

# Das Allgemeine Lineare Modell

Johannes Andres

Das Allgemeine Lineare Modell (ALM) ist ein abstraktes statistisches Modell, aus dem durch Spezifizierung wichtige Verfahren hervorgehen, wie z.B. die Varianzanalyse, die Regressionsanalyse oder die Kovarianzanalyse. Allgemein geht es um die Beziehung einer abhängigen Variablen (AV)  $Y$  zu gewissen qualitativen und/oder quantitativen unabhängigen Variablen (UV). Charakteristisch für das ALM sind zwei Annahmen. Die erste besagt, daß sich für jede Kombination von Ausprägungen der UVn der Erwartungswert von  $Y$  in einfacher Weise (nämlich „linear“) aus den Werten der UVn ergibt. Die zweite Annahme betrifft die Abweichung von  $Y$  vom zugehörigen Erwartungswert, den sogenannten „Fehler“. In der hier zu besprechenden stärksten („klassischen“) Version des ALM wird gefordert, daß die (theoretische) Verteilung dieses Fehlers für jeweils alle möglichen Kombinationen von Ausprägungen der UVn eine Normalverteilung ist, deren Varianz unabhängig von der speziellen Kombination ist. Darüber hinaus sollen die Realisierungen der AV und damit die Fehler unabhängig sein.

Die Annahmen sollen im ersten Abschnitt genauer gefaßt werden; ferner wird an Beispielen gezeigt, daß das ALM tatsächlich ein Obermodell für die genannten Verfahren ist, wobei auch die Grenzen zu anderen Modellen abgesteckt werden. Im zweiten Abschnitt geht es darum, wie Fragestellungen in bezug auf Parameter und Hypothesen in diesem allgemeinen Rahmen beantwortet werden können.

Die mathematischen Zusammenhänge können hier nur oberflächlich angedeutet werden. Genaue Auskunft zu allen behandelten Fragen findet man z.B. in Schach und Schäfer (1978); als Referenz für elementarstatistische Begriffe und Aussagen sei generell auf die einschlägigen Passagen in Hays (1988) hingewiesen. Die wichtigsten Begriffe aus der linearen Algebra sind knapp eingeführt in dem Artikel über Grundbegriffe der multivariaten Datenanalyse (vgl. Andres, in diesem Band).

## 1 Grundbegriffe und Beispiele

Es werden immer mehrere Realisierungen der AV  $Y$  betrachtet, meistens unter verschiedenen Bedingungen, die durch unterschiedliche Festlegungen gewisser UVn definiert werden. Die Beobachtungen seien mit  $y_1$  bis  $y_n$  bezeichnet (ihre Anzahl ist also gleich  $n$ ) und werden als Zufallsvariablen (ZVAn) aufgefaßt<sup>1</sup> (es geht ja allgemein um einen Versuch einer bestimmten Art und nicht um eine konkrete Versuchsdurchführung). Für die Erwartungswerte dieser ZVAn in den verschiedenen Bedingungen soll die folgende grundlegende *Modellgleichung* gelten:

---

<sup>1</sup>Wie vielfach üblich werden ZVAn und mögliche konkrete Werte der ZVAn mit der gleichen Symbolik notiert. Was jeweils gemeint ist, erschließt sich aus dem Kontext.

$$E(y_i) = \sum_{j=1}^k x_{ij}\beta_j . \quad (1)$$

Dabei bezeichnet  $E(y_i)$  den Erwartungswert von  $y_i$ , und die Formel gibt an, wie dieser sich zusammensetzt aus zwei Gruppen von Größen, nämlich den mit den UVn zusammenhängenden  $x_{ij}$  und gewissen *Parametern*  $\beta_j$ . Die  $x_{ij}$  spiegeln dabei die Werte wider, die die UVn bei der  $i$ -ten Beobachtung annehmen, die  $k$  Parameter  $\beta_j$  hängen hingegen nicht von der speziellen Beobachtung ab und können für ein erstes Verständnis als Gewichte gelten, mit denen die UVn die Ausprägung der  $y_i$  beeinflussen. Die Zahl  $k$  hat dabei nur indirekt mit der Zahl der UVn zu tun, wie man an den später folgenden Beispielen genauer sehen kann. Die Werte der  $\beta_j$  sind unbekannt, und ein zentrales Ziel ist es, sie bei Unterstellung der Richtigkeit des Modells zu schätzen.

Übrigens sieht man hier schon, daß die Sprechweise von *einer* Variablen  $Y$  eigentlich ungenau ist; es geht immer um die einzelnen Beobachtungen  $y_i$ , die bei ganz bestimmten – und festen – Ausprägungen der UVn zustande kommen, sowie um deren Verteilungen. Die UVn werden im Gegensatz zu den  $y_i$  nicht als ZVAN betrachtet, sondern als gesetzt, was bei manchen Anwendungen schwer nachzuvollziehen ist (vgl. die Anmerkungen zur linearen Regression weiter unten). Die Bezeichnungen „unabhängige“ und „abhängige“ Variable werden hier übrigens nicht in einem experimentellen Sinn gebraucht, sie beziehen sich nur darauf, daß unterstellt wird, daß die Erwartungswerte der abhängigen aus den Werten der unabhängigen in der angegebenen Weise bestimmbar sind.

Die Art und Weise, in der die Parameter  $\beta_j$  die Erwartungswerte der  $y_i$  bestimmen, ist mathematisch von großer Einfachheit: Die Parameter werden mit gewissen Koeffizienten (den  $x_{ij}$ ) multipliziert und die entstehenden Produkte aufaddiert. Solche Funktionen nennt man auch *linear*, der Name „Allgemeines Lineares Modell“ hat hier seinen Ursprung. Die Koeffizienten  $x_{ij}$  faßt man übersichtlich in einer Matrix  $\mathbf{X}$  zusammen, deren  $i$ -te Zeile nacheinander die Koeffizienten für die  $i$ -te Beobachtung enthält; dementsprechend läßt sich in der  $j$ -ten Spalte dieser Matrix ablesen, mit welchem Koeffizienten der Parameter  $\beta_j$  in die einzelnen Erwartungswerte eingeht (die Indizes  $i$  und  $j$  geben also die „Zeilen- und Spaltenadresse“ des Koeffizienten  $x_{ij}$  an). Die Matrix  $\mathbf{X}$  heißt auch *Designmatrix*.

Nun bestimmen aber die UVn die AV nicht deterministisch, sondern nur bis auf einen Fehler, der mit  $e_i$  bezeichnet werden soll; es gilt dann

$$y_i = \sum_{j=1}^k x_{ij}\beta_j + e_i, \quad (2)$$

wobei wir in der Summe  $\sum x_{ij}\beta_j$  den Erwartungswert wiedererkennen. Es folgt unmittelbar, daß alle Fehler den Erwartungswert 0 haben.

Für diese Fehler soll die Voraussetzung gelten, daß sie unabhängig sind und normalverteilt mit einer unbekanntem Varianz  $\sigma^2$ , die für alle Fehler gleich groß sein soll. Die letzte Forderung nennt man auch die nach der Varianzhomogenität.

Damit sind alle Voraussetzungen des ALM formuliert: Lineare funktionale Abhängigkeit der Erwartungswerte von den Parametern, Unabhängigkeit, Normalver-

teiltheit und Varianzhomogenität der Fehler; auf Abschwächungen soll später eingegangen werden.

Wir sehen, daß die gemeinsame Verteilung der  $y_i$  durch die (leider unbekannt) Parameter  $\beta_j$  und  $\sigma^2$  festgelegt ist; diesen Parametern gilt nun das Hauptinteresse. Die Frage nach ihrer Größe wird statistisch beantwortet durch die Angabe von vernünftigen Punktschätzern und Bereichsschätzern. Daneben sind Fragen der Art wichtig, ob ein Parameter, eine ganze Gruppe von Parametern oder – noch allgemeiner – eine Funktion der Parameter Null ist. Hier sind statistische Tests zu entwickeln, mit denen entsprechend formulierte Hypothesen entschieden werden können.

Der Vorzug der allgemeinen Formulierung des ALM liegt darin, daß die statistischen Techniken zur Behandlung solcher Fragen zunächst unabhängig von speziellen Problemen entwickelt werden können; die auf diese Weise erhaltenen allgemeinen Sätze und Formeln stehen dann für die jeweilige besondere Anwendung zur Verfügung und müssen nur noch spezifiziert werden.

Zur Verdeutlichung sollen nun einige wichtige Modelle dem ALM als Spezialfälle untergeordnet werden. Die möglichen statistischen Fragestellungen in den Beispielen zeigen die Flexibilität der Methoden zum ALM. Andererseits sollen auch die Grenzen der Anwendbarkeit des ALM deutlich werden. Die Beispiele sollen ferner zum Anlaß genommen werden, weitere wichtige Begriffe einzuführen. Zur späteren Bezugnahme werden die Beispiele numeriert.

