestimmtheitsmaß $R^2_{y(12)}$	R^2-Komponenten $r_{y1}b_1^*$	$r_{y2}b_2^*$	Überlappung $R^2_{y(12)} - r^2_{y(1-2)} - r^2_{y(2-1)}$
−.400	.84	.95	.60	.76	.97	.39	.58	.38
.000	.46	.61	.21	.37	.59	.21	.37	.00
.683	.08	.56	.00	.16	.38	.04	.34	.21
.757	.00	.61	.00	.16	.37	.00	.37	.21
.950	−1.21	1.77	.14	.30	.52	− .56	1.08	.07
.984	−4.39	4.93	.61	.77	.99	−2.03	3.02	−.40

Suppressorvariablen, deren Regressionskoeffizient b_k negativ ist. Für den Fall zweier Prädiktoren folgt daraus die Bedingung $r_{y2}r_{1(1\text{-}y)} < r_{y1}r_{2(1\text{-}y)}$, wenn x_2 eine Suppressorvariable sein soll. Gaensslen & Schubö (1976, p. 117, 123) sprechen dann von einer Suppressorvariablen, wenn sie für sich mit dem Kriterium praktisch nicht korreliert, aber im Zusammenhang mit den übrigen Prädiktoren den insgesamt erklärten Varianzanteil erhöht. Im Beispiel nennen sie x_2 eine Suppressorvariable, wenn $r^2_{y(2\text{-}1)}$ groß gegenüber r^2_{y2} bzw. wenn $|r_{y2}r_{12}|$ groß gegenüber $|r_{y1}|$ ist. Cohen & Cohen (1975, p. 91) nennen x_2 dann eine Suppressorvariable, wenn b^*_2 außerhalb des Intervalls zwischen 0 und r_{y2} liegt. Velicer (1978), der Conger kritisiert, spricht ähnlich wie Gaensslen & Schubö (1976) dann von Suppression, wenn $r^2_{y(2\text{-}1)}$ größer als r^2_{y2}, ist. Holling (1980) schließlich verwendet als Kriterium $r_{y(2\text{-}1)} > r_{y2}$ bzw. gleichwertig $r_{\hat{y}(2-1)} > r_{\hat{y}2}$, wenn x_2 so gepolt ist, daß r_{y2} positiv ist.

Unter *Kollinearität* oder Multikollinearität versteht man, daß die Matrix X'X singulär oder fast singulär ist. Den Fall der exakten Kollinearität hatten wir durch Annahme A5 in Kap. 1.2 ausgeschlossen; er soll für das weitere ausgeschlossen bleiben. Eine Verbesserung der Schätzungen für die Regressionskoeffizienten war unter dem Titel Ridge-Regression in Kap. 1.6 vorgeschlagen worden. Hier sollen noch praktische Auswirkungen der Kollinearität angedeutet werden.

Stellt man einen gemeinsamen Konfidenzbereich aus den Daten des Beispiels für die beiden Regressionskoeffizienten für standardisierte Daten

$$(110) \qquad \beta^*_1 = \frac{s_1}{s_y}\,\beta_1, \qquad \beta^*_2 = \frac{s_2}{s_y}\,\beta_2$$

auf, so erhält man aus Gl. (22) unter Annahme des klassischen Modells die Ungleichung

$$(111) \qquad \sum_{k=1}^{K}\sum_{l=1}^{K} r_{kl}\,(\beta^*_k - b^*_k)\,(\beta^*_l - b^*_l) \leq \frac{K(1-R^2)}{N-K-1}\,F_\alpha\,(K,N-K-1)\,,$$

welche das Innere einer Ellipse mit Mittelpunkt (b^*_1, b^*_2) beschreibt. In Abb. 5 sind die Konfidenzbereiche zu r_{y1} = .463, r_{y2} = .612 und verschiedene, aus Tab. 3 ausgewählte Werte für r_{12} abgebildet. Wir wollen hier davon absehen, daß die Mittelpunkte der Ellipsen in Abhängigkeit von r_{12} variieren. Was jetzt vor allem interessieren soll, ist, daß die Ellipsen um so eher gestreckt sind, je größer der Betrag von r_{12} ist. Für r_{12}=0 erhält man den Grenzfall eines Kreises als Konfidenzbereich, was davon herrührt, daß die Schätzungen b^*_1 und b^*_2 unkorreliert sind. Wenn r_{12} hoch ist, ist auch die negative Kovarianz der Schätzungen hoch; dies bedeutet, daß bei zufälliger Überschätzung des einen Regressionskoeffizienten der andere Regressionskoeffizient eher unterschätzt wird. Jeder Regressionskoeffizient kann dann für sich nicht zuverlässig ge-

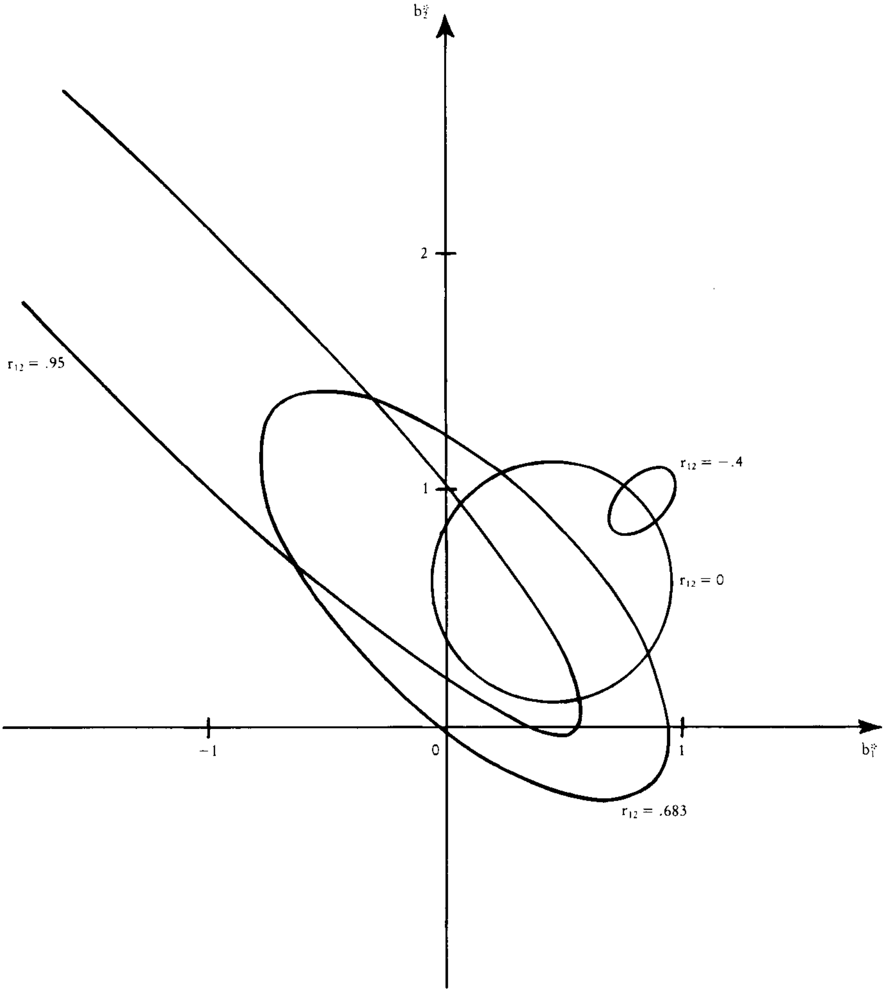

Abb. 5: Konfidenzbereiche der Regressionskoeffizienten für $r_{y1} = .463$, $r_{y2} = .612$ und verschiedene Werte der Prädiktoreninterkorrelation r_{12}.

schätzt werden. Die Zuverlässigkeit einer Prognose durch die Regressionsgleichung als Ganzes ist durch die Kollinearität der Prädiktoren jedoch nicht beeinträchtigt. über Möglichkeiten zur Orthogonalisierung von Prädiktoren berichtet Silvey (1969); Probleme der kollinearen Testauswertung diskutieren Schubö & Hentschel (1978) und Schubö (1980); über Schwierigkeiten bei Signifikanztests informiert Gordon (1968).

1.11 Schrittweise Regression

In vielen Bereichen der empirischen Forschung weiß man zwar, daß Zusammenhänge zwischen einem Kriterium und einer größeren Menge von Prädiktorvariablen bestehen, doch fallt eine Entscheidung darüber schwer, welche Untermenge der Prädiktorvariablen den Zusammenhang vermittelt. Eine Reduktion der Anzahl der Prädiktorvariablen ist aus Gründen der sparsameren Beschreibung wünschenswert. Wenn für das Forschungsgebiet keine Erfahrungen oder theoretische Hinweise vorliegen, wenn also keine a priori Reduktion der Prädiktoren möglich ist, ist man darauf angewiesen, die Variablenreduktion aufgrund der Zusammenhänge durchzuführen, die sich im vorliegenden Datensatz zeigen. Die in Kap. 1.10 angeführten Maße für die Wichtigkeit der Prädiktoren werden auch zu dem Zweck berechnet, Richtlinien für die Variablenreduktion zu erhalten. In diesem Kapitel sollen einige Methoden zur Auswahl der Prädiktoren diskutiert werden, die z.B. in Draper & Smith (1966, p. 163), Cohen & Cohen (1975, p. 102) und Graybill (1976, p. 439) vorgeschlagen werden.

Da die Potenzmenge der Menge der Prädiktoren endlich ist, steht im Prinzip dem nichts im Wege, der Reihe nach jede der 2^K-l möglichen nichtleeren Untermengen der Prädiktoren für Regressionsanalysen mit dem Kriterium zu verwenden. Diese Vorgehensweise wird als *Untermengen-Methode* (subset regression, all regressions method) bezeichnet und ist rechnerisch sehr aufwendig. Die „beste“ Einzelregression ist diejenige, die einerseits einen großen Varianzanteil erklärt, andererseits jedoch nur wenige Prädiktoren verwendet. Die Einzelregression mit dem besten Kompromiß zwischen diesen beiden gegensätzlichen Forderungen kann z.B. in einer ähnlichen Weise wie beim Scree-Test der Faktorenanalyse (z.B. Gaensslen & Schubö, 1976, p. 226) dadurch bestimmt werden, daß in einem Kreuzdiagramm das maximale Bestimmtheitsmaß für eine feste Zahl k von Prädiktoren gegen die Zahl k aufgetragen wird. Andere Kriterien und schnelle Algorithmen findet man bei Garside (1965), Hocking & Leslie (1967), LaMotte & Hocking (1970), Furnival (1971), Hocking (1972), Morgan & Tatar (1972) oder auch in IMSL (1979).

In sozialwissenschaftlichen Anwendungen ist allerdings für die Einschätzung der gefundenen besten Regression größte Vorsicht erforderlich. Eine Kreuzvalidierungsstudie zur Überprüfung der Regression (s. Kap. 1.5: Prognosetest) mit einer unabhängigen Stichprobe ist obligat. Es genügt keinesfalls, für die gefundene Regression einen Signifikanztest durchzuführen, weil die tatsächliche Irrtumswahrscheinlichkeit erheblich das gewählte Signifikanzniveau übersteigt. Aus Schubö (1982) wurde Tab. 4 entnommen, die für das Signifikanzniveau 5% die effektive Irrtumswahrscheinlichkeit angibt. Beispielsweise ist die Wahrscheinlichkeit dafür, daß der Zusammenhang zwischen den drei besten Prädiktoren aus insgesamt 13 Prädiktoren auf dem 5%-Niveau signifikant wird, gleich 58%, wenn in Wirklichkeit keinerlei Zusammenhang zwischen dem Kriterium und den Prädiktoren und zwischen den Prädiktoren untereinander besteht.

Tabelle 4: Gesamt-Irrtumswahrscheinlichkeit bei der empirischen Auswahl der Prädiktoren mit dem stärksten Zusammenhang zum Kriterium (Irrtumswahrscheinlichkeit α = .05 für den Einzeltest).

Gesamtzahl der Prädiktoren K	Anzahl verwendeter Prädiktoren										
	1	2	3	4	5	6	7	8	9	10	beliebig
2	.10	.05	-	-	-	-	-	-	-	-	.12
3	.14	.10	.05	-	-	-	-	-	-	-	.17
4	.19	.14	.09	.05	-	-	-	-	-	-	.22
5	.23	.19	.14	.08	.05	-	-	-	-	-	.27
6	.26	.26	.19	.14	.09	.05	-	-	-	-	.32
7	.30	.31	.24	.17	.13	.08	.05	-	-	-	.38
8	.34	.36	.32	.25	.18	.13	.08	.05	-	-	.43
9	.37	.39	.36	.30	.24	.17	.12	.08	.05	-	.46
10	.40	.43	.41	.37	.31	.23	.18	.12	.08	.05	.51
11	.43	.50	.46	.42	.36	.29	.22	.16	.12	.09	.56
12	.46	.53	.52	.49	.44	.37	.30	.23	.17	.12	.61
13	.49	.56	.58	.55	.51	.45	.38	.30	.23	.17	.66
14	.51	.60	.62	.61	.56	.51	.43	.36	.29	.22	.72
15	.54	.64	.67	.65	.61	.56	.49	.43	.36	.29	.75
16	.56	.68	.71	.70	.67	.63	.58	.52	.43	.35	.77
17	.58	.69	.72	.73	.71	.67	.62	.57	.49	.43	.80
18	.60	.73	.77	.76	.75	.71	.67	.61	.56	.50	.83
19	.62	.76	.80	.80	.79	.75	.72	.67	.62	.55	.85
20	.64	.78	.82	.83	.82	.81	.78	.75	.70	.64	.88
25	.72	.84	.90	.91	.91	.89	.87	.85	.83	.80	.96
29	.77	.91	.93	.96	.97	.96	.95	.94	.93	.92	.98

Dieselbe Problematik besteht bei drei weiteren Methoden zur empirischen Auswahl von Prädiktoren, der Vorwärts-Methode, der Rückwärts-Methode und der Austausch-Methode. Alle diese Methoden sollen mit geringerem Aufwand eine beinahe so gute Untermenge der Prädiktoren finden, wie sie die Untermengen-Methode liefert. Bei allen Methoden werden schrittweise Prädiktoren in die Regressionsgleichung aufgenommen bzw. wieder ausgeschieden. Keine der drei Methoden führt mit Sicherheit zur definiert besten Untermenge der Prädiktoren im zuvor genannten Sinn.

