

Multivariate Verfahren

Johannes Andres

Der Bereich der multivariaten Statistik umfaßt eine große Klasse von Verfahren, bei denen es grob gesprochen um Beziehungen zwischen mehreren Variablen geht, im Gegensatz zu den univariaten Verfahren, die im wesentlichen nur eine oder zwei Variablen behandeln. Bisweilen werden bei einer solchen Einteilung nur „abhängige Variablen“ (AV) einbezogen, bisweilen auch „unabhängige“ (UV), weshalb einzelne Verfahren, wie zum Beispiel die multiple Regression, manchmal der univariaten Statistik und manchmal der multivariaten zugerechnet werden.

Dieses Kapitel soll eine einleitende Beschreibung einiger der wichtigsten Verfahren geben. Zunächst sollen multivariate Verallgemeinerungen univariater Verfahren besprochen werden, nämlich des t -Tests und der Varianzanalyse. In Situationen, in denen diese Verfahren zur Anwendung kommen können, gibt es die Alternative univariater Auswertungen, unter Umständen auch mehrere multivariate Möglichkeiten. Neben der Behandlung dieses Aspektes wird als ein Konstruktionsprinzip für multivariate Tests das *Union-Intersection*-Prinzip vorgestellt und seine Beziehung zu simultanen Konfidenzbereichen erläutert. Schließlich werden mit Diskriminanzanalyse und kanonischer Korrelation zwei Verfahren ohne univariate Analoga angesprochen. Weitere Verfahren wie die Faktorenanalyse werden in eigenen Kapiteln behandelt.

Viele hier als bekannt vorausgesetzte Sachverhalte der linearen Algebra werden knapp in dem Kapitel über Grundbegriffe der multivariaten Datenanalyse (Andres, in diesem Band) erläutert, auf das auch bezüglich der Notation verwiesen sei. Genaueres zu allen Themen findet man z.B. in Mardia, Kent und Bibby (1993).

1 Hotellings T^2

Viele univariate Tests können für die multivariate Situation verallgemeinert werden. Als ein erstes Beispiel soll der Einstichproben- t -Test dienen, bei dessen Verallgemeinerung es um die Frage geht, ob der Erwartungswert μ eines p -dimensionalen Zufallsvektors \mathbf{x} gleich einem vorgegebenen Vektor μ_0 ist (Nullhypothese) oder nicht (dies heißt, daß man die entsprechenden Fragen für alle p Komponenten gleichzeitig stellt). Von \mathbf{x} wird dabei Multinormalverteilt mit einer unbekannt regulären Kovarianzmatrix Σ vorausgesetzt. Beispielsweise könnten die Komponenten von \mathbf{x} die Werte von mehreren physiologischen Variablen sein, deren Erwartungswerte man unter Normalbedingungen kennt, und von denen man wissen möchte, ob ihre Erwartungswerte sich bei einem neuen Entspannungsverfahren ändern, genauer, ob dies bei mindestens einer der Fall ist. Der Versuch bestehe nun darin, daß man n mal unabhängig \mathbf{x} realisiert (im Beispiel also die Variablen an n zufällig ausgewählten Versuchspersonen mißt), was zu Ergebnisvektoren $\mathbf{x}_1, \dots, \mathbf{x}_n$ führt, die als unabhängig

$N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ -verteilt vorausgesetzt werden sollen. Der Mittelwertsvektor sei $\bar{\mathbf{x}}$ und die Kovarianzmatrix \mathbf{S} . In dieser Situation bildet man die Statistik $T^2 := (n-1)\mathbf{d}'\mathbf{S}^{-1}\mathbf{d}$ mit $\mathbf{d} := \bar{\mathbf{x}} - \boldsymbol{\mu}_0$ und verwirft H_0 für große Werte. Unter H_0 hat diese „Hotellings T^2 “ genannte Statistik eine sogenannte $T^2(p, n-1)$ -Verteilung und das $(n-p)/(p(n-1))$ -fache davon, also $((n-p)/p)\mathbf{d}'\mathbf{S}^{-1}\mathbf{d}$, eine $F_{p, n-p}$ -Verteilung, woraus sich der kritische Wert c zum Niveau α als $c = (p(n-1)/(n-p))F_{p, n-p; \alpha}$ berechnet ($F_{p, n-p; \alpha}$ ist das α -Fraktile der $F_{p, n-p}$ -Verteilung, also der Wert, der bei dieser Verteilung rechts α abschneidet); H_0 wird also verworfen, falls $T^2 \geq c$ ist.

An diesem in der Praxis sicher nur selten anwendbaren Test sollen wegen seiner Einfachheit einige Charakteristika der multivariaten Herangehensweise verdeutlicht werden. Als erstes wird man sicher nach der Plausibilität des Vorgehens fragen. Ersetzt man, motiviert durch die Tatsache, daß $E(\mathbf{S}) = ((n-1)/n)\boldsymbol{\Sigma}$ ist, in der Teststatistik $(n-1)\mathbf{S}^{-1}$ durch $n\boldsymbol{\Sigma}^{-1}$, so wird daraus die quadrierte Mahalanobis-Distanz von $\bar{\mathbf{x}}$ und $\boldsymbol{\mu}_0$, bezogen auf die Kovarianzmatrix $(1/n)\boldsymbol{\Sigma}$ von $\bar{\mathbf{x}}$. Dies macht die Testprozedur plausibel, ebenso wie die Beobachtung, daß sich im Falle $p = 1$ der quadrierte univariate t -Wert ergibt. Als nächstes liegt die Frage nahe, wie man zu dieser Teststatistik gelangt. Interessanterweise kann man die Teststatistik in diesem Fall durch zwei verschiedene, in der multivariaten Statistik sehr oft zur Konstruktion von Tests angewendete Prinzipien gewinnen, die in anderen Fällen auch zu verschiedenen Resultaten führen können, nämlich als *Likelihood-Quotienten-Test* und durch Anwendung des *Union-Intersection-Prinzips* („UI-Prinzip“).

Bei der *Likelihood-Quotienten-Methode* werden zunächst unter H_0 und H_1 die Verteilungsparameter jeweils so geschätzt, daß die Wahrscheinlichkeit (bzw. Wahrscheinlichkeitsdichte) der Daten maximal wird. Die jeweiligen Maxima werden dann zueinander ins Verhältnis gesetzt, wobei der auftretende Quotient oft noch einer geeigneten monotonen Transformation (die also die Größer-Relation erhält oder umkehrt) unterworfen wird (dies ist auch bei T^2 der Fall); ist der Quotient des Maximums unter H_1 und des Maximums unter H_0 groß (die Daten sind dann – etwas vergrößernd – unter H_1 viel wahrscheinlicher als unter H_0), so wird das als Beleg für H_1 gewertet. Eine Anwendung des UI-Prinzips beginnt damit, daß man zu einer komplexen (z.B. multivariaten) Nullhypothese H_0 eine geeignete Menge einfacherer (z.B. univariater) Nullhypothesen H_{0i} (i aus einer geeigneten Indexmenge I) findet, deren gemeinsame Gültigkeit zu der von H_0 äquivalent ist. In Analogie zur Wahrscheinlichkeitstheoretischen Bedeutung des Durchschnitts von Ereignissen als gemeinsamem Eintreten aller Ereignisse kann man auch vom Durchschnitt der H_{0i} sprechen und damit ihre gemeinsame Gültigkeit meinen – daher der Bestandteil „*Intersection*“. Anschließend testet man dann alle H_{0i} , wobei α natürlich zu adjustieren ist, und verwirft H_0 genau dann, wenn mindestens eine der H_{0i} verworfen werden kann. Bezeichnet A_i das Ereignis „Verwerfen von H_{0i} “, so ist das Ereignis „Zurückweisung von H_0 “ die Vereinigung der A_i (also das Ereignis, daß mindestens ein A_i eintritt) – daher der Bestandteil „*Union*“. So jedenfalls ist das Konstruktionsprinzip. Im konkreten Fall sollte es gelingen, die Tests der H_{0i} mit Hilfe einer geeigneten Statistik „auf einen Schlag“ durchzuführen.

