

3. Kapitel

Uni- und multivariate Varianzanalyse mit festen Parametern

Helfried Moosbrugger und Rolf Steyer

1. Einführung und Überblick

Eine der am häufigsten auftretenden Fragen in der bio- und sozialwissenschaftlichen Forschung ist die, ob Mittelwerteunterschiede zwischen zwei oder mehr als zwei Gruppen bezüglich eines bestimmten „gemessenen“ Merkmals bestehen. Diese Frage kann durch einfaches Vergleichen der Mittelwerte in den Gruppen nicht zufriedenstellend entschieden werden, wenn nur eine *Stichprobe* von Gruppenmitgliedern untersucht wurde, aber dennoch eine Aussage für die *Population*, d.h. für *alle* Gruppenmitglieder getroffen werden soll.

Die erste exakte Lösung für den Fall zweier Gruppen wurde 1908 von Gosset unter dem Pseudonym „Student“ (vgl. z.B. Pearson & Wishart, 1943) veröffentlicht und ist unter der Bezeichnung *t-Test* auch heute noch eines der am häufigsten verwendeten inferenzstatistischen Verfahren.

Für den Fall mehrerer Gruppen hat Fisher (1925, 1935) die „grundlegenden Prinzipien der Varianzanalyse und ihre wichtigsten Techniken ausgearbeitet und publiziert“ (Stanley, 1978, S. 541, Übersetzung durch die Autoren). Schließlich wurde von Roy (z.B. 1939, 1946, 1957) die *multivariate Varianzanalyse* entwickelt (vgl. Stanley, 1978, S. 553), mit der man Mittelwerteunterschiede mehrerer Gruppen hinsichtlich *mehrerer* „gemessener“ Merkmale untersuchen kann.

Seit Cohens (1968) Artikel über „Multiple Regression als ein allgemeines System zur Datenanalyse“ (Übersetzung durch die Autoren) ist in den Fachzeitschriften der Psychologie eine Vielzahl von Arbeiten erschienen, die auf die allgemeine Theorie linearer Modelle eingehen, in welche die Varianzanalyse, wie z.B. auch die Regressionsanalyse, eingebettet werden kann (siehe z.B. Moosbrugger, 1978, oder Steyer, 1979, und die dort angegebene Literatur).

Die Theorie linearer Modelle war aber bereits vorher in der mathematisch-statistischen Literatur bekannt und weit entwickelt (siehe z.B. Anderson, 1958, Graybill, 1961, oder Scheffé, 1959).

Varianz- und Regressionsanalysen als Spezialfälle ein- und derselben Theorie zu behandeln wird dadurch möglich, daß sich Mittelwerteunterschiede einer Variablen in z.B. zwei Gruppen auch als Zusammenhang der betreffenden Variablen mit einer zweistufigen Kodiervariablen, welche die Gruppenzugehörigkeit anzeigt, darstellen lassen. Zusammenhänge zwischen zwei Variablen aber werden gewöhnlich mit Methoden der Regressions- und Korrelationsanalyse untersucht.

In dem vorliegenden Beitrag stellen wir die u.E. wichtigsten Begriffe der Theorie multivariater linearer Modelle mit festen Parametern dar und zeigen exemplarisch, wie auf dieser Basis verschiedene varianzanalytische Fragestellungen zu bearbeiten sind. Gegenüber der herkömmlichen Darstellung hat dies vor allem didaktische und konzeptuelle Vorteile; didaktische, weil nur wenige Grundbegriffe vermittelt werden müssen, und konzeptuelle, weil der Ansatz sehr allgemein ist.

Im zweiten Abschnitt stellen wir die *Grundbegriffe* der hier behandelten Theorie dar. Im dritten Abschnitt führen wir die *multivariate allgemeine lineare Hypothese* ein und erläutern deren Anwendung an einer Reihe verschiedener *Designs*. Dabei legen wir das „Zellenmittelwertemodell“ zugrunde, bei dem die Parametermatrix eines linearen Modells die Zellenmittelwerte der abhängigen Variablen als Komponenten enthält.

Im vierten Abschnitt behandeln wir dann Fragen der *Parameterschätzung* und im fünften schließlich solche der *Hypothesenbewertung*, wobei wir beide Themenkomplexe für den allgemeinen Fall abhandeln, bei dem die Matrix der unabhängigen Variablen nicht vollen Spaltenrang haben muß.

In allen Abschnitten setzen wir einige Grundkenntnisse der Matrixalgebra voraus, in welche z.B. Moosbrugger (1978), Searle (1966) oder Van de Geer (1971) einführen.

2. Multivariate lineare Modelle mit festen Parametern

2.1 Einleitung

In diesem zweiten Abschnitt wird diejenige Klasse von linearen statistischen Modellen formal charakterisiert, mit denen wir uns im weiteren Verlauf dieses Kapitels beschäftigen. Dabei werden wir zunächst in die grundlegenden Ideen

einführen und dann die Verbindung dieser Ideen mit Beobachtungseinheiten (z.B. Personen) und Stichproben herstellen.

Von vielen Autoren wird die hier dargestellte Theorie unter der Bezeichnung „Allgemeines (multivariates) lineares Modell“ abgehandelt (siehe z.B. Finn, 1974, Graybill, 1976, Horton, 1978, Mendenhall, 1968, Moosbrugger, 1978, Steyer, 1979), eine Bezeichnung, die Verwirrung stiften könnte, da sie die Annahme nahelegt, es handle sich hier um die allgemeinste Form linearer Modelle. Dies ist aber nicht der Fall, was man schon daran erkennen kann, daß etwa die „random“ und „mixed effects“ Modelle (vgl. z.B. Searle, 1971, S. 376ff.) nicht unter diese Kategorie fallen, genausowenig wie lineare Modelle mit stochastischen Prädiktoren (vgl. z.B. Graybill, 1976, Kap. 10).

2.2 Die grundlegenden Modellvorstellungen

In diesem Beitrag betrachten wir ausschließlich Modelle, bei denen für die stochastischen Vektoren $y = (y_1, y_2, \dots, y_p)$ von „abhängigen“ Variablen und $x = (x_1, x_2, \dots, x_Q)$ von „unabhängigen“ Variablen die Gleichung

$$(2.2.1) \quad E(y | x) = xB$$

(B ist ein großes griechisches Beta) gilt, welche die P einzelnen Gleichungen

$$(2.2.2) \quad E(y_p | x_1, x_2, \dots, x_Q) = x\beta_p = \beta_{p1}x_1 + \beta_{p2}x_2 + \dots + \beta_{pQ}x_Q$$

enthält. Demnach ist die bedingte Erwartung¹⁾ einer abhängigen Variablen y_p , $p \in \{1, 2, \dots, P\}$, unter x_q , $q \in \{1, 2, \dots, Q\}$, eine mit den festen Parametern β_{pq} gewichtete Summe der unabhängigen Variablen x_q . Die Koeffizienten β_{pq} , die auch partielle Regressionskoeffizienten genannt werden (vgl. z.B. Gaensslen & Schubö, 1973, S. 92) geben an, wie stark erwartungsgemäß zwei Werte von y_p differieren, wenn sich die zugehörigen Werte von x_q um den Betrag Eins unterscheiden, falls die Werte aller anderen unabhängigen Variablen x_{q^*} , $q^* \neq q$, gleich bleiben.

In der Varianzanalyse handelt es sich bei jedem y_p , $p \in \{1, 2, \dots, P\}$, um eine quantitative Variable und bei den x_q , $q \in \{1, 2, \dots, Q\}$, um qualitative Variablen, welche Gruppenzugehörigkeiten durch Zahlen wie z.B. 0 oder 1 anzeigen.

¹⁾ Man beachte, daß es sich bei der bedingten Erwartung um eine stochastische Variable handelt, im Unterschied zum bedingten Erwartungswert. Zur Definition dieser Begriffe der Wahrscheinlichkeitstheorie verweisen wir auf Bauer (1974) oder Müller (1975). Gleichungen für bedingte Erwartungen gelten nur mit Wahrscheinlichkeit 1, worauf wir im folgenden nicht weiter hinweisen.

Wir erläutern nun Gleichung 2.2.1 für einige bekannte statistische Modelle. Das einfachste aller denkbaren Modelle (mit $P=1$ und $Q=1$) ist

$$(2.2.3) \quad E(y|x) = \beta \cdot 1 = \beta,$$

welches auch in der Form

$$(2.2.4) \quad E(y) = \mu$$

geschrieben werden kann. Damit wird die Annahme formuliert, daß eine stochastische Variable y einen Erwartungswert hat, nämlich μ bzw. β . Die stochastische Variable x ist hier zu der Konstanten Eins „degeneriert“, weswegen in diesem Fall auch $E(y|x) = E(y)$ gilt. In einer Stichprobensituation ist hiermit oft die bekannte Fragestellung verbunden, ob ein Stichprobenmittelwert „signifikant“ von μ abweicht (vgl. z.B. Bortz, 1977, S. 156-160).

Ein ebenso bekanntes Modell mit $P=1$ und $Q=2$ liegt vor, wenn der stochastische Vektor x in Gleichung 2.2.5 nur die beiden Ausprägungen

$$x_1 = [1 \ 0] \text{ und } x_2 = [0 \ 1]$$

annehmen kann:

$$(2.2.5) \quad E(y|x) = x\beta = \beta_1 x_1 + \beta_2 x_2.$$

Es ist dann mit den beiden Gleichungen

$$(2.2.6) \quad E(y|x_{11}=1, x_{21}=0) = \beta_1 \cdot 1 + \beta_2 \cdot 0 = \beta_1$$

und

$$(2.2.7) \quad E(y|x_{12}=0, x_{22}=1) = \beta_1 \cdot 0 + \beta_2 \cdot 1 = \beta_2$$

gleichbedeutend.

Dieses Modell kann man auch mit den Parametern μ_1 statt β_1 und μ_2 statt β_2 schreiben, wobei μ_1 als Erwartungswert der abhängigen Variablen y in einer ersten und μ_2 als deren Erwartungswert in einer zweiten Gruppe zu interpretieren ist. In einer Stichprobensituation ist mit diesem Modell oft die Fragestellung verbunden, ob zwei Stichprobenmittelwerte „signifikant“ voneinander abweichen (vgl. z.B. Bortz, 1977, S. 160-164). Für $Q=3$ lautet die Modellgleichung

$$(2.2.8) \quad E(y|x) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

Im Fall dreier Gruppen hat der stochastische Vektor x nur die drei Ausprägungen

$$\begin{aligned} \mathbf{x}_1 &= [1 \ 0 \ 0], \\ \mathbf{x}_2 &= [0 \ 1 \ 0] \\ \text{und } \mathbf{x}_3 &= [0 \ 0 \ 1] \end{aligned}$$

Gleichung (2.2.7) ist dann mit den drei Gleichungen

$$E(y \mid x_{11}=1, x_{21}=0, x_{31}=0) = \beta_1 \cdot 1 + \beta_2 \cdot 0 + \beta_3 \cdot 0 = \beta_1$$

$$E(y \mid x_{12}=0, x_{22}=1, x_{32}=0) = \beta_1 \cdot 0 + \beta_2 \cdot 1 + \beta_3 \cdot 0 = \beta_2$$

$$E(y \mid x_{13}=0, x_{23}=0, x_{33}=1) = \beta_1 \cdot 0 + \beta_2 \cdot 0 + \beta_3 \cdot 1 = \beta_3$$

oder

$$(2.2.9) \quad E(y \mid \mathbf{x}=\mathbf{x}_q) = \beta_q, \quad q \in \{1,2,3\},$$

gleichbedeutend.

Eine verbreitetere Schreibweise dafür ist jedoch

$$(2.2.10) \quad E(y \mid \mathbf{x}=\mathbf{x}_q) = \mu_q = \mu + (\mu_q - \mu) = \mu + \alpha_q, \quad q \in \{1,2,3\},$$

wobei $\alpha_q := \mu_q - \mu$. Mit diesem Modell ist in Stichprobensituationen oft die *varianzanalytische Fragestellung* verbunden, ob zwischen den $Q=3$ Gruppen „signifikante“ Mittelwerteunterschiede bestehen.

Nachdem wir mit diesen einfachen Beispielen die Gleichung 2.2.1 erläutert haben, kehren wir zur Darstellung der allgemeinen Theorie zurück.

Aus der Gleichung 2.2.1 lassen sich eine Reihe von Eigenschaften eines linearen Modells ableiten. Betrachten wir z.B. den *Residualvektor* vom Typ $1 \times P$

$$(2.2.11) \quad \boldsymbol{\varepsilon} := \mathbf{y} - E(\mathbf{y} \mid \mathbf{x})$$

($\boldsymbol{\varepsilon}$ ist ein kleines griechisches Epsilon), so gelten dafür nach den Regeln 4, 5 und 10 des Anhangs C von Moosbrugger (1982), in diesem Band,

$$(2.2.12) \quad E(\boldsymbol{\varepsilon} \mid \mathbf{x}) = \mathbf{0},$$

nach Regel 1 jenes Anhangs

$$(2.2.13) \quad E(\boldsymbol{\varepsilon}) = \mathbf{0}$$

und nach Regel 11 jenes Anhangs

$$(2.2.14) \quad E(\mathbf{x}'\boldsymbol{\varepsilon}) = C(\mathbf{x}, \boldsymbol{\varepsilon}) = \mathbf{0}$$

wobei \mathbf{x}' ein Spaltenvektor vom Typ $Q \times 1$ ist.

Nach Gleichung 2.2.12 ist die bedingte Erwartung von $\boldsymbol{\varepsilon}$ unter $\mathbf{x} = (x_1, x_2, \dots, x_Q)$ gleich Null, wie auch der unbedingte Erwartungswert von $\boldsymbol{\varepsilon}$ (Gleichung 2.2.13). Wenn die bedingte Erwartung von $\boldsymbol{\varepsilon}$ gegeben x_1, x_2, \dots, x_Q und damit auch die bedingten Erwartungswerte für alle Ausprägungskombinationen der x_q gleich Null sind, dann kann auch keine Residualvariable ε_p mit einem x_q kovariieren (Gleichung 2.2.14) oder korrelieren.

Aus der Gleichung 2.2.1 folgt, wenn $E(\mathbf{x}'\mathbf{x})$ nonsingulär ist, die *Identifikationsgleichung* für die Parametermatrix \mathbf{B} (vgl. Moosbrugger, 1982, in diesem Band, Kap. 2.2.3).

$$(2.2.15) \quad \mathbf{B} = E(\mathbf{x}'\mathbf{x})^{-1} E(\mathbf{x}'\mathbf{y}).$$

Von praktischem Interesse sind neben den Parametern in der Matrix \mathbf{B} , die u.a. zur *Prädiktion* dienen können, auch die durch die x_1, x_2, \dots, x_Q , also durch \mathbf{x} bestimmte Varianz $V[E(y_p | \mathbf{x})]$ einer „abhängigen“ Variablen y_p , für die nach Gleichung 2.2.2

$$(2.2.16) \quad V[E(y_p | \mathbf{x})] = C(\mathbf{x}\boldsymbol{\beta}_p, \mathbf{x}\boldsymbol{\beta}_p) =$$

$$(2.2.17) \quad = \boldsymbol{\beta}'_p C(\mathbf{x}, \mathbf{x}) \boldsymbol{\beta}_p =$$

$$(2.2.18) \quad = V(y_p) - V(\varepsilon_p)$$

gilt (siehe Gleichung B.5 und Regel 2 des Anhangs B von Moosbrugger, 1982, in diesem Band). Sie ist ein Kennwert für die *Güte der Prädiktion* oder für die *Stärke des Zusammenhangs* zwischen y_p und \mathbf{x} , ebenso wie der durch \mathbf{x} bestimmte Varianzanteil von y_p , der *multiple Determinationskoeffizient*

$$(2.2.19) \quad R^2_{y_p|\mathbf{x}} := \frac{V[E(y_p | \mathbf{x})]}{V(y_p)} = \frac{V(y_p) - V(\varepsilon_p)}{V(y_p)}.$$

Dieser ist von großem praktischen Interesse, weil er ein *normiertes Maß* für die *praktische Bedeutsamkeit* (vgl. z.B. Bredenkamp, 1970) eines statistischen Zusammenhangs zwischen y_p und \mathbf{x} ist, insofern, als $R^2_{y_p|\mathbf{x}}$ 100% den durch \mathbf{x} bestimmten Prozentanteil der Varianz von y_p angibt.

Die positive Wurzel von $R^2_{y_p|\mathbf{x}}$ heißt *multipler Korrelationskoeffizient*, der ebenso wie der multiple Determinationskoeffizient Werte zwischen Null und Eins annehmen kann. Der Wert Eins steht dabei für einen perfekten Zusammenhang, der eine fehlerfreie Vorhersage von y_p aufgrund der Werte der x_q ermöglicht, und $R^2_{y_p|\mathbf{x}}$ gleich Null besagt, daß die Kenntnis der Werte der x_q keine bessere Vorhersage von y_p ermöglicht als deren Erwartungswert $E(y_p)$. Dies ist immer dann der Fall, wenn alle Parameter β_{pq} , $q \in \{1, 2, \dots, Q\}$, gleich Null sind, und kein Unterschied zwischen der bedingten Erwartung $E(y_p | \mathbf{x})$ und dem (unbedingten) Erwartungswert $E(y_p)$ besteht.