Bei der *Vorwärts-Methode* wird zunächst die einzelne, mit dem Kriterium am stärksten korrelierende Variable aufgenommen, dann als zweiter Prädiktor die Variable, die den größten Zugewinn des Zusammenhangs bewirkt, usw. Bei der *Rückwärts-Methode* werden aus dem Regressionsansatz mit allen Prädiktoren jeweils die Variablen ausgeschieden, deren Entfernen die kleinste Verminderung des Zusammenhangs zur Folge hat. Bei der *Austausch-Methode* wird im Prinzip wie bei der Vorwärts-Methode vorgegangen, doch wird nach der Aufnahme neuer Prädiktoren in den Regressionsansatz jeweils geprüft, ob andere Prädiktoren aus ihm entfernt werden können, ohne daß der Gesamtzusammenhang erheblich zurückgeht. Gleichzeitig kann bei allen Methoden die Aufnahme näherungsweise kollinearer Prädiktoren ausgeschlossen werden. Für den endgültigen Aufbau der Regressionsgleichung ist es im Fall der Kollinearität zweier oder mehrerer Prädiktoren meist dem Zufall vorbehalten, welche Prädiktoren herangezogen werden. Beinahe alle Computerprogramme ermöglichen die Prädiktorenauswahl nach wenigstens einer dieser drei Methoden, wobei leider für die endgültige Prädiktorenauswahl der Signifikanztest mit falscher Irrtumswahrscheinlichkeit berechnet wird. Der Anwender dieser Programme wird zwar subjektiv glücklicher, weil er einen signifikanten Zusammenhang berichten kann, doch besteht sein unbemerktes Unglück darin, daß er mit aller Wahrscheinlichkeit eine nicht gerechtfertigte Entscheidung trifft.

Eine erheblich geringere Verfälschung der Irrtumswahrscheinlichkeit ist durch eine Vorgehensweise für Prädiktorengruppen gegeben, wie sie von Cohen & Cohen (1975, p. 123) ausführlich beschrieben wird. Dabei werden die Prädiktoren a priori in disjunkte Variablengruppen A, B, usw. eingeteilt. Die Einteilung erfolgt entweder nach strukturellen oder nach funktionalen Gesichtspunkten. Ein struktureller Gesichtspunkt ist etwa, wenn die K_A = g-l Kodiervariablen (z.B. Dummyvariablen) zu einer qualitativen Variablen mit g Ausprägungen in eine Gruppe A zusammengefaßt wird, oder wenn nichtlineare Effekte einer quantitativen, standardnormalverteilten Variablen z durch eine Gruppe B von $K_B = 3$ orthogonalen Hermiteschen Polynomen z, z^2-1, z^3-3z berücksichtigt werden. Nach funktionalen Gesichtspunkten könnte man Prädiktoren des Studienerfolgs etwa in die Gruppen Reifezeugnisnoten (A), Leistungstest-variablen (B), motivationale Variablen (C) und Lebenslauf-Daten (D) einteilen.

Um die Vorgehensweise übersichtlich darstellen zu können, müssen wir zunächst eine kompakte Schreibeweise für die Behandlung von Variablengruppen einführen. Wenn die Gruppe A von Prädiktorvariablen

$$A = \{ x_1, x_2, \ldots, x_{K_A} \}$$

ist, schreibt man für die multiple Korrelation kurz

$$R_{y(A)} = R_{y(12\ldots K_A)}$$

und, wenn C die Vereinigung von A und B ist,

$$R_{y(AB)} = R_{y(C)}.$$

Die quadrierte semipartielle multiple Korrelation

(112) $$R^2_{y(A-B)} = R^2_{y(AB)} - R^2_{y(B)}$$

ist dann der Varianzanteil des Kriteriums y, der durch die Hinzunahme von A zusätzlich zu B erklärt wird. Für den Fall dreier Gruppen von Prädiktoren kann in dieser Begriffsbildung eine hierarchische Zerlegung

(113) $$R^2_{y(ABC)} = R_{y(A)} + R^2_{y(B-A)} + R^2_{y(C-AB)}$$

der erklärten Varianz vorgenommen werden. Ob die zu den Variablen der Gruppe A gehörenden Regressionskoeffizienten alle Null sind

$$\beta_1 = \beta_2 = \ldots = \beta_{K_A} = 0,$$

kann durch die Entscheidungsregel zum Verwerfen dieser Nullhypothese

(114) $$\frac{(N-K_A-K_B-1)\,(R^2_{y(AB)}-R^2_{y(B)})}{K_A(1-R^2{}_{y(AB)})} > F_\alpha(K_A, N-K_A-K_B-1)$$

entschieden werden, wenn die Annahmen des klassischen Modells für den Ansatz mit den zwei Sätzen A und B von Prädiktoren erfüllt sind. Für den Fall, daß A nur eine Variable enthält, entspricht dies genau der Entscheidungsregel (37).

Welche Strategien sich für die Signifikanztests bei Variablengruppen anbieten, soll nun am bereits erwähnten Beispiel der Kriteriumsvariablen Studienerfolg demonstriert werden. Einerseits kann im Rahmen des vollen Satzes von Prädiktoren ABCD jeweils geprüft werden, ob die einzelnen Gruppen zusätzlich zu den drei übrigen Gruppen signifikante Varianzanteile erklären. Beispielsweise wäre der Test dafür, ob die motivationalen Variablen C zusätzlich zu den Reifezeugnis-Noten A, den Leistungstest-Variablen B und den Lebenslauf-Daten D Varianzanteile erklären, durch

(115) $$\frac{(N-K-1)\,(R^2{}_{y(ABCD)}-R^2{}_{y(ABC)})}{K_C(1-R^2{}_{y(ABCD)})} > F_\alpha(K_C, N-K-1)$$

gegeben, wobei die Gesamtzahl der Prädiktoren

$$K = K_A + K_B + K_C + K_D$$

ist. Diese erste Strategie entspricht der ersten Stufe der Rückwärtsmethode.

Entsprechend der Vorwärtsmethode wird die zweite Strategie gewählt, mit dem Unterschied, daß nach inhaltlichen Gesichtspunkten a priori eine Hierarchie für die Aufnahme in die Regression festgelegt wird. Beispielsweise könnte es vernünftig erscheinen zu untersuchen, ob erstens die Reifezeugnisnoten A für sich mit dem Studienerfolg zusammenhängen, ob zweitens die Leistungstest-Variablen B eine Verbesserung der Stärke des Zusammenhangs mit sich bringen, ob drittens die motivationalen Variablen C zusätzlich zu A und B beitragen, und ob schließlich die Aufnahme von Lebenslauf-Daten D sinnvoll oder erforderlich ist. Wieder sei die Entscheidungsregel für die Aufnahme der motivationalen Variablen C als Beispiel angeführt:

$$(116) \qquad \frac{(N-K_A-K_B-K_C-1)\,(R^2_{y(ABC)}-R^2_{y(AB)})}{K_C(1-R^2_{y(ABC)})} > F_\alpha(K_c, N-K_A-K_B-K_C-1)$$

Bei beiden Vorgehensweisen werden zunächst nur vier Signifikanztests durchgeführt, wodurch sich die Gesamt-Irrtumswahrscheinlichkeit nicht in dem Maße wie bei der Untermengen-, Vorwärts-, Rückwärts- oder Austausch-Methode erhöht. Wenn man es wünscht, können weitere Signifikanztests in Form von geschützten Tests innerhalb der „signifikanten" Variablengruppen für die einzelnen Variablen angeschlossen werden. Man kann dabei hoffen, daß die endgültige Variablenauswahl nicht in dem Maße zufallsabhängig ist, wie dies bei den zuerst genannten Auswahlstrategien der Fall ist. Doch stehen zuverlässige Untersuchungen dazu noch aus.

1.12 Teststärke

Bisher haben wir im Zusammenhang mit der Prüfung statistischer Hypothesen immer nur von der Wahrscheinlichkeit, einen Fehler 1. Art zu begehen, gesprochen. Beim Prüfen statistischer Hypothesen können jedoch grundsätzlich zwei Fehlentscheidungen getroffen werden. Einmal kann eine tatsächlich richtige Hypothese aufgrund eines Tests verworfen werden, andererseits kann eine tatsächlich falsche Hypothese nicht verworfen werden. Solange man jedoch nur eine einzige Hypothese betrachtet, lassen sich nur Aussagen über die Wahrscheinlichkeit des Fehlers 1. Art, das vorzugebende Signifikanzniveau *a*, machen. Dies ist auch in dem von Fisher ursprünglich konzipierten Signifikanztest die einzige bestimmende Größe (vgl. Haagen & Seifert, 1979, Kap. 9, insbesondere p. 198-199). Solange man jedoch nur eine einzige Hypothese prüft, lassen sich keine Aussagen über die Wahrscheinlichkeit eines Fehlers 2. Art machen, und damit auch keine Aussagen über die „Güte" eines Tests, wenn man die Güte eines Tests durch die Wahrscheinlichkeit, richtig zu entscheiden, mißt. Zur Beurteilung der Güte eines Tests ist es also notwendig, neben der Nullhypothese auch die Alternativhypothese zu spezifizieren, wie dies auch im Rahmen der Neyman-Pearson-Testtheorie geschieht. Die *Güte-*

funktion (Teststärke, Trennschärfe) eines Tests gibt für den Bereich der Alternativhypothese die Wahrscheinlichkeit dafür an, die Nullhypothese zu verwerfen, d.h. bei Gültigkeit der Alternativhypothese eine richtige Entscheidung zu treffen (vgl. Haagen & Seifert, 1979, p. 209). Bezeichnen wir mit β[1]) die Wahrscheinlichkeit, einen Fehler 2. Art zu begehen, so ergibt sich für die Teststärke 1 - ß. Beim Prüfen statistischer Hypothesen ist man nun speziell an solchen Tests interessiert, die bei gegebenem Signifikanzniveau eine möglichst große Teststärke aufweisen (vgl. dazu auch die Ausführungen in Kap. 1.5).

Um die Teststärke bei der Prüfung der multiplen Korrelation zu berechnen, müssen wir also neben dem Stichprobenumfang N und der Zahl der Prädiktoren K noch die Größe der multiplen Korrelation R in der Grundgesamtheit festlegen. Durch diese Größen ist über die Verteilungsannahme A5 die Verteilung der Prüfgröße F, die nichtzentrale F-Verteilung (siehe z.B. Tang, 1938; Laubscher, 1960) vollständig bestimmt. Man nennt R oder meist die daraus abgeleitete Größe

$$f^2 = \frac{R^2}{1-R^2}$$

auch die Effektgröße. Im allgemeinen wird man die Effektgröße allerdings nicht kennen. Doch hat man häufig aufgrund von Ergebnissen ähnlicher Studien Vorstellungen darüber, in welchem Intervall die Effektgröße liegen wird. Sicherlich wird es jedenfalls möglich sein, eine Mindest-Effektgröße anzugeben, bei der man noch von einem bedeutsamen Effekt oder Zusammenhang sprechen könnte. In der Persönlichkeitsforschung etwa könnte man als absolutes Minimum von bedeutsamen Zusammenhängen eine Korrelation R = .2 (4% erklärte Varianz), also eine Effektstärke f^2 = .042 annehmen.

Allgemein läßt sich sagen, daß die Teststärke bei der Prüfung der multiplen Korrelation um so größer wird,

a) je größer das Signifikanzniveau a ist,
b) je größer der Stichprobenumfang N ist,
c) je größer die Effektstärke ist und
d) je kleiner die Anzahl der Prädiktoren ist.

---

[1]) Da die Wahrscheinlichkeit für einen Fehler 2. Art unter sonst gleichen Umständen von den Parametern des Parameterraumes der Alternativhypothese (es werden hier nur parametrisierbare Wahrscheinlichkeitsverteilungen (vgl. Haagen & Seifert, 1979, p. 206) betrachtet), müßten wir genauer ß als abhängig von diesen Parametern symbolisieren. Aus schreibtechnischen Gründen verzichten wir jedoch darauf.

Diese Einflußgrößen hängen in der Regel von äußeren Bedingungen ab; am ehesten scheinen noch der Stichprobenumfang und die Zahl der Prädiktoren relativ frei wählbar zu sein.

Ohne ausreichende Teststärke besteht kaum Aussicht, eine Entscheidung für oder gegen die Nullhypothese zu fallen. Bei der Planung einer Untersuchung sollte daher eine Abschätzung der Teststärke Pflicht sein. Als eine Art Standard kann eine Teststärke von .80 angesehen werden.

Eine ausführliche Diskussion der Teststärken aller im Zusammenhang mit der multiplen Regression üblichen Tests findet sich in Cohen & Cohen (1975). Mit noch besserer Genauigkeit werden Teststärken in Laubscher (1969), Lehmer (1944) und Tang (1938) angegeben.

1.12.1 Teststärke für die Prüfung der multiplen Regression

Anstatt für vorgegebenes Signifikanzniveau *a*, Stichprobenumfang N, Effektstärke f^2 und Anzahl der Prädiktoren K die Teststärke 1 - ß zu bestimmen, soll hier von einer vorgegebenen Teststärke und dem vorgegebenen Signifikanzniveau α ausgegangen werden; sind zwei der übrigen drei Größen gegeben, kann dann die dritte so bestimmt werden, daß der Test gerade mit der vorgegebenen Teststärke durchgeführt werden kann.

1.12.2 Bestimmen des erforderlichen Stichprobenumfangs N

Der Stichprobenumfang kann mit der Näherungsformel

(117) $$N = \frac{L}{f^2} + K + 1$$

berechnet werden, wobei

(118) $$L = c + \sqrt{dK - e}$$

in den in Tab. 5 aufgeführten Koeffizienten die Abhängigkeit vom Signifikanzniveau α und der Teststärke 1 - ß enthält und die Effektstärke

(119) $$f^2 = \frac{R^2}{1 - R^2}$$

von der vermuteten Populationskorrelation abhängt. Zusammengefaßt ergibt sich

(120) $$N = \frac{1 - R^2}{R^2}\left(c + \sqrt{dK - e}\right) + K + 1.$$

Tabelle 5: Koeffizienten zur näherungsweisen Berechnung der L-Werte

Teststärke $1-\beta$	Signifikanzniveau $\alpha = .05$ c	d	e	Signifikanzniveau $\alpha = .01$ c	d	e
.10	-.16	.26	-.09	.16	2.12	-.29
.30	.65	2.47	.42	1.95	6.38	1.02
.50	1.95	5.39	1.66	3.92	10.72	3.19
.60	2.87	7.17	3.08	5.09	13.23	4.66
.70	3.95	9.40	4.67	6.48	16.24	6.52
.75	4.61	10.78	5.60	7.31	18.05	7.60
.80	5.41	12.43	6.82	8.30	20.18	9.02
.85	6.52	14.37	9.20	9.52	22.81	10.90
.90	7.79	17.34	10.77	11.17	26.37	13.40
.95	10.05	22.07	14.69	13.85	32.15	17.83
.99	15.04	32.55	24.01	19.58	44.68	27.50

Obwohl die Auswertung dieser Formel im Einzelfall nicht schwierig ist, wurden die Ergebnisse für den vielleicht wichtigsten Fall $\alpha = .05$ und $1 - ß = .80$ in Tab. 6 mit etwas besserer Genauigkeit (relativer Fehler kleiner 1%) aufgenommen. Im Falle der einfachen Korrelation (K = 1) ergaben sich allerdings aus den Tabellen G.2 und E.2 bei Cohen & Cohen (1975) und ebenso aus den Tabellen 3.4.1 und 8.4.4 bei Cohen (1969) Abweichungen, die etwas über den von den Autoren angegebenen Fehlergrenzen liegen; unsere Tab. 6 enthält die jeweils größeren Forderungen für den Stichprobenumfang.