Das UI-Prinzip soll nun am Beispiel von Hotellings Einstichproben- T^2 genauer erläutert werden. Die univariaten Hypothesen erhält man hier dadurch, daß man die multivariate Hypothese auf alle möglichen Arten mit Hilfe von Linearkombinationen

ins Univariate übersetzt. Ist \mathbf{a} ein p -Vektor, der nicht gleich $\mathbf{0}$ ist, so können die Komponenten von \mathbf{a} als Koeffizienten zur Bildung einer Linearkombination benutzt werden, nämlich der Linearkombination, die einem Vektor \mathbf{x} den Wert $\mathbf{a}'\mathbf{x}$ zuordnet. Für den Zufallsvektor \mathbf{x} ist dann $\mathbf{a}'\mathbf{x}$ eine eindimensionale $N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$ -verteilte Zufallsvariable. Ist H_0 richtig, also $\boldsymbol{\mu} = \boldsymbol{\mu}_0$, so muß der Erwartungswert von $\mathbf{a}'\mathbf{x}$ gleich $\mathbf{a}'\boldsymbol{\mu}_0$ sein; diese eindimensionale Nullhypothese sei mit $H_{0\mathbf{a}}$ bezeichnet. Aus der Gültigkeit aller $H_{0\mathbf{a}}$ folgt umgekehrt H_0 – diese folgt ja offenbar schon dann, wenn man statt aller \mathbf{a} nur die Einheitsvektoren nimmt, denn die entsprechenden Nullhypothesen besagen gerade, daß alle Komponenten von $\boldsymbol{\mu}$ und $\boldsymbol{\mu}_0$ übereinstimmen. Die eindimensionale $H_{0\mathbf{a}}$ würde man mit einem zweiseitigen t -Test testen und dann verwerfen, wenn die zugehörige t -Statistik einen kritischen Wert \tilde{c} betragsmäßig übersteigt. Dabei gehört \tilde{c} zu einem Signifikanzniveau $\tilde{\alpha}$, das nun so zu adjustieren ist, daß bei Gültigkeit von H_0 die Wahrscheinlichkeit, auch nur irgendein $H_{0\mathbf{a}}$ zu verwerfen, höchstens gleich dem gesetzten Ausgangsniveau α ist; das Niveau $\tilde{\alpha}$ soll für alle univariaten Hypothesen dasselbe sein. Die t -Statistik für $H_{0\mathbf{a}}$ errechnet man leicht zu $\frac{\sqrt{n-1}(\mathbf{a}'\bar{\mathbf{x}} - \mathbf{a}'\boldsymbol{\mu}_0)}{\sqrt{\mathbf{a}'\mathbf{S}^{-1}\mathbf{a}}}$ und $H_{0\mathbf{a}}$ wird genau dann verworfen, wenn das Quadrat dieses Ausdrucks $\geq \tilde{c}^2$ ist. Mindestens eine der univariaten Nullhypothesen wird also genau dann verworfen, wenn irgendeines dieser Quadrate $\geq \tilde{c}^2$ ist. Nun existiert das Maximum der Werte $(n-1)(\mathbf{a}'(\bar{\mathbf{x}} - \boldsymbol{\mu}_0))^2 / \mathbf{a}'\mathbf{S}^{-1}\mathbf{a} = (n-1)\mathbf{a}'(\mathbf{d}\mathbf{d}')\mathbf{a} / \mathbf{a}'\mathbf{S}^{-1}\mathbf{a}$ und berechnet sich zu $(n-1)\mathbf{d}'\mathbf{S}^{-1}\mathbf{d}$, also zu Hotellings T^2 . Damit wird also mindestens eine der $H_{0\mathbf{a}}$ genau dann verworfen, wenn $T^2 \geq \tilde{c}^2$ gilt. Zum Schluß ist nur noch \tilde{c}^2 so zu bestimmen, daß dieses Ereignis bei Richtigkeit von H_0 mit Wahrscheinlichkeit α eintritt, und daraus folgt unter Berücksichtigung der oben erwähnten Verteilung von T^2 , daß $\tilde{c}^2 = (p(n-1)/(n-p))F_{p,n-p;\alpha}$ sein muß. Man hat so in der Tat alle univariaten Hypothesen auf einmal getestet, wobei noch nicht einmal das adjustierte $\tilde{\alpha}$ bestimmt werden mußte. Dieses kann natürlich nun aus \tilde{c} ermittelt werden. Wichtiger ist jedoch die Kenntnis des zugehörigen kritischen Wertes \tilde{c} für die Einzeltests.

Wird nun beim Testen H_0 auf α -Niveau verworfen, so bedeutet das, daß damit auch mindestens eine $H_{0\mathbf{a}}$ beim Testen auf $\tilde{\alpha}$ -Niveau verworfen werden kann (der kritische Wert des zugehörigen t -Test ist \tilde{c}). Man wird dann geneigt sein, sich nicht nur gegen H_0 , sondern auch gegen alle derartigen $H_{0\mathbf{a}}$ zu entscheiden. Eine Zusatzüberlegung zeigt, daß die Wahrscheinlichkeit einer Fehlentscheidung bei diesem Vorgehen $\leq \alpha$ ist, wobei mit „Fehlentscheidung“ hier die Entscheidung gegen mindestens eine richtige $H_{0\mathbf{a}}$ gemeint ist. In diesem Sinne ist der T^2 -Test auch als simultane Testung aller $H_{0\mathbf{a}}$ bei impliziter Adjustierung des Signifikanzniveaus interpretierbar. Einer der großen Vorzüge dieses Tests liegt also darin, daß man bei einem signifikanten Gesamtergebnis ohne weitere Adjustierprobleme sofort Ergebnisse für die Linearkombinationen erhält, denen in vielen Fällen das eigentliche Interesse gilt. Dies ist übrigens ähnlich bei den Kontrasten in der Varianzanalyse, und in der Tat ist auch der F -Test als UI-Test bezüglich aller Kontraste konstruierbar.

Gelegentlich kann es zu der Situation kommen, daß das Testergebnis signifikant wurde, daß aber bei der anschließenden Untersuchung der Einzelvariablen nirgends ein signifikanter Unterschied auftritt, so daß man sozusagen den entdeckten Unterschied keiner der p Ausgangsvariablen zuschreiben kann. Dies ist zwar für die Interpretation unangenehm, zeigt aber deutlich den Vorteil der Einbeziehung der Linearkombinationen in die Betrachtung: Die Unterschiede in den Einzelvariablen sind

in solchen Fällen, informell gesprochen, zu klein, um bemerkt zu werden; erst durch eine geeignete Linearkombination gelingt es, sie so zu bündeln, daß sie sichtbar werden (mindestens bei einer Linearkombination kann man sich ja in der anschließenden Einzeluntersuchung für die Alternativhypothese entscheiden).

Bei der Konstruktion des Tests wurden alle denkbaren Linearkombinationen betrachtet. Die Frage liegt nahe, ob man hier vielleicht nicht des Guten zuviel getan hat, indem man neben interessanten Linearkombinationen (zum Beispiel den Variablen selber) eine unübersehbare Fülle uninteressanter berücksichtigt hat. Die Alternative läge in der Testung aller zu interessierenden Linearkombinationen gehörenden univariaten Hypothesen bei entsprechender Adjustierung des Signifikanzniveaus für die Einzeltests. In vielen Fällen wird tatsächlich diese Alternative vorzuziehen sein; die Entscheidung darüber trifft man, indem man die kritischen Werte der t -Tests bei adjustiertem Signifikanzniveau mit dem oben angegebenen \tilde{c} vergleicht und sich dann für das Verfahren entscheidet, das eher zu signifikanten Ergebnissen führt. Bei genauen inhaltlichen Vorstellungen und konkreten Hypothesen über bestimmte Linearkombinationen wird man also nicht automatisch den T^2 -Test verwenden, sondern vorher die Alternativen prüfen – eine weitere Parallele zur Varianzanalyse.