Skalare Kennwerte der praktischen Bedeutsamkeit für den Fall *mehrerer* abhängiger Variablen ($P \geq 2$) sind weniger eindeutig anzugeben, da zunächst nicht nur ein einziger Kennwert für die Streubreite einer abhängigen Variablen y_p , nämlich die Varianz $V(y_p)$, vorliegt, sondern eine symmetrische $P \times P$ -Kovarianzmatrix der Komponenten von y . Analog zum univariaten Fall kann man zunächst jedoch *die durch x bestimmte Kovarianzmatrix von y*

$$(2.2.20) \quad C[E(y|x), E(y|x)] = C[xB, xB] = B' C(x, x) B =$$

$$(2.2.21) \quad = C(y, y) - C(\epsilon, \epsilon)$$

angeben (siehe Gleichung B.5 und Regel 2 des Anhangs B von Moosbrugger, 1982, in diesem Band). Ebenso wie im univariaten Fall können wir diese Kovarianzmatrizen auf die Kovarianzmatrix $C(y, y)$ der abhängigen Variablen normieren, indem wir die Gleichung 2.2.21 mit $C(y, y)^{-1}$ nachmultiplizieren, wodurch wir bei Nonsingularität von $C(y, y)$ die zu 2.2.19 analoge Gleichung

$$(2.2.22) \quad C[E(y|x), E(y|x)] C(y, y)^{-1} = [C(y, y) - C(\epsilon, \epsilon)] C(y, y)^{-1}$$

erhalten. Ausmultiplizieren und Umstellen ergibt

$$(2.2.23) \quad C[E(y|x), E(y|x)] C(y, y)^{-1} + C(\epsilon, \epsilon) C(y, y)^{-1} = I,$$

woraus man ersieht, daß sich der durch x bestimmte Anteil der Kovarianzmatrix (der linke Summand) und deren unbestimmter Anteil (der rechte Summand) zur Einheitsmatrix I aufaddieren. Ausgehend von diesen Matrizen wurden eine Reihe von multivariaten Verallgemeinerungen des Determinationskoeffizienten vorgeschlagen, die z.B. bei Shaffer und Gillo (1974) diskutiert werden. Am plausibelsten erscheint uns jedoch der Vorschlag von Cramer und Nicewander (1979), die den Schätzer für den Kennwert

$$(2.2.24) \quad R_{y|x}^2 := \frac{\text{Spur}[C[E(y|x), E(y|x)] C(y, y)^{-1}]}{P}$$

als *multivariaten Determinationskoeffizienten* favorisieren. Dabei ist

$$(2.2.25) \quad P = \text{Spur}[C(y, y) C(y, y)^{-1}] = \text{Spur } I$$

die Anzahl der Komponenten von y und zugleich der Rang von $C(y, y)$. Dieser Kennwert ist invariant gegenüber linearen Skalentransformationen von y , liegt zwischen Null und Eins, reduziert sich im univariaten Fall auf den Determinationskoeffizienten $R_{y|x}^2$ (siehe Gleichung 2.2.19) und addiert sich mit

$$\frac{\text{Spur}[C(\epsilon, \epsilon) C(y, y)^{-1}]}{P}$$

zu Eins auf. Außerdem besteht ein sehr enger Zusammenhang dieses multivariaten Determinationskoeffizienten zum Pillai-Bartlett Spur-Kriterium, wel-

ches wir neben anderen multivariaten Testkriterien im Abschnitt 5 behandeln werden.

2.3 Stichprobenmodelle

In Abschnitt 2.2 haben wir die lineare Modellgleichung

$$(2.3.1) \quad E(\mathbf{y} | \mathbf{x}) = \mathbf{x}\mathbf{B}$$

formuliert. Es stellt sich nun die Frage, wie man die in der Regel unbekanntes Parameter in der Matrix \mathbf{B} schätzen kann. In der psychologischen Forschung geben die Werte der stochastischen Variablen in \mathbf{y} und \mathbf{x} Merkmale von Beobachtungseinheiten, z.B. Versuchspersonen, wieder. Bei den abhängigen Variablen handelt es sich um Messungen einer ($P=1$) oder mehrerer ($P>1$) Eigenschaften einer Beobachtungseinheit, und bei den unabhängigen Variablen z.B. um Indikatoren der Zugehörigkeit zu einer Gruppe (z. B. Experimental- oder Kontrollgruppe).

Um ein Verfahren zur Schätzung der Parametermatrix \mathbf{B} entwickeln zu können, müssen wir unsere Vorstellungen über die uns vorliegenden N Beobachtungseinheiten präzisieren. Die Beobachtungseinheiten können entweder verschiedene V_{pn} sein oder Person-Zeit-Kombinationen, nämlich Personen zu verschiedenen Meßzeitpunkten. Zunächst legen wir fest, daß die Gleichung 2.3.1 für jede der N Beobachtungseinheiten gelten soll:

$$(2.3.2) \quad E(\mathbf{y}_n | \mathbf{x}_n) = \mathbf{x}_n \mathbf{B}, \quad \text{für alle } n \in \{1, 2, \dots, N\}.$$

Für varianzanalytische Fragestellungen liegt typischerweise mit jeder Beobachtungseinheit auch deren Gruppenzugehörigkeit fest und somit auch die Realisierung \mathbf{x}_n des Vektors der unabhängigen Variablen \mathbf{x}_n .²⁾ Bei den abhängigen Variablen in \mathbf{y}_n hingegen liegen in der Varianzanalyse bei gegebener Beobachtungseinheit n die Realisierungen \mathbf{y}_n des stochastischen Vektors \mathbf{y}_n noch nicht fest, d.h. auch bei gegebener Beobachtungseinheit ließe sich prinzipiell eine Variation der betreffenden Variablen feststellen.

Da bei gegebener Beobachtungseinheit n der stochastische Vektor \mathbf{x}_n mit seiner Realisierung \mathbf{x}_n identisch ist, gilt außer $E(\mathbf{y}_n | \mathbf{x}_n) = \mathbf{x}_n \mathbf{B}$ auch

$$(2.3.3) \quad E(\mathbf{y}_n | \mathbf{x}_n) = E(\mathbf{y}_n | \mathbf{x}_n = \mathbf{x}_n) = E(\mathbf{y}_n)$$

und

$$(2.3.4) \quad E(\mathbf{y}_n) = E[E(\mathbf{y}_n | \mathbf{x}_n)] = E(\mathbf{x}_n \mathbf{B}) = \mathbf{x}_n \mathbf{B}, \quad n \in \{1, 2, \dots, N\}$$

²⁾ Diese Eigenschaft unterscheidet die hier behandelten Modelle von solchen mit „stochastischen Regressoren“, über welche z.B. Graybill, 1976, S. 373ff., informiert.

(siehe Regel 1 des Anhangs C und Regel 1 des Anhangs A von Moosbrugger, 1982, in diesem Band).

Aus Gleichung 2.3.3 und 2.3.4 ist ersichtlich, daß das Stichprobenmodell einfach wie folgt geschrieben werden kann:

$$(2.3.5) \quad E(\mathbf{y}_n) = \mathbf{x}_n \mathbf{B}, \quad \text{für alle } n \in \{1, 2, \dots, N\}.$$

Es ist nützlich diese N-Gleichungen mit einer Matrixgleichung zu notieren, wobei die folgenden Matrizen verwendet werden:

Y bestehend aus $N \times P$ stochastischen Variablen,

$$(2.3.6) \quad \mathbf{Y} = \begin{bmatrix} \mathcal{Y}_{11}, \dots, \mathcal{Y}_{1p}, \dots, \mathcal{Y}_{1P} \\ \vdots & & \vdots & & \vdots \\ \mathcal{Y}_{n1}, \dots, \mathcal{Y}_{np}, \dots, \mathcal{Y}_{nP} \\ \vdots & & \vdots & & \vdots \\ \mathcal{Y}_{N1}, \dots, \mathcal{Y}_{Np}, \dots, \mathcal{Y}_{NP} \end{bmatrix},$$

X bestehend aus $N \times Q$ festen reellen Zahlen

$$(2.3.7) \quad \mathbf{X} = \begin{bmatrix} \mathbf{x}_{11}, \dots, \mathbf{x}_{1q}, \dots, \mathbf{x}_{1Q} \\ \vdots & & \vdots & & \vdots \\ \mathbf{x}_{n1}, \dots, \mathbf{x}_{nq}, \dots, \mathbf{x}_{nQ} \\ \vdots & & \vdots & & \vdots \\ \mathbf{x}_{N1}, \dots, \mathbf{x}_{Nq}, \dots, \mathbf{x}_{NQ} \end{bmatrix}$$

und

B bestehend aus $Q \times P$ festen reellen Zahlen

$$(2.3.8) \quad \mathbf{B} = \begin{bmatrix} \beta_{11}, \dots, \beta_{1p}, \dots, \beta_{1P} \\ \vdots & & \vdots & & \vdots \\ \beta_{q1}, \dots, \beta_{qp}, \dots, \beta_{qP} \\ \vdots & & \vdots & & \vdots \\ \beta_{Q1}, \dots, \beta_{QP}, \dots, \beta_{QP} \end{bmatrix}.$$

Für alle N Beobachtungseinheiten zusammen lautet das Stichprobenmodell dann in Matrixschreibweise

$$(2.3.9) \quad \mathbf{E}(\mathbf{Y}) = \mathbf{X} \mathbf{B}.$$

Die Annahme der Gleichheit der Parametermatrix \mathbf{B} für alle Beobachtungseinheiten in Gleichung 2.3.9 wird noch durch eine Annahme über die Kovarianzmatrix der Residualvektoren

$$(2.3.10) \quad \boldsymbol{\varepsilon}_n := \mathbf{y}_n - \mathbf{x}_n \mathbf{B}$$

ergänzt, nämlich durch

$$(2.3.11) \quad C(\boldsymbol{\varepsilon}_n, \boldsymbol{\varepsilon}_{n^*}) = \begin{cases} \boldsymbol{\Sigma}, & \text{für } n = n^* \\ \mathbf{O}, & \text{für } n \neq n^* \end{cases} \quad \text{für alle } n, n^* \in \{1, 2, \dots, N\}.$$

Dies bedeutet, daß die $P \times P$ -Kovarianzmatrix $\boldsymbol{\Sigma}$ der Residuen der abhängigen Variablen für jede der N Beobachtungseinheiten gleich ist und daß die $P \times P$ Kovarianzmatrizen der Residuen der abhängigen Variablen verschiedener Beobachtungseinheiten als Nullmatrizen angenommen werden.³⁾

Ob es zweckmäßig ist, mit dem Index n eine Person oder eine Person-Zeit-Kombination zu bezeichnen, entscheiden wir in einem Anwendungsfall danach, ob dann Gleichung 2.3.11 angenommen werden kann oder nicht. Man kann annehmen, daß die Kovarianzmatrix $C(\boldsymbol{\varepsilon}_n, \boldsymbol{\varepsilon}_{n^*})$ für zwei verschiedene Beobachtungseinheiten $n \neq n^*$ z.B. dann gleich der Nullmatrix ist, wenn \mathbf{y}_n und \mathbf{y}_{n^*} , die abhängigen Variablen von zwei verschiedenen Personen enthält und eine gegenseitige Beeinflussung, wie z.B. Abschreiben bei einem Test, ausgeschlossen ist. Wenn \mathbf{y}_n und \mathbf{y}_{n^*} Beobachtungen der gleichen Person zu verschiedenen Zeitpunkten enthält, kann diese Annahme u. U. auch als erfüllt angesehen werden. Bei den herkömmlichen Designs mit wiederholten Messungen (vgl. z.B. Hays, 1973, S. 568ff.) beispielsweise, macht man diese Annahme.

Sind wir bereit, die Annahme der Gleichung 2.3.11 zu akzeptieren, wenn n eine Person-Zeit-Kombination indiziert, dann schreiben wir die Zeilenvektoren der abhängigen Variablen für die verschiedenen Zeitpunkte der Messungen einer Person untereinander, d.h. der Index n weist dann nicht auf eine Person, sondern auf die Person zum Zeitpunkt t hin.

Können wir aus theoretischen Überlegungen oder auf Grund eines statistischen Tests die Gleichung 2.3.11 für Person-Zeit-Kombinationen n nicht akzeptieren, so fassen wir die an den Personen wiederholten Messungen als Realisationen von (formal gesehen) weiteren abhängigen Variablen auf. Hätten

³⁾ Im univariaten Fall ($P=1$) reduziert sich die Matrix $\boldsymbol{\Sigma}$ zu dem Skalar $\sigma^2 = \sigma_{11}$.

wir zunächst 2 abhängige Variablen dreimal gemessen, so hätten wir nach der Umstrukturierung $2 \times 3 = 6$ abhängige Variablen vorliegen. Gemäß Gleichung 2.3.11 dürfen die Residuen dieser 6 abhängigen Variablen miteinander korrelieren, wobei Σ deren Kovarianzmatrix ist.

Die Gleichungen 2.3.9 bis 2.3.11 werden in der Literatur häufig als „*multivariate Gauss-Markoff setup*“ und Gleichung 2.3.9 als „*multivariates allgemeines lineares Modell*“ bezeichnet. In diesem Artikel werden wir uns ausschließlich mit Modellen befassen, welche diese Gleichungen erfüllen. Für einige andere Modelle geben wir Literaturhinweise.

2.4 Zusammenfassende Bemerkungen

In diesem zweiten Abschnitt haben wir die grundlegenden Vorstellungen der hier behandelten Klasse von linearen statistischen Modellen dargestellt. Dabei handelt es sich um die lineare Modellgleichung, die Identifikationsgleichung, die Begriffe bestimmte Varianz und Determinationskoeffizient sowie um deren Verallgemeinerung im multivariaten Fall, bei dem mehrere abhängige Variablen vorliegen.

Auf der Stichprobenebene sind die wesentlichen Annahmen, daß mit einer Beobachtungseinheit auch die Realisation x des stochastischen Vektors x der unabhängigen Variablen festliegt („nicht-stochastische Regressoren“), daß die Parametermatrix B nicht-stochastische Koeffizienten enthält („fixed-effects“), die für alle Beobachtungseinheiten gleichermaßen gelten, daß die Kovarianzmatrix der Residuen der P abhängigen Variablen bei gegebener Beobachtungseinheit n für alle N Beobachtungseinheiten gleich ist (siehe Gleichung 2.3.11), und daß diese Residuen für unterschiedliche Beobachtungseinheiten unkorreliert sind.

Weiterführende Literatur: Die Literatur über lineare Modelle oder das „Allgemeine lineare Modell“ ist im letzten Jahrzehnt fast unübersehbar geworden. Zunächst seien ohne Anspruch auf Vollständigkeit einige Monographien genannt, die zur Einführung in die hier behandelte stochastische Theorie dienen können: Bock (1975), Cohen und Cohen (1975), Edwards (1976), Ezekiel und Fox (1959), Finn (1974), Fraser (1979), Gaensslen und Schubö (1976), Graybill (1976), Horton (1978), Krafft (1978), Lindeman et al. (1980), Mendenhall (1968), Moosbrugger (1978), Namboodiri et al. (1975), Overall und Klett (1972), Rao (1973), Schach und Schäfer (1978), Searle (1971), Seber (1980), Tatsuoka (1971), Timm (1975), Ward und Jennings (1973) sowie Winer (1971).

Zu Einführung in lineare Modelle mit stochastischen Parametern („random effects“) seien Ahrens und Läuter (1981), Hild (1977) und Searle (1971, Kap. 9-11), genannt, Modelle mit stochastischen Regressoren behandeln u.a. Graybill (1976) oder Johnston (1971).

3. Hypothesenformulierung in verschiedenen Designs

3.1 Einleitung

Die Schätzung von Parametern, Berechnung von Teststatistiken, Überschreitungswahrscheinlichkeiten etc., zur hier dargestellten allgemeinen Theorie sind heute, sofern Rechenanlagen verfügbar sind, weitgehend automatisiert durchführbar und für den Psychologen daher nicht von *primärem* Interesse. Dieser muß lediglich entscheiden, ob die dabei vorausgesetzten Annahmen erfüllt sind. Unabdingbar wichtig ist für jeden Psychologen jedoch das Wissen, welche Hypothesen er in welcher Situation untersuchen soll. Im folgenden werden wir daher zunächst auf die Hypothesenformulierung für verschiedene Designs eingehen, womit zugleich ein Eindruck vermittelt werden soll, wie weit gespannt der Anwendungsbereich der hier dargestellten Theorie ist, aber auch, wo dessen Grenzen liegen. Die Formulierung und Formalisierung von Hypothesen ist außerdem jener Bereich, der am wenigstens automatisiert werden kann und sollte.

Bei der Behandlung der Designs werden wir vom *Zellenmittelwertmodell* ausgehen, bei dem die Parameter in der Matrix B als bedingte Erwartungswerte der abhängigen Variablen in den Zellen des Designs interpretiert werden können. Die Besonderheiten dieses Modells werden wir im Abschnitt 3.2 behandeln. Im Abschnitt 3.3 wird die allgemeine multivariate lineare Hypothese eingeführt, und ihre Anwendung in den Abschnitten 3.4 bis 3.9 an verschiedenen Designs erläutert. Dabei gehen wir auf kreuzfaktorielle Designs, solche mit hierarchischen Faktoren und auf lateinische Quadrate ein. Bei allen Designs setzen wir gleiche Zellenfrequenzen voraus. Für nonorthogonale Designs sei auf die entsprechende Spezialliteratur, z.B. Appelbaum und Cramer (1974), Herr und Gaebelein (1978), Overall und Spiegel (1973a,b), Overall, Spiegel und Cohen (1975), Rawlings (1972, 1973) oder Steyer (1979), hingewiesen.

3.2 Das Zellenmittelwertmodell

Bei allen im folgenden behandelten Designs werden wir das „Zellenmittelwertmodell“ zugrunde legen, in welchem die Parameter β_{qp} in der Matrix B nichts anderes als die Erwartungswerte μ_{qp} der Variablen y_p in der q-ten Zelle eines Designs sind.⁴⁾

⁴⁾ Timm und Carlson (1975) benutzen für diesen Ansatz die Bezeichnung „Modell vollen Ranges“, die wir aber für ungeeignet halten, da noch viele andere Modelle existieren, bei denen die Matrix X vollen Rang hat, bei denen also die Anzahl der Spalten von X gleich dem Rang von X ist, ohne die genannten Eigenschaften zu besitzen.

Die Regel für die Kodierung der Gruppenzugehörigkeit einer Beobachtungseinheit (Be) als Vektor der unabhängigen Variablen x_q ist dann besonders einfach. Für jede Zelle benötigen wir dann eine unabhängige Variable x_q , die folgendermaßen definiert ist:

$$x_q = \begin{cases} 1, & \text{wenn die Be zur } q\text{-ten Zelle gehört} \\ 0, & \text{andernfalls.} \end{cases}$$

Wir überzeugen uns davon, daß dann β_{qp} der bedingte Erwartungswert von y_p in der q -ten Zelle ist. Dazu gehen wir von der Identifikationsgleichung 2.2.15

$$B = E(x'x)^{-1} E(x'y)$$

aus. Für $E(x'x)$ gilt in diesem Fall

$$(3.2.1) \quad E(x'x) = \begin{pmatrix} P(x_1=1) & & & 0 \\ & \cdot & & \\ & & \cdot & \\ 0 & & & P(x_Q=1) \end{pmatrix},$$

da die stochastische Variable $x_q \cdot x_{q^*}$, $q, q^* \in \{1, 2, \dots, Q\}$, nur dann einen Wert ungleich Null annehmen kann, wenn $q = q^*$. In diesem Fall nimmt $x_q \cdot x_{q^*} = x_q \cdot x_q$ den Wert Eins an, und aus der Definition des Erwartungswertes (siehe Gleichung A.1 des Anhangs A von Moosbrugger in diesem Band) folgt $E(x_q \cdot x_q) = 1 \cdot P(x_q \cdot x_q = 1) + 0 \cdot P(x_q \cdot x_q = 0) = P(x_q = 1)$.