Die Näherung (117) stammt von Cohen & Cohen (1975, p. 118). Die Beziehung (118) ist eine Approximation, für die die in Tab. 5 aufgeführten Koeffizienten durch eine Regressionsanalyse mit dem Kriterium K und den Prädiktoren L und L^2 aus den Tabellen von Cohen & Cohen (1975) bestimmt wurden. Um den Approximationsfehler für kleine Werte von K niedrig zu halten, wurde für K = 1 zehnfach und für K = 2 bis 5 jeweils fünffach gewichtet. Der relative Fehler der Näherung (118) gegenüber den tabellierten Werten von L beträgt höchstens 1% mit Ausnahme des Falles K = 1, wo relative Fehler bis zu 2,5% auftreten (in Tab. E2 bei Cohen & Cohen, 1975, ist für $1 - ß = .50$ und K = 24 ein Druckfehler enthalten: statt 13.02 muß L = 13.25 stehen). Diese nicht sehr erheblichen numerischen Fehler werden dadurch aufgewogen, daß es durch die Näherung (118) möglich wird, in Gl. (120) eine algebraische Beziehung zwischen dem Stichprobenumfang N, der Populationskorrelation R und der Prädiktorenzahl K für eine bestimmte Irrtumswahrscheinlichkeit und Teststärke aufzustellen. Diese Beziehung kann nach jeder der drei Größen aufgelöst werden, was im folgenden Verwendung findet.

Tabelle 6: Mindestens erforderlicher Stichprobenumfang N in Abhängigkeit von der Anzahl der Prädiktoren und der Größe der multiplen Korrelation in der Grundgesamtheit für den Signifikanztest der multiplen Korrelation aus der Stichprobe mit einem Signifikanzniveau von $\alpha = 0.05$ und einer Teststärke $1 - ß = 0.80$

Anzahl der Prädik-toren	Multiple Korrelation in der Grundgesamtheit .01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99
1	78812	3145	783	193	84	46	28	18	12	9	6	4	3
2	96544	3856	959	235	101	54	32	21	14	9	6	5	4
3	109269	4365	1086	267	115	62	37	24	16	11	7	6	5
4	119567	4776	1189	292	126	68	41	27	18	12	8	7	6
5	128453	5132	1278	315	136	74	45	29	20	14	10	8	7
6	136385	5450	1358	335	145	79	48	32	22	15	11	9	8
7	143618	5739	1430	353	154	84	52	34	23	17	12	10	9
8	150308	6007	1498	370	161	88	55	36	25	18	13	11	10
9	156563	6258	1561	386	169	93	57	38	27	19	14	12	11
10	162457	6494	1620	401	176	97	60	40	28	21	15	13	12
11	168047	6718	1676	416	182	101	63	42	30	22	16	14	13
12	173376	6931	1730	430	189	105	66	44	32	23	18	15	14
14	183376	7332	1831	456	201	112	71	48	35	26	20	17	16
16	192656	7705	1925	480	212	119	75	52	38	28	22	20	18
18	201354	8054	2013	503	223	125	80	55	40	31	24	22	20
20	209565	8383	2096	524	233	132	84	59	43	33	26	24	22
25	228411	9140	2288	575	257	146	95	67	50	39	32	29	27
30	245409	9823	2461	620	280	160	105	75	57	45	37	34	32
35	261017	10451	2620	663	300	174	115	83	64	51	43	39	37
40	275528	11035	2769	703	320	186	124	90	70	57	48	44	42
45	289146	11583	2909	740	339	198	133	98	77	63	53	50	47
50	302017	12101	3041	776	357	210	142	105	83	68	59	55	52
60	325938	13065	3288	844	391	233	159	119	95	80	69	65	62
70	347919	13952	3516	906	423	254	176	133	108	91	80	75	72
80	368368	14778	3728	965	454	275	192	147	120	102	90	85	82
90	387566	15553	3928	1022	483	295	208	160	132	113	101	96	92
100	405718	16287	4118	1075	512	314	223	174	144	124	111	106	102

Tabelle 7: Erforderliche Populationskorrelation R für einen Signifikanztest mit Signifikanzniveau α = .05. und Teststärke 1 - ß = .80 in Abhängigkeit vom Stichprobenumfang N und der Anzahl der Prädiktoren K.

Stichprobenumfang N	Anzahl der Prädiktoren K: 1	2	3	4	5	6	7	8	9	10	15	20	30	50	75	100
3	.99	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	.96	.96	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	.92	.92	.96	-	-	-	-	-	-	-	-	-	-	-	-	-
6	.88	.88	.92	.97	-	-	-	-	-	-	-	-	-	-	-	-
7	.85	.85	.89	.93	.97	-	-	-	-	-	-	-	-	-	-	-
8	.82	.82	.86	.90	.94	.97	-	-	-	-	-	-	-	-	-	-
9	.79	.79	.83	.87	.91	.94	.97	-	-	-	-	-	-	-	-	-
10	.77	.77	.81	.84	.88	.91	.94	.97	-	-	-	-	-	-	-	-
12	.71	.72	.76	.80	.83	.86	.89	.92	.95	.98	-	-	-	-	-	-
14	.67	.69	.73	.76	.79	.82	.84	.87	.90	.92	-	-	-	-	-	-
16	.64	.66	.70	.73	.75	.78	.81	.83	.86	.88	-	-	-	-	-	-
18	.61	.63	.67	.70	.72	.75	.77	.80	.82	.84	.96	-	-	-	-	-
20	.58	.61	.64	.67	.70	.72	.74	.76	.79	.81	.91	-	-	-	-	-
25	.53	.56	.59	.62	.64	.66	.68	.70	.72	.74	.83	.92	-	-	-	-
30	.49	.52	.55	.57	.60	.62	.63	.65	.67	.68	.76	.84	-	-	-	-
35	.45	.49	.52	.54	.56	.58	.59	.61	.63	.64	.71	.78	.93	-	-	-
40	.43	.46	.49	.51	.53	.55	.56	.58	.59	.60	.67	.73	.86	-	-	-
45	.40	.44	.46	.48	.50	.52	.53	.55	.56	.57	.63	.69	.80	-	-	-
50	.38	.42	.44	.46	.48	.50	.51	.52	.54	.55	.60	.65	.76	-	-	-
60	.35	.39	.41	.43	.44	.46	.47	.48	.49	.50	.55	.60	.68	.88	-	-
70	.33	.36	.38	.40	.41	.43	.44	.45	.46	.47	.51	.55	.63	.79	-	-
80	.31	.34	.36	.38	.39	.40	.41	.42	.43	.44	.48	.52	.58	.72	.95	-
100	.28	.31	.32	.34	.35	.36	.37	.38	.39	.40	.43	.46	.52	.62	.78	-
125	.25	.28	.29	.31	.32	.33	.34	.34	.35	.36	.39	.41	.46	.54	.65	.80
150	.23	.25	.27	.28	.29	.30	.31	.32	.32	.33	.36	.38	.42	.49	.58	.68
175	.21	.24	.25	.26	.27	.28	.29	.29	.30	.31	.33	.35	.39	.45	.52	.60
200	.20	.22	.23	.25	.25	.26	.27	.28	.28	.29	.31	.33	.36	.42	.48	.54
300	.16	.18	.19	.20	.21	.22	.22	.23	.23	.24	.25	.27	.29	.33	.38	.42
400	.14	.16	.17	.18	.18	.19	.19	.20	.20	.21	.22	.23	.25	.29	.32	.35
500	.13	.14	.15	.16	.16	.17	.17	.18	.18	.18	.20	.21	.23	.26	.28	.31
750	.11	.12	.13	.13	.14	.14	.14	.15	.15	.15	.16	.17	.19	.21	.23	.25
1000	.09	.10	.11	.11	.12	.12	.12	.13	.13	.13	.14	.15	.16	.18	.20	.21
2000	.07	.07	.08	.08	.09	.09	.09	.09	.09	.10	.10	.11	.12	.13	.14	.15

1.12.3 *Die erforderliche Populationskorrelation R*

Sind Signifikanzniveau α, Teststärke 1 - ß, Anzahl der Prädiktoren K und der Stichprobenumfang N gegeben, dann kann mit folgender Umformung der Gl. (120)

$$R = 1 \Big/ \sqrt{\frac{N - K - 1}{c + \sqrt{d\,K - e}} + 1} \tag{121}$$

die für diesen Signifikanztest mindestens erforderliche Größe der multiplen Korrelation in der Grundgesamtheit bestimmt werden. Die Ergebnisse für $\alpha = .05$ und 1 - ß = .80 werden in Tab. 7 dargestellt, wobei wiederum für K = 1 abweichend von Gl. (121) die ungünstigsten Angaben aus Cohen (1969) herangezogen wurden.

1.12.4 *Die höchstens sinnvolle Prädiktorenzahl K*

In der Praxis tritt schließlich häufig das Problem auf, bei bekannter Abschätzung der Stärke eines Zusammenhangs zwischen einem Kriterium und einem Satz von Prädiktorvariablen, die höchstens sinnvolle Anzahl von Prädiktorvariablen festzulegen. Löst man Gl. (120) nach K auf, erhält man

$$K = N - 1 - f^2\left(c + \left|\sqrt{f^4\,(d/2)^2 + d(N-1) - f^2cd - e} - f^2\,(d/2)\right|\right) \tag{122}$$

$$\text{mit } f^2 = \frac{R^2}{1 - R^2}\ .$$

Auch für diesen Fall werden die Ergebnisse in Tab. 8 dargestellt, wobei für K = 1 die besprochene Korrektur wieder vorgenommen wurde. Hier wurden die Werte stets abgerundet, während sie bei den Tabellen 6 und 7 stets nach oben aufgerundet wurden, damit die behauptete Teststärke auf alle Fälle erreicht wird.

1.12.5 *Teststärke für die übrigen Signifikanztests bei* der *Regressionsanalyse*

Bisher wurde nur berichtet, wie die Teststärke für den Signifikanztest der multiplen Korrelation, d.h. für die Nullhypothese $\beta_1 = \beta_2 \ldots = \beta_K = 0$, berechnet werden kann. Die Teststärke aller übrigen Signifikanztests, die sich auf eine unter der Nullhypothese F-verteilte Prüfgröße stützen, erhält man auf entsprechende Weise. Auch bei diesen Signifikanztests gehört die Verteilung der Prüfgröße F unter einer alternativen Hypothese zur Klasse der nichtzentralen F-Verteilungen. Entsprechend können die Gleichungen (117), (118) und (119) in etwas abgeänderter Form eingesetzt werden.

Tabelle 8: Größte Anzahl K von Prädiktoren, für die die Teststärke $1 - \beta$ mindestens .80 beträgt bei einem Signifikanzniveau des Signifikanztests für die multiple Korrelation von α = .05.
(Wenn die Teststärke .80 stets unterschritten wird, wird dies durch - symbolisiert; das Symbol + steht, wenn bei 100 Prädiktoren die Teststärke .80 überschritten wird.)

Stichproben-Umfang N	Multiple Korrelation in der Population R .05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	N
3	-	-	-	-	-	-	-	-	-	-	-	1	3
4	-	-	-	-	-	-	-	-	-	-	1	2	4
5	-	-	-	-	-	-	-	-	-	-	2	3	5
6	-	-	-	-	-	-	-	-	-	2	3	4	6
7	-	-	-	-	-	-	-	-	-	3	4	5	7
8	-	-	-	-	-	-	-	-	-	4	5	6	8
9	-	-	-	-	-	-	-	-	2	4	6	7	9
10	-	-	-	-	-	-	-	-	2	5	7	8	10
12	-	-	-	-	-	-	-	1	4	7	9	10	12
14	-	-	-	-	-	-	-	2	5	9	11	12	14
16	-	-	-	-	-	-	-	3	6	11	13	14	16
18	-	-	-	-	-	-	1	4	8	12	14	16	18
20	-	-	-	-	-	-	1	5	9	14	16	18	20
25	-	-	-	-	-	-	3	8	13	19	21	23	25
30	-	-	-	-	-	1	5	11	17	23	26	28	30
35	-	-	-	-	-	2	7	14	21	28	31	33	35
40	-	-	-	-	-	3	10	18	25	33	36	38	40
45	-	-	-	-	-	5	12	21	30	37	40	43	45
50	-	-	-	-	1	6	15	25	34	42	45	48	50
60	-	-	-	-	2	10	21	32	43	51	55	58	60
70	-	-	-	-	4	13	27	40	51	61	65	68	70
80	-	-	-	-	6	18	33	48	60	70	75	78	80
100	-	-	-	1	10	27	46	64	78	89	94	98	100
125	-	-	-	3	18	40	64	84	+	+	+	+	125
150	-	-	-	6	26	54	82	+	+	+	+	+	150
175	-	-	-	9	35	69	+	+	+	+	+	+	175
200	-	-	1	13	45	85	+	+	+	+	+	+	200
300	-	-	4	35	92	+	+	+	+	+	+	+	300
400	-	-	9	62	+	+	+	+	+	+	+	+	400
500	-	-	17	96	+	+	+	+	+	+	+	+	500
750	-	-	46	+	+	+	+	+	+	+	+	+	750
1000	-	2	86	+	+	+	+	+	+	+	+	+	1000
2000	-	17	+	+	+	+	+	+	+	+	+	+	2000

Für den Signifikanztest eines einzelnen Regressionskoeffizienten b_1 (Nullhypothese $\beta_1 = 0$, zweiseitige Alternativhypothese), der zugleich der Signifikanztest der partiellen Korrelation $r_{y1 \cdot 23\ldots K}$ und der semipartiellen Korrelation $r_{y(1\text{-}23\ldots K)}$ ist, wird nach Cohen & Cohen (1975) als Effektstärke die für die Gesamtheit abgeschätzte Größe

(123) $$f^2 = \frac{r^2_{y(1-23\ldots K)}}{1 - R^2} = \frac{R^2_{y(12\ldots K)} - R^2_{y(23\ldots K)}}{1 - R^2_{y(12\ldots K)}}$$

verwendet. Gleichung (117) bleibt unverändert, doch muß in Gl. (118) der Wert 1 für K eingesetzt werden:

(124) $$L = c + \sqrt{d - e}.$$

Soll der zusätzliche Beitrag einer Variablengruppe B zur Schätzung des Kriteriums y, zusätzlich zu einer anderen Variablengruppe A, geprüft werden, so entspricht dies der Prüfung der Nullhypothese, daß alle Regressionskoeffizienten der Variablen der Gruppe B in der Grundgesamtheit Null sind. Dies ist gleichzeitig der Test für die multiple semipartielle Korrelation

(125) $$R^2_{y(B-A)} = R^2_{y(AB)} - R^2_{y(A)}.$$

Dementsprechend wird die für die Alternativhypothese postulierte Effektstärke mit

(126) $$f^2 = \frac{R^2_{y(AB)} - R^2_{y(A)}}{1 - R^2_{y(AB)}}$$

bestimmt. Die Gleichungen (117) und (118) müssen abgeändert werden zu

(127) $$N = \frac{L}{f^2} + K_{AB} + 1$$

und

(128) $$L = c + \sqrt{d\,K_B - e}$$

wobei K_B die Zahl der Variablen der Gruppe B und K_{AB} die Gesamtzahl der Variablen in beiden Gruppen A und B sind. In Gleichung (128) wird stets als Wert K die Zahl der Freiheitsgrade für den Zähler von Gleichung (126) eingesetzt, hingegen in Gleichung (127) die Zahl der Variablen, die bei der Bildung des Fehlervarianz-Anteils im Nenner von Gleichung (126) beteiligt sind.