Das erste Beispiel (B1) ist die *einfache lineare Regression*. Hier soll der Erwartungswert der AV  $Y$  eine lineare Funktion der Werte einer UV  $U$  sein, also die Form

$$E(y_i) = au_i + b \quad (3)$$

besitzen; die unbekannt Parameter sind die Geradensteigung  $a$  und der  $y$ -Achsenabschnitt  $b$  (in der allgemeinen Notation müßten die Parameter  $\beta_1$  und  $\beta_2$  heißen, es sollen aber die traditionellen Bezeichnungen verwendet werden, ebenso in den folgenden Beispielen). Wählen wir  $a$  als ersten Parameter und  $b$  als zweiten, so besteht die Matrix  $\mathbf{X}$  aus zwei Spalten, deren erste die  $u_i$  und deren zweite lauter Einsen enthält;  $k$  ist also hier gleich 2. Hat die UV beispielsweise in drei Beobachtungen ( $n = 3$ ) die Werte 0, 1 und 2, so sieht die Designmatrix  $\mathbf{X}$  folgendermaßen aus:

$$\mathbf{X} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix}$$

Die Schätzer, die das ALM für die Parameter  $a$  und  $b$  liefert, sind die allgemein bekannten; Tests könnten zum Beispiel die (Null-)Hypothesen betreffen, daß die Geradensteigung oder der Achsenabschnitt oder beides Null ist. Ebenso wie in den übrigen Beispielen kann die Nullhypothese auch andere fest vorgegebene Werte für  $a$  und/oder  $b$  unterstellen.

Eine wichtige Frage betrifft hier die UV: Kann man ihre Werte fest vorgeben, variieren sie ebenfalls zufällig oder sind sie gar mit Meßfehlern behaftet? Im ersten Fall ist die Anwendung des ALM bei geeigneter Wahl der Werte unproblematisch, im zweiten Fall jedoch können Probleme auftreten: Alle Meßwerte von  $U$  könnten z.B. gleich sein. In diesem Fall bekommt man keine eindeutigen Schätzungen der Parameter nach dem noch zu erläuternden Verfahren, das mit der bekannten Methode

der kleinsten Quadrate übereinstimmt; anschaulich ist dies plausibel, passen doch alle Geraden, die in diesem Fall für den einzigen Wert von  $U$  durch den Mittelwert von  $Y$  gehen, gleich gut in dem Sinn, daß sie die Summe der Abweichungsquadrate minimieren. In diesem Fall haben wir es mit einer „entarteten“ Matrix  $\mathbf{X}$  zu tun, und es ist tröstlich zu erfahren, daß unter der Voraussetzung, daß die Realisierungen von  $U$  in den einzelnen Beobachtungen stochastisch unabhängig sind mit der gleichen stetigen Verteilung, die Wahrscheinlichkeit gleicher Werte von  $U$  Null ist, daß man also „fast sicher“ diesen unerfreulichen Fall ausschließen kann.

Während man sich also bei einem derartigen „stochastischen Regressor“ unter den angegebenen Voraussetzungen wenig Gedanken zu machen braucht, sind die Verfahren des ALM im letzten Fall, nämlich bei Meßfehlern in der UV, nicht mehr anwendbar, da sie zum Beispiel zu verzerrten (also nicht erwartungstreuen) Parameterschätzungen führen. Hier gilt auch schon die Grundgleichung nicht mehr, da man ja wohl unterstellen will, daß die Erwartungswerte von  $Y$  von den „wahren Werten“ von  $U$  und nicht etwa von den fehlerbehafteten Messungen in linearer Weise abhängen. Zur Untersuchung derartiger Zusammenhänge könnte man beispielsweise die Methoden der linearen Strukturgleichungsmodelle heranziehen.

Das nächste Beispiel (B2) ist die *polynomiale Regression*. Das letzte Beispiel könnte den Gedanken nahelegen, daß der Name des ALM etwas damit zu tun hat, daß die Beziehung zwischen der UV und der AV (bis auf den Fehler) linear ist; das folgende Beispiel zeigt jedoch, daß die unterstellte funktionale Beziehung zwischen UV und AV auch von anderer Gestalt sein kann. Ausgehend von einer UV  $U$  wird diesmal angenommen, daß die Erwartungswerte von  $Y$  sich aus den Werten von  $U$  durch Anwendung eines Polynoms z.B. dritten Grades (mit allerdings unbekanntem Koeffizienten) errechnen lassen; hier gilt also

$$E(y_i) = \beta_1 \cdot u_i^3 + \beta_2 \cdot u_i^2 + \beta_3 \cdot u_i + \beta_4. \quad (4)$$

Man erkennt, daß sich in der Tat die Erwartungswerte in der gewünschten Weise schreiben lassen; nimmt  $U$  in 5 Beobachtungen die Werte 0, 1, 2, 3 und 5 an, so ist

$$\mathbf{X} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 8 & 4 & 2 & 1 \\ 27 & 9 & 3 & 1 \\ 125 & 25 & 5 & 1 \end{pmatrix}.$$

Natürlich sind auch funktionale Beziehungen denkbar, in denen anstelle der Potenzen von  $U$  andere Funktionen (Logarithmen etc.) verwendet werden; sie lassen sich in ähnlicher Weise behandeln. Man erkennt hier übrigens deutlich, daß die Zahl  $k$  der Spalten der Matrix nichts mit der Zahl der UVn zu tun hat, sondern die Anzahl der unbekanntem Parameter wiedergibt.

Hier wird eine weitere generelle Voraussetzung für eine sinnvolle Anwendung des ALM einsichtig, nämlich, daß  $n$  größer als  $k$  sein soll. Im betrachteten Fall könnte man nämlich beim Vorliegen von nur 4 Beobachtungen mit verschiedenen Werten für  $U$  immer die Parameter so bestimmen, daß die zugehörigen Erwartungswerte mit den  $Y$ -Werten exakt übereinstimmen, das Modell also exakt paßt. Im Falle der

linearen Regression ist dies noch klarer: Durch zwei Punkte läßt sich immer eine Gerade legen. Eine Schätzung der Fehlervarianz ist dann unmöglich.

In diesem Beispiel könnte es sein, daß man sich vorzugsweise für die Parameter  $\beta_1$  und  $\beta_2$  bei  $u^3$  und  $u^2$  interessiert, also für die Abweichung von einer linearen Vorhersagefunktion. Dabei geht es oft um die gleichzeitige und gleichberechtigte Schätzung, wobei man auch die Größe des möglichen Fehlers berücksichtigen will. Dies führt zur Forderung nach der Konstruktion vernünftiger Konfidenzbereiche, einer Verallgemeinerung von Konfidenzintervallen auf diese zweidimensionale Fragestellung.

Mögliche Hypothesen können hier ebenfalls mehrere Parameter betreffen. Neben der Nullhypothese, daß es sich bei der „Vorhersagefunktion“ um eine Gerade handelt, daß also  $\beta_1$  und  $\beta_2$  gleich Null sind, ist z.B. auch die möglich, daß diese Funktion symmetrisch zur  $y$ -Achse ist ( $\beta_1 = \beta_3 = 0$ ) oder punktsymmetrisch zum Ursprung ( $\beta_2 = \beta_4 = 0$ ).

Das nächste Beispiel (B3) ist die *multiple Regression*. Hier ist der Erwartungswert der AV nicht die Funktion von einer, sondern von mehreren UVn („Prädiktoren“), die mit  $U_1, \dots, U_l$  bezeichnet seien. Die Grundgleichung schreibt sich so:

$$E(y_i) = \sum_{j=1}^l u_{ij} b_j + a. \quad (5)$$

Dabei ist  $u_{ij}$  der Wert, den die  $j$ -te UV bei der  $i$ -ten Beobachtung annimmt, die  $b_j$  sind die Regressionsgewichte, und  $a$  ist eine additive Konstante. Die unbekannt Parameter sind hier die  $b_j$  und  $a$ , es gilt also  $k = l + 1$ . Für einen Fall mit zwei Prädiktoren könnte die Matrix  $\mathbf{X}$  bei 4 Beobachtungen so aussehen:

$$\mathbf{X} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \\ 3 & 3 & 1 \\ 4 & 5 & 1 \end{pmatrix}.$$

Dabei stehen in den ersten beiden Spalten die Werte, die die Prädiktoren in den vier Beobachtungen nacheinander annehmen. Fragestellungen könnten hier beispielsweise sein: Ist mindestens ein Regressionsgewicht, also  $b_1$  oder  $b_2$ , von Null verschieden, ist  $b_1$  von Null verschieden, sind die beiden Gewichte gleich groß, ist  $a$  gleich Null oder gleich einer anderen vorgegebenen Zahl? Für alle diese Fragestellungen läßt sich mit den Methoden des ALM ein geeigneter Test finden.