Die Inverse einer solchen Diagonalmatrix ist wiederum eine Diagonalmatrix mit reziproken Werten der Ausgangsmatrix als Komponenten (vgl. z.B. Moosbrugger, 1978, S. 32).

$$(3.2.2) \quad E(x'x)^{-1} = \begin{pmatrix} \frac{1}{P(x_1=1)} & & & 0 \\ & \cdot & & \\ & & \cdot & \\ 0 & & & \frac{1}{P(x_Q=1)} \end{pmatrix}.$$

Für die Matrix $E(x'y)$ schließlich erhalten wir

$$(3.2.3) \quad E(\mathbf{x}'\mathbf{y}) = \begin{bmatrix} E(y_1 | x_1=1) \cdot P(x_1=1) & \dots & E(y_P | x_1=1) \cdot P(x_1=1) \\ \vdots & & \vdots \\ E(y_1 | x_Q=1) \cdot P(x_Q=1) & \dots & E(y_P | x_Q=1) \cdot P(x_Q=1) \end{bmatrix},$$

wobei wir für jede Komponente dieser Matrix die Regel

$$(x \cdot y) = \sum_{i=1}^I x_i \cdot E(y | x=x_i) \cdot P(x=x_i).$$

benutzen. Dabei sind x und y stochastische Variablen, wobei x jedoch diskret ist mit den 1 Realisationen x_i .

Ausmultiplizieren von $E(\mathbf{x}'\mathbf{x})^{-1} E(\mathbf{x}'\mathbf{y})$ schließlich führt zu

$$(3.2.4) \quad \mathbf{B} = \begin{bmatrix} E(y_1 | x_1=1) =: \mu_{11} & \dots & E(y_P | x_1=1) =: \mu_{1P} \\ \vdots & & \vdots \\ E(y_1 | x_Q=1) =: \mu_{Q1} & \dots & E(y_P | x_Q=1) =: \mu_{QP} \end{bmatrix}.$$

Diese Kodierungsweise nach dem Zellenmittelwertemodell hat nicht nur den Vorteil, daß die β_{qp} einfach zu interpretieren sind, sondern auch die rechen-technischen Vorzüge, daß nämlich ein Schätzer $\hat{\mathbf{B}}$ von \mathbf{B} ohne faktische Kodierung der Matrix \mathbf{X} sofort berechnet werden kann, daß die Inverse von $\mathbf{X}'\mathbf{X}$ immer existiert, es sei denn, in einer Zelle liegen keine Beobachtungen vor, und daß die Inverse von $\mathbf{X}'\mathbf{X}$ sehr leicht, auch ohne Rechner, zu bilden ist, indem man einfach den reziproken Wert der jeweiligen Komponente der Matrix $\mathbf{X}'\mathbf{X}$ nimmt, die ja selbst eine Diagonalmatrix der Zellenfrequenzen ist.

Bei den im folgenden zu behandelnden Designs können wir also immer davon ausgehen, daß \mathbf{B} die bedingten Erwartungswerte der P abhängigen Variablen in den Q Zellen enthält, also die Zellenmittelwerte μ_{qp} . Die Residuen der P -dimensionalen Vektoren der abhängigen Variablen sind weder innerhalb einer solchen Zelle noch zwischen diesen Zellen voneinander abhängig (siehe Gleichung 2.3.11).

3.3 Die multivariate allgemeine lineare Hypothese

Bei allen in diesem Abschnitt 3 behandelten Designs lassen sich die Hypothesen über die Parameter in der Matrix B in der Form der *multivariaten allgemeinen linearen Hypothese*

$$(3.3.1) \quad H_0: C B A = A$$

formulieren, wobei C eine $M \times Q$ -Matrix mit M linear unabhängigen⁵⁾ Zeilen ist, B die $Q \times P$ -Parametermatrix, A eine $P \times K$ -Matrix mit K linear unabhängigen Spalten und A eine $M \times K$ -Matrix. Die Matrizen C, A und A sind gemäß der jeweiligen Hypothese vom Anwender zu spezifizieren, wobei mit C Hypothesen über Parameter formuliert werden, die sich in jeweils einer Spalte von B befinden, und mit A Hypothesen über Parameter, die sich jeweils in einer Zeile von B befinden. Auf Einzelheiten werden wir in den nachfolgenden Abschnitten ausführlich eingehen.

Einen Zeilenvektor $\psi_m = c_m B I$ nennen wir einen *einfachen P-variaten Vergleich der unabhängigen Variablen*. Gilt dabei für den Zeilenvektor c_m auch

$$(3.3.2) \quad c_m \mathbf{1} = 0,$$

(1 ist ein passender Spalteneinheitsvektor), so sprechen wir auch von einem *Kontrast* anstatt von einem Vergleich.

Einen Spaltenvektor $\psi_k = I B a_k$ nennen wir einen *einfachen Q-variaten Vergleich der abhängigen Variablen*, bzw. *Kontrast*, wenn dabei für den Spaltenvektor a_k

$$(3.3.3) \quad \mathbf{1}' a_k = 0$$

gilt.

Mehrere solcher Zeilen- oder Spaltenvektoren heißen *globaler Vergleich* bzw. *Kontrast*.

Die Koeffizientenvektoren zweier Vergleiche bzw. Kontraste ψ_m und ψ_{m^*} oder ψ_k und ψ_{k^*} heißen *orthogonal*, wenn für ihr Skalarprodukt

$$(3.3.4) \quad c_m c_{m^*}' = 0$$

oder

$$(3.3.5) \quad a_k' a_{k^*} = 0$$

⁵⁾ Endlich viele Vektoren v_1, \dots, v_M eines Vektorraums heißen linear *unabhängig*, wenn aus $c_1 v_1 + \dots + c_M v_M = 0$ immer $c_1 = \dots = c_M = 0$ folgt (vgl. Kowalsky, 1969, S. 33). Sind v_1, \dots, v_M linear abhängig, so läßt sich daher immer mindestens einer von ihnen als Linearkombination der anderen darstellen (vgl. z.B. auch Moosbrugger, 1978, S. 35).

gilt. Bei gleichen Zellenfrequenzen sind dann auch die Vergleiche bzw. Kontraste selbst orthogonal (d.h. unkorreliert).

Im folgenden werden wir die Spezifizierung der Gleichung 3.3.1 bei verschiedenen Designs behandeln.

3.4 Gekreuzte Faktoren über den unabhängigen Variablen

Wir betrachten nun ein Design, bei dem bei jeder der N Beobachtungseinheiten genau P abhängige Variablen y_1, \dots, y_p erhoben werden, und bei dem jede dieser Beobachtungseinheiten in genau eine von z.B. 6 Zellen gehört, die durch das Kreuzen eines Faktors A mit zwei Stufen und eines Faktors B mit drei Stufen entstehen (siehe Tabelle 3.1). Dabei kann es sich um experimentelle Faktoren handeln, oder aber auch um solche, die man durch „reines Beobachten“⁶⁾ gewinnt, wie z.B. „Geschlecht“, „Bildungsklasse“ etc.

Wir gehen zunächst von einer eher *exploratorischen* Situation aus, in der man *keine gezielten* Hypothesen hat. Bei einem solchen Design ist dann zunächst von Interesse, ob in einer Stichprobe überhaupt „signifikante“, d.h. überzufällige Mittelwerteunterschiede zwischen den 6 Zellen vorliegen. Die verbal formulierte Nullhypothese besagt dann, daß alle sechs Erwartungswertvektoren $\beta_q = (\beta_{q1}, \dots, \beta_{qp})$, $q \in \{1, 2, \dots, 6\}$, gleich sind, oder mit anderen Worten, daß weder ein Haupteffekt von Faktor A , noch ein Haupteffekt von Faktor B noch ein Wechselwirkungseffekt der Faktorstufenkombinationen AB besteht. Die Bedeutung dieser Begriffe erläutert z.B. Moosbrugger (1982) in diesem Band, S. 18-19.

Prinzipiell lassen sich nun viele verschiedene Möglichkeiten denken, die *globale Hypothese*, daß keine Unterschiede zwischen den sechs Zeilenvektoren β_q bestehen, in der Form der multivariaten allgemeinen linearen Hypothese zu formulieren. Diese unterscheiden sich jedoch nicht in den resultierenden Quadratsummenmatrizen, mit denen man die Hypothesenbewertung durchführt (siehe Abschnitt 5).

Eine der einfachsten Möglichkeiten besteht darin, die Hypothese so zu formulieren, daß alle benachbarten Zeilenvektoren β_q bzw. μ_q gleich sind, d.h. daß deren Differenz jeweils gleich dem Nullvektor ist:

$$(3.4.1) \quad H_{0_{A,B,AB}}: \begin{array}{l} \mu_1 - \mu_2 = 0 \\ \mu_2 - \mu_3 = 0 \\ \mu_3 - \mu_4 = 0 \\ \mu_4 - \mu_5 = 0 \\ \mu_5 - \mu_6 = 0. \end{array}$$

⁶⁾ Edwards (1971, S. 8f.) führt in dieser Kategorie z.B. „organismische“ Variablen an.

Tabelle 3.1: *Gekreuzte Faktoren A und B als Gliederung der Q unabhängigen Variablen x_q , welche hier an insgesamt $N = N_1 + \dots + N_6$ Beobachtungseinheiten erhoben wurden. Beispiel für Faktor A: Geschlecht (A_1, A_2), für Faktor B: Altersstufen (B_1, B_2, B_3). Das Datenschema zeigt die $P = 4$ abhängigen Variablen y_p pro Beobachtungseinheit und deren bedingte Erwartungswerte $\hat{\mu}_{qp}$.*

			Variablennummer p				
			1	2	3	4	
Gliederung der unabhängigen Variablen	A ₁	B ₁	1	y_{111}	y_{112}	y_{113}	y_{114}
				\vdots	\vdots	\vdots	\vdots
		$y_{N_1,11}$	$y_{N_1,12}$	$y_{N_1,13}$	$y_{N_1,14}$		
		$\hat{\mu}_{11}$	$\hat{\mu}_{12}$	$\hat{\mu}_{13}$	$\hat{\mu}_{14}$		
		B ₂	2	y_{121}	y_{122}	y_{123}	y_{124}
				\vdots	\vdots	\vdots	\vdots
	$y_{N_2,21}$	$y_{N_2,22}$	$y_{N_2,23}$	$y_{N_2,24}$			
	$\hat{\mu}_{21}$	$\hat{\mu}_{22}$	$\hat{\mu}_{23}$	$\hat{\mu}_{24}$			
	B ₃	3	y_{131}	y_{132}	y_{133}	y_{134}	
\vdots			\vdots	\vdots	\vdots		
$y_{N_3,31}$	$y_{N_3,32}$	$y_{N_3,33}$	$y_{N_3,34}$				
$\hat{\mu}_{31}$	$\hat{\mu}_{32}$	$\hat{\mu}_{33}$	$\hat{\mu}_{34}$				
A ₂	B ₁	4	y_{141}	y_{142}	y_{143}	y_{144}	
			\vdots	\vdots	\vdots	\vdots	
	$y_{N_4,41}$	$y_{N_4,42}$	$y_{N_4,43}$	$y_{N_4,44}$			
	$\hat{\mu}_{41}$	$\hat{\mu}_{42}$	$\hat{\mu}_{43}$	$\hat{\mu}_{44}$			
	B ₂	5	y_{151}	y_{152}	y_{153}	y_{154}	
			\vdots	\vdots	\vdots	\vdots	
$y_{N_5,51}$	$y_{N_5,52}$	$y_{N_5,53}$	$y_{N_5,54}$				
$\hat{\mu}_{51}$	$\hat{\mu}_{52}$	$\hat{\mu}_{53}$	$\hat{\mu}_{54}$				
B ₃	6	y_{161}	y_{162}	y_{163}	y_{164}		
		\vdots	\vdots	\vdots	\vdots		
$y_{N_6,61}$	$y_{N_6,62}$	$y_{N_6,63}$	$y_{N_6,64}$				
$\hat{\mu}_{61}$	$\hat{\mu}_{62}$	$\hat{\mu}_{63}$	$\hat{\mu}_{64}$				

Da gemäß Gleichung 3.2.4

$$(3.4.2) \quad \mathbf{B} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ \vdots \\ \boldsymbol{\mu}_6 \end{bmatrix}$$

gilt, wählen wir die folgende Hypothesenmatrix

$$(3.4.3) \quad \mathbf{C}_{A,B,AB} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix},$$

wobei wir, falls nicht explizit anders angegeben, gleiche Zellenfrequenzen voraussetzen.

Die Matrix \mathbf{A} in der allgemeinen linearen Hypothese $\mathbf{C} \mathbf{B} \mathbf{A} = \mathbf{A}$ erlaubt, die P abhängigen Variablen beliebig linear zusammenzufassen. Man kann damit praktisch neue abhängige Variablen bilden. In anderen Designs kann \mathbf{A} jedoch auch die „Hauptrolle“ beim Formulieren von Hypothesen spielen, wie wir weiter unten zeigen werden.

$\mathbf{A}_{A,B,AB}$ sei hier die $P \times P$ -Einheitsmatrix \mathbf{I} , und $\boldsymbol{\Delta}_{A,B,AB}$ eine $5 \times P$ -Nullmatrix. Ausmultiplizieren von $\mathbf{C}_{A,B,AB} \mathbf{B} \mathbf{A}_{A,B,AB} = \boldsymbol{\Delta}_{A,B,AB}$ führt dann zu den Gleichungen 3.4.1. Die so formulierte Hypothese läßt sich mit den in Abschnitt 5 angegebenen Verfahren bewerten. Falls sich dabei ergibt, daß man die Nullhypothese beibehalten sollte, kann man das Verfahren hier abbrechen. Andernfalls jedoch kann man mit der nächsten Hypothese fortfahren.

Es ist dabei u.E. zu empfehlen, als nächstes die Nullhypothese zu untersuchen, daß keine *Interaktion (Wechselwirkung)* zwischen den Faktoren A und B bestehen, nämlich, daß die Unterschiede zwischen den Erwartungswertvektoren der Stufen A_1 und A_2 auf allen Stufen von Faktor B gleich sind:

$$(3.4.4) \quad \mathbf{H}_{0_{AB}}: \begin{aligned} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_4) - (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_5) &= \mathbf{0} \\ (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_5) - (\boldsymbol{\mu}_3 - \boldsymbol{\mu}_6) &= \mathbf{0}. \end{aligned}$$

In der Form $\mathbf{C} \mathbf{B} \mathbf{A} = \mathbf{A}$ wählen wir

$$(3.4.5) \quad \mathbf{C}_{AB} = \begin{bmatrix} 1 & -1 & 0 & -1 & 1 & 0 \\ 0 & 1 & -1 & 0 & -1 & 1 \end{bmatrix},$$

$A_{AB} = 1$ und $\Delta_{AB} = 0$. Multipliziert man $C_{AB} B A_{AB} = \Delta_{AB}$ aus, so erhält man die Gleichungen 3.4.4. Die Bewertung führt man wiederum mit den in Abschnitt 5 angegebenen Verfahren durch.

Sollte man zu dem Ergebnis kommen, daß wohl keine Interaktionen vorliegen, so kann man die *Haupteffekte* der Faktoren A bzw. B mit den Nullhypothesen untersuchen, daß erstens die Summen der Erwartungswertevektoren auf den Stufen von Faktor A bzw. zweitens auf den Stufen von Faktor B gleich sind. Es ist u.E. auch dann sinnvoll, diese Hypothesen zu untersuchen, wenn eine Interaktion AxB besteht, falls man am durchschnittlichen Effekt z.B. des Faktors A über die Stufen von B hinweg (bzw. umgekehrt) interessiert ist.

Die Nullhypothese für Faktor A kann man so formulieren:

$$(3.4.6) \quad H_{0_A}: (\mu_1 + \mu_2 + \mu_3) - (\mu_4 + \mu_5 + \mu_6) = 0.$$

in der Form $C B A = A$ wählen wir

$$(3.4.7) \quad C_A = [1 \quad 1 \quad 1 \quad -1 \quad -1 \quad -1],$$

sowie $A_A = I$ und $\Delta_A = 0$.

Die Nullhypothese für Faktor B ist

$$(3.4.8) \quad H_{0_B}: \begin{array}{l} (\mu_1 + \mu_4) - (\mu_2 + \mu_5) = 0 \\ (\mu_2 + \mu_5) - (\mu_3 + \mu_6) = 0 \end{array},$$

und in der Form $C B A = A$ ist dann

$$(3.4.9) \quad C_B = \begin{bmatrix} 1 & -1 & 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 & 1 & -1 \end{bmatrix},$$

$A_B = I$ und $\Delta_B = 0$ zu wählen. Die Hypothesen für die Haupteffekte (vgl. auch Moosbrugger, 1978, S. 139-143) lassen sich mit den Verfahren bewerten, die in Abschnitt 5 angegeben sind.

Wie ist nun die Hypothesenmatrix C für den Fall zu wählen, daß man z.B. den „Effekt“ von Faktor A auf der Stufe 1 des Faktors B untersuchen will? Dazu wäre einfach

$$(3.4.10) \quad C_{A|B_1} = [1 \quad 0 \quad 0 \quad -1 \quad 0 \quad 0]$$

zu setzen, was zusammen mit $A_{A|B_1} = I$ und $\Delta_{A|B_1} = 0$ der Nullhypothese

$$H_{0_{A|B_1}}: \quad \mu_1 - \mu_4 = 0$$

entspricht.

Man beachte, daß jede Zeile der Matrix C einer eigenen (Sub-) Hypothese entspricht, die man nach den gleichen allgemeinen Methoden des Abschnitts 5 untersuchen kann. Solche gezielten Einzelvergleiche sind legitim, falls gezielte Hypothesen vor der Datenerhebung bestehen⁷⁾.

Die Verallgemeinerung des hier behandelten Designs auf beliebig viele Faktoren mit beliebig vielen Stufen sollte keine Schwierigkeiten bereiten. Bei der Anordnung der Daten ist lediglich darauf zu achten, daß die Beobachtungseinheiten n untereinander angeordnet werden, sofern man bereit ist, die Unabhängigkeitsannahme (2.3.11) zu machen.

Es wurde bereits weiter oben darauf hingewiesen, daß die hier gewählten Hypothesen nicht die einzig möglichen und richtigen sind. Man könnte z.B. an der Frage interessiert sein, ob der Durchschnitt der Erwartungswerte in den Zellen 1, 2, 4 und 5 gleich dem Durchschnitt der Erwartungswerte in den Zellen 3 und 6 ist. Allgemein sind auch Hypothesen der Form $C B A = A \neq 0$ möglich.

