Diskriminanzanalyse

Helfried Moosbrugger und Tobias Richter

1 Grundlegendes

1.1 Problemstellungen der Diskriminanzanalyse

Die Diskriminanzanalyse ist ein multivariates statistisches Verfahren zur Analyse und Prüfung von Gruppenunterschieden. Ausgangspunkt ist in der Regel die folgende Datensituation: Aus k verschiedenen Populationen wurden Zufallsstichproben gezogen; für jedes der N Untersuchungsobjekte liegen Messwerte auf p verschiedenen Merkmalsvariablen vor (vgl. den Index der verwendeten Symbole und ihrer Bedeutungen im Anhang). Anhand dieser Informationen lassen sich mit der Diskriminanzanalyse Fragestellungen aus zwei verschiedenen, wenngleich logisch zusammenhängenden Problembereichen bearbeiten (zu dieser Unterscheidung vgl. z.B. Krzanowski & Marriott, 1995):

- *Diskrimination* zwischen Gruppen. Mit der hier ausschließlich behandelten linearen Diskriminanzanalyse (s.u.) kann eine geordnete Folge von *Diskriminanzfunktionen* bestimmt werden, die eine detaillierte Analyse multivariater Unterschiede zwischen Gruppen erlauben. Diskriminanzfunktionen sind diejenigen – voneinander unabhängigen – Dimensionen, die im Variablenraum eine optimale Trennung (Diskrimination) zwischen den Gruppen erzielen. Das diskriminatorische Potential der (in der Regel) intervallskalierten Merkmalsvariablen wird also auf eine geringere Anzahl von Dimensionen verdichtet; jede von ihnen repräsentiert diskriminatorische Information, die in den übrigen Diskriminanzfunktionen nicht enthalten ist. Weiterhin kann der Beitrag der ursprünglichen Merkmalsvariablen zu den einzelnen Diskriminanzfunktionen ermittelt werden, wodurch eine Beschreibung und gegebenenfalls eine Erklärung multivariater Unterschiede möglich wird. Umgekehrt lassen sich auf diesem Weg diejenigen Merkmalsvariablen seligieren, die zur Diskrimination zwischen den Gruppen am besten geeignet sind.

- *Klassifikation* von Objekten.[1] Ein weiteres Anwendungsgebiet der Diskriminanzanalyse besteht in der Formulierung von *Klassifikationsregeln*, anhand derer eine möglichst fehlerfreie Zuordnung „neuer" Objekte zu einer der definierten Gruppen erfolgen kann. Bei diesen „neuen" Objekten sind lediglich die Ausprägungen auf den p Merkmalsvariablen bekannt, nicht aber ihre Gruppenzugehörigkeit. Geschieht die Formulierung von Klassifikationsregeln auf parametrischer Basis, müssen die relevanten Parameter bekannt sein bzw. aus Stichprobendaten geschätzt werden und Annahmen über die Verteilungsform der Merkmalsvariablen in den Populationen gemacht werden. Die Verwendung von Diskriminanzfunktionen zu Klassifikationszwecken kann die Güte der Klassifikation erhöhen.

Die Bezeichnung „Diskriminanzanalyse" wird üblicherweise für eine Vielzahl von heterogenen Verfahren verwendet, die meist auf Klassifikationsprobleme bezogen sind, und die sich etwa hinsichtlich des Skalenniveaus der zugrundeliegenden Merkmalsvariablen, der zugrunde gelegten Verteilungsmodelle, der Einordnung als parametrische versus nicht-parametrische Verfahren oder der von einer Klassifikationsregel zu optimierenden Zielkriterien unterscheiden (für einen Überblick s. z.B. Das Gupta, 1973; Krauth, 1983). Im Mittelpunkt der folgenden Ausführungen zum Diskriminationsproblem steht die *lineare Diskriminanzanalyse*, die in ihrer Formulierung für den Fall von zwei Gruppen auf R. A. Fisher (1936) zurückgeht.[2] Ihre Ver-