Das Beispiel der multiplen Regression, es ist eines der typischsten für das ALM, sei zum Anlaß genommen, einige wichtige Begriffe einzuführen. Dazu nehmen wir zunächst an, daß in der Matrix  $\mathbf{X}$  der dritte Wert der zweiten UV nicht 3, sondern 4 ist. Die Spalten der neuen Matrix weisen nun eine Abhängigkeit auf: Addiert man die erste und die dritte Spalte komponentenweise, so erhält man die zweite. Dies hat gravierende Konsequenzen. Nehmen wir nämlich eine beliebige Belegung der drei Parameter  $b_1, b_2$  und  $a$  mit konkreten Zahlen und ändern sie ab, indem wir eine feste Zahl zu  $b_2$  hinzuaddieren und gleichzeitig von  $b_1$  und  $a$  abziehen (also z.B. erste Belegung:  $b_1 = 3, b_2 = 4, a = 5$  und zweite Belegung  $b_1 = 1, b_2 = 6, a = 3$ ), so sind die Erwartungswerte der  $y_i$  für beide Belegungen die gleichen (im Beispiel 16, 23, 30 und 37). Damit ist für beide mögliche Belegungen die gemeinsame Verteilung

der  $y_i$  die gleiche. Daher ist auch die Verteilung jeder Statistik, die sich aus den Werten der  $y_i$  berechnen läßt, für beide Belegungen die gleiche, so daß sich prinzipiell keine Möglichkeit bietet, auf statistischem Wege zwischen ihnen zu unterscheiden. Etwas vergrößernd ausgedrückt: Alle denkbaren Daten aus diesem Versuch passen genauso gut (oder genauso schlecht) zu der einen Belegung wie zu der anderen, empirisch sind die beiden Belegungen ununterscheidbar. Daher ist auch die Frage, welche Parameter die „wahren“ sind, hier im Grunde sinnlos. Man nennt in solchen Fällen die Parameter „nicht identifizierbar“, wobei Parameter *identifizierbar* heißen, wenn es möglich ist, ihre Größe bei Kenntnis der „wahren“ (theoretischen) Verteilung der beobachteten Variablen zu ermitteln.

Wenn auch die Parameter in diesem Fall nicht identifizierbar sind, so sind dennoch gewisse sinnvolle Aussagen über sie möglich: So könnte eine inhaltlich sinnvolle Hypothese lauten, daß der Einfluß der beiden UVn gleich groß, aber entgegengesetzt ist, daß also  $b_1 = -b_2$  gilt, anders formuliert, daß  $b_1 + b_2 = 0$  ist. Interessanterweise läßt sich diese Summe aus der Kenntnis der Erwartungswerte der  $y_i$  bestimmen: So ist  $E(y_2) - E(y_1) = 2b_1 + 3b_2 + a - (b_1 + 2b_2 + a) = b_1 + b_2$ . Erweitert man den Begriff der Identifizierbarkeit in naheliegender Weise auf Funktionen der Parameter, so ist jetzt die Funktion  $b_1 + b_2$  identifizierbar. Während also Aussagen über die Größe der Parameter  $b_1$  und  $b_2$  im Zusammenhang mit diesem Versuch sinnlos sind, gilt dies nicht für die Größe der Summe. Die Hypothese, daß diese Summe Null ist, ist eine sinnvolle und auch testbare Hypothese, während z.B. die Hypothese  $b_1 = 0$  wegen der fehlenden Identifizierbarkeit von  $b_1$  nicht testbar und auch nicht sinnvoll ist.

Die Funktion  $b_1 + b_2$  ist ein Beispiel für eine lineare Funktion der Parameter, wobei wir allgemein unter linearen Funktionen der Parameter  $\beta_1, \dots, \beta_k$  solche der Gestalt  $\psi = \sum_{j=1}^k c_j \beta_j$  mit gewissen Koeffizienten  $c_j$  verstehen wollen. (Im Beispiel ist  $c_1 = c_2 = 1$  und  $c_3 = 0$ .) Solche Funktionen sollen kurz *parametrische Funktionen* genannt werden. Eine Rückschau auf die bisherigen Beispiele ergibt, daß alle dort genannten Hypothesen über die Parameter  $\beta_j$  sich als Aussagen über parametrische Funktionen formulieren lassen. So ist die Hypothese, daß zwei Parameter gleich groß sind, gleichbedeutend damit, daß ihre Differenz Null ist. Natürlich sind auch die Parameter selbst parametrische Funktionen, nämlich solche, in denen alle Koeffizienten gleich Null sind bis auf jeweils eine Eins. Sehr viele Aussagen und Fragestellungen beziehen sich damit allgemein formuliert auf parametrische Funktionen, sind aber offenbar nur dann sinnvoll, wenn sie identifizierbare Funktionen betreffen.

So wird die Frage nach der Identifizierbarkeit eine zentrale, damit auch die Frage nach Bedingungen für Identifizierbarkeit. Im Beispiel war die fehlende Identifizierbarkeit eine Folge der „Abhängigkeitsstruktur“ der Spalten der abgeänderten Matrix  $\mathbf{X}$ . Die zweite Spalte war die Summe der ersten und dritten Spalte. Die Spalten sind also linear abhängig, weshalb der Rang der Designmatrix  $\mathbf{X}$ , er sei im folgenden mit  $r$  bezeichnet, nicht maximal ist. In unserem Beispiel ist  $r$  gleich 2. Nun kann eine Bedingung für die Identifizierbarkeit formuliert werden: Die Parameter eines Modells sind genau dann alle identifizierbar, wenn die Spalten von  $\mathbf{X}$  linear unabhängig sind, anders ausgedrückt, wenn der Rang  $r$  von  $\mathbf{X}$  gleich  $k$  ist. Eine parametrische Funktion ist genau dann identifizierbar, wenn ihre zu einer Zeile zusammengestellten Koeffizienten eine Linearkombination der Zeilen von  $\mathbf{X}$  bilden. Ist der Rang von  $\mathbf{X}$  gleich  $k$ , so ist jede parametrische Funktion identifizierbar.

Wie im Fall der einfachen linearen Regression besteht nun ein Problem darin, zu gewährleisten, daß die Matrix  $\mathbf{X}$  Rang  $k$  hat. Kann man die Werte der UVn beliebig setzen, so kann man auch über den Rang von  $\mathbf{X}$  verfügen. Sind jedoch die UVn Zufallsvariablen, so bleibt nur die Hoffnung, daß sie sich im Experiment geeignet realisieren. Immerhin gilt: Sind die UVn stetig verteilt, besitzt ihre gemeinsame Verteilung eine Dichte und hat ihre Kovarianzmatrix den (maximalen) Rang  $l$ , so ist der Rang  $r$  von  $\mathbf{X}$  für  $n \geq l + 1$  (was wegen  $n > k$  immer gilt) fast sicher gleich  $k$  (d.h. mit Wahrscheinlichkeit 1). Natürlich gilt auch hier, daß bei der Möglichkeit von Meßfehlern in den UVn das ALM nicht angewendet werden kann, falls die Erwartungswerte der  $y_i$  in der beschriebenen Weise von den „wahren Werten“ der  $U$ -Variablen abhängen sollen.

Beispiel B4 ist die *einfache Varianzanalyse*. Hier wird die AV  $Y$  unter mehreren Stufen eines unabhängigen Faktors untersucht. Hat dieser Faktor  $k$  Stufen, ist die Zahl der Beobachtungen unter der  $j$ -ten Stufe gleich  $n_j$  ( $\sum n_j = n$ ) und ist  $y_{ij}$  die  $i$ -te Beobachtung in der  $j$ -ten Faktorstufe, so soll gelten:

$$E(y_{ij}) = \mu_j, \quad (6)$$

wobei  $\mu_j$  der Erwartungswert für die  $j$ -te Stufe des Faktors ist. Ordnet man die  $n$  Beobachtungen so untereinander an, daß zuerst alle unter der ersten Stufe des Faktors kommen, dann alle unter der zweiten Stufe, etc., so ergibt sich mit  $\beta_j = \mu_j$  eine besonders einfache Matrix  $\mathbf{X}$  mit  $k$  Spalten, bei der die  $j$ -te Spalte nur aus Nullen besteht außer in den Zeilen, die zu den Beobachtungen unter der  $j$ -ten Faktorstufe gehören. Ist z.B.  $k = 3$  und werden unter den ersten beiden Faktorstufen je zwei Beobachtungen gemacht und unter der dritten eine, so ist

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Mögliche Fragen können hier die Größen einzelner Parameter betreffen, aber auch Vergleiche zwischen den Parametern; die üblicherweise gestellten laufen auf ganz bestimmte Vergleiche hinaus, nämlich solche, bei denen die Koeffizienten der entsprechenden parametrischen Funktionen sich zu Null addieren. Dies sind die sogenannten *Kontraste*. Der Vergleich von  $\mu_1$  und  $\mu_2$  geschieht mit der parametrischen Funktion  $\mu_1 - \mu_2$ , deren Koeffizienten  $(1, -1, 0)$  aufsummiert Null ergeben; möchte man wissen, ob sich  $\mu_1$  vom Mittelwert von  $\mu_2$  und  $\mu_3$  unterscheidet, so wählt man den Kontrast  $\mu_1 - (\mu_2 + \mu_3)/2$ . Offenbar zeichnen sich die Kontraste gerade dadurch aus, daß ihr Wert sich nicht ändert, wenn man zu allen Parametern dieselbe beliebig gewählte Konstante hinzuaddiert. In diesem Sinne sprechen sie auf Unterschiede an, nicht aber auf den durchschnittlichen Wert der  $\mu_j$ .