Das Gesagte soll an einem Beispiel weiter verdeutlicht werden. Bei einem *treatment* hält man es für möglich, daß es auf zwei Variablen Einfluß nimmt im Sinne einer Änderung der Erwartungswerte, von denen bekannt sei, daß sie ohne *treatment* beide gleich Null sind. Die Voraussetzungen zur Anwendung von Hotellings T^2 seien erfüllt. Hier ist H_0 die Hypothese $\boldsymbol{\mu} = \mathbf{0}$, und H_1 behauptet, daß mindestens einer der beiden Erwartungswerte $\neq 0$ ist. Für einen nicht mit multivariaten Methoden Vertrauten bietet es sich hier an, für beide Variablen die Hypothese, daß der Erwartungswert Null ist, mit einem t -Test auf $\alpha/2$ -Niveau zu testen und die Nullhypothese dann zu verwerfen, wenn mindestens einer der beiden Tests signifikant wird - in diesem Fall kann man sich sogar für die entsprechende univariate Alternativhypothese entscheiden. Die multivariate Alternative zu diesem Vorgehen ist der T^2 -Test. Zur Frage, ob ein Test dem anderen vorzuziehen sei, hat man die *power* zu betrachten (vgl. Willmes sowie Buchner, Erdfelder & Faul, in diesem Band). Dies sei für den Fall von 10 untersuchten Personen und $\alpha = .05$ für zwei mögliche Parameterkonstellationen getan: In beiden Fällen sei $\boldsymbol{\Sigma}$ die Einheitsmatrix, im ersten Fall sei $\mu_1 = 1$ und $\mu_2 = 0$, im zweiten Fall seien diese Erwartungswerte .6 und .8. Die *power* des T^2 -Tests ist für beide Fälle gleich .64, bei der Kombination der Einzeltests ist die *power* im ersten Fall gleich .69 und im zweiten Fall .62. Wie man sieht, ist keine der beiden Möglichkeiten der anderen grundsätzlich überlegen, es gibt also hier kein Patentrezept. Der T^2 -Test ist in diesem speziellen Fall dann überlegen, wenn beide Erwartungswerte etwa gleich stark von Null abweichen. Hier wird der Vorteil deutlich, den die Betrachtung nicht nur der Variablen, sondern auch aller Linearkombinationen hat: Die Unterschiede treten hier bei der Summenbildung stärker hervor. Übrigens hängt die *power* von den Verteilungsparametern nur über die quadrierte Mahalanobis-Distanz $(\boldsymbol{\mu} - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)$ ab, was man so deuten kann, daß keine Richtung der Abweichung bevorzugt wird; dies spricht dafür, Hotellings T^2 dann anzuwenden, wenn man weder Vorstellungen über $(\boldsymbol{\mu} - \boldsymbol{\mu}_0)$ noch über $\boldsymbol{\Sigma}$ hat.

An dieser Stelle ist eine kurze Bemerkung über die α -Adjustierung angebracht. Auf die Notwendigkeit entsprechender Überlegungen wird generell hingewiesen, ob-

wohl sie in der Praxis bei vielen Verfahren, man denke an die vielen Tests aller möglichen Haupteffekte und Interaktionen einer mehrfaktoriellen Varianzanalyse, traditionell unberücksichtigt bleiben. Als ein Vorzug vieler multivariater Verfahren gilt, daß man sich um derartige Adjustierung nicht zu kümmern braucht. Vor diesem Hintergrund soll in der weiteren Diskussion die Forderung nach der Beschränkung der Wahrscheinlichkeit eines Fehlers erster Art auch beim Testen mehrerer Hypothesen ernst genommen und immer wieder zum Thema gemacht werden.

Tests hängen oft eng mit Konfidenzbereichen zusammen, so auch hier. Um der einfacheren Formulierung willen soll die Entscheidungsregel für den Einstichproben- T^2 -Test unwesentlich abgeändert werden in der Weise, daß H_0 nur zu verwerfen ist, wenn T^2 größer als der kritische Wert ist (die Wahrscheinlichkeit, daß sich exakt der kritische Wert ergibt, ist sowieso gleich Null). Setzt man nun für μ_0 den wahren Erwartungswert μ ein, so wird die (dann richtige) H_0 genau dann verworfen, wenn $\mathbf{d}'\mathbf{S}^{-1}\mathbf{d} > k^2$ ist, wobei $k = \{(p/(n-p))F_{p,n-p;\alpha}\}^{1/2}$ ist. Dies ist äquivalent dazu, daß μ nicht in $\mathcal{E}(\mathbf{S}, \bar{\mathbf{x}}, k)$ liegt. Da die Wahrscheinlichkeit, H_0 zu verwerfen, gleich α ist, ist die Wahrscheinlichkeit des Gegenteils gleich $1 - \alpha$, und dieses Gegenteil ist äquivalent dazu, daß μ in $\mathcal{E}(\mathbf{S}, \bar{\mathbf{x}}, k)$ liegt. Damit ist das Ellipsoid $\mathcal{E}(\mathbf{S}, \bar{\mathbf{x}}, k)$ ein Konfidenzbereich zum Niveau $1 - \alpha$. Wäre zum Beispiel die Stichprobe aus Abbildung 2 des Kapitels über Grundbegriffe der multivariaten Datenanalyse (vgl. Andres, in diesem Band) eine Zufallsstichprobe aus einer Population, in der man die beiden Variablen als gemeinsam normalverteilt voraussetzen kann, so wäre die um den Faktor $\{(2/3)F_{2,3;.05}\}^{1/2} = 2.52$ vergrößerte Ellipse in dieser Abbildung der Konfidenzbereich für μ zum 95%-Niveau. Allgemein ist der Zusammenhang des Konfidenzbereichs mit dem Test von $H_0: \mu = \mu_0$ wie im Univariaten der, daß H_0 genau dann verworfen wird, wenn μ_0 nicht in dem Konfidenzbereich für μ liegt.

Ist p größer als 3, so ist der Konfidenzbereich K für μ nicht mehr anschaulich faßbar. Ob ein bestimmter Punkt in K liegt oder nicht, hat eine Rechnung zu zeigen. Es wäre wünschenswert, auch für die Komponenten von μ , also die Erwartungswerte der Einzelvariablen, Konfidenzintervalle zu besitzen, vielleicht auch Konfidenzbereiche für Paare von Einzelvariablen. Ist allgemein \mathbf{A} eine $(q \times p)$ -Matrix mit maximalem Rang und \mathbf{b} ein q -Vektor ($q \leq p$), so folgt aus der Bedingung, daß $\mathcal{E}(\mathbf{S}, \bar{\mathbf{x}}, k)$ den Parameter μ enthält, daß dann auch $\mathbf{A}\mu + \mathbf{b}$ im Bild dieses Ellipsoids unter der affinen Abbildung $\mathbf{x} \mapsto \mathbf{A}\mathbf{x} + \mathbf{b}$ liegt. Dies Bild ist aber gerade das Ellipsoid $\mathcal{E}(\mathbf{A}\mathbf{S}\mathbf{A}', \mathbf{A}\bar{\mathbf{x}} + \mathbf{b}, k)$. Damit ist die Wahrscheinlichkeit, daß alle Ellipsoide $\mathcal{E}(\mathbf{A}\mathbf{S}\mathbf{A}', \mathbf{A}\bar{\mathbf{x}} + \mathbf{b}, k)$ (k jetzt wie oben) das zugehörige $\mathbf{A}\mu + \mathbf{b}$ enthalten, mindestens $1 - \alpha$. Man hat damit, wie man sagt, eine Familie von *simultanen Konfidenzbereichen* für die $\mathbf{A}\mu + \mathbf{b}$ zum Niveau $1 - \alpha$ gefunden, anders ausgedrückt (unendlich) viele Konfidenzbereiche bei gleichzeitiger impliziter Adjustierung der Irrtumswahrscheinlichkeit. Darunter sind z.B. Konfidenzintervalle für die Komponenten von μ , die natürlich größer sind als die entsprechenden Intervalle bei einer Einzelschätzung ohne Adjustierung (für \mathbf{A} wählt man hier den entsprechenden transponierten Einheitsvektor und für \mathbf{b} die Zahl Null). Geometrisch sind z.B. im Zweidimensionalen diese Konfidenzintervalle die Projektionen der Konfidenzellipse auf die Achsen.