---

[1] Der Begriff „Klassifikation" wird hier im Sinne der Zuordnung von Objekten zu bereits bekannten wohldefinierten Gruppen gebraucht (vgl. Schweizer, 1996). Die Diskriminanzanalyse ist damit von der Art klassifikatorischer Probleme abzugrenzen, wie sie von der Clusteranalyse bearbeitet werden. Von manchen Autoren (z.B. Kendall, 1966; Krauth, 1983) wird daher für diskriminanzanalytische Klassifikationsprobleme der Terminus „Diskrimination" verwendet, während „Klassifikation" allein die clusteranalytische Zielvorgabe bezeichnen soll, für eine Menge von Untersuchungsobjekten anhand bestimmter Merkmale eine sinnvolle Gruppeneinteilung erst zu bestimmen. Wir wollen demgegenüber an der oben ausgeführten Explikation der Begriffe „Diskrimination" und „Klassifikation" festhalten, um die beiden grundlegenden Anwendungsbereiche der Diskriminanzanalyse klar unterscheiden zu können. Eine forschungspraktische Beziehung zwischen Diskriminanz- und Clusteranalyse besteht übrigens darin, dass clusteranalytisch ermittelte Gruppierungen anhand eines neuen Variablensatzes diskriminanzanalytisch überprüft werden können (vgl. z.B. Moosbrugger & Frank, 1992).

[2] Im Folgenden wird deshalb meist abkürzend nur noch von „Diskriminanzanalyse" die Rede sein, auch wenn die lineare Diskriminanzanalyse im Speziellen gemeint ist.

allgemeinerung auf den Mehr-Gruppen-Fall, mithin die Begründung der *multiplen (linearen) Diskriminanzanalyse*, ist im Wesentlichen durch Rao (1948) und Bryan (1951) geleistet worden. Neben ihrer weiten Verbreitung[3] sprechen eine Reihe inhaltlicher Gründe für eine bevorzugte Darstellung der linearen Diskriminanzanalyse (vgl. Erb, 1990):

- *Breite der Erkenntnismöglichkeiten.* Bei der linearen Diskriminanzanalyse sind im Unterschied zu anderen diskriminanzanalytischen Verfahren die Problembereiche der Diskrimination und der Klassifikation eng aufeinander bezogen, so dass sich neben verbesserten Möglichkeiten der Formulierung von Klassifikationsregeln auch Lösungen für den Problembereich der Diskrimination zwischen Gruppen ergeben.
- *Robustheit.* Die Anwendung der linearen Diskriminanzanalyse setzt das Vorliegen von vergleichsweise restriktiven Bedingungen bezüglich des zugrunde gelegten Datensatzes voraus, insbesondere das Vorliegen intervallskalierter Merkmalsvariablen, die multivariate Normalverteilung der Merkmalsvariablen in den untersuchten Populationen und die näherungsweise Gleichheit der Varianz-Kovarianzmatrizen in den Gruppen. Trotzdem führt das Verfahren auch bei teilweiser Verletzung dieser Voraussetzungen zu befriedigenden Resultaten: Nominalskalierte Merkmalsvariablen können problemlos in dichotome Dummy-Variablen umkodiert werden, die sich rechnerisch wie intervallskalierte Variablen behandeln lassen. Von einer Verletzung der multivariaten Normalverteilung der Merkmalsvariablen ist die Anwendung von Signifikanztests sowie die Klassifikation von Objekten auf der Basis von Wahrscheinlichkeiten der Gruppenzugehörigkeit betroffen. Zumindest die Möglichkeit einer reliablen Klassifikation bleibt jedoch auch bei nur ungenauer Bestimmung der Wahrscheinlichkeiten bestehen, wenn die Gruppencentroide hinreichend getrennt sind (vgl. Lachenbruch, 1975). Liegen schließlich innerhalb der Gruppen stark ungleiche Streuungen der Merkmalsvariablen vor, kann eines der beiden Ziele der linearen Diskriminanzanalyse, die Diskrimination zwischen Gruppen, nicht verfolgt werden, da keine verlässlichen Schätzungen der wahren Streuungen in den einzelnen Gruppen möglich sind. Im-