Die am häufigsten getestete mögliche Hypothese in diesem Modell ist wohl die, daß alle Erwartungswerte  $\mu_j$  gleich sind; äquivalent dazu ist die, daß alle Kontraste gleich Null sind, auch diese Hypothese läßt sich also in eine über parametrische Funktionen überführen. Übrigens ist die Hypothese auch schon äquivalent dazu, daß eine endliche Anzahl geeignet gewählter Kontraste Null ist, z.B.  $\mu_2 - \mu_1, \mu_3 - \mu_2$ ,

$\dots, \mu_k - \mu_{k-1}$ . Der Standard-Test, den die Spezialisierung der Theorie des ALM für diese (Null-)Hypothese liefert, ist erwartungsgemäß der bekannte  $F$ -Test.

Bisweilen schreibt man das varianzanalytische Modell auch mit einer anderen Parametrisierung, nämlich als

$$E(y_{ij}) = \mu + \alpha_j. \quad (7)$$

Hier sieht man sofort, daß die (jetzt  $k + 1$ ) Parameter nicht identifizierbar sind, weshalb man die Identifizierbarkeit durch eine Nebenbedingung erzwingt, im allgemeinen die, daß  $\sum n_j \alpha_j = 0$  gelten soll. Dann stellt sich der neue künstliche Parameter  $\mu$  als ein gewichtetes Mittel der  $\mu_j$  heraus, wobei die Gewichtungsfaktoren gerade  $n_j/n$  sind und die Größe der Zellbesetzungen widerspiegeln. Die  $\alpha_j$  sind dann die Abweichungen von diesem künstlichen Mittel und werden auch als *Effekte* bezeichnet. Bei geeigneter Wahl der  $n_j$  kann man  $\mu$  manchmal als Erwartungswert von  $Y$  in einer natürlichen Population interpretieren, im allgemeinen Fall hat dieser Parameter jedoch nur eine formale Bedeutung.

In ähnlicher Weise sind auch viele mehrfaktorielle varianzanalytische Modelle Spezialfälle des ALM. Eine Parametrisierung mit künstlichen Effekt-Parametern samt zugehörigen Nebenbedingungen ist von den Fragestellungen her oft geboten (wenn man z.B. Haupt- und Interaktionseffekte getrennt testen will, wobei gegebenenfalls eine  $\alpha$ -Adjustierung vorzunehmen ist). Die im Rahmen des ALM bereitgestellten Techniken erweisen sich auch als hilfreich zur Konstruktion geeigneter Verfahren bei verwandten Problemstellungen, die sich nicht dem ALM unterordnen, so bei der Varianzanalyse mit Zufallsfaktoren oder bei „gemischten Modellen“, bei denen Gruppen von Parametern durch unabhängige Realisierungen einer ZVA ersetzt werden. Viele Meßwiederholungsdesigns fallen unter diese Rubrik (der Unterschied zu den stochastischen Regressoren liegt darin, daß hier die Parameter zu ZVAn werden, während dort die Matrix  $\mathbf{X}$  „stochastisch“ ist).

Das letzte Beispiel (B5) sind *Mischformen von Varianz- und Regressionsanalyse*. Der einfachste Fall ist der zweier UVn, eines „qualitativen Faktors“ und eines „quantitativen Regressors“  $U$ , in dem der Erwartungswert der AV auf jeder Stufe des Faktors eine lineare Funktion des Regressors ist, wobei die Geradensteigung immer die gleiche sein soll, während der  $y$ -Achsenabschnitt für verschiedene Faktorstufen verschieden sein darf. Die Grundgleichung ist hier

$$E(y_{ij}) = au_{ij} + \mu_j, \quad (8)$$

wobei der Index  $j$  für die Faktorstufe steht und  $i$  die Beobachtungen innerhalb einer Faktorstufe durchnummeriert;  $a$  ist die Geradensteigung und  $\mu_j$  der Achsenabschnitt für Faktorstufe  $j$ . Gibt es z.B. zwei Faktorstufen, werden in der ersten zwei und in der zweiten drei Beobachtungen gemacht und sind die Werte der UV  $U$  für die erste Stufe 1 und 9 und für die zweite 1, 5 und 3, so sieht die Matrix  $\mathbf{X}$  bei der aus der Varianzanalyse nahegelegten Reihenfolge der  $y$ -Werte und bei der Reihenfolge  $a, \mu_1, \mu_2$  der Parameter so aus:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 \\ 9 & 1 & 0 \\ 1 & 0 & 1 \\ 5 & 0 & 1 \\ 3 & 0 & 1 \end{pmatrix}.$$

Mögliche Fragestellungen betreffen zunächst die Größe der Parameter, vielleicht auch die Differenz von  $\mu_1$  und  $\mu_2$  (eine parametrische Funktion). Eine mögliche (Null-)Hypothese könnte die der Gleichheit der  $\mu_j$  sein; in diesem Fall spricht man von einer *Kovarianzanalyse*, das Interesse ist ein „varianzanalytisches“. Eine andere – eher „regressionsanalytische“ – Hypothese könnte besagen, daß die Geradensteigung  $a$  gleich Null ist. Etwas allgemeiner ist es auch möglich, für die verschiedenen Faktorstufen verschiedene Geradensteigungen zuzulassen (was in unserem Beispiel die Zahl der Parameter auf 4 erhöhen würde); dann ist eine andere denkbare Hypothese die, daß die beiden Geradensteigungen gleich sind.

Natürlich läßt sich dieses Beispiel beliebig verallgemeinern, auf mehrere Faktoren mit Haupt- und Interaktionseffekten, auf mehrere Regressoren, die auch in Form von Polynomen in die Grundgleichung eingehen können, etc.

Bisweilen sagt man, daß die letzten beiden Spalten der Matrix  $\mathbf{X}$  sogenannten Dummy-Variablen entsprechen; der Grund dafür liegt darin, daß man formal so tun kann, als ob die Nullen und Einsen in diesen Spalten Werte von „Regressoren“ wären. Man kann dann zur Behandlung des Modells Computerprogramme für die multiple Regression heranziehen, die am ehesten die dem ALM angemessene Flexibilität besitzen. Auch eine Varianzanalyse läßt sich so „regressionsanalytisch“ rechnen.

## 2 Statistische Behandlung

Nachdem der letzte Abschnitt die Flexibilität des ALM und die Vielfalt möglicher Fragestellungen und Hypothesen deutlich gemacht hat, sollen in diesem Abschnitt die statistischen Methoden der Behandlung angedeutet werden; anstelle einer den Rahmen sprengenden Darstellung dieser Methoden sollen an einem einfachen Beispiel die geometrischen Aspekte beschrieben werden. Es soll um vier Hauptthemen gehen: Schätzung der Modellparameter, Punkt- und Bereichsschätzer identifizierbarer parametrischer Funktionen und Hypothesentesten. Das Beispiel ist die einfache lineare Regression (B1) mit drei Beobachtungen.

Der erste Schritt ist die Wahl einer geeigneten geometrischen Darstellung der Daten: Wenn die Anzahl der  $Y$ -Werte gleich  $n$  ist (die Matrix  $\mathbf{X}$  also  $n$  Zeilen hat), so faßt man diese Werte zu einem  $n$ -Tupel  $\mathbf{y} := (y_1, \dots, y_n)'$  zusammen, das man auch als einen Punkt in einem  $n$ -dimensionalen Koordinatensystem (oder einem „ $n$ -dimensionalen Raum“, dem *Personenraum*) auffassen kann. Haben sich zu den Werten 0, 1, und 2 der UV die  $y$ -Werte 1,  $-2$  und 5 ergeben, so können diese durch den Punkt  $\mathbf{y}$  mit den Koordinaten  $(1, -2, 5)$  in einem dreidimensionalen Koordinatensystem repräsentiert werden, wie in Abbildung 1a, von der zunächst nur das  $y$ -Koordinatensystem und der Punkt  $\mathbf{y}$  wichtig sind.