Wenn es auch scheinen mag, als sei der Einstichproben- T^2 -Test nur von geringer praktischer Bedeutung, so hat er doch eine wichtige Anwendung bei wiederholten Messungen. Beispielsweise interessiere man sich dafür, ob sich der Erwartungswert

einer Variable mit wachsendem Alter oder infolge mehrerer Interventionen ändert. Man wird dann die Variable zu den interessierenden p Zeitpunkten an n Personen erheben. Die Auswertung mit einer Varianzanalyse für wiederholte Messungen ist dann problematisch, wenn man an der Homogenität der Kovarianzmatrix zweifelt (also daran, daß die – theoretischen – Varianzen zu allen Meßzeitpunkten gleich groß sind und ebenso alle Kovarianzen zwischen je zwei Zeitpunkten). In diesem Fall kann eine multivariate Betrachtung weiterhelfen. Hier faßt man zunächst die p Zufallsvariablen, die die betrachtete Variable zu den einzelnen Meßzeitpunkten modellieren, zu einem Zufallsvektor \mathbf{x} zusammen, von dem Multinormalverteiltheit vorauszusetzen ist. Die üblicherweise hier betrachtete Nullhypothese, daß alle Komponenten von \mathbf{x} den gleichen Erwartungswert haben, ist jedoch nicht von der Form der Nullhypothese des Einstichprobenproblems (hier behauptet die Nullhypothese nicht, daß $\boldsymbol{\mu}$ gleich einem bestimmten $\boldsymbol{\mu}_0$ ist, sondern daß alle Komponenten von $\boldsymbol{\mu}$ gleich sind, daß $\boldsymbol{\mu}$ also ein Vielfaches des Vektors $(1, \dots, 1)'$ ist). Wie beim t -Test für abhängige Stichproben kann man nun jedoch Differenzen zwischen je zwei Komponenten bilden und fragen, ob diese Null sind. Genauer bildet man zum Beispiel $p - 1$ neue Zufallsvariablen $y_1 := x_2 - x_1, \dots, y_{p-1} := x_p - x_{p-1}$, deren Erwartungswerte bei Gültigkeit der Nullhypothese alle Null sein müßten, woraus umgekehrt auch die Nullhypothese folgt. Diese neuen Variablen faßt man nun zu einem Zufallsvektor $\mathbf{y} = \mathbf{R}\mathbf{x}$ zusammen (die Zeilen der zugehörigen Matrix \mathbf{R} haben alle die Form $(0, \dots, 0, -1, 1, 0, \dots, 0)$). Als lineare Transformation von \mathbf{x} ist auch \mathbf{y} multinormalverteilt. Die ursprüngliche Nullhypothese ist äquivalent dazu, daß der Erwartungswert von \mathbf{y} der Nullvektor ist, und dies läßt sich mit Hotellings T^2 testen. Die Linearkombinationen der y -Variablen entsprechen im übrigen gerade den Kontrasten einer entsprechenden Varianzanalyse, Tests und Konfidenzintervalle ergeben sich als Nebenprodukt. Der Verlust einer Dimension beim Übergang von \mathbf{x} zu \mathbf{y} wirkt sich verkleinernd auf die Konfidenzintervalle aus; diesen Vorteil bezahlt man damit, daß man ohne Zusatzüberlegungen zur α -Adjustierung keine Aussagen mehr über den Durchschnitt der Erwartungswerte machen kann (ebenso verhält es sich ja bei der Varianzanalyse mit dem „grand mean“). Interessiert man sich nicht für alle Erwartungswertdifferenzen, sondern nur für Linearkombinationen einer bestimmten Form, so kann man analog zur Technik der geplanten Vergleiche unter Umständen die Matrix \mathbf{R} durch eine mit weniger Zeilen ersetzen, was zu einem weiteren Gewinn an *power* führt. Diese Technik kann natürlich allgemein beim Einstichprobenproblem angewandt werden.

2 Multivariate Varianzanalyse

Neben dem Einstichproben- t -Test hat auch der Zweistichproben- t -Test eine multivariate Verallgemeinerung; es soll jedoch als nächstes Beispiel gleich die multivariate Varianzanalyse betrachtet werden. Hier könnte die Ausgangsfrage lauten, ob sich k Populationen im Hinblick nicht nur auf eine Variable wie im Univariaten, sondern auf p Variablen Y_1, \dots, Y_p unterscheiden, oder ob ein *treatment* mit k Stufen Auswirkungen auf nicht nur eine, sondern p Variablen hat. Zur Entscheidung plant man eine Untersuchung, in der man aus jeder Population Versuchspersonen (V_{pn}) zieht und bei jeder V_p die Werte in den p Variablen ermittelt; die Personen aus der j -ten

Population faßt man zu einer Zelle zusammen, die dann n_j Personen enthalte. Dabei sei $n := \sum n_j$ die Gesamtzahl der Personen. Den möglichen Ergebnisvektor \mathbf{y}_{ij} der i -ten Person in der j -ten Zelle setzt man als $N_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma})$ -verteilt voraus („Normalverteilungsannahme“), wobei das unbekannte $\boldsymbol{\Sigma}$ regulär und für alle Gruppen dasselbe sein soll („Homogenität der Kovarianzmatrizen“), ferner sollen die \mathbf{y}_{ij} gemeinsam unabhängig sein („Unabhängigkeitsannahme“). Die Nullhypothese ist wie im Univariaten die der Gleichheit aller $\boldsymbol{\mu}_j$, die Alternativhypothese behauptet unspezifisch, daß sich mindestens zwei dieser Vektoren unterscheiden. Wie beim Einstichproben- T^2 -Test wird man sich bei genaueren Vorstellungen über die erwarteten Unterschiede und die Kovarianzmatrix $\boldsymbol{\Sigma}$ überlegen, ob man nicht nach α -Adjustierung univariate Varianzanalysen für interessierende Linearkombinationen oder, vielleicht noch vorteilhafter, univariate Tests von bestimmten Kontrasten interessierender Linearkombinationen durchführt, statt die dabei oft größere *power* für ein eher nichtssagendes globales Ergebnis multivariater Tests zu verschenken.

Hat man sich für eine multivariate Varianzanalyse entschieden, so stehen immer noch mehrere Verfahren zur Verfügung; zum Beispiel führt die *Likelihood-Quotienten-Methode* hier im allgemeinen zu einem anderen Test als die *UI-Methode*, anders als beim Ein- und Zweistichprobenproblem. In den Tests spielen drei Matrizen eine wichtige Rolle, die die natürlichen Verallgemeinerungen der univariaten Quadratsummen sind, nämlich die Matrizen \mathbf{T} , \mathbf{B} und \mathbf{W} (die Bezeichnungen stehen für *total*, *between* und *within*). Die Matrix \mathbf{T} ist dabei die mit n multiplizierte Kovarianzmatrix aller Beobachtungen ohne Berücksichtigung der Zellzugehörigkeit, die Matrix \mathbf{B} erhält man, indem man für jede Person den Ergebnisvektor durch den Mittelwertsvektor der entsprechenden Zelle ersetzt und dann die zugehörige Kovarianzmatrix aller so modifizierten Ergebnisvektoren mit n multipliziert, und die Matrix \mathbf{W} ist die Summe der mit dem jeweiligen n_j multiplizierten getrennt für die Daten jeder einzelnen Zelle berechneten Kovarianzmatrizen. Damit spiegeln in der Tat diese Matrizen die Gesamtvariation der Daten, die Variation der Mittelwertsvektoren und die Variation innerhalb der Zellen wider, und analog zur Quadratsummenzerlegung gilt hier: $\mathbf{T} = \mathbf{B} + \mathbf{W}$. Die Matrizen \mathbf{T}/n , \mathbf{B}/n und \mathbf{W}/n seien naheliegenderweise kurz als „*Total*–“, „*Between*–“ und „*Within*-Kovarianzmatrix“ bezeichnet.