Zum Abschluß dieses Abschnitts sei davor gewarnt, zu viele Einzelvergleiche durchzuführen. Dies führt zu einer Akkumulation des α -Fehlers. Wenn bei einem Signifikanztest die Irrtumswahrscheinlichkeit für die fälschliche Ablehnung der Nullhypothese $\alpha = 0.05$ beträgt, so liegt sie bei fünf unabhängigen Signifikanztests auf diesem Niveau schon bei $\alpha_5 = 1 - (1 - \alpha)^5 = 1 - 0.77$, d.h. die Wahrscheinlichkeit bei fünf unabhängigen Signifikanztests fälschlicherweise einen „signifikanten“ zu finden, beträgt $\alpha_5 = 0.23$. Sind viele Signifikanztests durchzuführen, sollte man daher auf die Verfahren „simultaner Mittelwertvergleiche“ zurückgreifen (vgl. Aitkin, 1969, z.B. Miller, 1966, oder Gabriel, 1968, 1969a, b).

3.5 Gekreuzte Faktoren über den abhängigen Variablen

Wir betrachten nun die P abhängigen Variablen genauer und nehmen an, daß sie sich ebenfalls faktoriell gliedern lassen (siehe Tabelle 3.2), so, wie sich die Q unabhängigen Gruppen im vorangegangenen Abschnitt strukturieren ließen. Die Aufteilung der Gruppen wollen wir hier zunächst nicht weiter beachten, da wir sie ja bereits im vorangegangenen Abschnitt behandelt haben.

⁷⁾ Andernfalls handelt es sich nicht um Hypothesentesten, sondern um die Exploration von Daten, und die dabei gefundenen „signifikanten“ Tests sollte man nicht als signifikant, sondern als statistisch auffällig bezeichnen, da die Irrtumswahrscheinlichkeitsangaben (α -Niveau) dann sehr irreführend sein können.

Tabelle 3.2: *Gekreuzte Faktoren U und V als Gliederung der P abhängigen Variablen y_p* , welche an insgesamt $N = N_1 + \dots + N_6$ Beobachtungseinheiten (Ben) erhoben wurden. Dabei wird die Unabhängigkeitsannahme, die in Gleichung 2.3.11 enthalten ist, vorausgesetzt. Beispiel für Faktor U: Schulleistungstests (U_1, U_2), für Faktor V: Vortest (V_1), Nachtest (V_2). Das Datenschema zeigt die $P = 4$ abhängigen Variablen y_p pro Be und deren bedingte Erwartungswerte $\hat{\mu}_{qp}$.

Gliederung der abhängigen Variablen

Zellen- nummer q	U ₁		U ₂		Faktor U
	V ₁	V ₂	V ₁	V ₂	Faktor V
	1	2	3	4	Variablen- nummer p
1	y_{111}	y_{112}	y_{113}	y_{114}	
	\vdots	\vdots	\vdots	\vdots	
	$y_{N,11}$	$y_{N,12}$	$y_{N,13}$	$y_{N,14}$	
	$\hat{\mu}_{11}$	$\hat{\mu}_{12}$	$\hat{\mu}_{13}$	$\hat{\mu}_{14}$	
2	y_{121}	y_{122}	y_{123}	y_{124}	
	\vdots	\vdots	\vdots	\vdots	
	$y_{N,21}$	$y_{N,22}$	$y_{N,23}$	$y_{N,24}$	
	$\hat{\mu}_{21}$	$\hat{\mu}_{22}$	$\hat{\mu}_{23}$	$\hat{\mu}_{24}$	
3	y_{131}	y_{132}	y_{133}	y_{134}	
	\vdots	\vdots	\vdots	\vdots	
	$y_{N,31}$	$y_{N,32}$	$y_{N,33}$	$y_{N,34}$	
	$\hat{\mu}_{31}$	$\hat{\mu}_{32}$	$\hat{\mu}_{33}$	$\hat{\mu}_{34}$	
4	y_{141}	y_{142}	y_{143}	y_{144}	
	\vdots	\vdots	\vdots	\vdots	
	$y_{N,41}$	$y_{N,42}$	$y_{N,43}$	$y_{N,44}$	
	$\hat{\mu}_{41}$	$\hat{\mu}_{42}$	$\hat{\mu}_{43}$	$\hat{\mu}_{44}$	
5	y_{151}	y_{152}	y_{153}	y_{154}	
	\vdots	\vdots	\vdots	\vdots	
	$y_{N,51}$	$y_{N,52}$	$y_{N,53}$	$y_{N,54}$	
	$\hat{\mu}_{51}$	$\hat{\mu}_{52}$	$\hat{\mu}_{53}$	$\hat{\mu}_{54}$	
6	y_{161}	y_{162}	y_{163}	y_{164}	
	\vdots	\vdots	\vdots	\vdots	
	$y_{N,61}$	$y_{N,62}$	$y_{N,63}$	$y_{N,64}$	
	$\hat{\mu}_{61}$	$\hat{\mu}_{62}$	$\hat{\mu}_{63}$	$\hat{\mu}_{64}$	

Die Faktoren U und V können in verschiedenen Forschungssituationen entstehen: Falls man nicht bereit ist, die in Gleichung 2.3.11 enthaltene Unabhängigkeitsannahme zu machen, kann z.B. V_1 Vor- und V_2 Nachtest von 2 inhaltlich verschiedenen abhängigen Variablen (U_1, U_2) sein, oder umgekehrt können U_1 und U_2 wiederholte Messungen von 2 inhaltlich verschiedenen abhängigen Variablen sein (V_1 und V_2). Genausogut kann es sich jedoch auch bei U_1 und U_2 und bei V_1 und V_2 um wiederholte Messungen handeln, wobei U_1, U_2 aufeinanderfolgende experimentelle Versuchsbedingungen sind. Kann man hingegen die Annahme $C(\boldsymbol{\varepsilon}_n, \boldsymbol{\varepsilon}_{n^*}) = 0$ bei $n \neq n^*$, für die Residuenvektoren $\boldsymbol{\varepsilon}_n = y_n - E(y_n | x_n)$ machen, wobei y_n und y_{n^*} die abhängigen Variablen derselben Person zu verschiedenen Zeitpunkten enthält, so würde man besser die Meßwiederholungen „untereinander“ anordnen, wodurch der Index n dann nicht mehr die Person, sondern eine Person-Zeit-Kombination bezeichnet. Wottawa (1981) hat ein Verfahren angegeben, wie die Annahme der Unabhängigkeit der Residuen überprüft werden kann.

Formal betrachtet haben wir von jeder Beobachtungseinheit hier insgesamt $P = 2 \cdot 2$ abhängige Variablen vorliegen. Die Gleichungen 2.3.8 bis 2.3.11 werden also weiterhin vorausgesetzt, wobei hier $N = N_1 + \dots + N_4$ gilt, und Σ die Dimension $P \times P = 4 \times 4$ hat.

Da wir auch hier ein Zellenmittelwertmodell zugrundelegen, sind die Parameter in der Matrix B wiederum die Erwartungswerte der $P=4$ abhängigen Variablen in den $Q=6$ Zellen, deren Anordnung in Tabelle 3.2 wiedergegeben ist.

In formaler Hinsicht unterscheidet sich das Design der Tabelle 3.2 nicht von dem der Tabelle 3.1. Der einzige Unterschied besteht darin, daß hier die $P = 2 \cdot 2$ **abhängigen** Variablen untergliedert wurden, wobei die Untergliederungen der Q Gruppen nach den Faktoren A und B, die wir in Tabelle 3.1 vorgenommen hatten, außer acht gelassen wird.

Die Hypothesen über die Faktoren U, V und deren Interaktion $U \times V$ können nach der gleichen Methode wie im vorangegangenen Abschnitt formuliert werden. Der einzige Unterschied besteht darin, daß hier die Matrix A der multivariaten allgemeinen linearen Hypothese $C B A = A$ Verwendung findet und C jeweils die $Q \times Q$ -Einheitsmatrix ist. Die einzelnen Hypothesenmatrizen A können dabei so gewählt werden, wie die Hypothesenmatrizen C der entsprechenden Hypothesen in Abschnitt 3.4, aber in transponierter Form.

Bei einem Design wie dem vorliegenden können zusätzlich noch Interaktionen zwischen den „unabhängigen Faktoren“ A und B (entsprechend Tabelle 3.1) und den „abhängigen Faktoren“ U und V (entsprechend Tabelle 3.2) geprüft werden. So kann man z.B. die Nullhypothese formulieren, daß keine Interaktion zwischen den Q Gruppen und den P abhängigen Variablen bestehen:

$$(3.5.1) \quad H_{0_{(A,B,AB)} \times (U,V,UV)}: \\
(\mu_{11} - \mu_{12}) = (\mu_{21} - \mu_{22}) = \dots = (\mu_{61} - \mu_{62}) \\
(\mu_{12} - \mu_{13}) = (\mu_{22} - \mu_{23}) = \dots = (\mu_{62} - \mu_{63}) \\
(\mu_{13} - \mu_{14}) = (\mu_{23} - \mu_{24}) = \dots = (\mu_{63} - \mu_{64}).$$

Anschaulich bedeutet das, daß die Differenzen zwischen den Erwartungswerten der P=4 abhängigen Variablen von den Q=6 Zellen nicht moderiert werden (vgl. Moosbrugger, 1982, in diesem Band, S. 19). Dabei können diese Differenzen zwischen den P=4 Variablen wohl verschieden sein. Diese Nullhypothese können wir auch in Form von 15 Differenzen anschreiben, welche Null sein müssen:

$$(3.5.2) \quad H_{0_{(A,B,AB)} \times (U,V,UV)}: \\
(\mu_{11} - \mu_{12}) - (\mu_{21} - \mu_{22}) = (\mu_{12} - \mu_{13}) - (\mu_{22} - \mu_{23}) = (\mu_{13} - \mu_{14}) - (\mu_{23} - \mu_{24}) = 0 \\
(\mu_{21} - \mu_{22}) - (\mu_{31} - \mu_{32}) = (\mu_{22} - \mu_{23}) - (\mu_{32} - \mu_{33}) = (\mu_{23} - \mu_{24}) - (\mu_{33} - \mu_{34}) = 0 \\
\vdots \\
(\mu_{51} - \mu_{52}) - (\mu_{61} - \mu_{62}) = (\mu_{52} - \mu_{53}) - (\mu_{62} - \mu_{63}) = (\mu_{53} - \mu_{54}) - (\mu_{63} - \mu_{64}) = 0$$

Zur Formulierung dieser Wechselwirkungshypothese zwischen den unabhängigen und abhängigen Faktoren in der Form C B A = A wählen wir

$$(3.5.3) \quad C_{A,B,AB} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

und zugleich

$$(3.5.4) \quad A_{U,V,UV} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}$$

sowie $\Delta_{(A,B,AB) \times (U,V,UV)} = 0$. Ebenso gut ließen sich aber auch Hypothesen formulieren, daß keine Interaktionen A x U, A x V etc. bestehen. Nachdem einmal die Arbeit der Formulierung der Hypothesen in der Form C B A = A geleistet ist, können diese nach den Verfahren untersucht werden, die wir in Abschnitt 5 behandeln.

Die Verallgemeinerung eines solchen Designs auch für beliebig viele Faktoren über den abhängigen Variablen ist offensichtlich. Bei der Festlegung des Designs ist lediglich darauf zu achten, daß für jede Be (z.B. Person oder Person-Zeit-Kombination) die P abhängigen Variablen jeweils nebeneinander und die Beobachtungseinheiten untereinander angeordnet werden, so daß die Gleichungen 2.3.6 bis 2.3.11 erfüllt sind.

3.6 Hierarchische Faktoren über den unabhängigen Variablen

Wir betrachten nun ein Design, bei dem ein fünfstufiger Faktor B einem zweistufigen Faktor A hierarchisch untergeordnet ist, bei dem also jede Stufe von B nur mit einer Stufe des Faktors A zusammen vorkommt. Beide Faktoren strukturieren dabei die $Q=5$ Zellen, innerhalb derer sich dann insgesamt $N = N_1 + \dots + N_5$ P-dimensionale Vektoren abhängiger Variablen befinden. Ein typischer Fall wäre, daß man P abhängige Variablen bei zwei verschiedenen Lehrmethoden (Faktor A) untersuchen will, wobei man die Schulkassen B_1 bis B_5 einbezieht.

Wir setzen wieder voraus, daß die Zellenmittelwertekodierung für X angewandt wurde, so daß B die $5 \times P$ -Matrix der bedingten Erwartungswerte der abhängigen Variablen in den fünf Zellen ist, und zwar in der gleichen Anordnung wie diese in Tabelle 3.3 dargestellt sind.

Bei einem solchen Design sind mehrere Hypothesen von Interesse. Eine erste globale Hypothese besagt, daß keine Unterschiede zwischen den Gruppen B_1 bis B_5 bezüglich der Erwartungswerte der P abhängigen Variablen bestehen.

Diese Hypothese läßt sich in der Form der allgemeinen multivariaten linearen Hypothese $C B A = A$ darstellen, indem wir z.B.

$$(3.6.1) \quad C_{A,B} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

$A_{A,B} = I$ und $\Delta_{A,B} = 0$ wählen. Falls die Bewertung nach den in Abschnitt 5 angegebenen Verfahren ergibt, daß diese Hypothese beizubehalten sei, kann das Verfahren abgebrochen werden. Andernfalls überprüfen wir als zweite Hypothese, daß sich die Erwartungswerte der P abhängigen Variablen zwischen den Gruppen A_1 und A_2 nicht unterscheiden. Diese läßt sich in der Form $C B A = A$ formulieren, indem wir z.B.

$$(3.6.2) \quad C_A = [1/2 \quad 1/2 \quad -1/3 \quad -1/3 \quad -1/3],$$

$A_A = I$ und $\Delta_A = 0$ wählen.

Der Faktor B ist unter Faktor A „genestet“. Folglich können Hypothesen über Faktor B nur auf den Stufen von Faktor A geprüft werden. Wir formulieren die dritte bzw. vierte Hypothese, daß sich die Erwartungswerte der P abhängigen Variablen zwischen den Gruppen B_1 und B_2 bzw. B_3 bis B_5 auf der Stufe A_1 bzw. A_2 nicht unterscheiden, in der Form $C B A = A$, indem wir z.B.

Tabelle 3.3 *Hierarchische Faktoren A und B als Gliederung der Q unabhängigen Variablen x_q , welche an insgesamt $N = N_1 + \dots + N_5$ Beobachtungseinheiten erhoben wurden. Beispiel für Faktor A: Lehrmethoden (A_1, A_2), für Faktor B: Schulklassen (B_1 bis B_5). Faktor B ist dem Faktor A untergeordnet („genestet“), weil jede Stufe von B nur unter einer Stufe von A vorkommt. Das Datenschema zeigt die $P=5$ abhängigen Variablen y_p pro Beobachtungseinheit und deren bedingte Erwartungswerte $\hat{\mu}_{qp}$.*

			Variablennummer p					
			1	2	3	4	5	
Gliederung der unabhängigen Variablen	Faktor A	Faktor B	Zellennummer q					
	A_1	B_1	1	y_{111}	y_{112}	y_{113}	y_{114}	y_{115}
				\vdots	\vdots	\vdots	\vdots	\vdots
				$y_{N_1,11}$	$y_{N_1,12}$	$y_{N_1,13}$	$y_{N_1,14}$	$y_{N_1,15}$
				$\hat{\mu}_{11}$	$\hat{\mu}_{12}$	$\hat{\mu}_{13}$	$\hat{\mu}_{14}$	$\hat{\mu}_{15}$
A_2	B_2	2	y_{121}	y_{122}	y_{123}	y_{124}	y_{125}	
			\vdots	\vdots	\vdots	\vdots	\vdots	
			$y_{N_2,21}$	$y_{N_2,22}$	$y_{N_2,23}$	$y_{N_2,24}$	$y_{N_2,25}$	
			$\hat{\mu}_{21}$	$\hat{\mu}_{22}$	$\hat{\mu}_{23}$	$\hat{\mu}_{24}$	$\hat{\mu}_{25}$	
A_2	B_3	3	y_{131}	y_{132}	y_{133}	y_{134}	y_{135}	
			\vdots	\vdots	\vdots	\vdots	\vdots	
			$y_{N_3,31}$	$y_{N_3,32}$	$y_{N_3,33}$	$y_{N_3,34}$	$y_{N_3,35}$	
			$\hat{\mu}_{31}$	$\hat{\mu}_{32}$	$\hat{\mu}_{33}$	$\hat{\mu}_{34}$	$\hat{\mu}_{35}$	
A_2	B_4	4	y_{141}	y_{142}	y_{143}	y_{144}	y_{145}	
			\vdots	\vdots	\vdots	\vdots	\vdots	
			$y_{N_4,41}$	$y_{N_4,42}$	$y_{N_4,43}$	$y_{N_4,44}$	$y_{N_4,45}$	
			$\hat{\mu}_{41}$	$\hat{\mu}_{42}$	$\hat{\mu}_{43}$	$\hat{\mu}_{44}$	$\hat{\mu}_{45}$	
A_2	B_5	5	y_{151}	y_{152}	y_{153}	y_{154}	y_{155}	
			\vdots	\vdots	\vdots	\vdots	\vdots	
			$y_{N_5,51}$	$y_{N_5,52}$	$y_{N_5,53}$	$y_{N_5,54}$	$y_{N_5,55}$	
			$\hat{\mu}_{51}$	$\hat{\mu}_{52}$	$\hat{\mu}_{53}$	$\hat{\mu}_{54}$	$\hat{\mu}_{55}$	

$$(3.6.3) \quad \mathbf{C}_{B|A_1} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \end{bmatrix}$$

bei $\mathbf{A}_{B|A} = \mathbf{I}$ und $\mathbf{\Delta}_{B|A} = 0$ bzw.

$$(3.6.4) \quad \mathbf{C}_{B|A_2} = \begin{bmatrix} 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

bei $\mathbf{A}_{B|A} = \mathbf{I}$ und $\mathbf{\Delta}_{B|A} = 0$ wählen.

Die beiden Hypothesen über Faktor B können auch gemeinsam formuliert und überprüft werden.

Interaktionen zwischen den Faktoren A und B lassen sich bei einem hierarchischen Design nicht überprüfen.

3.7 Hierarchische Faktoren über den abhängigen Variablen

Das im vorangegangenen Abschnitt dargestellte Prinzip der Strukturierung der Beobachtungseinheiten durch hierarchische Faktoren läßt sich auch auf die abhängigen Variablen übertragen. So ist z.B. denkbar, daß mit jeder Person fünf Versuchsphasen (V_1 bis V_5) durchgeführt werden, die sich aber in zwei Hauptphasen (U_1 und U_2) aufgliedern lassen, wobei V_1 bis V_3 der Stufe U_1 untergeordnet sind und V_4 bis V_5 der Stufe U_2 . Sind wir aus theoretischen Gründen (z.B. wegen Existenz von „Lerneffekten“) nicht in der Lage, die Gleichung 2.3.11 vorauszusetzen, wenn n eine Person-Zeit-Kombination indiziert, so können wir die fünf Meßwiederholungen als Realisationen von fünf (formal verschiedenen) abhängigen Variablen auffassen, so daß n nun die Person angibt, von der jeweils 5 Meßwerte vorliegen. So umstrukturiert wären die Gleichungen 2.3.8 bis 2.3.11 wieder gültig, wenn Abhängigkeiten zwischen den Personen ausgeschlossen werden können.