Bei einer gegebenen Konstellation von Parametern kann man in dieser Darstellung den Punkt eintragen, der sich ergeben würde, wenn alle Fehler Null wären, das heißt den Punkt, dessen Koordinaten die Erwartungswerte der  $y_i$  bei der speziellen Parameterkonstellation sind. Man nennt ihn naheliegenderweise den Erwartungswert von  $\mathbf{y}$  für die betrachtete Parameterkonstellation und kann seine  $i$ -te Koordinate als „Vorhersage“ von  $y_i$  ansehen. Im Beispiel erhält man z.B. für  $a = 1$  und  $b = 1$  die Vorhersage  $(1, 2, 3)'$ . Man kann nun diese Erwartungswerte von  $\mathbf{y}$  für alle möglichen Parameterbelegungen bilden und faßt die sich ergebenden Punkte zweckmäßigerweise

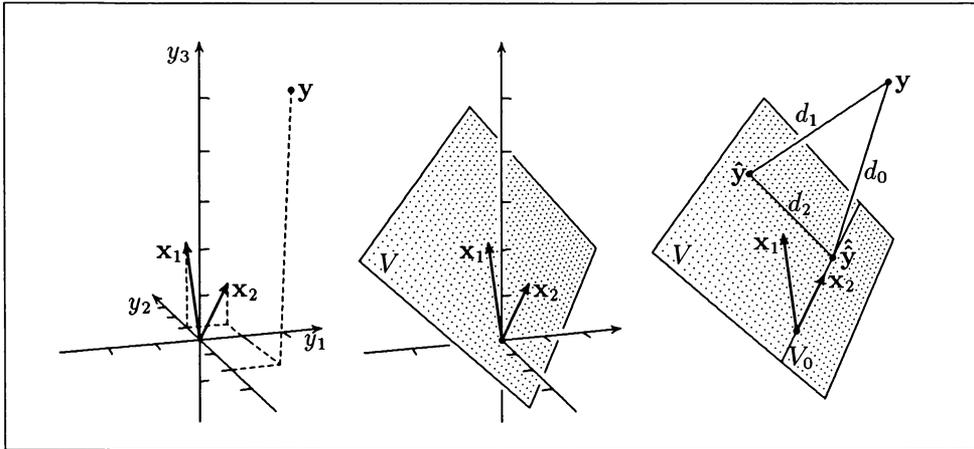


ABBILDUNG 1. Geometrische Darstellung des Beispiels, a: Personenraum, b: Ausschnitt aus dem Unterraum  $V$ , c: Veranschaulichung zur Parameterschätzung und zum Hypothesentesten; weitere Erläuterungen im Text.

zu einer Menge  $V$  zusammen.

Im Beispiel ist  $V$  eine Ebene, aus der ein Ausschnitt in Abbildung 1b wiedergegeben ist. Daß  $V$  hier tatsächlich eine Ebene ist, erkennt man daran, daß die Vorhersage für feste Werte  $a$  und  $b$  die Linearkombination der Spalten  $x_1$  und  $x_2$  von  $X$  mit den Koeffizienten  $a$  und  $b$  ist. Die Menge aller derartiger Linearkombinationen ist gerade die durch diese beiden Spalten „aufgespannte“ Ebene (vgl. Abbildung 1b).

Allgemein ist die Menge  $V$  immer ein linearer Unterraum des Personenraums, und zwar der durch die Spalten der Matrix  $X$  aufgespannte. Seine Dimension ist der Rang  $r$  von  $X$ , stimmt also mit der Spaltenzahl  $k$  dann überein, wenn die Spalten linear unabhängig sind.

Im Beispiel ist jeder Punkt aus  $V$  der Erwartungswert zu genau einer Belegung der Parameter  $a$  und  $b$  und entspricht damit einer möglichen Vorhersagegeraden (nämlich der mit Steigung  $a$  und Achsenabschnitt  $b$ ). Die zu einem Punkt gehörenden Parameterwerte von  $a$  und  $b$  sind dabei gerade die Koordinaten dieses Punktes in dem Koordinatensystem, das auf  $V$  durch die Vektoren  $x_1$  und  $x_2$  etabliert wird. Die  $y$ -Koordinaten des Punktes sind andererseits gerade die Vorhersagen zu den Werten der UV. Analoge Aussagen gelten immer dann, wenn der Rang von  $X$  gleich  $k$  ist.

Das erste Problem ist nun die *Schätzung der Modellparameter*. Das geometrische Bild legt nahe, diese so zu schätzen, daß die Vorhersage durch die geschätzten Modellparameter möglichst nahe an den Daten liegt, genauer, daß der Abstand von  $y$  zu dem zu den geschätzten Parametern gehörenden Punkt von  $V$  minimal wird. Da man durch iterierte Anwendung des Satzes des Pythagoras den quadrierten Abstand von zwei Punkten im Personenraum als Summe der quadrierten Koordinatendifferenzen erhält, kann man diese Forderung auch so formulieren: „Wähle die Parameter so, daß die Summe der quadrierten Abweichungen der Beobachtungen von den durch die Parameter vorhergesagten Werten minimal wird.“ Die Schätzung geschieht also nach der schon aus der einfachen linearen Regression bekannten *Methode der kleinsten*

*Quadrate*, und die dort zu minimierende Summe erhält jetzt eine neue Bedeutung als ein quadrierter Abstand von zwei Punkten im Personenraum.

Geometrisch ist klar, daß der Punkt von  $V$ , der am nächsten an  $\mathbf{y}$  gelegen ist, der Fußpunkt des Lotes von  $\mathbf{y}$  auf  $V$  sein muß, der hier mit  $\hat{\mathbf{y}}$  bezeichnet sei. Lot und Fußpunkt sind in Abbildung 1c eingezeichnet. Die Parameter, die zu  $\hat{\mathbf{y}}$  gehören, werden durch Lösung der sogenannten *Normalgleichungen* ermittelt. Diese bilden ein lineares Gleichungssystem mit  $k$  Gleichungen in den unbekanntem Modellparametern  $\beta_1, \dots, \beta_k$ , das sich in Matrixschreibweise knapp als

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y} \quad (9)$$

formulieren läßt, wobei  $\boldsymbol{\beta}$  der Vektor der Modellparameter ist. (Der mit Matrizenrechnung etwas vertraute Leser erkennt in der Umformulierung  $\mathbf{X}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 0$  leicht die Forderung, daß die Strecke von der Vorhersage  $\mathbf{X}\boldsymbol{\beta}$  nach  $\mathbf{y}$  senkrecht auf allen Spalten von  $\mathbf{X}$  und damit auf dem von diesen Spalten aufgespannten  $V$  stehen soll; daher auch der Name – normal bedeutet nämlich in manchen Zusammenhängen soviel wie senkrecht.) Die  $i$ -te Normalgleichung lautet ausführlich formuliert:

$$\sum_{j=1}^k a_{ij}\beta_j = b_i \quad \text{mit} \quad a_{ij} = \sum_{l=1}^n x_{li}x_{lj} \quad \text{und} \quad b_i = \sum_{l=1}^n x_{li}y_l. \quad (10)$$

Die Normalgleichungen im Beispiel sind  $5a + 3b = 8$  und  $3a + 3b = 4$ , was als Lösung  $\hat{a} = 2$  und  $\hat{b} = -2/3$  liefert (wie in B1 angedeutet, also die Parameter der üblichen Regressionsgeraden). Dies sind, wie man auch in der (zentralperspektivischen) Darstellung von Abbildung 1c nachvollziehen kann, gerade die Koordinaten von  $\hat{\mathbf{y}}$  in dem  $\mathbf{x}_1$ - $\mathbf{x}_2$ -Koordinatensystem von  $V$ . Ein weiteres Beispiel: In B4 (einfaktorielle Varianzanalyse) sind die Lösungen der Normalgleichungen für die  $\mu_j$  die Mittelwerte der Beobachtungen in den entsprechenden Zellen.

Die Normalgleichungen besitzen immer eine Lösung, die allerdings nur dann eindeutig ist, wenn der Rang  $r$  der Matrix  $\mathbf{X}$  gleich  $k$  ist, wenn also die Parameter identifizierbar sind. Sind die Parameter identifizierbar, die Normalgleichungen also eindeutig lösbar, so sind die Lösungen auch erwartungstreue Schätzer der Modellparameter – nicht identifizierbare Parameter hingegen sind naheliegenderweise auch nicht schätzbar.