Die folgenden Vorüberlegungen sollen helfen, den in den Tests verwendeten Statistiken eine anschauliche Deutung zu geben. Der F -Test in der univariaten Varianzanalyse läuft grob gesehen darauf hinaus, daß man die Varianz zwischen den Gruppen durch die Varianz innerhalb teilt und H_0 dann verwirft, wenn die in dieser Weise standardisierte Varianz hinreichend groß ist. Im multivariaten Fall liegt es nahe, eine ähnliche Standardisierung durch eine lineare Transformation zu erreichen, die bewirkt, daß dadurch die „*Within*-Kovarianzmatrix“ zur Einheitsmatrix wird (das die Variabilität innerhalb der Zellen beschreibende Ellipsoid wird dadurch zu einer „Kugel“), und danach die „*Between*-Kovarianzmatrix“ der transformierten Daten zu untersuchen, genauer, deren Eigenwerte, in denen sich ja die Variabilität zwischen den Zellen widerspiegelt. Eine mögliche Transformation ist die Mahalanobis-Transformation bezüglich der „*Within*-Kovarianzmatrix“, die als transformierte „*Between*-Kovarianzmatrix“ die Matrix $\mathbf{W}^{-1/2}\mathbf{B}\mathbf{W}^{-1/2}$ liefert. Die Eigenwerte dieser Matrix sind die gleichen wie die der einfacher zu berechnenden Matrix $\mathbf{W}^{-1}\mathbf{B}$ und seien (der Größe nach) mit $\lambda_1, \dots, \lambda_p$ bezeichnet. Alternativ bietet

sich die Standardisierung an der Gesamtvarianz an, diesmal mit der Mahalanobis-Transformation bezüglich der „Total-Kovarianzmatrix“, die dazu führt, daß die Kovarianzmatrix aller Daten (ohne Rücksicht auf die Zellen berechnet) zur Einheitsmatrix wird. Danach sind die neuen „Within-“ und „Between-Kovarianzmatrizen“ gleich $\mathbf{T}^{-1/2}\mathbf{W}\mathbf{T}^{-1/2}$ und $\mathbf{T}^{-1/2}\mathbf{B}\mathbf{T}^{-1/2}$ und ihre Eigenwerte die gleichen wie die der Matrizen $\mathbf{T}^{-1}\mathbf{W}$ und $\mathbf{T}^{-1}\mathbf{B}$. Diese Eigenwerte hängen eng mit den λ_i zusammen, sie sind nämlich $1/(1 + \lambda_i)$ für $\mathbf{T}^{-1}\mathbf{W}$ und $\lambda_i/(1 + \lambda_i)$ für $\mathbf{T}^{-1}\mathbf{B}$ ($i = 1, \dots, p$). Sie reflektieren die Variabilität innerhalb und zwischen den Zellen nach Standardisierung.

Die *Likelihood*-Quotienten-Methode liefert nach einer monotonen Transformation als Teststatistik den als „Wilks' Λ “ bezeichneten Quotienten der Determinanten von \mathbf{W} und \mathbf{T} ; die Nullhypothese ist dabei für kleine Werte zu verwerfen. Dieser Quotient ist gleichzeitig die Determinante von $\mathbf{T}^{-1/2}\mathbf{W}\mathbf{T}^{-1/2}$ und kann, da die Determinante einer Kovarianzmatrix eine der möglichen multivariaten Verallgemeinerungen der Varianz ist, als Maß für die Variabilität innerhalb der Zellen nach Standardisierung mit der „Total-Kovarianzmatrix“ interpretiert werden. Bemerkenswert ist, daß die Teststatistik gleich dem Produkt der $1/(1 + \lambda_i)$ ist und damit nur von den Eigenwerten von $\mathbf{W}^{-1}\mathbf{B}$ abhängt. Die Verteilung der Teststatistik unter der Nullhypothese ist die sogenannte Wilks'- Λ -Verteilung mit den Parametern p , $n - k$ und $k - 1$ (analog zu den Freiheitsgraden der F -Statistik).

Die UI-Methode reduziert die multivariate Fragestellung wieder auf die entsprechenden univariaten Fragestellungen für alle Linearkombinationen der AV_n , ob sich die Zellenerwartungswerte irgendeiner solchen Linearkombination unterscheiden. Die Strategie, die multivariate Nullhypothese dann zu verwerfen, wenn der F -Bruch einer der zugehörigen univariaten Varianzanalysen hinreichend groß wird, führt dazu, als Teststatistik den maximalen derartigen F -Bruch zu wählen. Bis auf einen von den Stichprobengrößen abhängigen Faktor ist dies der größte Eigenwert λ_1 der Matrix $\mathbf{W}^{-1}\mathbf{B}$ (die zugehörige Linearkombination wird mit den Komponenten des zugehörigen Eigenvektors gebildet). Vertafelt sind meistens kritische Werte für $\theta := \lambda_1/(1 + \lambda_1)$ unter der Nullhypothese; wie bei Wilks' Λ ist die Verteilung von θ unter der Nullhypothese von den drei Parametern p , $n - k$ und $k - 1$ abhängig (die Bezeichnung ist leider uneinheitlich). Die Statistiken λ_1 und θ können als größte Eigenwerte der Matrizen $\mathbf{W}^{-1/2}\mathbf{B}\mathbf{W}^{-1/2}$ und $\mathbf{T}^{-1/2}\mathbf{B}\mathbf{T}^{-1/2}$ als Maße der Variabilität der Zellmittelwerte nach den beiden beschriebenen Standardisierungen gedeutet werden. Für θ findet man auch die Bezeichnung „Roy's largest root“ nach dem Begründer der UI-Methode (die Eigenwerte einer Matrix sind gleichzeitig die Nullstellen – „Wurzeln“ – ihres charakteristischen Polynoms).

Der aus der UI-Methode resultierende Test der multivariaten Nullhypothese ist dann wieder gleichwertig mit dem simultanen Test aller univariaten Nullhypothesen für die Linearkombinationen mit α -Adjustierung und, da die Varianzanalyse ihrerseits als UI-Test aufgefaßt werden kann, weitergehend gleichwertig mit dem simultanen Test aller Kontraste beliebiger Linearkombinationen der AV_n mit α -Adjustierung. Hieraus ergibt sich für den UI-Test der Vorteil, daß ohne weitere Überlegungen über Verletzung des Signifikanzniveaus post hoc Aussagen über solche Kontraste gemacht werden können. Nach dem Signifikantwerden der multivariaten Varianzanalyse kann sich z.B. die Frage stellen, ob ein durch einen Kontrastvektor \mathbf{c} gegebener Kontrast einer durch die Komponenten eines Vektors \mathbf{a} gegebener

nen Linearkombination der AVn sich von Null unterscheidet. Zur Beantwortung bildet man mit den Mittelwertvektoren $\bar{\mathbf{y}}_j$ der Zellen den standardisierten Kontrast $\sum c_j \mathbf{a}' \bar{\mathbf{y}}_j / \{ \sum (c_j^2 / n_j) \mathbf{a}' \mathbf{W} \mathbf{a} / (n - k) \}^{1/2}$ und vergleicht diesen mit dem kritischen Wert $\{(n - k) \theta_\alpha / (1 - \theta_\alpha)\}^{1/2}$, wo θ_α das α -Fraktile der zugehörigen θ -Verteilung ist. Ist der Betrag des Kontrasts größer als der kritische Wert, so kann man die Hypothese, daß der entsprechende theoretische Kontrast gleich Null ist, zurückweisen. Der Vergleich des angegebenen kritischen Wertes mit dem $\alpha/2$ -Fraktile der t_{n-k} -Verteilung, die man benutzen würde, wenn man ausschließlich diesen Kontrast testen wollte, macht den Verlust an *power* für die simultane Testung deutlich. Interessiert man sich z.B. bei fünf AVn und sieben Gruppen mit je 6 Vpn dafür, ob sich in der Summe der Variablen (z.B. dem Gesamtwert eines aus fünf Untertests bestehenden Intelligenztests) der Durchschnitt der Erwartungswerte der ersten beiden Gruppen von dem Durchschnitt der Erwartungswerte der letzten beiden unterscheidet, so ergibt sich für $\alpha = .05$ bei einem θ_α von .315 für den Post-hoc-Test ein kritischer Wert von 4.01 und für den Einzeltest ein kritischer Wert von 2.03.

Mit dem UI-Prinzip kann man auch simultane Konfidenzbereiche konstruieren. So ergibt sich für einen durch \mathbf{c} gegebenen Kontrast einer durch \mathbf{a} gegebenen Linearkombination als simultanes $1 - \alpha$ -Vertrauensintervall das Intervall um den empirischen Kontrast $\sum c_j \mathbf{a}' \bar{\mathbf{y}}_j$ mit halber Länge $\{ \sum (c_j^2 / n_j) \mathbf{a}' \mathbf{W} \mathbf{a} \theta_\alpha / (1 - \theta_\alpha) \}^{1/2}$. Im Beispiel wäre $\mathbf{a} = (1, 1, 1, 1, 1)'$ und $\mathbf{c} = (.5, .5, 0, 0, 0, -.5, -.5)'$, daher das Vertrauensintervall zu 95% das Intervall um die Differenz der Durchschnitte der ersten beiden und der letzten beiden Mittelwerte der Summe der Variablen mit der halben Länge $\{(1/6) \cdot \mathbf{a}' \mathbf{W} \mathbf{a} \cdot .46\}^{1/2}$. Auch hier ist dieses Intervall etwa doppelt so groß wie das entsprechende ausschließlich für diesen Kontrast konstruierte.