In einer ersten globalen Nullhypothese behaupten wir, daß sich die Erwartungswerte der abhängigen Variablen zwischen den fünf Versuchsphasen nicht unterscheiden. Dazu würde man $\mathbf{C}_{U,V} = \mathbf{I}$, $\mathbf{\Delta}_{U,V} = 0$ und

$$(3.7.1) \quad \mathbf{A}_{U,V} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

wählen. Dabei können sehr wohl Unterschiede zwischen den Q Zellen bestehen.

Tabelle 3.4: *Hierarchische Faktoren U und V als Gliederung der P abhängigen Variablen y_p , welche hier an insgesamt $N = N_1 + \dots + N_5$ Beobachtungseinheiten erhoben wurden. Beispiel für Faktor U: vormittags (U_1), nachmittags (U_2), für Faktor V: 5 Versuchsphasen (V_1 bis V_5). Faktor V ist Faktor U hierarchisch untergeordnet („genestet“), weil jede Stufe von V nur unter einer Stufe von U vorkommt. Das Datenschema zeigt die $P = 5$ abhängigen Variablen y_p pro Beobachtungseinheit und deren bedingte Erwartungswerte $\hat{\mu}_{qp}$.*

Zellen- nummer q		Gliederung der abhängigen Variablen					Faktor U
		U_1			U_2		
		V_1	V_2	V_3	V_4	V_5	
	1	2	3	4	5		
1	y_{111}	y_{112}	y_{113}	y_{114}	y_{115}		
	\vdots	\vdots	\vdots	\vdots	\vdots		
	$y_{N_1,11}$	$y_{N_1,12}$	$y_{N_1,13}$	$y_{N_1,14}$	$y_{N_1,15}$		
	$\hat{\mu}_{11}$	$\hat{\mu}_{12}$	$\hat{\mu}_{13}$	$\hat{\mu}_{14}$	$\hat{\mu}_{15}$		
2	y_{121}	y_{122}	y_{123}	y_{124}	y_{125}		
	\vdots	\vdots	\vdots	\vdots	\vdots		
	$y_{N_2,21}$	$y_{N_2,22}$	$y_{N_2,23}$	$y_{N_2,24}$	$y_{N_2,25}$		
	$\hat{\mu}_{21}$	$\hat{\mu}_{22}$	$\hat{\mu}_{23}$	$\hat{\mu}_{24}$	$\hat{\mu}_{25}$		
3	y_{131}	y_{132}	y_{133}	y_{134}	y_{135}		
	\vdots	\vdots	\vdots	\vdots	\vdots		
	$y_{N_3,31}$	$y_{N_3,32}$	$y_{N_3,33}$	$y_{N_3,34}$	$y_{N_3,35}$		
	$\hat{\mu}_{31}$	$\hat{\mu}_{32}$	$\hat{\mu}_{33}$	$\hat{\mu}_{34}$	$\hat{\mu}_{35}$		
4	y_{141}	y_{142}	y_{143}	y_{144}	y_{145}		
	\vdots	\vdots	\vdots	\vdots	\vdots		
	$y_{N_4,41}$	$y_{N_4,42}$	$y_{N_4,43}$	$y_{N_4,44}$	$y_{N_4,45}$		
	$\hat{\mu}_{41}$	$\hat{\mu}_{42}$	$\hat{\mu}_{43}$	$\hat{\mu}_{44}$	$\hat{\mu}_{45}$		
5	y_{151}	y_{152}	y_{153}	y_{154}	y_{155}		
	\vdots	\vdots	\vdots	\vdots	\vdots		
	$y_{N_5,51}$	$y_{N_5,52}$	$y_{N_5,53}$	$y_{N_5,54}$	$y_{N_5,55}$		
	$\hat{\mu}_{51}$	$\hat{\mu}_{52}$	$\hat{\mu}_{53}$	$\hat{\mu}_{54}$	$\hat{\mu}_{55}$		

Als zweite Hypothese prüfen wir, ob sich die Erwartungswerte der abhängigen Variablen zwischen den Faktorstufen U_1 und U_2 nicht unterscheiden. Dazu wählen wir $C_U = I$, $\Delta_U = 0$ und

$$(3.7.2) \quad \mathbf{A}_U = \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \\ -1/2 \\ -1/2 \end{pmatrix}.$$

Der Faktor V ist unter den Faktor U „genestet“. Hypothesen über Faktor V sind folglich nur auf den Stufen von Faktor U überprüfbar:

Wir formulieren die dritte bzw. vierte Hypothese, daß sich die Erwartungswerte der P abhängigen Variablen zwischen den Stufen V_1 bis V_3 auf Stufe U_1 bzw. V_4 und V_5 auf Stufe U_2 nicht unterscheiden, in der Form $C \ B \ A = \Delta$, indem wir z.B.

$$(3.7.3) \quad \mathbf{A}_{V|U_1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

bei $C_{V|U} = I$ und $\Delta_{V|U_1} = 0$ bzw.

$$(3.7.4) \quad \mathbf{A}_{V|U_2} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}$$

bei $C_{V|U} = I$ und $\Delta_{V|U_2} = 0$ wählen.

Die beiden Hypothesen über Faktor V können auch gemeinsam formuliert und überprüft werden. Hypothesen über eine Interaktion $U \times V$ lassen sich bei einem solchen Design nicht prüfen. Jedoch können durchaus Hypothesen über die Interaktion von „abhängigen“ und „unabhängigen“ Faktoren geprüft werden (vgl. Abschnitt 3.5).

Es sei noch bemerkt, daß die Q Zellen beliebig, d.h. kreuzfaktoriell, hierarchisch oder sonstwie, strukturiert sein können. Eine weitere Verallgemeinerung dieses Designs liegt in der Möglichkeit, zu jeder Versuchsphase mehrere abhängige Variablen zugleich zu erheben, so daß dieser neue Faktor mit U und mit V gekreuzt wäre. Prinzipiell läßt sich jede Strukturierung der abhängigen Variablen mit jeder Strukturierung der unabhängigen Beobachtungseinheiten kombinieren, womit jeweils ein neues Design konstruiert wäre.

3.8 Lateinisches Quadrat über den unabhängigen Variablen

Eine weitere Möglichkeit, die unabhängigen Beobachtungseinheiten zu strukturieren, ist das Prinzip des „lateinischen Quadrats“. Bei diesem Designtyp hat man mehrere Faktoren mit gleicher Stufenzahl und P abhängige Variablen vorliegen, die ihrerseits durch ein beliebiges Design strukturiert sein können.

Tabelle 3.5: *Faktoren A, B und C im „lateinischen Quadrat“ als Gliederung der Q unabhängigen Variablen x_q , welche insgesamt an $N = N_1 + \dots + N_4$ Beobachtungseinheiten erhoben wurden. Beispiel für Faktor A: Geschlecht (A_1, A_2), für Faktor B: Sozialstatus (B_1, B_2), Faktor C: Unterrichtsmethoden (C_1, C_2). Das Datenschema zeigt die $P=4$ abhängigen Variablen y_p pro Beobachtungseinheit und deren bedingte Erwartungswerte $\hat{\mu}_{qp}$.*

				Variablennummer p				
Faktor A	Faktor B	Faktor C	Zellen- nummer q	1	2	3	4	
Gliederung der unabhängigen Variablen	A ₁	B ₁	C ₁	1	y_{111}	y_{112}	y_{113}	y_{114}
					\vdots	\vdots	\vdots	\vdots
					$y_{N_1,11}$	$y_{N_1,12}$	$y_{N_1,13}$	$y_{N_1,14}$
					$\hat{\mu}_{11}$	$\hat{\mu}_{12}$	$\hat{\mu}_{13}$	$\hat{\mu}_{14}$
	A ₁	B ₂	C ₂	2	y_{121}	y_{122}	y_{123}	y_{124}
\vdots					\vdots	\vdots	\vdots	
					$y_{N_2,21}$	$y_{N_2,22}$	$y_{N_2,23}$	$y_{N_2,24}$
					$\hat{\mu}_{21}$	$\hat{\mu}_{22}$	$\hat{\mu}_{23}$	$\hat{\mu}_{24}$
	A ₂	B ₁	C ₂	3	y_{131}	y_{132}	y_{133}	y_{134}
\vdots					\vdots	\vdots	\vdots	
					$y_{N_3,31}$	$y_{N_3,32}$	$y_{N_3,33}$	$y_{N_3,34}$
					$\hat{\mu}_{31}$	$\hat{\mu}_{32}$	$\hat{\mu}_{33}$	$\hat{\mu}_{34}$
	A ₂	B ₂	C ₁	4	y_{141}	y_{142}	y_{143}	y_{144}
\vdots					\vdots	\vdots	\vdots	
					$y_{N_4,41}$	$y_{N_4,42}$	$y_{N_4,43}$	$y_{N_4,44}$
					$\hat{\mu}_{41}$	$\hat{\mu}_{42}$	$\hat{\mu}_{43}$	$\hat{\mu}_{44}$

Bei einem dreifaktoriellen Design beispielsweise, ordnet man die ersten beiden Faktoren A und B wie bei einem kreuzfaktoriellen Design an, und dann den dritten Faktor C so, daß jede seiner Stufen jeweils nur auf einer Stufe der anderen beiden Faktoren vorkommt (siehe Tabelle 3.5). Bei zweistufigen Faktoren führt das dazu, daß man statt $2 \cdot 2 \cdot 2 = 8$ Zellen im kreuzfaktoriellen Design im lateinischen Quadrat nur $2 \cdot 2 = 4$ Zellen mit Beobachtungseinheiten zu füllen hat. Der Nachteil besteht darin, daß keine Interaktionshypothesen geprüft werden können.

Zu prüfen ist zunächst die Hypothese, daß überhaupt keine Unterschiede in den Erwartungswertevektoren der P abhängigen Variablen zwischen den vier Zellen bestehen. In der Form $C B A = \Delta$ wählen wir dafür z.B. $A_{A,B,C} = I$, $\Delta_{A,B,C} = \mathbf{O}$ und

$$(3.8.1) \quad C_{A,B,C} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}.$$

Die zweite Hypothese, daß keine Unterschiede in den Erwartungswertevektoren der P abhängigen Variablen zwischen den Stufen A_1 und A_2 bestehen, formulieren wir durch $A_A = I$, $\Delta_A = 0$ und

$$(3.8.2) \quad C_A = [1 \quad 1 \quad -1 \quad -1],$$

die entsprechende Hypothese für B_1 und B_2 durch $A_B = I$, $\Delta_B = 0$ und

$$(3.8.3) \quad C_B = [1 \quad -1 \quad 1 \quad -1],$$

sowie für C_1 und C_2 durch $A_C = I$, $\Delta_C = 0$ und

$$(3.8.4) \quad C_C = [1 \quad -1 \quad -1 \quad 1].$$

3.9 Lateinisches Quadrat über den abhängigen Variablen

Auch die abhängigen Variablen können im „lateinischen Quadrat“ strukturiert werden, sofern sie in mehrere Faktoren U, V und W mit gleicher Stufenzahl gegliedert werden können (siehe Tabelle 3.6). Die Ökonomie des Verfahrens besteht darin, daß wir bei z.B. zweistufigen Faktoren nur $P = 2 \cdot 2 = 4$ abhängige Variablen erheben müssen anstelle von $P = 2 \cdot 2 \cdot 2 = 8$ abhängigen Variablen im vergleichbaren kreuzfaktoriellen Design. Allerdings können auch hier (wie in Abschnitt 3.8) keine Interaktionshypothesen über die abhängigen Variablen selbst geprüft werden, wohl aber über Wechselwirkungen zwischen den unabhängigen und den abhängigen Variablen (vgl. Abschnitt 3.5).

Tabelle 3.6.: Faktoren U, V und W im „lateinischen Quadrat“ als Gliederung der P abhängigen Variablen y_p , welche insgesamt an $N = N_1 + \dots + N_4$ Beobachtungseinheiten erhoben wurden. Beispiel für Faktor U: Meßmethode 1 (U_1) Meßmethode 2 (U_2); Faktor V: vormittags (V_1), nachmittags (V_2); Faktor W: Geschwindigkeit (W_1), Genauigkeit (W_2). Das Datenschema zeigt die $P = 4$ abhängigen Variablen y_p pro Beobachtungseinheit und deren bedingte Erwartungswerte $\hat{\mu}_{qp}$.

Gliederung der abhängigen Variablen

Zellen- nummer q	U ₁		U ₂		Faktor U
	V ₁	V ₂	V ₁	V ₂	Faktor V
	W ₁	W ₂	W ₂	W ₁	Faktor W
	1	2	3	4	Variablen- nummer p
1	y_{111}	y_{112}	y_{113}	y_{114}	
	\vdots	\vdots	\vdots	\vdots	
2	$y_{N_2,11}$	$y_{N_2,12}$	$y_{N_2,13}$	$y_{N_2,14}$	
	$\hat{\mu}_{11}$	$\hat{\mu}_{12}$	$\hat{\mu}_{13}$	$\hat{\mu}_{14}$	
3	y_{121}	y_{122}	y_{123}	y_{124}	
	\vdots	\vdots	\vdots	\vdots	
4	$y_{N_2,21}$	$y_{N_2,22}$	$y_{N_2,23}$	$y_{N_2,24}$	
	$\hat{\mu}_{21}$	$\hat{\mu}_{22}$	$\hat{\mu}_{23}$	$\hat{\mu}_{24}$	
5	y_{131}	y_{132}	y_{133}	y_{134}	
	\vdots	\vdots	\vdots	\vdots	
6	$y_{N_3,31}$	$y_{N_3,32}$	$y_{N_3,33}$	$y_{N_3,34}$	
	$\hat{\mu}_{31}$	$\hat{\mu}_{32}$	$\hat{\mu}_{33}$	$\hat{\mu}_{34}$	
7	y_{141}	y_{142}	y_{143}	y_{144}	
	\vdots	\vdots	\vdots	\vdots	
8	$y_{N_4,41}$	$y_{N_4,42}$	$y_{N_4,43}$	$y_{N_4,44}$	
	$\hat{\mu}_{41}$	$\hat{\mu}_{42}$	$\hat{\mu}_{43}$	$\hat{\mu}_{44}$	

Wir prüfen zunächst die globale Hypothese, daß in jeder Zelle die bedingten Erwartungswerte der P abhängigen Variablen gleich sind. In der Form $\mathbf{C} \mathbf{B} \mathbf{A} = \mathbf{\Delta}$ wählen wir z.B. $\mathbf{C}_{U,V,W} = \mathbf{I}$, $\mathbf{\Delta}_{U,V,W} = \mathbf{O}$ und

$$(3.9.1) \quad \mathbf{A}_{U,V,W} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}.$$

Als zweite Hypothese prüfen wir die Gleichheit der bedingten Erwartungswerte der Stufen U_1 und U_2 mit

$$(3.9.2) \quad \mathbf{A}_U = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}$$

bei $\mathbf{C}_U = \mathbf{I}$ und $\mathbf{\Delta}_U = \mathbf{0}$. Die entsprechenden Hypothesen für V_1 und V_2 bzw. W_1 und W_2 prüfen wir mit

$$(3.9.3) \quad \mathbf{A}_V = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}$$

bzw.

$$(3.9.4) \quad \mathbf{A}_W = \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}$$

bei $\mathbf{C}_V = \mathbf{C}_W = \mathbf{I}$ und $\mathbf{\Delta}_V = \mathbf{\Delta}_W = \mathbf{0}$.

3.10 Zusammenfassende Bemerkungen

In diesem dritten Abschnitt wurde die Hypothesenformulierung in verschiedenen Designs für das Zellenmittelwertmodell bei gleichen Zellenfrequenzen behandelt. Dabei wurden exemplarisch drei Prinzipien zur Strukturierung von Variablen dargestellt, nämlich die kreuzfaktorielle Strukturierung, die hierarchische, und die nach dem „lateinischen Quadrat“. Es wurde gezeigt, wie diese Strukturierungsgesichtspunkte sowohl für die unabhängigen als auch für die abhängigen Variablen verwendet werden können. Dabei wurde z.B. eine wiederholt gemessene Variable formal als weitere abhängige Variable betrachtet, was dann angezeigt ist, wenn die Residuen der wiederholt gemessenen Varia-

blen zwischen den Meßpunkten nicht als unabhängig angenommen werden können (siehe Gleichung 2.3.11).

Weiterführende Literatur: Zur Klassifikation von Designs s. z.B. Bredenkamp (1980), Campbell und Stanley (1966) oder Kirk (1972, 1978). Winer (1971) gibt für viele Designs Rechenformeln an. Andere Arten von Vergleichen oder Kontrasten stellen fast alle der im Abschnitt 2 angegebenen Monographien dar, genauso wie auch andere Arten von Modellen, die anstelle des Zellenmittelwertmodells verwendet werden können.

4. Parameterschätzung

4.1 Einleitung

Die Parameter in der $Q \times P$ -Matrix B eines multivariaten linearen Stichprobenmodells, welches die Gleichungen

$$(4.1.1) \quad E(Y) = X B$$

$$(4.1.2) \quad C(\boldsymbol{\varepsilon}_n, \boldsymbol{\varepsilon}_{n^*}) = C(\mathbf{y}_n, \mathbf{y}_{n^*}) = \begin{cases} \boldsymbol{\Sigma}, & \text{für } n = n^* \\ \mathbf{O}, & \text{für } n \neq n^* \end{cases} \quad n, n^* \in \{1, 2, \dots, N\}$$

erfüllt, wobei $\boldsymbol{\varepsilon}_n := \mathbf{y}_n - \mathbf{x}_n B$, sowie die Parameter in der $P \times P$ -Kovarianzmatrix $\boldsymbol{\Sigma}$ können prinzipiell auf verschiedene Arten geschätzt werden. Keine der verschiedenen Schätzmethoden ist für alle Situationen optimal, und es würde den Rahmen des Artikels sprengen, wenn wir versuchen würden, alle Schätzmethoden darzustellen und kritisch zu bewerten. Statt dessen beschränken wir uns auf die Darstellung der Schätzung nach dem *Kriterium der kleinsten Quadrate ohne* (Abschnitt 4.2) *und mit Nebenbedingungen* (Abschnitt 4.3) sowie auf die *Maximum-Likelihood-Schätzung* unter der Annahme, daß die $\boldsymbol{\varepsilon}_n$ jeweils P -variater normalverteilt sind, mit gleicher Kovarianzmatrix $\boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}$ für alle $n \in \{1, 2, \dots, N\}$ (Abschnitt 4.4). Die Erwartungswerte und die Kovarianzmatrix der Schätzer von B und $\boldsymbol{\Psi} := CBA$ sind Gegenstand des Abschnitts 4.5, ebenso wie die Schätzung der Kovarianzmatrix $\boldsymbol{\Sigma}$. Abschließend geben wir einige Hinweise auf weiterführende Literatur zu anderen Schätzmethoden und einigen Problemen, die in Zusammenhang mit der Parameterschätzung stehen. Die Schätztheorie, die in diesem Abschnitt dargestellt wird, ist nicht auf das Zellenmittelwertmodell beschränkt. Sie ist für alle Modelle anwendbar, welche die Gleichungen 4.1.1 und 4.1.2 erfüllen. Darunter fallen auch *Regressionsmodelle* mit „nichtstochastischen Regressoren“.