Auch der quadrierte Abstand  $d^2 := \sum (y_i - \hat{y}_i)^2$  von  $\mathbf{y}$  zu  $\hat{\mathbf{y}}$  (der ja, da von  $\mathbf{y}$  abhängig, eine ZVA ist) hat eine statistische Bedeutung: sein Erwartungswert ist das  $(n - r)$ -fache der Fehlervarianz  $\sigma^2$ . Mit  $s^2 := d^2/(n - r)$  erhält man also einen erwartungstreuen Schätzer dieses Modellparameters. Teilt man hingegen  $d^2$  durch  $\sigma^2$ , so ist das Ergebnis  $\chi^2$ -verteilt mit  $(n - r)$  Freiheitsgraden („df“), worauf beim Hypothesentesten zurückgegriffen wird. Anschaulich könnte man die Anzahl  $(n - r)$  der Freiheitsgrade so deuten: Zunächst kann der Fehler in  $n$  Dimensionen variieren, die Variation in den  $r$  Dimensionen von  $V$  geht jedoch in die Parameterschätzung ein, so bleiben bei der Fehlerschätzung nur noch  $n - r$  Dimensionen übrig, was auch den Namen „Freiheitsgrade“ etwas verständlicher macht.

Neben den Modellparametern sind, wie die Beispiele zeigen, auch identifizierbare parametrische Funktionen der Parameter interessant; für eine solche Funktion  $\psi$  bekommt man einen erwartungstreuen Schätzer  $\hat{\psi}$ , wenn man in  $\psi$  anstelle der Parameter eine Lösung der Normalgleichungen einsetzt. Die so entstehenden Schätzer sind

lineare Funktionen der  $y_i$  und zeichnen sich dadurch aus, daß sie unter allen erwartungstreuen Schätzern, die lineare Funktionen in den  $y_i$  sind, minimale Varianz besitzen; sie heißen auch *Gauß-Markoff-Schätzer* („GM-Schätzer“). Beispielsweise erhält man den GM-Schätzer für einen Kontrast in der einfaktoriellen Varianzanalyse, indem man für die  $\mu_j$  die entsprechenden Zellenmittelwerte einsetzt. Oder: In der Variante zu B3 interessierte die parametrische Funktion  $\psi = b_1 + b_2$ . Die dortigen Erörterungen zeigen, daß  $y_2 - y_1$  ein erwartungstreuer Schätzer von  $\psi$  ist, der zudem linear in den  $y_i$  ist. Als GM-Schätzer findet man hier jedoch  $\hat{\psi} = -.3y_1 - .1y_2 + .1y_3 + .3y_4$ , dessen Varianz offensichtlich kleiner ist.

Oft ist man beim Parameterschätzen nicht mit einer erwartungstreuen Schätzung zufrieden, sondern möchte darüber hinaus Konfidenzintervalle angeben. Solche Intervalle kann man für die GM-Schätzer konstruieren: Ist  $\hat{\psi} = \sum a_i y_i$  ein GM-Schätzer für eine parametrische Funktion  $\psi$ , so ist die Varianz von  $\hat{\psi}$  gleich  $\sigma^2 \sum a_i^2$ . Da  $\sigma^2$  unbekannt ist, liegt es nahe, dafür die Schätzung  $s^2$  zu nehmen. Es zeigt sich, daß das normalverteilte  $\hat{\psi}$  und  $s^2$  unabhängig sind (was sich geometrisch in der Orthogonalität der Verbindung von  $\mathbf{y}$  und  $\hat{\mathbf{y}}$  (deren quadrierte Länge  $(n-r)s^2$  ist) zu  $V$  widerspiegelt), und daher ist  $(\hat{\psi} - \psi) / \sqrt{s^2 \sum a_i^2}$   $t$ -verteilt mit  $(n-r)$  df, woraus sich unmittelbar das gewünschte Konfidenzintervall errechnet.

In der Regel interessieren mehrere identifizierbare parametrische Funktionen und nicht nur eine. Würde man für alle einzeln Konfidenzintervalle angeben, so würde man bekanntlich den Fehler der Kumulierung des  $\alpha$ -Niveaus begehen. Dem kann man auf unterschiedliche Arten ausweichen. Eine Möglichkeit ist die  $\alpha$ -Adjustierung, eine andere die Konstruktion mehrdimensionaler Konfidenzbereiche, die für mit linearer Algebra vertraute Leser kurz skizziert sei. Dazu sei für die  $q$  interessierenden parametrischen Funktionen  $\psi_j$  lineare Unabhängigkeit vorausgesetzt. Der Spaltenvektor  $\boldsymbol{\psi}$  enthalte als Komponenten die Werte dieser Funktionen und  $\hat{\boldsymbol{\psi}}$  deren GM-Schätzer. Die Koeffizienten der GM-Schätzer seien in einer Matrix  $\mathbf{A}$  so zusammengefaßt, daß die  $j$ -te Zeile die Koeffizienten für  $\psi_j$  enthält. Mit  $\mathbf{B} := \mathbf{A}\mathbf{A}'$  ist dann  $((\hat{\boldsymbol{\psi}} - \boldsymbol{\psi})' \mathbf{B}^{-1} (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) / q) / s^2$   $F$ -verteilt mit  $q$  und  $(n-r)$  df. Hieraus erhält man als Konfidenzbereich zum Niveau  $1 - \alpha$  ein „Ellipsoid“ mit Mittelpunkt  $\hat{\boldsymbol{\psi}}$  und Hauptachsen in Richtung der Eigenvektoren von  $\mathbf{B}$ , deren halbe Länge gleich der Wurzel aus dem Produkt des zugehörigen Eigenwertes mit  $s^2$ ,  $q$  und  $F_{q,n-r;\alpha}$  ist ( $F_{q,n-r;\alpha}$  ist  $\alpha$ -Fraktile der  $F_{q,n-r}$ -Verteilung, also der Wert, der dort rechts  $\alpha$  abschneidet).

Leider sind die so konstruierten Konfidenzbereiche für  $q > 3$  der Anschauung nicht mehr zugänglich. Dennoch erhält man mit ihrer Hilfe wichtige Informationen über die betrachteten  $q$  parametrischen Funktionen sowie über alle ihre Linearkombinationen in Form *simultaner Konfidenzintervalle*. Ist nämlich  $\phi = \sum c_j \psi_j$  irgendeine Linearkombination der  $\psi_j$ , so ist es plausibel, als Konfidenzbereich für  $\phi$  alle diejenigen Werte zu nehmen, die zu irgendeinem Element des Konfidenzbereiches für  $\boldsymbol{\psi}$  gehören (anders gesagt, das Bild des Konfidenzbereiches für  $\boldsymbol{\psi}$  unter der Abbildung, die jedem möglichen  $\boldsymbol{\psi}$  das zugehörige  $\phi$  zuordnet). Es ergeben sich so symmetrische Konfidenzintervalle um die GM-Schätzung  $\hat{\phi}$ , deren halbe Länge gleich  $\sqrt{q s^2 \sum a_i^2 F_{q,n-r;\alpha}}$  ist, wobei die  $a_i$  wieder die Koeffizienten des GM-Schätzers  $\hat{\phi}$  sind. Die Intervalle sind also um den Faktor  $\sqrt{q F_{q,n-r;\alpha} / t_{n-r;\alpha/2}}$  größer als die aus der Einzelschätzung ohne Adjustierung – dies ist der Preis, den man dafür zu zahlen hat, daß man mehr als nur eine Aussage machen möchte. Diese Intervalle, nach

Scheffé auch als S-Intervalle benannt, haben folgende Eigenschaft: Die Wahrscheinlichkeit, ein  $\mathbf{y}$  zu erhalten, so daß alle Linearkombinationen der Komponenten von  $\boldsymbol{\psi}$  durch die zugehörigen Intervalle überdeckt werden, ist  $1 - \alpha$ . Ein Beispiel für die Nützlichkeit simultaner Intervalle liefert die einfache lineare Regression, bei der man sich oft für den Erwartungswert von  $Y$  bei gegebenen möglichen Werten  $u$  des Prädiktors interessiert. Dieser Erwartungswert ist gleich  $au + b$ , also eine parametrische Funktion für jedes  $u$ , und deshalb liefern die S-Intervalle Konfidenzintervalle für diese Erwartungswerte simultan für alle Werte  $u$ .

Neben diesen S-Intervallen gibt es für den Spezialfall der Varianzanalyse mit gleichen Zellbesetzungen als Alternative noch die nach Tukey benannten T-Intervalle. Eine allgemeine Regel dafür, ob S-Intervalle, T-Intervalle oder Einzelintervalle mit  $\alpha$ -Adjustierung vorzuziehen sind, ist schwer angebar – dies hängt von der Zahl und Art der interessierenden parametrischen Funktionen ab, kann aber im Einzelfall (natürlich vor der Auswertung!) geprüft werden.