Diese Art der Bestimmung der Teststärke ist nicht auf quantitativ gemessene Prädiktorvariablen beschränkt, sie kann ebenso für Prädiktoren erfolgen, durch die Kategorien von qualitativen Variablen dargestellt werden. So kann auch für die univariate Varianzanalyse mit dem selben Formalismus die Teststärke bzw. der erforderliche Stichprobenumfang für eine bestimmte Teststärke angegeben werden. Wird die Varianzanalyse nicht über die Nomenklatur der multiplen Regression sondern mit der traditionell auf Mittelwertsdifferenzen aufbauenden Symbolik beschrieben, dann tritt lediglich an die Stelle der quadrierten multiplen Korrelation R^2 das Verhältnis η^2 der Varianz zwischen den Gruppen zur Gesamtvarianz.

2. Kanonische Korrelation

2.1 Einführung

Die kanonische Korrelation kann als Verallgemeinerung der multiplen Regression im korrelationsanalytischen Modell (vgl. Kap. 1.7) gesehen werden. Bei der multiplen Regression besteht das Problem darin, eine Kriteriumsvariable mit Hilfe von zwei oder mehr Prädiktorvariablen linear zu erklären, so daß eine Funktion des Vorhersagefehlers möglichst klein wird. Dies ist äquivalent damit, die Koeffizienten der Linearkombination der Prädiktorvariablen so zu bestimmen, daß die Korrelation zwischen der Kriteriumsvariable und der Linearkombination der Prädiktorvariablen möglichst groß ist. Die kanonische Korrelation stellt insofern eine Verallgemeinerung der multiplen Regression dar, als gleichzeitig mehr als eine Kriteriumsvariable in die Analyse einbezogen wird.

Allgemein läßt sich die Problemstellung der kanonischen Korrelation folgendermaßen beschreiben: Gegeben sind eine Menge von mindestens zwei Kriteriumsvariablen und eine Menge von Prädiktorvariablen, die ebenfalls zwei oder mehr Variablen enthält. Ziel ist es, den Zusammenhang zwischen den beiden Variablenmengen mit Hilfe von möglichst wenigen und möglichst einfachen aus den beiden Variablenmengen abgeleiteten Variablen, bei möglichst kleiner Fehlervarianz - oder, was auf dasselbe hinausläuft, mit möglichst hoher Korrelation - zu beschreiben. Man konstruiert zu diesem Zweck jeweils Linearkombinationen aus den Elementen der beiden Variablenmengen, so daß die Korrelation zwischen den beiden Linearkombinationen maximal ist.

Üblicherweise wird bei dem Modell der kanonischen Korrelation davon ausgegangen, daß beide Variablenmengen Zufallsgrößen enthalten. Das bedeutet, daß in diesem Modell beide Variablenmengen als „gleichwertig“ angesehen werden, d.h. im statistischen Modell keine Variablenmenge a priori als Kriterium bzw. Prädiktor festgelegt ist. Die Beziehungen zwischen den beiden Variablenmengen werden also als ungerichtet bzw. symmetrisch angesehen. Wir werden auch hier diese Darstellungsform wählen. Analog zur klassischen Korrelation kann man die kanonische Korrelation aber auch so formulieren, daß nur die Variablen der einen Variablenmenge (Menge der Kriteriumsvariablen) als zufällige Größen aufgefaßt werden, während die Werte der Prädiktorvariablen fest vorgegebene Größen sind. Hier wird also im Modell bereits festgelegt, welche Variablenmenge als Kriterium und welche als Prädiktor anzusehen ist. Der Zusammenhang zwischen den beiden Modellen läßt sich dadurch herstellen, daß das Modell mit fest vorgegebenen Prädiktorwerten als bedingte Betrachtungsweise des Modells mit zwei Mengen von zufälligen Größen angesehen werden kann. Bei Anderson (1958, p. 296-297) wird der Zusammenhang zwischen den beiden Modellen dargestellt.

Grundlegend für die kanonische Korrelation als statistisches Modell ist die Arbeit von Hotelling (1936), der schon zuvor (1935) ein Anwendungsbeispiel aus der Psychologie beschrieb. Weitere Anwendungsbeispiele der kanonischen Korrelation etwa aus dem Bereich der Pädagogik finden sich bei Barnett und Lewis (1963). Tintner (1946) und Bartlett (1948) zeigen Anwendungen der kanonischen Korrelation in der Ökonomie. Ein Überblick über den Zusammenhang zwischen kanonischer Korrelation, Faktorenanalyse und Diskriminanzanalyse findet sich bei McKeon (1966).

In neuerer Zeit (Vinograde, 1950) wurde die kanonische Korrelation auf mehr als zwei Variablenmengen verallgemeinert. Kettenring (1971) hat in seinem Artikel die verschiedenen Modelle der verallgemeinerten kanonischen Korrelation zusammenfassend dargestellt.

In dem Buch von Gnanadesikan (1977) findet sich eine auch für den Anwender dieser Verfahren lesbare, kompakte Darstellung der Arbeit von Kettenring, deren Nomenklatur wir in Kap. 2.6 übernehmen.

Weitere grundlegende Arbeiten zu dieser Verallgemeinerung sind die Artikel von Steel (1951), Horst (1965) und Carroll (1968). Bei Lohmöller & Oerter (1979) und Hanke et al. (1980) findet sich eine empirische Untersuchung aus dem Bereich der Pädagogik, in der die verallgemeinerte Korrelation als Analysetechnik verwendet wurde. Beiträge zur verallgemeinerten Korrelation sind in der Literatur auch unter dem Stichwort „multi set factor analysis“ zu finden.

2.2 Das Modell der kanonischen Korrelation für zwei Variablenmengen mit zufälligen Größen

Wir gehen von zwei Mengen von Zufallsgrößen aus, die wir in den beiden Zufallsvektoren

$$\mathbf{x}_1 = \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1p} \end{bmatrix} \quad \text{und } \mathbf{x}_2 = \begin{bmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2q} \end{bmatrix}$$

zusammenfassen, wobei wir annehmen, daß für die Anzahl der Komponenten in beiden Vektoren gilt

$$p \leq q.$$

Das Ziel der kanonischen Korrelation ist es, den Zusammenhang zwischen den beiden Variablenmengen durch Funktionen möglichst einfacher Bauart zu beschreiben, wobei ein Maximum an Korrelation zwischen den Variablenmengen erklärt werden soll. Man wählt dazu jeweils Linearkombinationen zwischen den Elementen der beiden Variablenmengen, also

(129) $$v = \boldsymbol{\alpha}'\mathbf{x}_1, \; w = \boldsymbol{\beta}'\mathbf{x}_2$$

wobei die p- bzw. q-komponentigen Koeffizientenvektoren so zu bestimmen sind, daß die Korrelation zwischen den Zufallsvariablen v und w maximal ist. Die so konstruierten Zufallsgrößen sollen *kanonische Variablen* und der Korrelationskoeffizient zwischen beiden *kanonischen Korrelation* heißen.

Da der Wert der Korrelation zwischen den Variablen v und w sich nicht ändert, wenn wir etwa $\boldsymbol{\alpha}$ mit einem beliebigen Skalar a und ß mit einem beliebigen Skalar b multiplizieren, müssen wir bei der Suche nach den optimalen $\boldsymbol{\alpha}$ und ß an diese Koeffizientenvektoren weitere Bedingungen stellen. Diese an $\boldsymbol{\alpha}$ und ß zusätzlich gestellten Normierungsbedingungen können wir auch als Nebenbedingungen an Parameter der Zufallsgrößen v und w formulieren. Das Kriterium zur Konstruktion der kanonischen Variablen ist also:

> Bestimme die Koeffizientenvektoren $\boldsymbol{\alpha}$ und ß der Linearkombinationen $v = \boldsymbol{\alpha}'\mathbf{x}_1$ und $w = \boldsymbol{\beta}'\mathbf{x}_2$ so, daß $\varrho(v,w)$ unter den Nebenbedingungen var(v) = var(w) = 1 die maximale Korrelation zwischen den beiden Zufallsgrößen v und w ergibt.

Bezeichnen wir mit

$\boldsymbol{\Sigma}_{11}$ die Kovarianzmatrix des Zufallsvektors x_1
$\boldsymbol{\Sigma}_{22}$ die Kovarianzmatrix des Zufallsvektors x_2
$\boldsymbol{\Sigma}_{12}$ die Kovarianzmatrix zwischen den Zufallsvektoren x_1 und x_2
$\boldsymbol{\Sigma}_{21}$ die Kovarianzmatrix zwischen den Zufallsvektoren x_2 und x_1 wobei $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}'$ gilt,

so können wir das Maximierungsproblem auch so formulieren:

Zu maximieren ist $\varrho(v,w) = \boldsymbol{\alpha}'\boldsymbol{\Sigma}_{12}\boldsymbol{\beta}$ unter den Nebenbedingungen $\mathrm{var}(v) = \boldsymbol{\alpha}'\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha} = 1$; $\mathrm{var}(w) = \boldsymbol{\beta}'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta} = 1$.

Aus der mathematischen Ableitung ergibt sich dann das Problem, die Gleichungssysteme

$$\mathbf{\Sigma}_{12}\boldsymbol{\beta} - \varrho\mathbf{\Sigma}_{11}\boldsymbol{\alpha} = 0$$
$$\mathbf{\Sigma}_{21}\boldsymbol{\alpha} - \varrho\mathbf{\Sigma}_{22}\boldsymbol{\beta} = 0$$

(130)
$$\boldsymbol{\alpha}'\mathbf{\Sigma}_{11}\boldsymbol{\alpha} = 1$$
$$\boldsymbol{\beta}'\mathbf{\Sigma}_{22}\boldsymbol{\beta} = 1$$

nach $\boldsymbol{\alpha}$ und ß aufzulösen, wobei ϱ abkürzend für den Wert des Korrelationskoeffizienten an der Stelle des Maximums steht.

Nach einigen algebraischen Umformungen dieser Gleichungssysteme (eine ausführliche Darstellung findet man bei Timm, 1975, p. 348-349 oder bei Anderson, 1958, p. 288-296) erhält man

(131)
$$(\mathbf{\Sigma}_{11}{}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}{}^{-1}\mathbf{\Sigma}_{21} - \varrho^2\mathbf{I})\ \boldsymbol{\alpha} = 0$$
$$(\mathbf{\Sigma}_{22}{}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}{}^{-1}\mathbf{\Sigma}_{12} - \varrho^2\mathbf{I})\ \boldsymbol{\beta} = 0.$$

Damit wurde das Auffinden der zur maximalen Korrelation gehörenden Koeffizienten $\boldsymbol{\alpha}$ und ß auf ein Eigenwertproblem zurückgeführt. Eine zusammenfassende Darstellung des Eigenwertproblems findet man in fast allen Statistikbüchern zur multivariaten Analyse und zur linearen Algebra (z.B. Oberhofer, 1979). Man beachte, daß hier die zusätzlichen Modellannahmen eingegangen sind, daß $\mathbf{\Sigma}_{11}$ und $\mathbf{\Sigma}_{22}$ nichtsingulär sind.

Bevor wir angeben, wie aus (131) die gesuchten optimalen $\boldsymbol{\alpha}$ und ß ermittelt werden, müssen wir einige allgemeine Bemerkungen zu diesem Eigenwertproblem machen. Für die beiden Eigenwertgleichungen ergeben sich jeweils gleich viele, identische, von Null verschiedene Eigenwerte. Die Anzahl der von Null verschiedenen Eigenwerte ist gleich dem Rang der Kovarianzmatrix $\mathbf{\Sigma}_{12}$ (Rao, 1973, p. 496; Anderson, 1958, p. 293-294). Nehmen wir an, daß der Rang von $\mathbf{\Sigma}_{12}$ gleich k ist, dann gibt es also zur ersten Eigenwertgleichung p - kund zur zweiten q - k viele Eigenwerte mit dem Wert Null.

Zur Bestimmung der gesuchten Koeffizientenvektoren $\boldsymbol{\alpha}$ und ß benötigen wir also nur das erste Gleichungssystem in (131). Man bestimmt den größten Eigenwert ϱ_1^2 (wir nehmen an, daß die Eigenwerte der Größe nach geordnet sind: $\varrho_1^2 \geqq \varrho_2^2 \geqq \ldots \geqq \varrho_k^2$); die positive Wurzel aus dem maximalen Eigenwert heißt *erste kanonische Korrelation*. Dann bestimmt man den zum maximalen Eigenwert gehörenden Eigenvektor $\boldsymbol{\alpha}_1$ und berechnet schließlich aus der Beziehung

(132)
$$\boldsymbol{\beta}_1 = \frac{1}{\varrho_1}\ \mathbf{\Sigma}_{22}{}^{-1}\mathbf{\Sigma}_{21}\boldsymbol{\alpha}_1$$

den Eigenvektor, der zum maximalen Eigenwert im zweiten Gleichungssystem in (131) gehört.