4.2 Kriterium der kleinsten Quadrate

Die Standardmethode, die Parametermatrix B zu schätzen, ist, den Schätzer \hat{B} so zu wählen, daß die *Spur der Fehlerquadratsummenmatrix*

$$(4.2.1) \quad \begin{aligned} f(\hat{B}) &= \text{Spur} [(Y - X \hat{B})' (Y - X \hat{B})] = \\ &= \text{Spur} [Y'Y - 2 \hat{B}'X'Y + \hat{B}'X'X \hat{B}] \end{aligned}$$

ein Minimum wird (vgl. z.B. Timm, 1975, S. 185). Dies erreichen wir durch Differenzieren der Funktion $f(B)$ nach B (s. Bock, 1975, S. 68-69)

$$(4.2.2) \quad \frac{\partial f(\hat{B})}{\partial \hat{B}} = -2 X'Y + 2 X'X \hat{B}$$

(vgl. z.B. Timm, 1975, S. 100), und Nullsetzen der resultierenden Gleichung:

$$(4.2.3) \quad \frac{\partial f(\hat{B})}{\partial \hat{B}} = O.$$

Einen Schätzer \hat{B} , der diese Bedingung erfüllt, notieren wir mit \hat{B} . Die Bildung der zweiten Ableitung zeigt, daß die Funktion $f(B)$ an der Stelle $\hat{B} = \hat{B}$ tatsächlich ein Minimum aufweist.

Aus den Gleichungen 4.2.2 und 4.2.3 erhalten wir dann die sogenannte Normalgleichung

$$(4.2.4) \quad X'X \hat{B} = X'Y,$$

die unter der Annahme der Nonsingularität von $X'X$ zum Schätzer

$$(4.2.5) \quad \hat{B} = (X'X)^{-1} X'Y$$

führt. Man beachte die analogen Strukturen dieser Gleichung zu der Identifikationsgleichung 2.2.15.

Ist $X'X$ jedoch singular, so kann man dennoch mit der allgemeinen Inversen (vgl. z.B. Rao & Mitra, 1971) $(X'X)^-$, welche die Eigenschaft $X'X (X'X)^- X'X = X'X$ erfüllt, eine Lösung für B angeben, nämlich

$$(4.2.6) \quad \hat{B} = (X'X)^- X'Y + (I - G) Z$$

wobei $G = (X'X)^- X'X$ und Z eine beliebige $Q \times P$ -Matrix ist (vgl. z.B. Timm, 1975, S. 185). Dieser Schätzer ist im allgemeinen jedoch nicht eindeutig, da man Z beliebig wählen kann, d.h. es gibt unendlich viele dieser Schätzer. Erst wenn $X'X$ nonsingular ist, d.h. wenn alle Spaltenvektoren von X linear unab-

hängig sind und damit der Rang R von X gleich der Anzahl Q der Spalten von X ist, wird $\check{\mathbf{B}}$ eindeutig, und Gleichung 4.2.5 ist mit Gleichung 4.2.6 äquivalent. Eine andere Möglichkeit, einen eindeutigen Schätzer zu erhalten, besteht in der Aufstellung von Nebenbedingungen (Restriktionen) über B .

4.3 Kriterium der kleinsten Quadrate unter Nebenbedingungen

In vielen Fällen will man die Parametermatrix \mathbf{B} unter bestimmten Nebenbedingungen schätzen, die man z.B. in der Form $\mathbf{H} \mathbf{B} = \mathbf{\Theta}$ aufschreiben kann.

Dabei setzen wir voraus, daß die Anzahl der Zeilen der Matrix H gleich ihrem Rang ist. Solche Nebenbedingungen können zum einen Restriktionen sein, die erst die Schätzbarkeit von B ermöglichen, wie z.B. $\sum_j \alpha_j = 0$ in der klassischen Varianzanalyse. Zum anderen kann es sich aber auch um solche Restriktionen handeln, die aus inhaltlich theoretischen Gründen gesetzt werden, wie etwa, daß alle Interaktionsparameter gleich Null sind, was auch ermöglicht, daß Interaktions- und Fehlervarianz zu einer neuen Fehlervarianzschätzung „ge-poolt“ werden können.

Zur Minimierung der Funktion $f(B)$ (siehe Gleichung 4.2.1) unter der Nebenbedingung $H B = \mathbf{\Theta}$ führt man eine Matrix $\mathbf{\Lambda}$ von Lagrange-Multiplikatoren ein, so daß die Funktion

$$(4.3.2) \quad g(\check{\mathbf{B}}, \check{\mathbf{\Lambda}}) = \text{Spur} [(\mathbf{Y} - \mathbf{X} \check{\mathbf{B}})' (\mathbf{Y} - \mathbf{X} \check{\mathbf{B}})] + 2 \text{Spur} [\check{\mathbf{\Lambda}}' (\mathbf{H} \check{\mathbf{B}} - \mathbf{\Theta})]$$

zu minimieren ist (vgl. z.B. Timm, 1975, S. 189).

Wir bilden zunächst die partiellen Ableitungen (s. Bock, 1975, S. 68-69)

$$(4.3.2) \quad \frac{\partial [g(\check{\mathbf{B}}, \check{\mathbf{\Lambda}})]}{\partial (\check{\mathbf{B}})} = -2 \mathbf{X}' \mathbf{Y} + 2 \mathbf{X}' \mathbf{X} \check{\mathbf{B}} + 2 \mathbf{H}' \check{\mathbf{\Lambda}}$$

und

$$(4.3.3) \quad \frac{\partial [g(\check{\mathbf{B}}, \check{\mathbf{\Lambda}})]}{\partial (\check{\mathbf{\Lambda}})} = \mathbf{H} \check{\mathbf{B}} - \mathbf{\Theta}.$$

Schätzer $\check{\mathbf{B}}$ und $\check{\mathbf{\Lambda}}$, welche die Bedingungen

$$(4.3.4) \quad \frac{\partial [g(\check{\mathbf{B}}, \check{\mathbf{\Lambda}})]}{\partial (\check{\mathbf{B}})} = \mathbf{O}$$

und

$$(4.3.5) \quad \frac{\partial g(\hat{\mathbf{B}}, \hat{\mathbf{\Lambda}})}{\partial (\hat{\mathbf{\Lambda}})} = \mathbf{0}$$

erfüllen, bezeichnen wir mit $\hat{\mathbf{B}}_H$ bzw. $\hat{\mathbf{\Lambda}}_H$. Die zweite Ableitung zeigt, daß die Funktion $g(\mathbf{B}, \mathbf{\Lambda})$ an der Stelle $\hat{\mathbf{B}} = \hat{\mathbf{B}}_H$ bzw. $\mathbf{\Lambda} = \hat{\mathbf{\Lambda}}_H$ tatsächlich ein Minimum aufweist.

Aus den Gleichungen 4.3.2 bis 4.3.5 erhalten wir

$$(4.3.6) \quad \mathbf{X}'\mathbf{X} \hat{\mathbf{B}}_H = \mathbf{X}'\mathbf{Y} - \mathbf{H}'\hat{\mathbf{\Lambda}}_H$$

und

$$(4.3.7) \quad \mathbf{H} \hat{\mathbf{B}}_H = \mathbf{\Theta}.$$

Man benutzt nun die für ein konsistentes Gleichungssystem $\mathbf{AB} = \mathbf{C}$ (mit der Matrix \mathbf{B} von Unbekannten) gültige allgemeine Lösung

$$(4.3.8) \quad \mathbf{B} = \mathbf{A}^{-1}\mathbf{C} + (\mathbf{I} - \mathbf{A}^{-1}\mathbf{A})\mathbf{Z},$$

wobei \mathbf{Z} eine beliebige Matrix passenden Typs ist (siehe z.B. Timm 1975, S. 55). Auf die Gleichung 4.3.6 angewandt erhält man daraus für $\hat{\mathbf{B}}_H$ die Lösung

$$(4.3.9) \quad \hat{\mathbf{B}}_H = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'\hat{\mathbf{\Lambda}}_H + [\mathbf{I} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}]\mathbf{Z}.$$

Unter Verwendung von Gleichung 4.2.6 folgt dann

$$(4.3.10) \quad \hat{\mathbf{B}}_H = \hat{\mathbf{B}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'\hat{\mathbf{\Lambda}}_H.$$

Vormultiplikation mit $[\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}']^{-1}\mathbf{H}$ und Einsetzen von Gleichung 4.3.7 ergibt

$$(4.3.11) \quad \hat{\mathbf{\Lambda}}_H = [\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}']^{-1}(\mathbf{H}\hat{\mathbf{B}} - \mathbf{\Theta}).$$

Aus den Gleichungen 4.3.10 und 4.3.11 erhalten wir dann

$$(4.3.12) \quad \hat{\mathbf{B}}_H = \hat{\mathbf{B}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'[\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}']^{-1}(\mathbf{H}\hat{\mathbf{B}} - \mathbf{\Theta}).$$

Hat Gleichung 4.3.12 eine eindeutige Lösung, so ist $\hat{\mathbf{B}}_H$ der Kleinste-Quadrate-Schätzer für die Parametermatrix \mathbf{B} unter der Nebenbedingung $\mathbf{HB} = \mathbf{\Theta}$ (vgl. z.B. Searle, 1971, S. 204ff. oder Timm, 1975, S. 178).

4.4 Maximum-Likelihood-Kriterium

Bei den Schätzern nach dem Kriterium der kleinsten Quadrate müssen außer der Gleichheit der Kovarianzmatrizen (siehe Gleichung 4.1.2) keine Verteilungsannahmen getroffen werden, es sei denn, zum Nachweis einiger Optimalitätseigenschaften. Ist man jedoch zum Beispiel in der Lage anzunehmen, daß die Dichtefunktion der N Vektoren $\boldsymbol{\varepsilon}_n$ der Residualvariablen eine P -variate Normalverteilung mit den $P \times P$ -Kovarianzmatrizen $\boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}$ ist, daß die $\boldsymbol{\varepsilon}_n$ stochastisch unabhängig sind, daß $N \geq P+Q$ und daß der Rang R von X gleich Q , der Anzahl ihrer Spalten, ist, dann führt die Maximierung der Likelihoodfunktion

$$L(\hat{\mathbf{B}}) = 2 \pi^{-(1/2)PN} |\boldsymbol{\Sigma}|^{-(1/2)N} \exp \left[-\frac{1}{2} \sum_{n=1}^N (\mathbf{y}_n - \mathbf{x}_n \hat{\mathbf{B}})' \boldsymbol{\Sigma}^{-1} (\mathbf{y}_n - \mathbf{x}_n \hat{\mathbf{B}}) \right]$$

bezüglich $\hat{\mathbf{B}}$ zu dem Maximum-Likelihood-Schätzer

$$(4.4.1) \quad \hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$$

(vgl. z.B. Anderson, 1958, S. 181), der also unter den genannten Bedingungen mit dem Kleinste-Quadrate-Schätzer identisch ist.

4.5 Erwartungswerte- und Kovarianzmatrix der Parametervektoren β_p und ψ_k

Wir zeigen zunächst, daß der Schätzer

$$(4.5.1) \quad \hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$$

ein *erwartungstreuer Schätzer* von \mathbf{B} ist:

$$(4.5.2) \quad \begin{aligned} E(\hat{\mathbf{B}}) &= E [(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}] = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'E(\mathbf{Y}) = \\ &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X} \mathbf{B} = \mathbf{B}, \end{aligned}$$

wobei wir lediglich $E(\mathbf{Y}) = \mathbf{X} \mathbf{B}$ (siehe Gleichung 4.1.1) verwendet haben. Entsprechend gilt für den Schätzer

$$(4.5.3) \quad \hat{\boldsymbol{\Psi}} = \mathbf{C}\hat{\mathbf{B}}\mathbf{A}$$

die Gleichung

$$(4.5.4) \quad E(\hat{\boldsymbol{\Psi}}) = E(\mathbf{C}\hat{\mathbf{B}}\mathbf{A}) = \mathbf{C}\mathbf{B}\mathbf{A}.$$

Die *Kovarianzmatrix* von einem Spaltenvektor β_p der Parametermatrix \hat{B} ist

$$(4.5.5) \quad \begin{aligned} C(\hat{\beta}_p, \hat{\beta}_p) &= C[(X'X)^{-1} X'y_p, (X'X)^{-1} X'y_p] = \\ &= (X'X)^{-1} X' C(y_p, y_p) X (X'X)^{-1} = \\ &= \sigma_p^2 (X'X)^{-1}, \end{aligned}$$

wobei y_p die p -te Spalte der Matrix Y ist. Die Kovarianzmatrix $C(y_p, y_p)$ ist nämlich die $N \times N$ -Diagonalmatrix $\sigma_p^2 I$ und $\sigma_p^2 = \sigma_{pp}$ die Varianz von y_p , der p -ten, Komponente des Vektors der abhängigen Variablen. Die Matrix $C(\hat{\beta}_p, \hat{\beta}_p)$ wird z.B. für die Bestimmung von Konfidenzintervallen der β_{qp} benötigt (siehe Abschnitt 5.6).

Für die *Kovarianzmatrix* eines $M \times 1$ -Spaltenvektors $\hat{\psi}_k = C\hat{B}a_k$ welche z.B. für die Angabe von Konfidenzintervallen bestimmter Linearkombinationen von B benötigt wird, gilt

$$(4.5.6) \quad \begin{aligned} C(\hat{\psi}_k, \hat{\psi}_k) &= C[C\hat{B}a_k, C\hat{B}a_k] = C C(\hat{B}a_k, \hat{B}a_k) C' = \\ &= C(X'X)^{-1} X' [C(Ya_k, Ya_k)] X (X'X)^{-1} C' = \\ &= C(X'X)^{-1} X' [V(y_{n a_k}) I] X (X'X)^{-1} C' = \\ &= V(y_{n a_k}) C(X'X)^{-1} C', \end{aligned}$$

wobei $V(y_{n a_k}) I$ eine $N \times N$ -Diagonalmatrix und $V(y_{n a_k})$ die Varianz von $y_{n a_k}$ ist, die wegen der Annahme 4.1.2 für alle N Zeilen von Y gleich ist. Für die N Varianzen $V(y_{n a_k})$ gilt dabei

$$(4.5.7) \quad V(y_{n a_k}) = C(y_{n a_k}, y_{n a_k}) = a'_k C(y_{n_1}, y_{n_1}) a_k = a'_k \Sigma a_k.$$

Die $P \times P$ -Kovarianzmatrix Σ kann durch die folgenden Schätzer erwartungstreu geschätzt werden:

$$(4.5.8) \quad \begin{aligned} \hat{\Sigma} &= \frac{1}{N-R} (Y - X\hat{B})'(Y - X\hat{B}) = \\ &= \frac{1}{N-R} Y'[I - X(X'X)^{-1}X']Y \end{aligned}$$

(siehe z.B. Timm, 1975, S. 186).

4.6 Zusammenfassende Bemerkungen

In diesem Abschnitt wurden die Standardschätzmethoden für die Parameter des Modells $E(Y) = X B$ dargestellt. Dabei ist insbesondere die in Abschnitt 4.3 angegebene Schätzmethode nach dem Kriterium der kleinsten Quadrate unter einer Nebenbedingung auch in solchen Fällen anwendbar, in denen die Matrix X keinen vollen Spaltenrang hat. Dies ist z.B. bei klassischen varianz-analytischen Modellen der Fall, bei denen man diesen „Mangel“ durch Nebenbedingungen wie z.B. $\sum \alpha_j = 0$ „beheben“ muß. Ein Einblick in die Komplexität des Themas „Schätzmethoden“ soll andeutungsweise in den Bemerkungen zur weiterführenden Literatur vermittelt werden.

Weiterführende Literatur: Grundlage für eine kritische Bewertung verschiedener Schätzmethoden sind Gütekriterien für Schätzer, wie Erwartungstreue, Effizienz, Robustheit etc. Eine gut lesbare Einführung ist z.B. Burkholder (1978, S. 251ff.). Als anspruchsvolle und ausführliche Arbeit sei dazu Zacks (1971) genannt. Andrews et al. (1972) geben einen Überblick zum Thema „Robustheit“.

Andere Schätzmethoden als die hier dargestellten werden z.B. in Bandemer et al. (1977) dargestellt. Eine wichtige Alternative ist z.B. die *verallgemeinerte Methode der kleinsten Quadrate*, die dann angewandt werden kann, wenn die Kovarianzmatrizen $C(y_n, y_n)$ nicht mehr für alle N Beobachtungseinheiten als gleich angenommen werden können (vgl. z.B. Searle, 1971, S. 87 oder Humak, 1977, S. 30). Eine weitere wichtige Schätzmethode ist die *Bayer-Schätzung*, die vor allem dann gewinnbringend verwendet werden kann, wenn Vorwissen über die Parameter besteht. Dazu verweisen wir auf Bandemer et al. (1977), Humak (1977), Leonhard (1975) sowie Lindley und Smith (1972).

Bei der *Stein'schen Schätzung* besteht der Vorteil darin, daß extreme Werte („Ausreißer“) den Schätzer weniger stark beeinflussen als beim Kriterium der kleinsten Quadrate (vgl. Bandemer et al., 1977, S. 111), was man auch mit der *L_p -Approximation* (vgl. z.B. Späth, 1973, 1974) erreichen kann, bei der das Minimierungskriterium nicht unbedingt die Summe der Abweichungsquadrate (L_2), sondern z.B. die der absoluten Beträge der Abweichungen (L_1) ist. Schließlich sei noch auf Harter (1974-1975) hingewiesen, der eine ausführliche Bibliographie über das „Kriterium der kleinsten Quadrate und einige Alternativen“ (Übersetzung durch die Autoren) liefert.