Abschließend zur Veranschaulichung ein Beispiel aus der einfachen linearen Regression: Für die Prädiktorwerte  $-1, 0, 1, 2, 3, 4, 5$  sollen von der Kriteriumsvariablen in der gleichen Reihenfolge folgende Werte vorliegen:  $2, -1, 2, 3, 4, 3, 8$ . Die Punktwolke und die geschätzte Regressionsgerade findet sich in Abbildung 2a. Die GM-Schätzungen für  $a$  und  $b$  sind dann beide gleich 1 und als Punktepaar in der Abbildung 2b eingetragen. Der Konfidenzbereich für  $a$  und  $b$  zum Niveau 95% ist das Innere der ebenfalls eingezeichneten Ellipse um diesen Punkt. Auf der  $a$ - und der  $b$ -Achse sind auch die Endpunkte von Konfidenzintervallen zum 95%-Niveau für diese Parameter angegeben, die mit E bezeichneten sind die für die Einzelparameterschätzung, die mit A und S bezeichneten sind die für die simultane Schätzung beider Parameter, wobei A für die  $\alpha$ -Adjustierung und S für die S-Intervalle steht. Man erkennt den Verlust, den man für die gleichzeitige Schätzung hinnehmen muß, ebenso, daß in diesem Fall die A-Intervalle kleiner sind als die S-Intervalle, was jedoch dadurch ausgeglichen wird, daß bei Adjustierung nicht ausgeschlossen werden kann, daß beide Parameter Null sind, während die zur S-Methode gehörende Ellipse diese Möglichkeit nicht enthält; ein Blick auf die Punktwolke macht es plausibel, daß nicht ausgeschlossen werden kann, daß für die „wahre“ Regressionsgerade Steigung oder Achsenabschnitt Null sind, wohl aber, daß diese beiden Parameter gleichzeitig Null sind. Ferner sieht man, daß die S-Intervalle die Projektionen der Ellipse auf die Achsen sind.

Was die *Hypothesenprüfung* angeht, so wurde schon in den Beispielen des 1. Teils deutlich, daß in einem speziellen Modell eine Vielzahl von Hypothesen interessant sein kann; alle dort betrachteten (Null-)Hypothesen ließen sich als Behauptungen über parametrische Funktionen formulieren, nämlich, daß eine gewisse Zahl  $q$  von linear unabhängigen identifizierbaren parametrischen Funktionen  $\psi_1, \dots, \psi_q$  bestimmte vorgegebene Werte annehmen. Der Fall beliebiger vorgegebener Werte läßt sich auf den reduzieren, daß alle Werte gleich 0 sind, und daher soll die hier betrachtete Nullhypothese lauten, daß die zu einer Spalte  $\boldsymbol{\psi}$  zusammengefaßten Funktionen  $\psi_1, \dots, \psi_q$  alle Null sind, kurz:  $H_0: \boldsymbol{\psi} = \mathbf{0}$  mit Alternativhypothese  $H_1: \boldsymbol{\psi} \neq \mathbf{0}$ .

Geometrisch bildeten die Erwartungswerte von  $\mathbf{y}$  bei allen möglichen Parameterbelegungen einen Unterraum  $V$ ; schränkt man sich auf diejenigen Parameterbelegungen ein, die mit der Nullhypothese verträglich sind, so erhält man eine Teilmenge

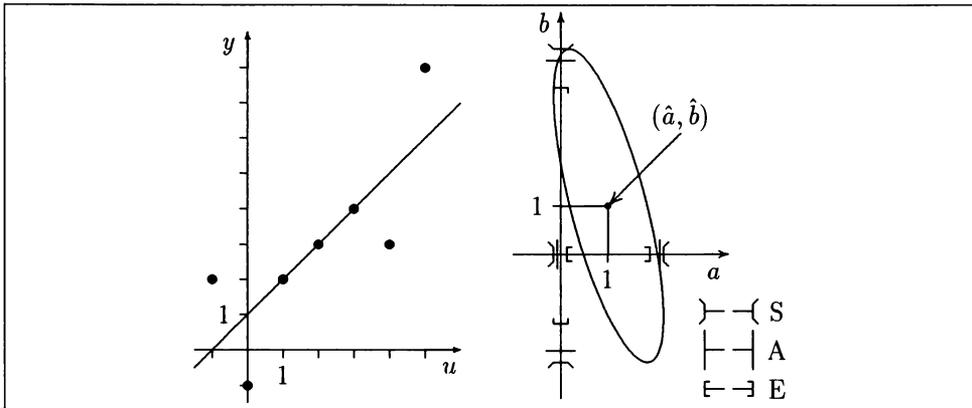


ABBILDUNG 2. Konfidenzbereiche und Konfidenzintervalle zum 95%-Niveau bei einer einfachen linearen Regression, a: Punktwolke und Regressionsgerade, b: Konfidenzellipse für  $a$  und  $b$ , auf den Achsen sind Konfidenzintervalle für  $a$  und  $b$  markiert. Dabei bedeuten S: Scheffé-Intervalle, A: Intervalle mit  $\alpha$ -Adjustierung, E: Intervalle ohne Adjustierung (zum Vergleich).

$V_0$  von  $V$ , die sich als Unterraum der Dimension  $r - q$  herausstellt (der Forderung, daß  $q$  Funktionen Null sein sollen, entsprechen  $q$  „verbotene“ Dimensionen).

Wenn im Beispiel getestet werden soll, ob die Geradensteigung  $a$  Null ist, so lautet die Nullhypothese  $a = 0$ , daher ist  $q = 1$  und die mit der Nullhypothese verträglichen Vorhersagen liegen auf einem eindimensionalen Unterraum  $V_0$ , also auf einer Geraden. Dies sieht man auch unmittelbar: bei Geradensteigung Null ist die Vorhersage für alle drei  $y$ -Werte gleich groß, nämlich  $b$ , die möglichen Vorhersagen im Personenraum sind also alle Punkte, deren drei Koordinaten übereinstimmen, und damit alle Vielfachen von  $\mathbf{x}_2$ , so daß  $V_0$  die Gerade ist, auf der  $\mathbf{x}_2$  liegt.

Unterwirft man das ursprüngliche Modell der Einschränkung, daß  $H_0$  gelten soll, so erhält man ein neues Modell, das „eingeschränkte Modell“. In gewisser Weise werden diese Modelle durch die Unterräume  $V$  und  $V_0$  (die Mengen der mit dem jeweiligen Modell verträglichen Erwartungswerte von  $\mathbf{y}$ ) repräsentiert. Wenn die Nullhypothese gilt, so sollte das ursprüngliche Modell nicht wesentlich besser zu den Daten passen als das eingeschränkte. Letzteres ist jedoch zu erwarten, wenn  $H_0$  nicht gilt. Zur Testung der Hypothese liegt es nahe, eine Parameterschätzung wie im ursprünglichen Modell auch im eingeschränkten Modell vorzunehmen, und dann zu prüfen, ob die hier vorhergesagten Werte wesentlich schlechter passen. Der geschätzte Erwartungswert von  $\mathbf{y}$  zu den unter  $H_0$  optimalen Parametern sei  $\hat{\mathbf{y}}$ , dies ist wieder der Fußpunkt des Lotes von  $\mathbf{y}$  auf  $V_0$ .

In Abbildung 1c ist  $\hat{\mathbf{y}}$  eingezeichnet. Zum Vergleich der Anpassungsgüte liegt es nahe, den Quotienten  $Q$  aus den Abständen  $d_0$  zwischen  $\mathbf{y}$  und  $\hat{\mathbf{y}}$  und  $d_1$  zwischen  $\mathbf{y}$  und  $\hat{\mathbf{y}}$  zu bilden und  $H_0$  dann zu verwerfen, wenn er hinreichend groß ist. Genau dies ist auch die übliche Vorgehensweise, wobei der Quotient jedoch noch einer monotonen Transformation unterworfen wird: Man betrachtet nämlich

$$F := \frac{n-r}{q} (Q^2 - 1) = \frac{(n-r)}{q} \frac{(d_0^2 - d_1^2)}{d_1^2}. \quad (11)$$

Bei Gültigkeit von  $H_0$  ist diese Statistik  $F$ -verteilt mit  $q$  und  $n - r$  Freiheitsgraden, und man verwirft  $H_0$ , falls  $F$  das  $\alpha$ -Fraktile  $F_{q,n-r;\alpha}$  übersteigt. Die Aussage über die Verteilung von  $F$  wird im nächsten Absatz plausibel gemacht, an dieser Stelle sei nochmals auf die geometrische Bedeutung der Anzahl der Freiheitsgrade („fehlende Dimensionen“) hingewiesen.