Die beiden beschriebenen Tests sind nicht die einzigen Möglichkeiten zum Testen der multivariaten Hypothese: Als weitere mögliche Teststatistiken finden sich unter anderem die „Pillai-Bartlett-Trace“ $\sum \lambda_i / (1 + \lambda_i)$ oder die „Hotelling-Lawley-Trace“ $\sum \lambda_i$, die die Spuren der Matrizen $\mathbf{T}^{-1/2} \mathbf{B} \mathbf{T}^{-1/2}$ bzw. $\mathbf{W}^{-1/2} \mathbf{B} \mathbf{W}^{-1/2}$ sind (gleichzeitig auch die Spuren von $\mathbf{T}^{-1} \mathbf{B}$ bzw. $\mathbf{W}^{-1} \mathbf{B}$). Beide Statistiken können daher als Maß für die Variabilität zwischen den Zellen nach einer Standardisierung bezüglich der „Total-Kovarianzmatrix“ bzw. der „Within-Kovarianzmatrix“ gedeutet werden. Die erste der beiden Teststatistiken kann man auch mit einer gewissen Berechtigung als ein Maß der durch die UV aufgeklärten Varianz ansehen. Allen hier besprochenen multivariaten Teststatistiken ist außerdem gemeinsam, daß sie von den Ergebnissen des Experiments nur über die Eigenwerte λ_i abhängen.

Für den Test der Nullhypothese der multivariaten Varianzanalyse stehen nun mehrere Statistiken zur Verfügung; nach welchen Kriterien soll man seine Auswahl treffen? Mögliche Kriterien sind die *power* und die Robustheit der Tests (ein Test ist umso robuster, je weniger sich Verletzungen der Verteilungsannahmen auf die Wahrscheinlichkeit des Fehlers 1. Art auswirken), daneben auch die Verträglichkeit mit Post-hoc-Tests. Das letzte Kriterium erfüllt der größte Eigenwert, er ist jedoch anfällig gegen Verletzungen der Verteilungsannahmen, während sich in vielen Situationen die Pillai-Bartlett-Trace-Statistik als der robusteste Test erweist (Olson, 1976). Zur Ermittlung der *power* benötigt man die Verteilungen der Teststatistiken unter möglichen Alternativhypothesen. Ist es im univariaten Fall nur ein Parameter, nämlich der Nonzentralitätsparameter der F -Verteilung, der die *power* bestimmt, so

sind in der multivariaten Situation mehrere Parameter entscheidend. Ersetzt man für alle Personen die Ergebnisvektoren durch die Erwartungswertvektoren der zugehörigen Zelle und bildet formal die Kovarianzmatrix Ψ dieser neuen Vektoren, so spiegelt diese Matrix gut die Variabilität der Erwartungswertvektoren wider (man beachte auch die Rolle der Stichprobengrößen) und entspricht damit (bis auf den Faktor $1/n$) der Matrix \mathbf{B} . Standardisiert man jetzt, wie oben beschrieben, an der wahren Kovarianzmatrix Σ der Fehler, so erhält man die Matrix $\Sigma^{-1/2}\Psi\Sigma^{-1/2}$, deren Eigenwerte $\gamma_1, \dots, \gamma_p$ (es sind übrigens wieder dieselben wie die Eigenwerte der Matrix $\Sigma^{-1}\Psi$) die Variabilität der standardisierten Erwartungswertvektoren charakterisieren. Sie entsprechen damit auf theoretischer Ebene den Eigenwerten λ_i von $\mathbf{B}^{-1}\mathbf{W}$, und da die Teststatistiken alle Funktionen der λ_i sind, ist es nicht verwunderlich, daß sich herausstellt, daß die *power* aller vier Tests nur abhängig ist von den Eigenwerten von $\Sigma^{-1}\Psi$, der Gesamtstichprobengröße und der Anzahl k der Zellen. Die γ_i spielen also die Rolle des univariaten Nonzentralitätsparameters. Die Anzahl der von Null verschiedenen Eigenwerte ist höchstens gleich dem Minimum s von p und $k - 1$ und hat geometrisch die Bedeutung der Dimension des kleinsten affinen Unterraums, der alle Erwartungswertvektoren enthält (entsprechendes gilt auch für die λ_i). Was die Größe der Eigenwerte angeht, so sind zwei Extremfälle denkbar, nämlich der, daß nur ein Eigenwert von Null verschieden ist, und der, daß s Eigenwerte von Null verschieden und untereinander gleich sind. Im ersten Fall liegen alle Erwartungswertvektoren auf einer Geraden, und man spricht von einer *konzentrierten Nonzentralitätsstruktur*. Im zweiten Fall ist sozusagen die Variabilität (nach Standardisierung) in allen Richtungen eines s -dimensionalen affinen Unterraums gleich groß und die Struktur gewissermaßen „maximal richtungslos“; dementsprechend nennt man eine solche Nonzentralitätsstruktur *diffus*. In mehreren Studien finden sich deutliche Hinweise darauf, daß die *power* im Falle näherungsweise konzentrierter Nonzentralitätsstrukturen bei Verwendung des größten Eigenwertes am höchsten ist, während bei eher diffusen Strukturen die drei anderen Tests größere *power* haben und sich eine Überlegenheit der Pillai-Bartlett-*Trace* andeutet. Zusammenfassend kommt Olson (1976) zu der Empfehlung, im allgemeinen diesen letzten Test zu benutzen.

Neben der im Grunde direkt aus dem Univariaten verallgemeinerten Fragestellung, ob sich die k Erwartungswertvektoren μ_j unterscheiden, taucht in der Situation der multivariaten Varianzanalyse gelegentlich eine weitere, typisch multivariate Frage auf, nämlich die nach der Dimensionalität der Menge der μ_j . Im Zusammenhang mit der Diskussion verschiedener Nonzentralitätsstrukturen hat sie sich bereits angedeutet. Es geht darum, wie groß die Dimension desjenigen affinen Unterraums ist, der von den Erwartungswertvektoren aufgespannt wird. Hat man zum Beispiel drei Gruppen, so werden die zugehörigen Vektoren in den meisten Fällen in einer Ebene (also einem affinen Unterraum der Dimension 2) liegen, es kann aber auch sein, daß die Situation dazu entartet, daß alle drei Vektoren auf einer Geraden liegen oder im Extremfall sogar zusammenfallen. Dann wäre die Dimension des aufgespannten affinen Unterraums 1 bzw. 0; der letzte Fall entspricht gerade der Nullhypothese der Varianzanalyse. Allgemein kann die Dimension dieses Raumes nicht größer sein als das Minimum der Zahlen p und $k - 1$, und es ist zu fragen, ob dieser Maximalwert wirklich erreicht wird. Wie zu erwarten, spielen beim Test der Hypothese, daß die Dimension des Unterraums größer als eine gerade untersuchte Zahl r ist, gegen

die Nullhypothese, daß die Dimension gleich r ist, die letzten $p - r$ betrachteten Eigenwerte $\lambda_{r+1}, \dots, \lambda_p$ von $\mathbf{W}^{-1}\mathbf{B}$ eine zentrale Rolle, in denen sich widerspiegelt, wie stark die standardisierten Mittelwertvektoren um denjenigen r -dimensionalen Unterraum streuen, der ihnen am nächsten liegt. Genauer benutzt man die Statistik

$$D_r^2 := (n - 1 - (p + k)/2) \sum_{i=r+1}^p \ln(1 + \lambda_i), \quad (1)$$

die unter der Nullhypothese asymptotisch $\chi_{(p-r)(k-r-1)}^2$ -verteilt ist.