Die beiden Linearkombinationen

(133) $$v_1 = \boldsymbol{\alpha}_1'\mathbf{x}_1, \quad w_1 = \boldsymbol{\beta}_1'\mathbf{x}_2$$

sind dann die gesuchten Variablen mit maximaler Korrelation,

Zur praktischen Lösung des obigen Eigenwertproblems führt man die Eigenwertgleichungen (131) in Eigenwertgleichungen mit symmetrischen Matrizen über. Dies geschieht mit Hilfe der Faktorisierung von symmetrischen Matrizen nach Cholesky, wonach sich jede symmetrische Matrix als Produkt einer unteren Dreiecksmatrix und ihrer Transponierten darstellen läßt. Setzen wir also für $\boldsymbol{\Sigma}_{11}$ = TT' und $\boldsymbol{\Sigma}_{22}$ = FF' so ergeben sich mit

(134) $$\boldsymbol{\gamma}_1 = \mathbf{T}'\boldsymbol{\alpha}_1 \text{ bzw. } \boldsymbol{\delta}_1 = \mathbf{F}'\boldsymbol{\beta}_1$$

anstelle von (131) die beiden Gleichungssysteme

(135) $$\begin{aligned}(\mathbf{T}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}{}^{-1}\boldsymbol{\Sigma}_{21}\,(\mathbf{T}')^{-1} - \varrho^2\mathbf{I})\boldsymbol{\gamma}_1 = 0.\\ (\mathbf{F}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}{}^{-1}\boldsymbol{\Sigma}_{12}\,(\mathbf{F}')^{-1} - \varrho^2\mathbf{I})\boldsymbol{\delta}_1 = 0.\end{aligned}$$

Mit der Ableitung der beiden kanonischen Variablen ist eigentlich die gestellte Aufgabe, zwei Linearkombinationen aus den beiden Variablenmengen mit maximaler Korrelation zu konstruieren, gelöst. Nun ergibt sich in praktischen Anwendungsfällen häufig das Problem, daß durch diese beiden Linearkombinationen nicht sehr viel der gemeinsamen Korrelation zwischen den beiden Variablenmengen erklärt werden kann. Das bedeutet, daß die konstruierten kanonischen Variablen nur einen geringen Erklärungswert für das Prädiktionsproblem haben. Nun haben wir oben festgestellt, daß es ingesamt k von Null verschiedene Eigenwerte und damit k von Null verschiedene Korrelationen gibt, die den Zusammenhang zwischen den beiden Variablenmengen beschreiben. Es liegt also nahe, weitere Linearkombinationen der Variablenmengen mit Hilfe der zu diesen Eigenwerten gehörenden Eigenvektoren zu konstruieren. Die auf diese Weise konstruierten Linearkombinationen $v_i = \boldsymbol{\alpha}_i'\mathbf{x}_1$, $w_j = \boldsymbol{\beta}_j'\mathbf{x}_2$ haben die Eigenschaften

$$\text{(I)}\quad \operatorname{cov}(v_i, v_j) = \begin{cases} 1 \text{ für } i = j \\ 0 \text{ für } i \neq j \end{cases}$$

$$\text{(II)}\quad \operatorname{cov}(w_i, w_j) = \begin{cases} 1 \text{ für } i = j \\ 0 \text{ für } i \neq j \end{cases}$$

$$\text{(III)}\quad \operatorname{cov}(v_i, w_i) = \begin{cases} \varrho_i \neq 0 & \text{für } i = 1,2, \ldots, k \\ 0 & \text{für } i > k \end{cases}$$

$$\text{(IV)}\quad \operatorname{cov}(v_i, w_j) = 0 \text{ für } i \neq j$$

(Zum Beweis: siehe z.B. Rao, 1973, p. 496). Insgesamt können wir dann k solcher Linearkombinationen konstruieren, die vollständig den korrelativen

Zusammenhang zwischen den beiden Variablenmengen beschreiben. Man nennt die zum i-ten größten Eigenwert gehörenden Linearkombinationen

$$v_i = \boldsymbol{\alpha}_i'\mathbf{x}_1, \quad w_i = \boldsymbol{\beta}_i'\mathbf{x}_2 \tag{136}$$

die *i-ten kanonischen Variablen.*

Zusammenfassung der Eigenschaften aller möglichen Linearkombinationen, die mit Hilfe aller Eigenvektoren $\boldsymbol{\alpha}_i$ (i = 1,2...,p) und $\boldsymbol{\beta}_j$ (j = 1,2,...,q) gebildet werden können:

Berücksichtigt man noch die zu den Eigenwerten mit dem Wert Null gehörenden Eigenvektoren und faßt man alle Eigenvektoren $\boldsymbol{\alpha}_i$ und $\boldsymbol{\beta}_j$ zu den Matrizen

$$\mathbf{A} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_p), \quad \mathbf{B} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_q) \tag{137}$$

zusammen[2]), so ergibt sich wegen der Unkorreliertheit der Linearkombinationen für die Kovarianzmatrizen von

$$\mathbf{A}'\mathbf{x}_1 \text{ und } \mathbf{B}'\mathbf{x}_2$$

$$\mathbf{A}'\boldsymbol{\Sigma}_{11}\mathbf{A} = \mathbf{I} \text{ und } \mathbf{B}'\boldsymbol{\Sigma}_{22}\mathbf{B} = \mathbf{I}.$$

Für die Kovarianzmatrix zwischen $A'x_1$ und $B'x_2$ erhalten wir eine Matrix der Form

$$L = \underbrace{\left.\begin{matrix} \rho_1 & 0 & \ldots & 0 & 0 & \ldots & 0 & \ldots & 0 \\ 0 & \rho_2 & \ldots & 0 & 0 & \ldots & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & & & \vdots & & \vdots \\ 0 & 0 & \ldots & \rho_k & 0 & \ldots & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 & \ldots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 & \ldots & 0 \end{matrix}\right\}}_{q} p$$

Die Transformation der ursprünglichen Variablenvektoren x_1 und x_2 mit Hilfe der oben gefundenen Eigenvektoren bewirkt also, daß die ursprüngliche Kovarianzmatrix

---

[2]) Wenn $\boldsymbol{\Sigma}_{12}$ den Rang k < p hat, dann sind die Koeffizientenmatrizen A und B nicht eindeutig bestimmt. Diese Unbestimmtheit beinhaltet, daß A und B nur bestimmt sind bis auf orthogonale Transformationen. Diese Unbestimmtheit kann durch zusätzliche Forderungen an die Matrizen A und B beseitigt werden. Dieses Problem ist analog dem *Rotationsproblem* im Modell der Faktorenanalyse.

(139) $$\begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix} \text{ transformiert wird in } \begin{pmatrix} \mathbf{I}_p & \mathbf{L} \\ \mathbf{L}' & \mathbf{I}_q \end{pmatrix}$$

2.3 Schätzung der kanonischen Korrelationen und der Koeffizientenvektoren der kanonischen Variablen

Wir nehmen an, daß von den Zufallsvektoren $\mathbf{x}_1$ und x_2 jeweils N Realisierungen vorliegen. Fassen wir diese Realisierungen in den Datenmatrizen $X_{1(Nxp)}$ und $X_{2(Nxq)}$ zusammen, so erhalten wir für die Stichprobenmatrix bei entsprechender Aufspaltung wie in (139)

(140) $$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix}$$

mit

$$\mathbf{S}_{11} = \frac{1}{N} \mathbf{X}_1' \left(\mathbf{I} - \frac{1}{N}\mathbf{J}\right) \mathbf{X}_1$$

(141) $$\mathbf{S}_{12} = \mathbf{S}_{21}{}' = \frac{1}{N} \mathbf{X}_1' \left(\mathbf{I} - \frac{1}{N}\mathbf{J}\right) \mathbf{X}_2$$

$$\mathbf{S}_{22} = \frac{1}{N} \mathbf{X}_2' \left(\mathbf{I} - \frac{1}{N}\mathbf{J}\right) \mathbf{X}_2$$

wobei J die Einsmatrix vom Typ (NxN) ist.

Unter der Annahme, daß die Zufallsvektoren x_1, x_2 gemeinsam normalverteilt sind, ist S der Maximum-Likelihood-Schätzer für die Kovarianzmatrix $\mathbf{\Sigma}$. Der Schätzer für **A,B** und für die kanonischen Korrelationen ϱ_i ergeben sich aus (131) bzw. (135), indem wir die Populationsparameter durch die jeweiligen Stichprobengrößen ersetzen und dieselben algebraischen Umformierungen vornehmen wie in Abschnitt 2.2. Unter Beachtung der Bemerkungen in Fußnote [2]) ergeben sich Schätzer für **A,B** und die ϱ_i, die zugleich Maximum-Likelihood-Schätzer für diese Parameter sind (vgl. dazu Anderson, 1958, p. 299). Die Verteilung dieser Schätzer wird bei Kshirsagar (1972, p. 261-277) dargestellt. Bezeichnen wir die Maximum-Likelihood-Schätzer für $\boldsymbol{\alpha}_i$, $\boldsymbol{\beta}_i$ und ϱ_i mit $\mathbf{a}_i$, $\mathbf{b}_i$ und $\mathbf{r}_i$, so gelten analog wie für die Parameter folgende Beziehungen.

(142) $$\begin{aligned} \mathbf{S}_{12}\mathbf{b}_j &= r_j\, \mathbf{S}_{11}\mathbf{a}_j \\ \mathbf{S}_{21}\mathbf{a}_j &= r_j\, \mathbf{S}_{22}\mathbf{b}_j \\ \mathbf{a}_j'\mathbf{S}_{11}\mathbf{a}_j &= 1 \\ \mathbf{b}_j'\mathbf{S}_{22}\mathbf{b}_j &= 1 \end{aligned}$$

Berechnet man die kanonischen Variablen aus der Stichprobenkorrelationsmatrix R, wobei R analog zur Stichprobenkovarianzmatrix S aufgespalten wird in

(143) $$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{pmatrix},$$

so gelten folgende Beziehungen

(144) $$\begin{aligned} \mathbf{R}_{12}\tilde{\mathbf{b}}_j &= r_j \mathbf{R}_{11}\tilde{\mathbf{a}}_j \\ \mathbf{R}_{21}\tilde{\mathbf{a}}_j &= r_j \mathbf{R}_{22}\tilde{\mathbf{b}}_j \\ \tilde{\mathbf{a}}_j' \mathbf{R}_{11}\tilde{\mathbf{a}}_j &= 1 \\ \tilde{\mathbf{b}}_j' \mathbf{R}_{22}\tilde{\mathbf{b}}_j &= 1 \end{aligned}$$

mit

(145) $$\tilde{\mathbf{a}}_j = [\mathrm{diag}(\mathbf{S}_{11})]^{\frac{1}{2}}\, a_j, \quad \tilde{\mathbf{b}}_j = [\mathrm{diag}(\mathbf{S}_{22})]^{\frac{1}{2}}\, \mathbf{b}_j.$$

2.4 Test zur Bestimmung der Anzahl der kanonischen Variablen

Wir hatten bereits mehrfach festgestellt, daß die maximale Anzahl der kanonischen Variablen gleich dem Rang der Kovarianzmatrix $\mathbf{\Sigma}_{12}$ ist. Wenn der Rang $(\mathbf{\Sigma}_{12}) = k < p$ ist, lassen sich also k kanonische Variablenpaare mit von Null verschiedenen kanonischen Korrelationen konstruieren. Wenn man in praktischen Fällen den Parameter k schätzen will, ergibt sich das folgende Problem. Auch wenn Rang $(\mathbf{\Sigma}_{12}) = k < p$ ist, hat die Matrix S_{12} mit Wahrscheinlichkeit 1 vollen Rang, also p. Das bedeutet, daß sich aufgrund dieses Rangkriteriums immer p Paare kanonischer Stichprobenvariablen ergeben. Bei der Entscheidung, wie viele kanonische Variablen aus den Stichprobenwerten konstruiert werden sollen, geht man nun anders vor; man prüft zunächst die Hypothese, daß zwischen den beiden Variablenmengen kein Zusammenhang besteht. Diese Hypothese besagt, daß $\mathbf{\Sigma}_{12} = 0$ ist. Zur Prüfung dieser Hypothese gibt es Tests von Hotelling (1951), Pillai (1955), Roy (1957, Kap. 14) und Wilks (1932). Bei Kshirsagar (1972, p. 331-334) findet man eine kurze Diskussion dieser Tests mit weiteren Literaturhinweisen insbesondere zu Untersuchungen über die Güte dieser Tests. Allerdings gibt es für die Anwendung dieser Tests kaum Entscheidungsregeln dafür, welches der verschiedenen Testkriterien vorzuziehen ist. Da für die Prüfgröße nach Wilks eine für praktische Fälle sehr gute Chiquadrat-Approximation von Bartlett (1938) existiert, hat sich diese Prüfgröße in der empirischen Anwendung durchgesetzt:

(146) $$\Lambda = \prod_{i=1}^{s} (1 - r_i^2); \quad s = \min(p,q).$$

Die Chiquadrat-Approximation von Bartlett dazu ist

(147) $$\chi_B^2 = -\left[(N-1) - \tfrac{1}{2}(p+q+1)\right] \log \Lambda,$$

wobei diese Prüfgröße annähernd χ^2-verteilt ist, mit $p \cdot q$ Freiheitsgraden. Danach wird die Hypothese

$$H_0\text{: } \mathbf{\Sigma}_{12} = 0 \text{ zugunsten von } H_1\text{: } \mathbf{\Sigma}_{12} \neq 0$$

immer dann verworfen, wenn der Prüfwert

$$\chi^2_B > \chi^2_\alpha(pq) \text{ ist,}$$

wobei a das Signifikanzniveau ist. Eine bessere Näherung als die von Bartlett an Wilks lambda-Kriterium stammt von Rao (1973, p. 472), deren Berechnung bei Gaensslen & Schubö (1976, p. 176-177) zu finden ist. Wenn die Nullhypothese H_0: $\mathbf{\Sigma}_{12} = 0$, d.h. daß kein Zusammenhang zwischen den beiden Variablenmengen besteht, zu verwerfen ist, eliminiert man den Beitrag der ersten kanonischen Korrelation von Λ und prüft, ob die Prüfgröße

$$\chi^2_B = -\left[(N-1) - \frac{1}{2}(p+q+1)\right] \log \Lambda^* \tag{148}$$

$$\text{mit} \quad \Lambda^* = \prod_{i=2}^{s} (1 - r_i^2)$$

einen signifikanten Zusammenhang zwischen den beiden Variablenmengen liefert. Die Prüfgröße ist in diesem Fall annähernd Chiquadrat-verteilt mit (p - 1)(q - 1) Freiheitsgraden. Dieses Vorgehen wird so lange fortgesetzt, bis die Nullhypothese nicht mehr zu verwerfen ist. Daraus ergibt sich dann die Anzahl der signifikanten kanonischen Korrelationen und damit die Anzahl der signifikanten kanonischen Variablen.