5. Hypothesenbewertung

5.1 Einleitung

In Abschnitt 3 haben wir ausführlich für verschiedene Designs gezeigt, wie man Hypothesen über die Parameter in der Matrix B des multivariaten linearen Stichprobenmodells

$$(5.1.1) \quad E(\mathbf{Y}) = \mathbf{XB}$$

formulieren kann, und zwar in der Form der allgemeinen multivariaten linearen Hypothese

$$(5.1.2) \quad H_0: \boldsymbol{\Psi} := \mathbf{CBA} = \boldsymbol{\Delta}$$

mit der $M \times Q$ -Matrix \mathbf{C} , der $Q \times P$ -Matrix \mathbf{B} , der $P \times K$ -Matrix \mathbf{A} und der $M \times K$ -Matrix $\boldsymbol{\Delta}$, wobei \mathbf{C} den Rang $M \leq R$ und \mathbf{A} den Rang $K \leq P \leq R$ hat und R der Rang der $N \times Q$ -Matrix \mathbf{X} ist. Dort hatten wir immer vorausgesetzt, daß die Parametermatrix \mathbf{B} die bedingten Erwartungswerte der P abhängigen Variablen in den Q Zellen enthält (siehe Abschnitt 3.2). Diese Voraussetzung brauchen wir hier nicht zu machen, d.h. die in diesem Abschnitt dargestellten Methoden zur Bewertung der Hypothese 5.1.2 gelten für alle Modelle, welche die Gleichung 5.1.1 und

$$(5.1.3) \quad C(y_n, y_{n^*}) = C(\boldsymbol{\varepsilon}_n, \boldsymbol{\varepsilon}_{n^*}) = \begin{cases} \boldsymbol{\Sigma}, & \text{für } n = n^* \\ \mathbf{O}, & \text{für } n \neq n^* \end{cases} \quad n, n^* \in \{1, 2, \dots, N\},$$

erfüllen, wobei $\boldsymbol{\varepsilon}_n = y_n - \mathbf{x}_n \mathbf{B}$, und bei denen $\boldsymbol{\Psi} := \mathbf{CBA}$ schätzbar ist (vgl. z.B. Timm, 1975, S. 173). Auch der Abschnitt über Parameterschätzung war nicht auf das Zellenmittelwertmodell beschränkt.

Für den multivariaten Fall wurden eine Reihe von Testkriterien vorgeschlagen, von denen wir die vier wichtigsten darstellen. Dabei folgen wir der Auswahl von Olson (1976), der auch eine ausführliche Diskussion über die Vor- und Nachteile der Kriterien in verschiedenen Situationen liefert (vgl. auch Stevens, 1979). Bei den Verteilungen aller vier Testkriterien wird die P -variante Normalverteilung der Zeilenvektoren y_n , $n \in \{1, 2, \dots, N\}$, und deren Unabhängigkeit voneinander (siehe Gleichung 5.1.3) vorausgesetzt, und bei allen vier Testkriterien spielen die beiden Matrizen Q_e und Q_h eine wesentliche Rolle, die wir im folgenden Absatz behandeln.

Die Matrix

$$(5.1.4) \quad \begin{aligned} Q_e &= (\mathbf{YA} - \mathbf{XB}\hat{\mathbf{A}})'(\mathbf{YA} - \mathbf{XB}\hat{\mathbf{A}}) = \\ &= \mathbf{A}'\mathbf{Y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{YA}, \end{aligned}$$

die wir *Fehlerquadratsummenmatrix* nennen, folgt unter den oben angegebenen Voraussetzungen der zentralen K -variaten Wishart-Verteilung $W_x(N-R, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$ (siehe z.B. Johnson & Kotz, 1972. S. 158).

Die Matrix

$$(5.1.5) \quad Q_h = (\mathbf{C}\hat{\mathbf{B}}\mathbf{A} - \boldsymbol{\Delta})' [\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1} (\mathbf{C}\hat{\mathbf{B}}\mathbf{A} - \boldsymbol{\Delta}),$$

die wir *Hypothesenquadratsummenmatrix* nennen, folgt ebenfalls der zentralen K-variaten Wishart-Verteilung $W_K(M, \mathbf{A}\Sigma\mathbf{A}')$, falls $\Psi = \mathbf{CBA} = \mathbf{\Delta}$, d.h. wenn die Nullhypothese richtig ist. Zur Ableitung dieser Gleichung siehe z.B. Timm (1975, S. 183).

5.2 Wilks' Lambda-Kriterium

Unter den in Abschnitt 5.1 angegebenen Voraussetzungen folgt das *Lambda-Kriterium* von Wilks (1932)

$$(5.2.1) \quad \Lambda := \frac{|\mathbf{Q}_c|}{|\mathbf{Q}_e + \mathbf{Q}_h|}$$

der zentralen Wilks' Lambda-Verteilung $\Lambda(K, M, N-R)$, falls $\mathbf{CBA} = \mathbf{\Delta}$. Zum Beweis dieser Behauptung verweisen wir auf Timm (1975, S. 189ff), bei dem man auch eine Tabelle für die Verteilung von Λ findet.

Eine häufig anstelle von Λ verwendete Statistik, die auf Rao zurückgeht (vgl. Rao, 1973, S. 556), ist

$$(5.2.2) \quad F = \frac{1 - \Lambda^{1/a}}{\Lambda^{1/a}} \cdot \frac{ab - 2c}{K \cdot M},$$

die unter den genannten Bedingungen approximativ der zentralen F-Verteilung folgt mit den Freiheitsgraden $K \cdot M$ und $ab - 2c$, wobei

$$(5.2.3) \quad a = \begin{cases} 1, & \text{falls } K \cdot M = 2 \\ \sqrt{\frac{K^2 \cdot M^2 - 4}{K^2 + M^2 - 5}}, & \text{andernfalls,} \end{cases}$$

sowie

$$(5.2.4) \quad b = N - R + M - \frac{K + M + 1}{2}$$

und

$$(5.2.5) \quad c = \frac{K \cdot M - 2}{4}$$

sind. Wenn $(ab - 2c)$ keine ganze Zahl ist, kann man den ganzzahligen Anteil verwenden, um „konservativer“ zu testen.

Über die Berechnung von Λ als Quotient von verallgemeinerten Varianzen, wesentliche Eigenschaften von Λ , sowie eine approximativ X-verteilte Prüfstatistik nach Bartlett (1947) berichtet Moosbrugger (1978, S. 114-124).

5.3 Roy's Eigenwert-Kriterium

Roy (z.B. 1957) hat den größten Eigenwert der Matrix $(Q_e + Q_h)^{-1} Q_h$, nämlich den Wert ϑ_{\max} , welcher die Determinantengleichung

$$(5.3.1) \quad \left| (Q_e + Q_h)^{-1} Q_h - \vartheta I \right| = 0$$

erfüllt (vgl. Moosbrugger, 1978, S. 45f.), als Testkriterium für die Nullhypothese

$$(5.3.2) \quad H_0: \bigcap_v (\mathbf{CBA}v = \Delta v)$$

vorgeschlagen, die nur dann wahr ist, wenn für jeden beliebigen K-dimensionalen Vektor v die Gleichung $\mathbf{CBA}v = \Delta v$ gilt (vgl. z.B. Timm, 1975, S. 192).

Die Verteilung dieses Testkriteriums ϑ_{\max} wurde von Heck (1960) und Pillai (1960) tabelliert mit den Argumenten

$$(5.3.3) \quad I = \min (K, M),$$

$$(5.3.4) \quad a = \frac{|K-M| - 1}{2}$$

und

$$(5.3.5) \quad b = \frac{N-R-K-1}{2} .$$

Dabei benutzt Pillai die Buchstaben s für I , m für a und n für b . Tabellen für dieses Testkriterium findet man aber auch bei Morrison (1967) und Timm (1975).

5.4 Hotelling-Lawley Spur Kriterium

Ein anderes Testkriterium, das auf Lawley (1938) und Hotelling (1951) zurückgeht, ist

$$(5.4.1) \quad U = \text{Spur} [Q_h Q_e^{-1}] = \sum_{i=1}^I \lambda_i ,$$

wobei $I = \min (K, M)$ und λ_i der i -te Eigenwert der Matrix $Q_h Q_e^{-1}$ ist. Pillai (1960) hat die zentrale Verteilung dieser Statistik abgeleitet und tabelliert, ebenfalls mit den Argumenten I , a und b (siehe Gleichungen 5.3.3 bis 5.3.5), wobei Pillai auch hier die Buchstaben s für I , m für a und n für B benutzt. Auch bei Timm (1975) findet man eine Tabelle für diese Statistik.

5.5 Pillai-Bartlett Spur Kriterium

Ein weiteres Testkriterium, das *Pillai-Bartlett-Spur-Kriterium*, ist definiert als

$$(5.5.1) \quad V = I - \text{Spur} [Q_e (Q_h + Q_e)^{-1}] = \sum_{i=1}^I (1 - \kappa_i),$$

wobei $I = \min(K, M)$ und κ_i der i -te Eigenwert der Matrix $Q_e (Q_h + Q_e)^{-1}$ ist. Eine ausführliche Tabelle dieser Statistik findet man in Mijares (1964). Weniger umfangreiche Tabellen sind in Pillai (1960), Pillai und Jayachandran (1967, 1970), Pillai und Mijares (1959) sowie Timm (1975) abgedruckt.

Pillai (1960, S. 19) hat außerdem eine F-Approximation von V angegeben, nämlich

$$(5.5.2) \quad F = \frac{(N-R-K+I) \cdot V}{T \cdot (I-V)},$$

wobei $T = \max(K, M)$. Diese Statistik ist annähernd zentral F-verteilt mit $J \cdot T$ und $I \cdot (N-R-K+I)$ Freiheitsgraden.

Nach Olson (1976, S. 584) ist diese Approximation hinreichend genau und nur bei kleinen Freiheitsgraden leicht konservativ.

5.6 Einfache Konfidenzintervalle

In diesem Abschnitt werden einfache Konfidenzintervalle behandelt, im Unterschied zu simultanen, über welche z.B. Gabriel (1968, 1969 a, b) informiert.

5.6.1 Konfidenzintervall für einen Parameter

Aus den in Abschnitt 5.1 aufgeführten Annahmen folgt, daß

$$(5.6.1) \quad \frac{\hat{\beta}_{qp} - \beta_{qp}}{\sqrt{V(\hat{\beta}_{qp})}} \sim N(0,1)$$

mit Erwartungswert Null und Varianz Eins normalverteilt ist, wobei die Varianz $V(\hat{\beta}_{qp})$ das q -te Element der Hauptdiagonalen der Kovarianzmatrix

$$(5.6.2) \quad C(\hat{\beta}_p, \hat{\beta}_p) = \sigma_p^2 (\mathbf{X}'\mathbf{X})^{-1}$$

(siehe Gleichung 4.5.5) ist (vgl. z.B. Searle, 1971, S. 107). Außerdem folgt aus diesen Annahmen, daß

$$(5.6.3) \quad \frac{\hat{\beta}_{qp} - \beta_{qp}}{\sqrt{\hat{V}(\hat{\beta}_{qp})}} \sim t(N-R)$$

t-verteilt ist mit N-R Freiheitsgraden, wobei R der Rang der Matrix X ist und $\hat{V}(\hat{\beta}_{qp})$ der Schätzer von $V(\beta_{qp})$. Den Schätzer $\hat{V}(\hat{\beta}_{qp})$ erhalten wir als q-tes Element der Hauptdiagonalen von

$$(5.6.4) \quad \hat{C}(\hat{\beta}_p, \hat{\beta}_p) = \hat{\sigma}_p^2 (\mathbf{X}'\mathbf{X})^{-1},$$

wobei $\hat{\sigma}_p^2$ das p-te diagonale Element der Matrix $\hat{\Sigma}$ (siehe Gleichung 4.5.8) ist.

Ein Konfidenzintervall für β_{qp} konstruieren wir auf folgende Weise: Es sei $\pm t(N-R, \alpha/2)$ der untere bzw. der obere Punkt der t-Verteilung, für welche

$$(5.6.5) \quad P[-t(N-R, \alpha/2) \leq t \leq t(N-R, \alpha/2)] = 1 - \alpha$$

die Wahrscheinlichkeit ist, daß ein t-verteilter Wert in dem angegebenen Intervall liegt. Wendet man dies auf 5.6.3 an, so erhält man

$$(5.6.6) \quad P[-t(N-R, \alpha/2) \leq \frac{\hat{\beta}_{qp} - \beta_{qp}}{\sqrt{\hat{V}(\hat{\beta}_{qp})}} \leq t(N-R, \alpha/2)] = 1 - \alpha.$$

Daraus ergeben sich die Grenzen des symmetrischen Konfidenzintervalles

$$(5.6.7) \quad \begin{aligned} \hat{\beta}_{qp} - t(N-R, \alpha/2) \sqrt{\hat{V}(\hat{\beta}_{qp})} &\leq \beta_{qp} \leq \\ &\leq \hat{\beta}_{qp} + t(N-R, \alpha/2) \sqrt{\hat{V}(\hat{\beta}_{qp})} \end{aligned}$$

für den unbekannt Parameter β_{qp} , in dem der Parameter β_{qp} mit Wahrscheinlichkeit $1 - \alpha$ liegt.

Entsprechend kann auch für jede Linearkombination $\mathbf{c}'\hat{\beta}_p$ der Koeffizienten eines Vektors $\hat{\beta}_p$ (das ist eine Spalte der $Q \times P$ -Parametermatrix B) ein Konfidenzintervall angegeben werden, dessen Grenzen wir durch

$$(5.6.8) \quad \mathbf{c}'\hat{\beta}_p \pm t(N-R, \alpha/2) \cdot \sqrt{\hat{V}(\mathbf{c}'\hat{\beta}_p)},$$

finden, wobei

$$(5.6.9) \quad \hat{V}(\mathbf{c}'\hat{\beta}_p) = \hat{\sigma}_p^2 \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}$$

(siehe Gleichung 5.6.4).

Dabei kann \mathbf{c}' z.B. eine Zeile der Matrix C der allgemeinen multivariaten linearen Hypothese $H_0 : \mathbf{\Psi} = \mathbf{CBA} = \mathbf{\Delta}$ sein, oder auch eine Zeile \mathbf{x}_n in der Matrix X des Modells $\mathbf{E}(\mathbf{Y}) = \mathbf{XB}$. Im letzteren Fall ergibt 5.6.8 ein Konfidenzintervall für den Erwartungswert

$$(5.6.10) \quad E(\hat{y}_{np}) = E(\mathbf{x}_n \hat{\beta}_p)$$

des vorhergesagten Wertes \hat{y}_{np} der Beobachtungseinheit n auf der Variablen y_p .

5.6.2 Konfidenzintervall für Realisationen einer abhängigen Variablen y_p

Von größerem Interesse dürfte allerdings sein, in welchem Intervall eine Realisation der Beobachtungseinheit n in der abhängigen Variablen y_p liegt. Dieses Konfidenzintervall wird nach den folgenden Überlegungen konstruiert. Man geht aus von der Differenz

$$(5.6.11) \quad \hat{y}_{np} - y_{np} = \hat{y}_{np} - (\mathbf{x}_n \boldsymbol{\beta}_p + e_{np}),$$

welche die Varianz

$$(5.6.12) \quad V(\hat{y}_{np} - y_{np}) = V(\hat{y}_{np}) + V(e_{np})$$

hat, da $\mathbf{x}_n \boldsymbol{\beta}_p$ eine Konstante und $C(\hat{y}_{np}, e_{np}) = 0$ ist. Die stochastische Variable

$$(5.6.13) \quad \frac{\hat{y}_{np} - y_{np} - E(\hat{y}_{np} - y_{np})}{\sqrt{\hat{V}(\hat{y}_{np} - y_{np})}} = \frac{\hat{y}_{np} - y_{np}}{\sqrt{\hat{V}(\hat{y}_{np} - y_{np})}} \sim t(N-R)$$

ist wiederum t -verteilt mit $N-R$ Freiheitsgraden, wobei N die Anzahl der Beobachtungseinheiten und R der Rang der Matrix \mathbf{X} ist.

Analog zu dem oben für β_{qp} angestellten Überlegungen (siehe Gleichung 5.6.7) erhalten wir wieder

$$(5.6.14) \quad P\left[-t(N-R, \alpha/2) \leq \frac{\hat{y}_{np} - y_{np}}{\sqrt{\hat{V}(\hat{y}_{np} - y_{np})}} \leq +t(N-R, \alpha/2)\right] = 1 - \alpha.$$

Daraus ergibt sich das symmetrische Konfidenzintervall

$$(5.6.15) \quad \hat{y}_{np} - t(N-R, \alpha/2) \cdot \sqrt{\hat{V}(\hat{y}_{np} - y_{np})} \leq y_{np} \leq \hat{y}_{np} + t(N-R, \alpha/2) \cdot \sqrt{\hat{V}(\hat{y}_{np} - y_{np})},$$

in dem mit Wahrscheinlichkeit $1 - \alpha$ eine Realisation von y_{np} (ein noch unbeobachteter Wert der Beobachtungseinheit n auf der p -ten abhängigen Variablen) liegt.

Die dazu benötigte Varianzschätzung finden wir gemäß der Gleichung 5.6.12, wobei wir $V(\hat{y}_{np})$ nach Gleichung 5.6.9 (mit $c' = \mathbf{x}_n$) schätzen und $\mathbf{V}(e_{np})$ durch $\hat{\sigma}_p^2$, dem p -ten Element der Diagonalmatrix $\hat{\boldsymbol{\Sigma}}$ (siehe Gleichung 4.5.8), als

$$(5.6.16) \quad \hat{V}(\hat{y}_{np} - y_{np}) = \hat{\sigma}_p^2 [1 + \mathbf{x}_n (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}'_n].$$

5.7 Praktische Signifikanz

Der in Abschnitt 2.2 (Gleichung 2.2.24) angegebene multivariate Determinationskoeffizient kann als praktisches Signifikanzmaß dienen und, falls $C \mathbf{1} = 0$, $\mathbf{1}'\mathbf{A} = 0'$ und $\mathbf{\Delta} = 0$ sind mit $\mathbf{1}$ bzw. $\mathbf{1}'$ als „passenden“ Einheitsvektoren, durch

$$(5.7.1) \quad R_H^2 = \frac{\text{Spur}[\mathbf{Q}_H \mathbf{Q}_T^{-1}]}{P}$$

geschätzt werden, mit der Hypothesenquadratsummenmatrix

$$(5.7.2) \quad \mathbf{Q}_H = (\mathbf{C}\hat{\mathbf{B}}\mathbf{A} - \mathbf{\Delta})' [\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1} (\mathbf{C}\hat{\mathbf{B}}\mathbf{A} - \mathbf{\Delta})$$

und der Totalquadratsummenmatrix

$$(5.7.3) \quad \mathbf{Q}_T = (\mathbf{Y}\mathbf{A} - \bar{\mathbf{Y}}\mathbf{A})' (\mathbf{Y}\mathbf{A} - \bar{\mathbf{Y}}\mathbf{A}) = \mathbf{A}'\mathbf{Y}'\mathbf{Y}\mathbf{A} - \mathbf{A}'\bar{\mathbf{Y}}'\bar{\mathbf{Y}}\mathbf{A},$$

wobei

$$(5.7.4) \quad \bar{\mathbf{Y}} = \frac{1}{N} \mathbf{1} \mathbf{1}' \mathbf{Y}$$

eine $N \times P$ -Matrix ist, welche in jeder Zeile die Gesamtmittelwerte der P abhängigen Variablen über alle N Beobachtungseinheiten als Komponenten enthält. $\mathbf{1}$ bezeichnet dabei den Einheitsspaltenvektor, der aus N Einsen besteht.