Eine weitere naheliegende Frage ist die, „in welcher Richtung“ denn die Unterschiede zwischen den Erwartungswertvektoren am größten sind, in welcher Linearkombination der Variablen sich also die Unterschiede am deutlichsten zeigen. Will man gleichzeitig noch ein Maß für die Größe dieser Variationen (auf theoretischer Ebene) definieren, so liegt es nahe, Überlegungen wie bei der Definition von ω^2 in der univariaten Varianzanalyse anzustellen. Dazu stellt man sich ein zweistufiges Experiment vor, in dem in einem ersten Schritt zufällig eine der k Treatmentstufen ausgewählt wird und in einem zweiten Schritt dann eine Beobachtung in dieser Stufe realisiert wird. Die Auswahl der j -ten Stufe soll dabei – und darin liegt eine gewisse Künstlichkeit – mit Wahrscheinlichkeit n_j/n erfolgen. Bezeichnet man den sich so ergebenden Zufallsvektor mit \mathbf{y} , so ist der Erwartungswert von \mathbf{y} gleich $\boldsymbol{\mu} := \sum (n_j/n)\boldsymbol{\mu}_j$ und die Kovarianzmatrix gleich $\boldsymbol{\Psi} + \boldsymbol{\Sigma}$, worin sich die beiden „Variationsquellen“ des zusammengesetzten Experiments zeigen. Fragt man nun nach der Linearkombination der Komponenten von \mathbf{y} mit den deutlichsten Unterschieden zwischen den Gruppen, in dem Sinne, daß das univariate ω^2 dieser Linearkombination maximal wird, so sind deren Koeffizienten gerade die Komponenten des Eigenvektors zum größten Eigenwert γ_1 von $\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}$ und das zugehörige ω^2 ergibt sich als $\gamma_1/(1 + \gamma_1)$. Wie bei der Hauptkomponentenanalyse kann man weiter nach einer mit der ersten unkorrelierten Linearkombination fragen, die unter dieser Nebenbedingung ein maximales ω^2 besitzt, etc. Es ergeben sich dann die weiteren Eigenvektoren bei analogem Zusammenhang der Eigenwerte mit den jeweiligen ω^2 -Werten.

Standardisiert man diese p Linearkombinationen nun noch so, daß die Fehlervarianz jeweils gleich 1 wird, so erhält man die sogenannten *kanonischen Variaten* (auf theoretischer Ebene). Faßt man sie zu einer neuen p -dimensionalen Variablen \mathbf{z} zusammen, so ist \mathbf{z} eine lineare Transformation von \mathbf{y} mit bemerkenswerten Eigenschaften: Die Komponenten von \mathbf{z} sind unkorreliert, ebenso die „Fehleranteile“ in den Komponenten, die zusätzlich Varianz 1 haben. Wenn vorher die Dimension des durch die $\boldsymbol{\mu}_j$ aufgespannten affinen Unterrums gleich r war, so ist dies natürlich auch für die transformierten Erwartungswertvektoren der Fall, zusätzlich ist jedoch diese Ebene parallel zu der durch die ersten r Koordinatenvektoren aufgespannten, d.h. die letzten $p - r$ Koordinaten dieser Vektoren sind jeweils gleich, hier findet also keine treatmentbedingte Variation mehr statt. Die ersten Koordinaten haben außerdem die Eigenschaft, daß die durch das *treatment* bedingte Variation für sie sukzessive maximal im geschilderten Sinn ist. Diese übersichtliche Darstellung hat leider den Nachteil, daß die benötigten theoretischen Parameter meist nicht bekannt sind und man auf „Schätzungen“ angewiesen ist. Naheliegenderweise geht man dabei so vor, daß man die unbekanntenen Matrizen $\boldsymbol{\Psi}$ und $\boldsymbol{\Sigma}$ durch \mathbf{B} und \mathbf{W} ersetzt (eigentlich durch \mathbf{B}/n und \mathbf{W}/n , die plausible Schätzer für $\boldsymbol{\Psi}$ und $\boldsymbol{\Sigma}$ sind, der Faktor $1/n$

hebt sich jedoch in der weiteren Rechnung weg) und als Koeffizientenvektoren für die dann zu konstruierenden (empirischen) kanonischen Variaten die Eigenvektoren von $\mathbf{W}^{-1}\mathbf{B}$ wählt, die noch so normiert werden, daß die erwartungstreuen Schätzungen der Fehlervarianz in allen Komponenten gleich 1 sind. Analog gilt dann, daß die kanonischen Variaten unkorreliert sind, ebenso die zugehörigen Fehler, die zusätzlich gleiche Varianz haben. Darüber hinaus sind die Varianzen der Zellenmittelwerte in den neuen Koordinaten im obigen Sinn sukzessiv maximiert.

3 Diskriminanzanalyse

Hier wird schon ein weiterer verwandter Problemkreis angesprochen, nämlich der der Diskriminanzanalyse. Eine mögliche Ausgangssituation ist dabei die, daß man in k Populationen p Variablen erhoben hat und nun wissen möchte, wie man möglichst effektiv mit Hilfe dieser Variablen zwischen den Gruppen unterscheiden kann und wie man künftig die Zugehörigkeit von Individuen zu einer der Populationen allein mit Hilfe dieser Variablen mit möglichst hoher Treffsicherheit prognostizieren kann. Beispielsweise kann es darum gehen, aufgrund von Persönlichkeitstests psychisch Kranke psychiatrischen Kategorien zuzuordnen. Hier soll nur der einfachste Fall behandelt werden, in dem man voraussetzt, daß die betrachteten Variablen in den Populationen Normalverteilungen mit gleicher Kovarianzmatrix Σ , aber möglicherweise verschiedenen Erwartungswerten μ_j besitzen. Die Situation ist damit die der multivariaten Varianzanalyse. Zunächst soll diskutiert werden, wie man vorgehen würde, wenn Σ und die μ_j bekannt wären. Dann hätte man in den (theoretischen) kanonischen Variaten eine vielversprechende Transformation: die Kovarianzmatrix der transformierten Variablen ist für jede Population die Einheitsmatrix, die (relativen) Unterschiede zwischen den Erwartungswerten sind in der ersten Variate maximal, die damit in diesem Sinn am besten zwischen den Populationen unterscheidet, in den weiteren Variaten sind die Unterschiede jeweils maximal unter den angegebenen Nebenbedingungen und die Unterschiede verschwinden für alle Variaten mit einem Index, der größer ist als die Dimension r des durch die Erwartungswerte aufgespannten affinen Unterraums. Dies legt nahe, daß die ersten r Variaten alle zur Unterscheidung zwischen den Populationen nötige Information enthalten und daß man, wenn man aus Gründen der Übersichtlichkeit eine Beschreibung mit weniger Dimensionen anstrebt, immer die ersten Variaten wählen sollte. Will man ein weiteres Individuum aufgrund seiner Werte in den kanonischen Variaten klassifizieren, so liegt es nahe, nur die ersten r Variaten dazu heranzuziehen und das Individuum in diejenige Population einzuordnen, deren Erwartungswertvektor es in diesen ersten Variaten am nächsten kommt. Als Abstandsmaß würde man den normalen euklidischen Abstand wählen, da durch die Kovarianzstruktur keine Richtung bevorzugt ist. Ist zum Beispiel r gleich 2, so kann man die Situation in einer Ebene abbilden; jeder Population entspricht dann ein Punkt in der Ebene, nämlich der, dessen Koordinaten die Erwartungswerte der Population in diesen beiden Variaten sind, und ein durch einen weiteren Punkt repräsentiertes Individuum würde derjenigen Population zugeordnet, zu deren Erwartungswertvektor der Abstand minimal ist; man überzeugt sich leicht, daß die „Einzugsbereiche“ der Populationen dann konvexe Gebiete sind, die stückweise durch Strecken, Strahlen oder Geraden begrenzt sind. Die vorgeschla-

gene Regel kann anders auch so formuliert werden: Man ordnet ein Individuum derjenigen Population zu, zu deren Erwartungswertvektor die Mahalanobis-Distanz (bezüglich Σ) minimal ist. Die vorgeschlagene Zuordnungsregel ist tatsächlich optimal in dem Sinne, daß sie die Wahrscheinlichkeit einer Fehlklassifikation minimiert, falls die Apriori-Wahrscheinlichkeiten, daß das zufällig ausgewählte Individuum aus einer bestimmten Population stammt, für alle Populationen gleich groß sind (auch bei unterschiedlichen Apriori-Wahrscheinlichkeiten kann eine ähnliche optimale Zuordnungsregel gefunden werden, ebenso, falls die Fehlklassifizierungen zu unterschiedlichen Kosten führen und diese minimiert werden sollen).