Zu diesem Testvorgehen ist allerdings zu bemerken, daß bei der Durchführung mehrerer Tests mit Hilfe desselben Datenmaterials die tatsächliche Irrtums-Wahrscheinlichkeit größer als das vorgegebene Signifikanzniveau a ist (vgl. dazu Gaensslen & Schubö, 1976, p. 112 und Haagen & Seifert, 1979, Kap. 9).

2.5 Extraktions- und Redundanzmaße

Zur besseren Interpretation der kanonischen Variablen ist es von Interesse, einerseits die Korrelation jeder kanonischen Variablen einer Variablenmenge mit den einzelnen Variablen dieser Variablenmenge zu kennen. Daraus lassen sich Varianzanteile berechnen, die angeben, zu wieviel Prozent die Varianz innerhalb einer Variablenmenge der i-ten kanonischen Variablen zugerechnet werden kann. Andererseits ist man an sog. Redundanzmaßen interessiert, die angeben, wieviel Prozent der Variabilität einer Variablenmenge der Variabilität der kanonischen Variablen, die aus der anderen Variablenmenge konstruiert wurde, zugerechnet werden kann. Diese Redundanzmaße geben einen Index für das korrelative überlappen der beiden Variablenmengen an. Je größer der Redundanzindex, desto stärker ist die Überlappung der beiden Variablenmengen. Eine ausführliche Darstellung der hier behandelten Extraktions- und Redunzanzmaße einschließlich eines Anwendungsbeispiels findet man bei Gaensslen & Schubö (1976, p. 179-191).

Die *Strukturmatrix* ist diejenige Matrix, die die Korrelation zwischen den kanonischen Variablen einer Variablenmenge und den Elementen dieser Variablenmenge beschreibt. Für jede Variablenmenge läßt sich also eine zugehörige Strukturmatrix konstruieren.

Allgemein ergibt sich für den p-komponentigen Vektor mit den Korrelationen zwischen der kanonischen Variablen v_i und den Komponenten des Zufallsvektors x_1

$$\varrho(\mathbf{x}_1, v_i) = \varrho(\mathbf{x}_1, \boldsymbol{\alpha}_i'\mathbf{x}_1) = [\text{diag}\ (\boldsymbol{\Sigma}_{11})]^{-\frac{1}{2}}\ \boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}_i.$$

Entsprechend gilt für die Korrelationen zwischen der kanonischen Variablen w_i mit den Komponenten des Zufallsvektors x_2

$$\varrho(\mathbf{x}_2, w_i) = \varrho(\mathbf{x}_2, \boldsymbol{\beta}_i'\mathbf{x}_2) = [\text{diag}\ (\boldsymbol{\Sigma}_{22})]^{-\frac{1}{2}}\ \boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_i.$$

Die entsprechenden Korrelationen aus den Stichprobenwerten sind dann

$$r(\mathbf{X}_1, \hat{v}_i) = [\text{diag}(\mathbf{S}_{11})]^{-\frac{1}{2}}\ \mathbf{S}_{11}\mathbf{a}_i$$

bzw.

$$r(\mathbf{X}_2, \hat{w}_i) = [\text{diag}\ (\mathbf{S}_{22})]^{-\frac{1}{2}}\ \mathbf{S}_{22}\mathbf{b}_i$$

wobei hier mit X_1 und X_2 die Beobachtungswertmatrizen bezeichnet werden.

Für standardisierte Meßwerte Z_{x1}, Z_{x2} erhalten wir schließlich

(149) $$r\ (\mathbf{Z}_{x_1}, \hat{v}_i) = \mathbf{R}_{11}\tilde{\mathbf{a}}_i, \text{ mit } \tilde{\mathbf{a}}_i = [\text{diag}(\mathbf{S}_{11})]^{\frac{1}{2}}\ \mathbf{a}_1$$

bzw.

(150) $$r(\mathbf{Z}_{x_2}, \hat{w}_i) = \mathbf{R}_{22}\tilde{\mathbf{b}}_i\ , \text{ mit } \tilde{\mathbf{b}}_i = [\text{diag}(\mathbf{S}_{22})]^{\frac{1}{2}}\ \mathbf{b}_i$$

Die beiden Strukturmatrizen K_1 und K_2 sind also

(151) $$\begin{aligned}\mathbf{K}_1 &= (\mathbf{R}_{11}\tilde{\mathbf{a}}_1, \mathbf{R}_{11}\tilde{\mathbf{a}}_2, \ldots, \mathbf{R}_{11}\tilde{\mathbf{a}}_p) = \mathbf{R}_{11}(\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2, \ldots, \tilde{\mathbf{a}}_p) = \mathbf{R}_{11}\tilde{\mathbf{A}}\\ \mathbf{K}_2 &= (\mathbf{R}_{22}\tilde{\mathbf{b}}_1, \mathbf{R}_{22}\tilde{\mathbf{b}}_2, \ldots, \mathbf{R}_{22}\tilde{\mathbf{b}}_p) = \mathbf{R}_{22}(\tilde{\mathbf{b}}_1, \tilde{\mathbf{b}}_2, \ldots, \tilde{\mathbf{b}}_q) = \mathbf{R}_{22}\tilde{\mathbf{B}}.\end{aligned}$$

Das Skalarprodukt von $r(Z_{x1}, \hat{v}_i)$ kann interpretiert werden als Quadrat des multiplen Korrelationskoeffizienten zwischen $\hat{v}_i$ und den Realisierungen der standardisierten Zufallsvariablen x_{11}, x_{12},. . . ,x_{1p}. Da die Gesamtvarianz in der ersten Variablenmenge gleich p ist, kann

(152) $$v_{x_1}^2 = \frac{1}{p}[(r(\mathbf{Z}_{x_1}, \hat{v}_i)]'[r(\mathbf{Z}_{x1},\hat{v}_i)] = \frac{\tilde{\mathbf{a}}_i{}'\mathbf{R}_{11}^2\ \tilde{\mathbf{a}}_i}{p}$$

auch als der Varianzanteil der ersten Variablenmenge interpretiert werden, der der i-ten kanonischen Variablen dieser Variablenmenge zugerechnet werden kann.

Analog ist

(153) $$v_{x_2}^2 = \frac{1}{q}\ [r(\mathbf{Z}_{x_2}, \hat{w}_i)]'[r(\mathbf{Z}_{x_2}, \hat{w}_i)] = \frac{\tilde{\mathbf{b}}_i'\mathbf{R}_{22}^2\tilde{\mathbf{b}}_i}{q}$$

der Varianzanteil der zweiten Variablenmenge, der der i-ten kanonischen Variablen der zweiten Variablenmenge zugerechnet werden kann.

Von größerer Bedeutung für die Interpretation der kanonischen Variablen sind die sog. *Redundanzmaße,* die, wie bereits einleitend erwähnt, das Ausmaß der gegenseitigen Überlappung der beiden Variablenmengen beschreiben. Man berechnet dazu zunächst die Korrelationen zwischen den kanonischen Variablen v_i und den Komponenten des anderen Zufallsvektors x_2.

Dafür ergibt sich

$$\varrho(\mathbf{x}_2,v_i) = \varrho(\mathbf{x}_2,\boldsymbol{\alpha}_i'\ \mathbf{x}_1) = [\text{diag}\ (\boldsymbol{\Sigma}_{22})]^{-\frac{1}{2}}\ \boldsymbol{\Sigma}_{21}\boldsymbol{\alpha}_i$$

bzw. wegen $\varrho_i\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_i = \boldsymbol{\Sigma}_{21}\boldsymbol{\alpha}_i$

$$\varrho(\mathbf{x}_2,v_i) = \varrho_i \cdot [\text{diag}(\boldsymbol{\Sigma}_{22})]^{-\frac{1}{2}}\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_i$$

wobei $\varrho_i = \boldsymbol{\alpha}_i'\ \boldsymbol{\Sigma}_{12}\ \boldsymbol{\beta}_i$;

analog ist

$$\varrho(\mathbf{x}_1,w_i) = \varrho(\mathbf{x}_1,\boldsymbol{\beta}_i'\ \mathbf{x}_2) = [\text{diag}(\boldsymbol{\Sigma}_{11})]^{-\frac{1}{2}}\boldsymbol{\Sigma}_{12}\boldsymbol{\beta}_i$$
$$= \varrho_i\cdot[\text{diag}\ (\boldsymbol{\Sigma}_{11})]^{-\frac{1}{2}}\ \boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}_i.$$

Für die Korrelationen aus den Stichprobenwerten ergibt sich dann, wenn standardisierte Meßwerte zugrunde liegen,

(154) $$r\ (\mathbf{Z}_{x_2},\hat{v}_i) = \mathbf{R}_{21}\tilde{\mathbf{a}}_i = r_i\mathbf{R}_{22}\tilde{\mathbf{b}}_i$$

bzw.

(155) $$r\ (\mathbf{Z}_{x_1},\hat{w}_i) = \mathbf{R}_{12}\tilde{\mathbf{b}}_i = r_i\mathbf{R}_{11}\tilde{\mathbf{a}}_i$$

wobei r_i die i-te kanonische Korrelation der Stichprobe bezeichnet. Das Skalarprodukt von (154) kann interpretiert werden als Quadrat des multiplen

Korrelationskoeffizienten zwischen $\hat{v}_i$ und den Realisierungen der standardisierten Zufallsvariablen $x_{21},\ldots,x_{2q}$. Da die Gesamtvarianz dieser Variablenmenge gleich q ist, kann

$$(156)\qquad \nu^2_{x_2|v_i} = \frac{1}{q}[r(\mathbf{Z}_{x_2},\hat{v}_i)]'[r(\mathbf{Z}_{x_2},\hat{v}_i)] = \frac{1}{q}\tilde{\mathbf{a}}_i'\mathbf{R}_{12}\mathbf{R}_{21}\tilde{\mathbf{a}}_i = \frac{1}{q}r_i^2\,\tilde{\mathbf{b}}_i'\,\mathbf{R}^2_{22}\,\tilde{\mathbf{b}}_i$$

als Varianzanteil der *zweiten* Variablenmenge interpretiert werden, der der i-ten kanonischen Variablen der *ersten* Variablenmenge zugeordnet werden kann.

Analog erhalten wir:

$$(157)\qquad \omega^2_{x_1|w_i} = \frac{1}{p}[r(\mathbf{Z}_{x_1},\hat{w}_i)]'[r(\mathbf{Z}_{x_1},\hat{w}_i)] = \frac{1}{p}\tilde{\mathbf{b}}_i'\mathbf{R}_{21}\mathbf{R}_{12}\tilde{\mathbf{b}}_i = \frac{1}{p}r_i^2\,\tilde{\mathbf{a}}_i'\,\mathbf{R}^2_{11}\,\tilde{\mathbf{a}}_i$$

als den Varianzanteil der ersten Variablenmenge, der der i-ten kanonischen Variablen der *zweiten* Variablenmenge zugerechnet werden kann.

Summiert man die Redundanzindizes (156) bzw. (157) über alle signifikanten kanonischen Variablen, so nennt man diese Summen die *totale Redundanz* der Variablenmenge $\{x_{21},\ldots,x_{2q}\}$ bzw. $\{x_{11},\ldots,x_{1p}\}$.

Zwischen den oben definierten Extraktions- und Redundanzmaßen gelten folgende Zusammenhänge, die für die praktische Berechnung nützlich sind:

$$(158)\qquad \nu^2_{x_2|v_i} = r_i^2\,\omega^2_{x_2}$$

bzw.

$$(159)\qquad \omega^2_{x_1|w_i} = r_i^2\,\nu^2_{x_1}.$$

2.6 Verallgemeinerung der kanonischen Korrelation auf mehr als zwei Variablenmengen

Anstelle der bisherigen zwei Zufallsvektoren seien jetzt m Zufallsvektoren

$$\mathbf{x}'_j = (x_{j1}, x_{j2},\ldots,x_{jp_j}),\ j = 1,2,\ldots m$$

gegeben.

Die Zufallsvektoren seien so indiziert, daß gilt

$$p_1 \leqq p_2 \leqq \ldots \leqq p_m \text{ mit } p = \sum_{j=1}^{m} p_j.$$

Wiederum seien $\boldsymbol{\Sigma}_{jl}$ die Kovarianzmatrizen zwischen x_j und x_l.

Die Kovarianzmatrizen für j = 1 sollen nichtsingulär sein. Analog zum Fall m = 2 erfolgt die Konstruktion der kanonischen Variablen in p_1 (kleinste Anzahl der Komponenten unter den m Zufallsvektoren) Schritten. Auf jeder Stufe gibt es allerdings statt eines Paares jetzt m kanonische Variablen. Gesucht ist also auf der Stufe s ein Vektor

$$\mathbf{z}^{(s)\prime} = (z_1^{(s)}, z_2^{(s)}, \ldots, z_m^{(s)}),$$

wobei $z_j^{(s)}$ eine Linearkombination des Zufallsvektors x_j ist, die analog der Normierungsbedingung im gewöhnlichen Fall m = 2 die Varianz 1 haben soll.

Entsprechend der Vorgehensweise im Fall m = 2 werden im verallgemeinerten Modell die Linearkombinationen so bestimmt, daß bestimmte Funktionen der Korrelationsmatrix $\boldsymbol{\Phi}^{(s)}$ von $\mathbf{z}^{(s)}$ maximiert (bzw. minimiert) werden. Seien

$$z_j^{(1)} = \boldsymbol{\alpha}_j^{(1)\prime}\, \mathbf{x}_j,\; j = 1,2,\ldots,m \tag{160}$$

die gesuchten Linearkombinationen auf der ersten Stufe, so sind z.B. nach dem SUMCOR-Kriterium „sum of correlations method" (vgl. Kettenring, 1971, p. 435), die Koeffizientenvektoren $\boldsymbol{\alpha}_j^{(1)}$ so zu bestimmen, daß

$$\sum_{i=1}^{m} \sum_{j=i+1}^{m} \varphi_{ij}^{(1)} \overset{!}{=} \text{Maximum} \tag{161}$$

unter den Normierungsbedingungen $\boldsymbol{\alpha}_j^{(1)\prime}\boldsymbol{\Sigma}_{jj}\boldsymbol{\alpha}_j^{(1)} = 1$, wobei

$$\boldsymbol{\Phi}^{(1)} = \begin{bmatrix} 1 & \varphi_{12}^{(1)} & \ldots & \varphi_{1m}^{(1)} \\ \varphi_{12}^{(1)} & & & \cdot \\ \vdots & & & \vdots \\ \varphi_{1m}^{(1)} & \ldots & \ldots & 1 \end{bmatrix} \tag{162}$$

die Korrelationsmatrix aller Linearkombinationen auf der ersten Stufe ist. Bei Kettenring (1971) werden insgesamt fünf solcher Optimierungskriterien zur Konstruktion der kanonischen Variablen diskutiert, die alle für den Fall m = 2 mit der gewöhnlichen kanonischen Korrelation übereinstimmen.