$R_H^2 \cdot 100\%$ gibt bei konstant gehaltenem $\mathbf{A} = \mathbf{I}$ an, wieviel Varianzprozent der abhängigen Variablen durch die Variationsquelle, auf welche sich die betreffende Hypothese bezieht, determiniert sind. Man beachte, daß man mit $\mathbf{A} \neq \mathbf{I}$ in der multivariaten allgemeinen linearen Hypothese $\mathbf{C}\hat{\mathbf{B}}\mathbf{A} - \mathbf{\Delta} = 0$ einen neuen Vektor abhängiger Variablen

$$(5.7.5) \quad \mathbf{w} = \mathbf{y} \mathbf{A}$$

bildet.

5.8 Zusammenfassende Bemerkungen

In diesem Abschnitt wurden vier Kriterien für multivariate Signifikanztests dargestellt, die alle invariant bezüglich linearer Transformationen der Ausgangsdaten sind. Im Gegensatz zum univariaten Fall, in dem der F-Test der gleichmäßig beste (uniformly most powerful) ist, gibt es im multivariaten Fall keine Teststatistik mit dieser Optimalitätseigenschaft. Statt dessen hängt die Teststärke der Kriterien davon ab, ob sich die Mittelwerteunterschiede zwischen den Gruppen auf *eine* Dimension konzentrieren („concentrated noncen-

trality structure“ (Olson, 1976, S. 581)) oder ob sie über mehrere Dimensionen der abhängigen Variablen verstreut („diffuse noncentrality structure“, ebenda) sind. Im ersten Fall ist Roys Kriterium 6 des größten Eigenwertes und im letzten Fall das Pillai-Bartlett-Spur-Kriterium V die Statistik mit der größten Teststärke. Für eine ausführlichere Diskussion verweisen wir auf Olson (1976) und Stevens (1979).

Ein Signifikanztest sollte immer von der Berechnung eines Maßes der Effektgröße begleitet sein. Bredenkamp (1970) diskutiert univariate „Maße der praktischen Signifikanz“. Für den multivariaten Fall gaben wir den multivariaten Determinationskoeffizienten als praktisches Signifikanzmaß an, welcher auf den Überlegungen von Cramer und Nicewander (1979), Shaffer und Gillo (1974) sowie Stevens (1972) beruht. Die erörterten Konfidenzintervalle für die Modellparameter und für prädierte Werte der abhängigen Variablen liefern nützliche Informationen zur Bemessung der statistischen Bedeutsamkeit von Zusammenhängen und Unterschieden, die sich in den Modellparametern β_{qp} niederschlagen.

Weiterführende Literatur: Die multivariaten Testkriterien werden in den Standardwerken zur multivariaten Analyse dargestellt (vgl. z.B. Bock, 1975, Finn, 1974, oder Timm, 1975). Eine wichtige Ergänzung zu Signifikanztests sind Verfahren multipler Vergleiche. Wir verweisen dazu auf die Monographie von Miller (1966) sowie auf Carmer und Swanson (1973), Einot und Gabriel (1975), Gabriel (1964, 1969a,b), Garmes (1971a,b, 1973), Petrinovich und Hardyck (1969), Ryan (1959) und Scheffé (1953). Ein nützlicher Sammelband für alle mit der Varianzanalyse zusammenhängenden Probleme ist Krishnaiah (1980).

Literatur

- Ahrens & Läuter, 1981. Mehrdimensionale Varianzanalyse. Akademie-Verlag Berlin.
- Aitkin, M. A. 1969. Multiple comparisons in psychological experiments. *British Journal of Mathematical and Statistical Psychology*, 22, 193-198.
- Anderson, T. W. 1958. An introduction to multivariate statistical analysis. New York: Wiley.
- Andrews, D. F. et al. 1972. Robust estimates of location: Survey and advances: Princeton: University Press.
- Appelbaum, M. I. & Cramer, E. M. 1974. Some problems in the nonorthogonal analysis of variance. *Psychological Bulletin*, 81, 335-343.
- Bandemer, H. et al. 1977. Theorie und Anwendung der optimalen Versuchsplanung I. Berlin: Akademie-Verlag.
- Bartlett, M. S. 1947. Multivariate analysis. *Journal of the Royal Statistical Society. Serie B*, 9, 176-197.

- Bauer, H. 1974. *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie*. Berlin: de Gruyter.
- Bock, R. D. 1975. *Multivariate statistical methods in behavioral research*. New York: McGraw Hill.
- Bortz, J. 1977. *Lehrbuch der Statistik*. Berlin: Springer.
- Bredenkamp, J. 1970. über Maße der praktischen Signifikanz. *Zeitschrift für Psychologie*, 177, 275-285.
- Bredenkamp, J. 1980. *Theorie und Planung psychologischer Experimente*. Darmstadt: Steinkopff.
- Burkholder, D. L. 1978. Estimation: Point estimation. In: Kruskal, W. H. & Tanur, J. M. (Hrsg.), 251-259.
- Campbell, D. T. & Stanley, J. C. 1966. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Carmer, S. G. & Swanson, M. R. 1973. Evaluation of ten pairwise multiple comparison procedures by Monte-Carlo methods. *Journal of the American Statistical Association*, 68, 66-74.
- Cohen, J. 1968. Multiple regression as a general data analytic system. *Psychological Bulletin*, 70, 426-443.
- Cohen, J. & Cohen, P. 1975. *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, N. J.: Lawrence Erlbaum Associates.
- Cramer, E. M. & Nicewander, W. A. 1979. Some symmetric, invariant measures of multivariate association. *Psychometrika*, 44, 43-54.
- Edwards, A. L. 1971. *Versuchsplanung in der psychologischen Forschung*. Weinheim: Beltz.
- Edwards, A. L. 1976. *An introduction to linear regression and correlation*. San Francisco: Freeman.
- Einot, J. & Gabriel, K. R. 1975. A study of the powers of several methods of multiple comparisons. *Journal of the American Statistical Association*, 70, 574-583.
- Ezekiel, M. & Fox, K. A. 1959. *Methods of correlation and regression analysis*. New York: Wiley.
- Finn, J. D. 1974. *A general model for multivariate analysis*. New York: Holt, Rinehart and Winston.
- Fisher, R. A. 1925. *Statistical methods for research workers*. London: Oliver & Boyd.
- Fisher, R. A. 1935. *The design of experiments*. London: Oliver & Boyd.
- Fraser, D. A. S. 1979. *Inference and linear models*. New York: McGraw Hill.
- Gabriel, K. R. 1964. A procedure for testing the homogeneity of all sets of means in analysis of variance. *Biometrics*, 20, 459-477.
- Gabriel, K. R. 1968. Simultaneous test procedures in multivariate analysis of variance. *Biometrika*, 55, 489-504.

- Gabriel, K. R. 1969a. Simultaneous test procedures - some theory of multiple comparisons. *Annals of Mathematical Statistic*, 40, 224-250.
- Gabriel, K. R. 1969b. A comparison of some methods of simultaneous inference in MANOVA. In: Krishnaiah, P. R. (Hrsg.), *Multivariate Analysis II*. New York: Academic Press.
- Gaensslen, H. & Schubö, W. 1976. *Einfache und komplexe statistische Analyse*. München: Reinhardt.
- Garnes, P. A., 1971a. Multiple comparisons of means. *American Educational Research Journal*, 8, 531-65.
- Garnes, P. A., 1971b. Inverse relation between the risks of type I and type II errors and suggestions for the unequal n case in multiple comparisons. *Psychological Bulletin*, 75, 97-102.
- Garnes, P. A., 1973. Type IV errors revisited. *Psychological Bulletin*, 80, 304-307.
- Gocka, E. F. 1973. Regression analysis of proportional cell data. *Psychological Bulletin*, 80, 25-27.
- Graybill, F. A. 1961. *An introduction to linear statistical models*. Vol. I. New York: McGraw Hill.
- Graybill, F. A. 1976. *Theory and application of the linear model*. Belmont: Wadsworth.
- Harter, H. L. 1974a. The method of least squares and some alternatives. Part I. *International Statistical Review*, 42, 147-174.
- Harter, H. L. 1974b. The method of least squares and some alternatives. Part II. *International Statistical Review*, 42, 235-264.
- Harter, H. L. 1975a. The method of least squares and some alternatives. Part III. *International Statistical Review*, 43, 1-44.
- Harter, H. L. 1975b. The method of least squares and some alternatives. Part IV. *International Statistical Review*, 43, 125-190.
- Harter, H. L. 1975c. The method of least squares and some alternatives. Part V. *International Statistical Review*, 43, 269-278.
- Hays, W. L. 1973. *Statistics for the social sciences*. London: Holt, Rinehart and Winston.
- Heck, D. L. 1960. Charts of some upper percentage points of the distribution of the largest characteristic root. *Annals of Mathematical Statistics*, 31, 625-642.
- Herr, D. G. & Gaebelein, J. 1978. Nonorthogonal two-way analysis of variance. *Psychological Bulletin*, 85, 207-216.
- Hild, C. 1977. *Schätzen und Testen in einem Regressionsmodell mit stochastischen Koeffizienten*. Meisenheim: Hain.
- Horton, R. L. 1978. *The general linear model. Data analysis in the social and behavioral sciences*. New York: McGraw Hill.
- Hotelling, H. 1951. A generalized T-test and measure of multivariate dispersion. *Pro-*

ceedings of the Second Berkeley Symposium on Mathematics and Statistics, 23-41.

- Humak, K. M. S. 1977. Statistische Methoden der Modellbildung. Berlin: Akademie-Verlag.
- Johnson, N. L. & Kotz, S. 1972. Distributions in statistics: Continuous multivariate distributions. New York: Wiley.
- Johnston, J. 1971. Econometric methods. Tokyo: Kogakusha, McGraw Hill.
- Kirk, R. E. 1972. Classification of ANOVA designs. In: Kirk, R. E. 1972. Statistical issues: A reader for the behavioral sciences. Monterey: Wadsworth.
- Kirk, R. E. 1978. Experimental design: Procedures for the behavioral sciences. Belmont, Cal.: Wadsworth.
- Köpcke, W. & Überla, K. (Hrsg.). 1980. Biometrie - heute und morgen. Berlin: Springer.
- Kowalsky, H. J. 1969. Lineare Algebra. Berlin: de Gruyter.
- Krafft, O. 1978. Lineare statistische Modelle. Göttingen: Vandenhoeck und Ruprecht.
- Krishnaiah, R. 1980. Handbook of Statistics, Vol. 1, Analysis of Variance. Amsterdam: North Holland.
- Kruskal, W. H. & Tanur, J. M. (Hrsg.). 1978. International encyclopedia of statistics. New York: The Free Press.
- Lawley, D. N. 1938. A generalization of Fisher's z-test. Biometrika, 30, 180-187.
- Leonhard, T. 1975. A Bayesian approach to the linear model with unequal variances. Technometrics, 17, 95-102.
- Lindeman, R. H., Merenda, P. F. & Gold, R. Z. 1980. Introduction to bivariate and multivariate analysis. Glenview: Scott, Foresman & Company.
- Lindley, D. V. & Smith, A. F. M. 1972. Bayes estimates for the linear model (with discussion). Journal of the Royal Statistical Society, Series B, 34, 1-41.
- Mendenhall, W. 1968. Introduction to linear models and the design and analysis of experiments. Belmont: Wadsworth.
- Mijares, T. A. 1964. Percentage points of the sum $V_1^{(s)}$ of s roots ($s = 1-50$). Manila: University of the Phillipines, Statistical Center.
- Miller, R. G. 1966. Simultaneous statistical inference. New York: McGraw Hill.
- Moosbrugger, H. 1978. Multivariate statistische Analyseverfahren. Stuttgart: Kohlhammer.
- Moosbrugger, H. 1982. Modelle zur Beschreibung statistischer Zusammenhänge in der psychologischen Forschung. In: Bredenkamp, J. & Feger, H. (Hrsg.), Enzyklopädie der Psychologie, Forschungsmethoden der Psychologie, Bd. 4, Göttingen: Hogrefe.
- Morrison, D. F. 1967. Multivariate statistical methods. New York: McGraw Hill.
- Müller, P. H. (Hrsg.). 1975. Lexikon der Stochastik. Berlin: Akademie-Verlag.

- Namboodiri, N. K., Carter, L. F. & Blalock, H. M. jr. 1975. Applied multivariate analysis and experimental designs. New York: McGraw Hill.
- Olson, C. L. 1976. On choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin*, 83, 579-586.
- Overall, J. E. & Klett, C. J. 1972. Applied multivariate analysis. New York: McGraw Hill.
- Overall, J. E. & Spiegel, D. K. 1973a. Comment on „Rawling's nonorthogonal analysis of variance“. *Psychological Bulletin*, 79, 164-167.
- Overall, J. E. & Spiegel, D. K. 1973b. Comment on „Regression analysis of proportionate cell data“. *Psychological Bulletin*, 80, 28-30.
- Overall, J. E., Spiegel, D. K. & Cohen, J. 1975. Equivalence of orthogonal and nonorthogonal analysis of variance. *Psychological Bulletin*, 82, 182-186.
- Pearson, E. S. & Wishart, J. 1943. „Student's“ collected papers. London: University College, Biometrika Office.
- Petrinovich, L. F. & Hardyck, C. D. 1969. Error rates for multiple comparison methods. *Psychological Bulletin*, 71, 43-54.
- Pillai, K. C. S. 1960. Statistical tables for tests of multivariate hypotheses. Manila: University of the Phillipines, Statistical Center.
- Pillai, K. C. S. & Mijares, T. A. 1959. On the moments of the trace of a matrix and approximations to its distribution. *Annals of Mathematical Statistics*, 30, 1135-1140.
- Pillai, K. C. S. & Jayachandran, K. 1967. Power comparisons of tests of two multivariate hypotheses based on four criteria. *Biometrika*, 54, 195-210.
- Pillai, K. C. S. & Jayachandran, K. 1970. On the exact distribution of Pillai's $V^{(s)}$ criterion. *Journal of the American Statistical Association*, 65, 447-454.
- Rao, C. R. 1973. Linear statistical inference and its applications. New York: Wiley.
- Rao, C. R. & Mitra, S. K. 1971. Generalized inverse of matrices and its application. New York: Wiley.
- Rawlings, R. R. jr. 1972. Note on nonorthogonal analysis of variance. *Psychological Bulletin*, 77, 373-374.
- Rawlings, R. R. jr. 1973. Comments on the Overall and Spiegel paper. *Psychological Bulletin*, 79, 168-169.
- Roy, S. N. 1939. p-statistics or some generalisations in analysis of variance appropriate to multivariate Problems. *Sankhya*, 4, 381-396.
- Roy, S. N. 1946. Multivariate analysis of variance: the sampling distribution of the p-statistics on the non-null hypothesis. *Sankhyā*, 8, 15-52.
- Roy, S. N. 1957. Some aspects of multivariate analysis. New York: Wiley.
- Ryan, T. A. 1959. Multiple comparisons in psychological Research. *Psychological Bulletin*, 56, 26-47.
- Schach, S. & Schäfer, T. 1978. Regressions- und Varianzanalyse. Berlin: Springer.

- Scheffé H. 1953. A method for judging all contrasts in the analysis of variance. *Biometrika*, 40, 87-104.
- Scheffé H. 1959. *The analysis of variance*. New York: Wiley.
- Searle, S. R. 1966. *Matrix algebra for the biological sciences*. New York: Wiley.
- Searle, S. R. 1971. *Linear models*. New York: Wiley.
- Seber, G. A. F. 1980. *The linear hypothesis: a general theory*. London: Griffin and Company.
- Shaffer, J. P. & Gillo, M. W. 1974. A multivariate extension of the correlation ratio. *Educational and Psychological Measurement*, 34, 521-524.
- Späth, H. 1973. *Algorithmen für elementare Ausgleichmodelle*. München: Oldenbourg.
- Späth, H. 1974. *Algorithmen für multivariable Ausgleichmodelle*. München: Oldenbourg.
- Stanley, J. C. 1978. Linear hypothesis: analysis of variance. In: Kruskal, W. H. & Tanur, J. M. (Hrsg.), 541-554.
- Stevens, J. P. 1972. Global measures of association in multivariate analysis of variance. *Multivariate Behavioral Research*, 7, 37-38.
- Stevens, J. 1979. Comment on Olson: choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin*, 86, 355-360.
- Steyer, R. 1979. *Untersuchungen zur nonorthogonalen Varianzanalyse*. Weinheim: Beltz.
- Student (Gosset, W. S.). 1908. The probable error of a mean. In: Pearson, E. S. & Wishart, J. 1943. „Student's“ collected papers. London: University College, Biometrika Office.
- Tatsuoka, M. M. 1971. *Multivariate analysis: techniques for educational and psychological research*. New York: Wiley.
- Timm, N. H. 1975. *Multivariate analysis with applications in education and psychology*. Monterey: Brooks/Cole.
- Timm, N. H. & Carlson, J. E. 1975. *Analysis of variance through full rank models*. *Multivariate Behavioral Research: Monograph* 75-1.
- Van de Geer, J. P. 1971. *Introduction to multivariate analysis for the social sciences*. San Francisco: Freeman.
- Ward, J. H. & Jennings, E. 1973. *Introduction to linear models*. Englewood Cliffs: Prentice-Hall.
- Wilks, S. S., 1932. Certain generalizations in the analysis of variance. *Biometrika*, 24, 471-494.
- Winer, B. J. 1971. *Statistical principles in experimental design*. New York: McGraw Hill.
- Wottawa, H. 1981. Probleme der Abstestung der Verteilungsvoraussetzung im allgemeinen linearen Modell. Vortrag auf der 23. Tagung experimentell arbeitender Psychologen, Berlin 1981.
- Zacks, S. 1971. *The theory of statistical inference*. New York: Wiley.