Leider sind in den seltensten Fällen die Verteilungsparameter μ_j und Σ bekannt. Dann liegt es nahe, die gleiche Konstruktion mit geeigneten Schätzungen durchzuführen, um der optimalen Entscheidungsregel wenigstens nahezukommen. Das führt dann zu den empirischen kanonischen Variaten. Die mit den Komponenten der Eigenvektoren von $\mathbf{W}^{-1}\mathbf{B}$ gebildeten Linearkombinationen heißen deshalb auch *Diskriminanzfunktionen*. Es sei daran erinnert, daß die erste Diskriminanzfunktion gerade diejenige Linearkombination ist, in der die Gruppenmittelwerte die größte Varianz haben in dem Sinne, daß der F -Bruch der zugehörigen Varianzanalyse maximiert wird. Analoges gilt für die weiteren Diskriminanzfunktionen. Bei der Frage, ob genügend Diskriminanzfunktionen gefunden sind, wird man die Statistik (1) zu Rate ziehen. Über die Zugehörigkeit eines neuen Individuums zu einer der Populationen wird man wieder auf der Grundlage der Distanzen zu den Mittelwertvektoren entscheiden, wobei die korrekte Standardisierung der Diskriminanzfunktionen (die „Fehlervarianzen“ müssen gleich sein) wesentlich ist. Die angegebene Konstruktion kann auch durchgeführt werden, wenn die Verteilungsvoraussetzungen nicht als erfüllt betrachtet werden können. Die konstruierten Diskriminanzfunktionen haben dann immer noch die Eigenschaft, auf Stichprobenebene sukzessive das Verhältnis von Zwischengruppenvarianz zu Innergruppenvarianz zu maximieren, jedoch werden die Zuordnungsregeln ihren optimalen Charakter im allgemeinen verlieren.

4 Kanonische Korrelation

Als letztes Verfahren sei noch die kanonische Korrelation erwähnt. Hier geht es um die Beschreibung des Zusammenhangs von zwei Variablengruppen. Vielleicht hat man in einem Experiment bei den V_{pn} mehrere physiologische Variablen einerseits und mehrere psychologische Variablen andererseits erhoben, die jeweils Ähnliches messen sollen. Die Frage nach dem Zusammenhang kann man nun in der Weise präzisieren, daß man jeweils Linearkombinationen der Variablengruppen sucht dergestalt, daß die Korrelation zwischen diesen Linearkombinationen maximiert wird. Man könnte das so interpretieren, daß man beide Gruppen geeignet so durch je eine neue Variable repräsentiert, daß diese Repräsentanten eine maximale Korrelation aufweisen. Hat man solche Variablen gefunden, so kann man weiter fragen, wie der Zusammenhang zwischen dem ist, was durch sie nicht erfaßt wurde, und zwei neue Linearkombinationen suchen, die zu der jeweiligen ersten Linearkombination unkorreliert sind und unter dieser Nebenbedingung maximal korrelieren. Der Prozeß kann daraufhin analog fortgeführt werden. Das vorgeschlagene Verfahren erweist sich als durchführbar, falls z.B. die Kovarianzmatrizen der beiden Variablengruppen regulär

sind. Es gelingt dann insgesamt, die beiden Gruppen von Variablen in jeweils eine Gruppe von gleichviel neuen Variablen linear so zu transformieren, daß die neuen Variablen einer Gruppe alle unkorreliert sind und Varianz 1 besitzen, und so, daß darüber hinaus Variablen der einen Gruppe mit allen der anderen Gruppe unkorreliert sind, außer wenn sie den gleichen Index haben. Die jeweils ersten neuen Variablen haben dabei maximale Korrelation unter allen Linearkombinationen der beiden Variablengruppen, die zweiten haben maximale Korrelation unter allen mit den ersten neuen Variablen unkorrelierten Linearkombinationen etc. Die neuen Variablen nennt man dann auch *kanonische Korrelationsvariablen* und die Korrelationen die *kanonischen Korrelationskoeffizienten*.

Als Technik hat die kanonische Korrelation vielfältige Anwendung, eine Quer-Verbindung besteht z.B. zur Varianzanalyse: Man kann in dem am Ende bei der Beschreibung der kanonischen Variaten betrachteten zweistufigen Experiment einen neuen $(k - 1)$ -dimensionalen Zufallsvektor definieren durch die Vorschrift, daß er in allen Komponenten Nullen besitzen soll, außer in derjenigen, deren Nummer gleich der ausgewählten Gruppe ist (es wird also die Gruppenzugehörigkeit kodiert, wobei die letzte Gruppe durch den Nullvektor repräsentiert wird). Die kanonische Korrelation von \mathbf{y} mit dieser neuen Kodiervariablen liefert dann als kanonische Korrelationsvariablen auf Seiten von \mathbf{y} bis auf unterschiedliche Normierung gerade die kanonischen Variaten; die quadrierten kanonischen Korrelationen sind dann die als Analoga zu ω^2 charakterisierten Funktionen $\gamma_i/(1 + \gamma_i)$ der Eigenwerte γ_i von $\Sigma^{-1}\Psi$. Entsprechendes gilt auch auf Stichprobenebene bei Einführung von die Gruppenzugehörigkeit kodierenden „Dummy-Variablen“. Wenn solche Überlegungen auch nicht unbedingt zu größerer begrifflicher Klarheit beitragen, so können sie doch dazu dienen, mit Computerprogrammen zur kanonischen Korrelation bequem zu sonst nicht leicht erhältlichen Rechenergebnissen für andere Zwecke zu kommen.

5 Weiterführende Literatur

Für den mathematisch etwas Vorgebildeten ist vor allem die in sich geschlossene Darstellung von Mardia, Kent und Bibby (1993) zu empfehlen. Wer weniger mathematische Vorkenntnisse besitzt, findet gute Darstellungen der statistischen Prinzipien und der benötigten linearen Algebra in van de Geer (1971). Stark anwendungsorientiert ist das Buch von Stevens (1986). Eine sehr schöne anschauliche beschreibende Einführung geben Flury und Riedwyl (1983).

Literaturverzeichnis

- Flury, B. & Riedwyl, H. (1983). *Angewandte multivariate Statistik*. Stuttgart: Fischer.
- van de Geer, J. P. (1971). *Introduction to multivariate analysis for the social sciences*. San Francisco: Freeman.
- Mardia, K. V., Kent, J. T. & Bibby, J. M. (1993) *Multivariate analysis* (9th ed.). London: Academic Press.
- Olson, C. L. (1976). On choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin*, 83, 579–586.
- Stevens, J. (1986). *Applied multivariate statistics for the social sciences*. Hillsdale: Erlbaum.

Corrigenda

S. 202, 2. Abschnitt, Zeile 6

Nicht: „... und $\boldsymbol{\mu}_o$, bezogen ...“ sondern: „... und $\boldsymbol{\mu}_0$, bezogen ...“

S. 202, 1. Abschnitt, Zeile 15f

Nicht: „... leicht zu $\sqrt{n-1}(\mathbf{a}'\bar{\mathbf{x}} - \mathbf{a}'\boldsymbol{\mu}_0)/\sqrt{\mathbf{a}'\mathbf{S}^{-1}\mathbf{a}}$ und ...“

sondern: „... leicht zu $\sqrt{n-1}(\mathbf{a}'\bar{\mathbf{x}} - \mathbf{a}'\boldsymbol{\mu}_0)/\sqrt{\mathbf{a}'\mathbf{S}\mathbf{a}}$ und ...“

S. 202, 1. Abschnitt, Zeile 19

Nicht: „... der Werte $(n-1)(\mathbf{a}'(\bar{\mathbf{x}} - \boldsymbol{\mu}_0))^2/\mathbf{a}'\mathbf{S}^{-1}\mathbf{a} = (n-1)\mathbf{a}'(\mathbf{d}\mathbf{d}')\mathbf{a}/\mathbf{a}'\mathbf{S}^{-1}\mathbf{a}$...“

sondern: „... der Werte $(n-1)(\mathbf{a}'(\bar{\mathbf{x}} - \boldsymbol{\mu}_0))^2/\mathbf{a}'\mathbf{S}\mathbf{a} = (n-1)\mathbf{a}'(\mathbf{d}\mathbf{d}')\mathbf{a}/\mathbf{a}'\mathbf{S}\mathbf{a}$...“

S. 209, 2. Abschnitt, Zeile 3

Nicht: „... als simultanes $1 - \alpha$ -Vertrauensintervall ...“

sondern: „... als simultanes $(1 - \alpha)$ -Vertrauensintervall ...“