Wir wollen hier nur auf die MAXVAR („maximum variance method") Methode, die von Horst (1961b, 1965) stammt, eingehen, da sich diese Methode auf die Lösung von Eigenwertproblemen direkt zurückführen läßt.

Fassen wir die Koeffizientenvektoren der Linearkombinationen auf der ersten Stufe zu folgender Blockdiagonalmatrix $\mathbf{D}_{\alpha(1)}$

$$\mathbf{D}_{\alpha(1)} = \begin{bmatrix} \boldsymbol{\alpha}^{(1)\prime} & 0' \ldots & 0' \\ 0' & \boldsymbol{\alpha}_2^{(1)\prime} \ldots & 0' \\ \vdots & & \vdots \\ 0' & \ldots\ldots & \boldsymbol{\alpha}_m^{(1)\prime} \end{bmatrix}$$

zusammen, dann gilt mit

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \ldots & \boldsymbol{\Sigma}_{1m} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \ldots & \vdots \\ \vdots & \vdots & \vdots \\ \boldsymbol{\Sigma}_{m1} & \boldsymbol{\Sigma}_{m2} \ldots & \boldsymbol{\Sigma}_{mm} \end{bmatrix}$$

für $\boldsymbol{\Phi}^{(1)}$

(163) $$\boldsymbol{\Phi}^{(1)} = \mathbf{D}_{\alpha(1)} \boldsymbol{\Sigma}\, \mathbf{D}'_{\alpha(1)}.$$

Wenn wir wie im gewöhnlichen Fall die Kovarianzmatrizen $\boldsymbol{\Sigma}_{jj}$ mit Hilfe der Cholesky-Zerlegung darstellen durch $\boldsymbol{\Sigma}_{jj} = T_jT'_j$ und die transformierten Variablen $y_j = T_j^{-1}x_j$ benützen, so ergibt sich für die Kovarianzmatrix der y_j

$$\boldsymbol{\Gamma} = \begin{bmatrix} \mathbf{I} & \boldsymbol{\Gamma}_{12} & \boldsymbol{\Gamma}_{1m} \\ \boldsymbol{\Gamma}_{21} & & \vdots \\ \vdots & & \vdots \\ \boldsymbol{\Gamma}_{m1} & \boldsymbol{\Gamma}_{m2} & \mathbf{I} \end{bmatrix}$$

mit $\boldsymbol{\Gamma}_{ij} = \mathbf{T}_i^{-1} \boldsymbol{\Sigma}_{ij} (\mathbf{T}_j^{-1})'$.

Für die kanonischen Variablen $z_j^{(1)}$ gilt dann wegen $x_j = T_jy_j$

(164) $$z_j^{(1)} = \boldsymbol{\alpha}_j^{(1)\prime}\mathbf{T}_j\mathbf{y}_j = \boldsymbol{\beta}_j^{(1)\prime}\mathbf{y}_j;\ \text{mit } \boldsymbol{\beta}_j^{(1)\prime} = \boldsymbol{\alpha}_j^{(1)\prime}\mathbf{T}_j.$$

Dann ergibt sich für $\boldsymbol{\Phi}^{(1)}$

(165) $$\boldsymbol{\Phi}^{(1)} = \mathbf{D}_{\beta(1)} \boldsymbol{\Gamma}\, \mathbf{D}'_{\beta(1)},$$

wobei $\mathbf{D}_{\beta(1)}$ analog als Blockdiagonalmatrix wie $\mathbf{D}_{\alpha(1)}$ definiert ist. In diesem Fall genügen die Koeffizientenvektoren den Nebenbedingungen $\boldsymbol{\beta}_j^{(1)\prime}\, \boldsymbol{\beta}_j^{(1)} = 1$. Nach dem MAXVAR-Kriterium von Horst (1961b, 1965) sind die Koeffizientenvektoren so zu bestimmen, daß die Varianz der ersten Hauptkomponente von $z^{(l)}$ maximal ist. Dieses Optimierungskriterium ist äquivalent damit, die zum maximalen Eigenwert von $\boldsymbol{\Phi}^{(1)}$ gehörenden Eigenvektoren zu bestimmen. Der Zusammenhang zwischen den Eigenwerten von $\boldsymbol{\Phi}^{(1)}$ und $\boldsymbol{\Gamma}$ ist in Lemma 2 bei Kettenring (1971) formuliert.

Bisher wurden die kanonischen Variablen für die Grundgesamtheit formuliert. Liegt eine Stichprobe vom Umfang N für jede der m Variablenmengen vor (man beachte, daß $N > p$ sein soll), so ersetzen wir für die Ableitung der kanonischen Variablen, die aus den Stichprobenwerten gewonnen werden, analog zum gewöhnlichen Fall, in den für die Grundgesamtheit geltenden Beziehungen die Grundgesamtheitsparameter durch die jeweiligen Stichprobenkennwerte.

$S = (S_{ij})$ entspricht dann der Kovarianzmatrix $\mathbf{\Sigma}$ der Variablen x_j,
$R = (R_{ij})$ entspricht dann der Korrelationsmatrix $\mathbf{\Gamma}$ der transformierten Variablen y_j,

wobei für die einzelnen Stichproben-Kovarianzmatrizen die Cholesky-Zerlegungen vorgenommen werden.

Sei

$$\mathbf{S}_{jj} = \mathbf{E}_j \mathbf{E}'_j,$$

dann gilt

(166) $$\mathbf{R}_{ij} = \mathbf{E}_i^{-1} \mathbf{S}_{ij} (\mathbf{E}_j^{-1})'.$$

Bei der Berechnung der kanonischen Koeffizientenvektoren aus den Stichprobenwerten geht man dann folgendermaßen vor: Man berechnet die erste Hauptkomponente der Variablen y_j; dazu bestimmt man den zum größten Eigenwert von $\mathbf{R}$ gehörenden Eigenvektor $v^{(1)}$. $v^{(1)}$ wird dann partitioniert in m Teilvektoren $v^{(1)\prime} = (v_1^{(1)\prime}, v_2^{(1)\prime}, \ldots, v_m^{(1)\prime})$, wobei jeder zur Länge Eins normiert wird. Daraus ergeben sich dann die Schätzer $b_j^{(1)}$ als

(167) $$\mathbf{b}_j^{(1)} = \frac{\mathbf{v}_j^{(1)}}{\| \mathbf{v}_j^{(1)} \|} .$$

Wir haben hier nur die Bestimmung der m kanonischen Variablen auf der ersten Stufe aufgezeigt, da die Interpretation zusätzlicher kanonischer Variablen auf den weiteren Stufen in praktischen Anwendungsfallen meist auf Schwierigkeiten stößt. Ein auf Horst (1965) zurückgehendes Anwendungsbeispiel der verallgemeinerten kanonischen Korrelation wird auch bei Kettenring (1971) und in dem Buch von Gnanadesikan (1977, p. 74-78) diskutiert.

Literatur

Aigner, D. J. & Goldberger, A. S. 1977. Latent variables in socio-economic models. Amsterdam: North Holland.

Anderson, O. D. 1976. Time series analysis and forecasting. The Box-Jenkins approach. London: Butterworths.

Anderson, T. W. 1958. An introduction to multivariate statistical analysis. New York: Wiley.

Bard, Y. 1974. Nonlinear parameter estimation. New York: Academic Press.

Barnett, V. D. & Lewis, T. 1963. A study of the relation between G. C. E. and degree results. Journal of the Royal Statistical Society, A, 126, 187-195.

Bartlett, M. S. 1938. Further aspects of the theory of multiple regression. Proceedings of the Cambridge Philosophical Society, 34, 33-40.

Bartlett, M. S. 1948a. A note on the statistical estimation of demand and supply relations from time series. Econometrica, 16, 323.

Bartlett, M. S. 1948b. Internal and external factor analysis. British Journal of Psychology (Stat. Sec.), 1, 73-81.

Beutel, P., Küffner, H. & Schubö, W. 1980. Statistik-Programm-System für die Sozialwissenschaften: SPSS 8. Stuttgart: Gustav Fischer.

Beutel, P. & Schubö, W. 1982. Statistik-Programm-System für die Sozialwissenschaften: SPSS 9. Stuttgart: Gustav Fischer.

Bortz, J. 1977. Lehrbuch der Statistik. Für Sozialwissenschaftler. Berlin: Springer.

Bowden, D. C. 1970. Simultaneous confidence bands for linear regression models. Journal of the American Statistical Association, 65, 413-421.

Box, G. E. P. 1974. Statistics and the environment. Journal of the Washington Academy of Sciences, 64 (2), 52-59.

Box, G. E. P. & Jenkins, G. H. 1970. Time series analysis. Forecasting and control. San Francisco: Holden Day.

Box, G. E. P., Jenkins, G. H. & Baton, D. W. 1967. Models for forecasting seasonal and non-seasonal time series. In: Harris, B. (ed.), Spectral analysis of time series. New York: Wiley.

Box, G. E. P. & Pierce, D. A. 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. Journal of the American Statistical Association, 65, 1509-1526.

Box, G. E. P. & Tiao, G. C. 1965. A change in level of a nonstationary time series. Biometrika, 52, 181-192.

Box, G. E. P. & Tiao, G. C. 1975. Intervention analysis with applications to economic and environmental problems. Journal of the American Statistical Association, 70, 70-79.

Garroll, J. D. 1968. Generalization of canonical correlation analysis to three or more sets of variables. Proceedings of the American Psychological Association, 76th Annual Convention, 227-228.

Chang, J. J. & Carroll, J. D. 1964. A general index of nonlinear correlation and its application to the program of relating physical and psychological dimensions. American Psychologist, 19, 540-541.

Cochran, W. G. 1970. Some effects of errors of measurement on multiple correlation. Journal of the American Statistical Association, 65, p. 22-34.

Cohen, J. & Cohen, P. 1975. Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale: Lawrence Erlbaum.

Conger, A. J. 1974. A revised definition for suppressor variables: A guide to their identification and interpretation. Educational and Psychological Measurement, 34, 35-46.

Conger, A. J. & Jackson, D. M. 1972. Suppressor variables, prediction, and the interpretation of psychological relationships. Educational and Psychological Measurement, 32, 579-599.

Cooley, W. W. & Lohnes, P. R. 1971. Multivariate data analysis. New York: Wiley.

Dahme, B. 1975. Zeitreihenanalyse des psychotherapeutischen Prozesses. In: Tack, W. H. (Hrsg.). Bericht über den 29. Kongreß der Deutschen Gesellschaft für Psychologie. Göttingen: Hogrefe.

Dahme, B. 1976. Beziehungen zwischen Modellen der mathematischen Lern- und Entscheidungstheorien und ARIMA-Modellen. Konsequenzen für die Psychotherapie-Prozeßforschung. Arbeitspapier Medizinische Psychologie, Universität Hamburg.

Darlington, R. B. 1968. Multiple regression in psychological research and practice. Psychological Bulletin, 69 (3), 161-182.

Dhrymes, Ph. 1978. Introductory Economics. New York: Springer.

Dixon, W. J. & Brown, M. B. 1981. BMDP: Biomedical Computer programs P-series. Berkeley: University of California Press.

Draper, N. R. & Smith, H. 1966. Applied regression analysis. New York: Wiley.

Durbin, J. & Watson G. 1950. Testing for serial correlation in least squares regression I. Biometrika, 37, 402-428.

Durbin, J. & Watson, G. 1951. Testing for serial correlation in least squares regression II. Biometrika, 38, 159-178.

Durbin, J. & Watson, G. 1971. Testing for serial correlation in least squares regression III. Biometrika, 58, 1-19.

Edwards, A. L. 1979. Multiple regression and the analysis of variance and covariance. San Francisco: Freeman.

Finn, J. D. 1974. A general model for multivariate analysis. New York: Holt, Rinehart and Winston.

Frane, J. W. 1976. Some simple procedures for handling missing data in multivariate analysis. Psychometrika, 41, 409-415.

Fuller, W. A. 1976. Introduction to statistical time series. New York: Wiley.

Furnival, G. M. 1971. All possible regressions with less computations. Technometrics, 13, 403-408.

Gaensslen, H. & Schubö, W. 1976. Einfache und komplexe statistische Analyse. München: Ernst Reinhardt.

Garside, M. J. 1965. The best subset in multiple regression analysis. Applied Statistics Journal of the Royal Statistical Society, Series C, 14, 196-200.

Gartside, P. S. 1972. A study of methods for comparing several variances. Journal of the American Statistical Association, 67, 342-346.

Gigerenzer, G. 1981. Messung und Modellbildung in der Psychologie. München: Ernst Reinhardt.

Glass, G. V., Willson, V. L. & Gottman, J. M. 1975. Design and analysis of time-series experiments. Boulder, Colorado: Colorado Associated University Press.

Gnanadesikan, R. 1977. Methods for statistical data analysis of multivariate observations. New York: Wiley.

Gordon, R. A. 1968. Issues in multiple regression. American Journal of Sociology, 73, 592-616.

Graybill, F. A. 1976. Theory and application of the linear model. North Scituate: Duxbury.

Gudat, U. & Revenstorff, D. 1976. Interventionseffekte in klinischen Zeitreihen. Archiv für Psychologie, 128, 16-44.

Haagen, K. & Pertler, R. 1976. Methoden der Statistik. Band I. Stuttgart: Kohlhammer.

Haagen, K. & Seifert, H. G. 1979. Methoden der Statistik für Psychologen. Band II. Stuttgart: Kohlhammer.

Hanke, B., Lohmöller, J. B. & Mandl, H. 1980. Schülerbeurteilung in der Grundschule. München: Oldenbourg.

Hannan, E. J. 1955. Exact tests for serial correlation. Biometrika, 42, 133-142.

Hannan, E. J. 1960. Time series analysis. London: Methuen.

Hart, B. I. & von Neumann, J. 1942. Tabulation of the probabilities for the ratio of the mean square successive difference to the variance. Annals of Mathematical Statistics, 13, 207-214, 445-447.

Hays, W. L. 1963. Statistics for psychologists. New York: Holt, Rinehart & Winston.

Hersen, M. & Barlow, D. 1976. Single case experimental designs: strategies for studying behavior change in the individual. New York: Pergamon.