Bezeichnet man den Abstand zwischen  $\hat{y}$  und  $\hat{y}$  mit  $d_2$ , so gilt wegen des rechten Winkels zwischen den Strecken von  $\hat{y}$  nach  $y$  und  $\hat{y}$  nach  $\hat{y}$  ( $\hat{y}$  ist Lotfußpunkt!) nach dem Satz des Pythagoras die Beziehung  $d_1^2 + d_2^2 = d_0^2$ , weshalb der  $F$ -Bruch als  $(d_2^2/q)/(d_1^2/(n-r))$  geschrieben werden kann. Der Quotient  $d_2^2/\sigma^2$  ist nichtzentral  $\chi^2$ -verteilt mit  $q$  df und einem Nonzentralitätsparameter<sup>2</sup> (NZP)  $\delta$ , der gleich dem  $1/\sigma$ -fachen Abstand des zur wahren Parameterkonstellation gehörigen Erwartungswertes von  $V_0$  ist; daß  $d_1^2/\sigma^2$  eine  $\chi^2$ -Verteilung mit  $n-r$  df besitzt, wurde schon erwähnt. Ferner sind  $d_1^2$  und  $d_2^2$  unabhängig, weshalb  $F = (d_2^2/(q\sigma^2))/(d_1^2/((n-r)\sigma^2))$  nichtzentral  $F$ -verteilt ist mit NZP  $\delta$ . Dabei trifft  $H_0$  genau dann zu, wenn der Erwartungswert in  $V_0$  liegt, wenn also  $\delta$  gleich Null und damit die  $F$ -Verteilung eine zentrale ist. Mit  $\delta$  kann die *power* des Tests bestimmt werden.

Falls  $V_0$  aus allen Punkten besteht, deren sämtliche Koordinaten übereinstimmen (z.B. bei der üblichen Nullhypothese der Einwegvarianzanalyse oder in der einfachen linearen Regression bei der Hypothese, daß die Geradensteigung gleich Null ist), so erhält die Zerlegung  $d_1^2 + d_2^2 = d_0^2$  eine zusätzliche Bedeutung als Varianzzerlegung: Bis auf einen Faktor  $1/n$  sind nämlich die  $d_i^2$  (empirische) Varianzen,  $d_0^2$  die Varianz der  $Y$ -Werte,  $d_2^2$  die der aufgrund der Schätzung vorhergesagten Werte, und  $d_1^2$  die der „Fehler“ (In der Einwegvarianzanalyse sind  $d_0^2$ ,  $d_1^2$  und  $d_2^2$  die Quadratsummen  $SS_{total}$ ,  $SS_{within}$  und  $SS_{between}$ ). Derartige Varianzzerlegungen lassen sich für mehrfaktorielle varianzanalytische Designs noch verfeinern, und auch dort können sie als Berechnungen von Distanzen nach dem Satz des Pythagoras interpretiert werden.

Zum  $F$ -Test ist zu bemerken, daß man ihn auch als monotone Transformation eines Likelihood-Quotienten-Tests erhalten kann, womit er als Likelihood-Quotienten-Test bezeichnet werden kann. Über Optimalitätseigenschaften hinsichtlich der *power* informiert z.B. Scheffé (1959).

Häufig weisen Tests und Konfidenzintervalle eine enge Verwandtschaft auf. So auch hier: Es gilt nämlich, daß die Nullhypothese  $\psi = \mathbf{0}$  genau dann auf dem  $\alpha$ -Niveau zurückgewiesen wird, wenn das Konfidenzellipseid für  $\psi$  zum Niveau  $1 - \alpha$  den Nullpunkt des Koordinatensystems nicht enthält. Darüber hinaus besteht ein enger Zusammenhang mit den S-Intervallen: Der  $F$ -Test wird genau dann signifikant, wenn das S-Intervall mindestens einer Linearkombination der Komponenten von  $\psi$  die Null nicht enthält. Mit einer gewissen Berechtigung kann man dann die entsprechende parametrische Funktion (bei der Varianzanalyse also einen speziellen Kontrast) für die Zurückweisung der Nullhypothese verantwortlich machen: Der  $F$ -Test erweist sich als äquivalent zur simultanen Testung aller auf die parametrischen Funktionen bezogenen Nullhypothesen. Das Beispiel zur Konstruktion von Konfi-

<sup>2</sup>Wegen fehlender Einheitlichkeit seien kurz die hier verwendeten Definitionen gegeben: Sind  $x_1, \dots, x_q$  unabhängig normalverteilt mit Varianz 1 und gilt  $E(x_1) = \delta \geq 0$  und  $E(x_2) = \dots = E(x_q) = 0$ , so heißt die Verteilung von  $\sum x_i^2$  eine nichtzentrale  $\chi^2$ -Verteilung mit  $q$  df und NZP  $\delta$ , kurz  $\chi_{q,\delta}^2$ -Verteilung; sind  $u$  und  $v$  unabhängig und hat  $u$  eine  $\chi_{n_1,\delta}^2$ -Verteilung und  $v$  eine  $\chi_{n_2}^2$ -Verteilung, so heißt die Verteilung von  $(u/n_1)/(v/n_2)$  eine nichtzentrale  $F$ -Verteilung mit  $n_1$  und  $n_2$  df und NZP  $\delta$ . Für  $\delta = 0$  erhält man die bekannten „zentralen“ Verteilungen als Grenzfall.

denzbereichen läßt jedoch auch vermuten, daß der  $F$ -Test nicht in allen Fällen der vorteilhafteste ist. In der Tat wird man in vielen Fällen, wenn es um einzelne parametrische Funktionen geht, vor allem, wenn darüber gerichtete Hypothesen vorliegen,  $t$ -Tests mit  $\alpha$ -Adjustierung dem globalen  $F$ -Test vorziehen.

Abschließend sei zunächst auf Verallgemeinerungsmöglichkeiten des ALM hingewiesen. Die erste betrifft die schon erwähnte Möglichkeit, auch (Null-)Hypothesen zu testen, die nicht von der Form  $\psi = \mathbf{0}$  sind, sondern von der Form  $\psi = \mathbf{c}$  für einen festen Spaltenvektor  $\mathbf{c}$ . Dies Problem läßt sich durch eine einfache Transformation auf Bekanntes reduzieren. Die Forderung nach Unabhängigkeit und Varianzhomogenität der Fehler läßt sich abschwächen zu der Forderung, daß die Fehler eine gemeinsame nicht ausgeartete Normalverteilung besitzen, deren Kovarianzmatrix bis auf einen gemeinsamen Faktor  $\sigma^2$  festliegt. Dies erlaubt es zum Beispiel bei der einfachen Regression, Fälle zu behandeln, wo die Fehlervarianzen proportional sind zu den Werten der UV etc. Nach einer geeigneten Transformation können auch solche Fragestellungen mit den beschriebenen Methoden behandelt werden. Was schließlich die Normalverteilungsannahme angeht, so läßt sich zeigen, daß auch bei Nichterfülltheit unter einigermaßen milden Voraussetzungen bei „wachsendem Stichprobenumfang“ (was genauer zu präzisieren ist) die Tests und Konfidenzbereiche aus der Theorie des klassischen ALM „asymptotisch“ ihr Niveau einhalten, weshalb sie mit einiger Berechtigung auch bei „hinreichend vielen“ Beobachtungen von nicht normalverteilten Variablen Anwendung finden.

### 3 Weiterführende Literatur

Ausführliche elementare Einführungen geben Ward und Jennings (1973) und Moosbrugger und Klutky (1987), etwas fortgeschrittener auch Draper und Smith (1967) und Horton (1978). Das Buch von Fox (1984) ist in der Darstellung schon recht anspruchsvoll und behandelt auch verwandte Gebiete. Mathematisch etwas schwieriger sind die Werke von Schach und Schäfer (1978), Christensen (1987) und Searle (1970). Klassisch auf dem Gebiet der Varianzanalyse kann schließlich das Buch von Scheffé (1959) genannt werden.

#### Literaturverzeichnis

- Christensen, R. (1987). *Plain answers to complex questions*. New York: Springer.
- Draper, N. R. & Smith, H. (1967). *Applied regression analysis*. New York: Wiley.
- Fox, J. (1984). *Linear statistical models and related methods*. New York: Wiley.
- Hays, W. L. (1988). *Statistics* (4th ed.). Fort Worth: Holt, Rinehart and Winston.
- Horton, R. L. (1978). *The general linear model*. New York: McGraw-Hill.
- Moosbrugger, H. & Klutky, N. (1987). *Regressions- und Varianzanalysen auf der Basis des allgemeinen linearen Modells*. Bern: Huber.
- Schach, S. & Schäfer, T. (1978). *Regressions- und Varianzanalyse*. Berlin: Springer.
- Scheffé, H. (1959). *The analysis of variance*. New York: Wiley.
- Searle, S. R. (1970). *Linear models*. New York: Wiley.
- Ward, J. H. & Jennings, E. (1973). *Introduction to linear models*. Englewood Cliffs: Prentice Hall.