Hoadley, B. 1970. A bayesian look at inverse linear regression. Journal of the American Statistical Association, 65, 356-369.

Hocking, R. R. 1972. Criteria for selection of a subset regression: Which one should be used? Technometrics, 14, 967-970.

Hocking, R. R. & Leslie, R. N. 1967. Selection of the best subset in regression analysis. Technometrics, 9, 531-540.

Hoerl, A. E. & Kennard, R. W. 1970. Ridge regression: biased estimation of nonorthogonal Problems. Technometrics, 12, 55-67.

Holling, H. 1980. Theoretische und empirische Analysen zum Suppressorkonzept. Berlin: Phil. Dissertation.

Holtzmann, W. 1963. Statistical models for the study of change in the single case. In: Harris, C. W. (ed.). Problems in measuring change. Madison, Wisc.: University of Wisconsin Press.

Hope, K. 1975. Methoden multivariater Analyse. Weinheim: Beltz.

Horst, P. 1961a. Relations among m sets of measures. Psychometrika, 26, 129-150.

Horst, P. 1961b. Generalized canonical correlations and their applications to experimental data. Journal of Clinical Psychology (Monograph Supplement), 14, 331-347.

Horst, P. 1965. Factor analysis of data matrices. New York: Holt, Rinehart & Winston.

Hotelling, H. 1935. The most predictable criterion. Journal of Educational Psychology, 26, 139-142.

Hotelling, H. 1936. Relations between two sets of variates. Biometrika, 28, 321-377.

Hotelling, H. 1951. A generalized T-test and measure of multivariate dispersion. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 23-41.

Huber, H. P. 1967. Zeitreihenanalyse im diagnostischen Einzelfall. In: Merz, F. (Hrsg.). Bericht über den 25. Kongreß der Deutschen Gesellschaft für Psychologie. Münster 1966. Göttingen: Hogrefe.

Huber, H. P. 1973. Psychometrische Einzelfalldiagnostik. Weinheim: Beltz.

Hummell, H. J. & Ziegler, R. 1976. Korrelation und Kausalität. Stuttgart: Enke.

IMSL. 1979. The IMSL library. Fortran subroutines in the areas of mathematics and statistics. Houston: IMSL.

Jahn, J. 1979. Zur Bewertung von Schätzungen bei der gewöhnlichen Regressionsanalyse. Darmstadt: Fachbereich Mathematik der Technischen Hochschule.

Jenkins, G. M. & Watts, D. G. 1968. Spectral analysis and its applications. San Francisco: Holden Day.

Jöreskog, K. G. 1973. Analyzing psychological data by structural analysis of covariance matrices. In: Goldberger, A. S. & Duncan, O. D. (ed.). Structural analysis of covariance matrices. New York: Seminar Press, 85-112.

Jöreskog, K. G. 1977. Structural equation models in the social sciences: Specification, estimation and testing. In: Krishnaiah, P. R. (ed.). Applications of statistics. Amsterdam: North Holland.

Jöreskog, K. G. 1978. Structural analysis of covariance and correlation matrices. Psychometrika, 43, 443-477.

Johnston, J. 1963. Econometric methods. New York: McGraw Hill.

Keeser, W. 1981. Zeitreihenanalyse in der klinischen Psychologie. Ein empirischer Beitrag zur Box-Jenkins Methodologie. Berlin: Springer.

Kendall, M. G. 1976. Time-series. London: Griffin.

Kendall, M. G. & Stuart, A. 1976. The advanced theory of statistics. Vol. 2. London: Griffin.

Kendall, M. G. & Stuart, A. 1966. The advanced theory of statistics. Vol. 3. New York: Hafner.

Kerlinger, F. N. & Pedhazur, E. J. 1973. Multiple regression in behavioral research. New York: Holt, Rinehart & Winston.

Kettenring, J. R. 1971. Canonical analysis of several sets of variables. Biometrika, 58, 483-451.

Klein, L. 1953. A textbook of econometrics. Evanston: Row, Peterson & Co.

Kmenta, J. 1971. Elements of econometrics. New York: Macmillan.

Koerts, J. & Abrahamse, A. P. 1969. On the theory and application of the general linear model. Rotterdam: University Press.

Kruskal, J. B. 1971. Monotone regression: continuity and differentiability properties. Psychometrika, 36, 57-62.

Krutchkoff, R. G. 1967. Classical and inverse regression - methods of calibration. Technometrics, 9, 425-439.

Kshirsagar, A. M. 1972. Multivariate analysis. Statistics: Textbooks and monographs, Vol. 2. New York: Marcel Dekker.

Küchler, M. 1979. Multivariate Analyseverfahren. Stuttgart: Teubner.

LaMotte, L. R. & Hocking, R. R. 1970. Computational efficiency in the selection of regression variables. Technometrics, 12, 83-93.

Leeuw, J. de. 1977. Correctness of Kruskal's algorithms for monotone regression with ties. Psychometrika, 42, 141-144.

Lienert, G. A. 1967. Testaufbau und Testanalyse. Weinheim: Beltz.

Lösel, F. & Wüstendörfer, W. 1974. Zum Problem unvollständiger Datenmatrizen in der empirischen Sozialforschung. Kölner Zeitschrift für Soziologie, 26, 342-357.

Lohmüller, J. B. 1979. Parameterschätzung für Linearstrukturbeziehungsmodelle unter partiellen Kleinstquadratkriterien. Forschungsbericht 79.01, Fachbereich Pädagogik der Hochschule der Bundeswehr München, Neubiberg.

Lohmöller, J. B. & Oerter, R. (Hrsg.). 1979. Medien in der Erziehungsausbildung. München: Oldenbourg.

Lovell, M. C. & Prescott, E. 1970. Multiple regression with inequality constraints: pretesting bias, hypothesis testing and efficiency. Journal of the American Statistical Association, 65, 913-925.

Maddala, G. S. 1977. Econometrics. New York: McGraw-Hill.

Malinvaud, E. 1966. Statistical methods of econometrics. Amsterdam: North Holland.

Marquardt, D. 1963. An algorithm for least-squares estimation of nonlinear parameters. Journal of SIAM, 11, 431-441.

McCleary, R. & Hay, R. A. 1980. Applied time series analysis for the social sciences. London: Sage Publications.

McDonald, R. P. 1974. Testing pattern hypotheses for covariance matrices. Psychometrika, 39, 189-201.

McGee, V. E. & Carleton, W. T. 1970. Piecewise regression. Journal of the American Statistical Association, 65, 1109-1124.

McKeon, J. J. 1966. Canonical analysis: Some relations between canonical correlation, factor analysis, discriminant analysis, and scaling theory. Psychometric Monograph, 13.

Milliken, G. A. & Graybill, F. A. 1970. Extensions of the general linear hypothesis model. Journal of the American Statistical Association, 65, 797-807.

Moosbrugger, H. 1978. Multivariate statistische Analyseverfahren. Stuttgart: Kohlhammer.

Moosbrugger, H. 1980a. Lineare statistische Modelle zur Beschreibung linearer und nichtlinearer multipler Variablenzusammenhänge. Frankfurt/Main: Institut für Psychologie der Johann-Wolfgang-Goethe-Universität.

Moosbrugger, H. 1980b. Zur differentiellen Prognostizierbarkeit bei nichtlinearen Test-Kriterium-Zusammenhängen. Frankfurt/Main: Institut für Psychologie der Johann-Wolfgang-Goethe-Universität.

Morgan, J. A. & Tatar, J. F. 1972. Calculation of the residual sum of squares for all possible regressions. Technometrics, 14, 317-325.

Morrison, D. F. 1976. Multivariate statistical methods. New York: McGraw-Hill.

Namboodiri, N. K., Carter, L. F. & Blalock, H. M. 1975. Applied multivariate analysis and experimental design. New York: McGraw-Hill.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrennet-, K. & Bent, D. H. 1975. SPSS: Statistical package for the social sciences. New York: McGraw-Hill.

Oberhofer, W. 1979. Lineare Algebra. München: Oldenbourg.

Olkin, I. & Pratt, J. W. 1958. Unbiased estimation of certain correlation coefficients. Annals of Mathematical Statistics, 29, 201-211.

Overall, J. E. & Klett, C. J. 1972. Applied multivariate analysis. New York: McGraw-Hill.

Petermann, F. (Hrsg.). 1977a. Psychotherapieforschung. Weinheim: Beltz.

Petermann, F. (Hrsg.). I977b. Methodische Grundlagen Klinischer Psychologie. Weinheim: Beltz.

Petermann, F. 1978. Veränderungsmessung. Stuttgart: Kohlhammer.

Pillai, K. C. S. 1955. Some new test criteria in multivariate analysis. Annals of Mathematical Statistics, 26, 117-121.

Ramsey, J. B. 1969. Tests for specification errors in classical linear least-squares regression analysis. Journal of the Royal Statistical Society, Series B, 31, 350-371.

Rao, C. R. 1973. Lineare statistische Methoden und ihre Anwendungen. Berlin: Akademie-Verlag.

Revenstorff, D. 1979. Zeitreihenanalyse für klinische Daten. Weinheim: Beltz.

Revenstorff, D. & Keeser, W. 1979. Zeitreihenanalyse von Therapieverläufen. In: Petermann, F. & Hehl, F. (Hrsg.). Einzelfallanalyse. München: Urban & Schwarzenberg.

Rock, D. A., Werts, C. E., Linn, R. L. & Jöreskog, K. G. 1977. A maximum likelihood solution to the errors in variables and errors in equations model. Multivariate Behavioral Research, 12, 187-197.

Roy, S. N. 1957. Some aspects of multivariate analysis. New York: Wiley.

Schach, S. & Schäfer, Th. 1978. Regressions- und Varianzanalyse. Berlin: Springer.

Schneeweiß, H. 1974. Ökonometrie. Würzburg: Physica-Verlag.

Schönfeld, P. 1969. Methoden der Ökonometrie. Band I. Berlin: Franz Vahlen.

Schönfeld, P. 1971. Methoden der Ökonometrie. Band II. Berlin: Franz Vahlen.

Schubö, W. 1980. Statistische Verlaufsanalyse am Beispiel eines Inferenztests. In: Hentschel, U. & Smith, G. (Hrsg.). Die Wahrnehmung als Zugang zu diagnostischen Problemen. Wiesbaden: Akademische Verlagsgesellschaft.

Schubö, W. 1982. Faktorenanalyse in der psychiatrischen Forschung. München: Phil. Dissertation.

Schubö, W. & Hentschel, U. 1978. Improved reliability estimates for the serial color-word test. Scandinavian Journal of Psychology, 19, 91-95.

Schuchard-Ficher, C., Backhaus, K., Humme, U., Lohrberg, W., Plinke, W. & Schreiner, W. 1980. Multivariate Analysemethoden; eine anwendungsorientierte Einführung. Berlin: Springer.

Seibel, H. D. & Nygreen, G. T. 1972. Pfadanalyse. Ein statistisches Verfahren zur Untersuchung linearer Kausalmodelle. Zeitschrift für Sozialpsychologie, 3, 5-12.

Seifert, H.-G. 1975. Statistische Methoden der Ökonometrie. Unveröffentlichtes Vorlesungsmanuskript. Regensburg: Lehrstuhl für Ökonometrie.

Sievers, W. 1977. über Dummy-Variablen-Kodierung in der Varianzanalyse. Psychologische Beiträge, 19, 3, 459-462.

Silvey, S. D. 1969. Multicollinearity and imprecise estimation. Journal of the Royal Statistical Society, Series B, 31, 539-552.

Steel, R. G. D. 1951. Minimum generalized variance for a set of linear functions. Annals of Mathematical Statistics, 22, 456-460.

Steward, D. & Love, W. 1968. A general canonical correlation index. Psychological Bulletin, 70, 160-163.

Steyer, R. 1979. Rechenregeln für bedingte Erwartungen. Frankfurt/Main: Institut für Psychologie der Johann-Wolfgang-Goethe-Universität.

Stiefel, E. 1961. Einführung in die numerische Mathematik. Stuttgart: Teubner.

Stroud, T. W. F. 1974. Comparing regressions when measurement error variances are known. Psychometrika, 39, 53-68.

Sulz, K.-D. 1980. Dimensionale Analyse kognitiver Konzepte. München: Phil. Dissertation.

Tatsuoka, M. M. 1971. Multivariate analysis: Techniques for educational and psychological research. New York: Wiley.

Theil, H. 1971. Principles of econometrics. New York: Wiley.

Theobald, C. M. 1974. Generalisation of mean square error applied to ridge regression. Journal of the Royal Statistical Society, Series B, 36, 103-106.

Thorndike, R. M. 1978. Correlational procedures for research. New York: Gardner Press.

Timm, N. H. 1970. The estimation of variance-covariance and correlation matrices from incomplete data. Psychometrika, 35, 417-437.

Timm, N. H. 1975. Multivariate analysis with applications in education and psycholo-. gy. Belmont, California: Wadsworth.

Tintner, G. 1946. Some applications of multivariate analysis in economic data. Journal of the American Statistical Association, 41, 472.

van de Geer, J. B. 1971. Introduction to multivariate analysis for the social sciences. San Francisco: Freeman.

Velicer, W. F. 1978. Suppressor variables and the semipartial correlation coefficient. Educational and Psychological Measurement, 38, 953-958.

Vetter, H. 1972. Dynamische und statistische Kausalanalyse. Zeitschrift für Sozialpsychologie, 3, 13-22.

Vinod, H. D. 1978. Ridge regression and related techniques: A survey. The Review of Economics and Statistics, 60, 121-131.

Vinograde, B. 1950. Canonical positive definite matrices under internal linear transformations. Proceedings of the American Mathematical Society, 1, 159-161.

Wainer, H. 1971. Piecewise regression: a simplified procedure. The British Journal of Mathematical and Statistical Psychology, 24, 83-92.

Wainer, H. & Thissen, D. 1976. Three steps towards robust regression. Psychometrika, 41, 9-34.

Wilks, S. S. 1932. Certain generalizations in the analysis of variance. Biometrika, 24, 471-494.

Wilson, K. V. 1973. Linear regression equations as behavior models. In: Royce, J. R. (ed.). Multivariate analysis and psychological theory. London: Academic Press.

Wonnacott, R. J. & Wonnacott, Th. H. 1970. Econometrics. New York: Wiley.

Young, F. W., de Leeuw, J. & Takane, Y. 1976. Regression with qualitative and quantitative variables: An alternating least squares method with optimal scaling features. Psychometrika, 41, 505-529.

Yule, G. U. 1927. On a method of investigating periodicities in disturbed series. With special reference to Wölfers' sunspot numbers. Philosophical Transactions of the Royal Society, Series A, 226, 267-298.