---

[3] Nach Krauth (1983) ist die (multiple) lineare Diskriminanzanalyse in der psychologischen Forschung das am häufigsten eingesetzte diskriminanzanalytische Verfahren. Zudem ist sie das einzige Verfahren, das in sämtlichen der gängigen statistischen Programmpakete (SPSS, SPSS Inc., 1996; SAS, SAS Institute Inc., 1998; BMDP, Dixon, 1992) verfügbar ist.

merhin lassen sich aber auch in diesem Fall unter Umgehung der Schätzung von Diskriminanzfunktionen Klassifikationsregeln direkt auf Basis der Merkmalsvariablen formulieren.

- *Interpretierbarkeit.* Aufgrund des linearen Charakters der Diskriminanzfunktionen liefert die lineare Diskriminanzanalyse Ergebnisse, die einer inhaltlichen Interpretation verhältnismäßig leicht zugänglich sind. Auch die Inspektion der Ergebnisse mittels graphischer Veranschaulichungen wird gegenüber nichtlinearen Verfahren wesentlich vereinfacht.

1.2 Beziehungen der Diskriminanzanalyse zu anderen multivariaten Analyseverfahren

Die engen Beziehungen der Diskriminanzanalyse zu einer Reihe anderer statistischer Analyseverfahren lassen sich an ihren mathematischen Grundlagen insbesondere dann gut ablesen, wenn man sich verdeutlicht, dass die Diskriminanzanalyse prinzipiell einen Beitrag zu *jedem* Forschungsproblem leisten kann, das den Vergleich von zwei oder mehr als zwei Gruppen hinsichtlich eines multivariaten Datensatzes beinhaltet (vgl. Huberty, 1975).[4] Eine kurze Skizze der Stellung der Diskriminanzanalyse zu anderen gängigen statistischen Verfahren kann diesen Anspruch verdeutlichen:

Mit der *multivariaten Varianzanalyse* (MANOVA) hat die Diskriminanzanalyse die Variablenkonstellation gemeinsam, die den Ausgangspunkt der Analyse bildet. Während sich die genuine Zielsetzung der MANOVA auf die Prüfung von Hypothesen beschränkt, welche Gesamtunterschiede von Populationen hinsichtlich eines über mehrere Merkmalsvariablen operationalisierten, komplexen Merkmals betreffen, können wir mit der Diskriminanzanalyse die weitergehende Fragestellung untersuchen, inwieweit die einzelnen Variablen – im multivariaten Zusammenhang – zum Zustandekommen des Gesamtunterschiedes beitragen. Hierzu stellt die Diskriminanzanalyse als konfirmatorisches Verfahren auch Mittel zu einer detaillierten inferenzstatistischen Prüfung zur Verfügung.

---

[4] Wenn jedoch – wie vielfach in einschlägigen Lehrbuchkapiteln (z.B. Hartung & Elpelt, 1984) – unter dem Stichwort "Diskriminanzanalyse" lediglich Klassifikationsverfahren betrachtet werden, wird die Einbettung der Diskriminanzanalyse in den Kontext der multivariaten Statistik häufig nicht sichtbar.

Ein diskriminanzanalytisch zu bearbeitendes Forschungsanliegen kann demnach darin bestehen, multivariate Gruppenunterschiede im Rückgriff auf die beteiligten Merkmalsvariablen zu erklären bzw. – sofern "neue" Objekte vorgegebenen Gruppen zugeordnet werden sollen – Gruppenzugehörigkeiten zu prognostizieren. Daran wird die Nähe zu *regressionsanalytischen* Fragestellungen deutlich, mit dem einzigen Unterschied, dass die dort typischerweise intervallskalierte, kontinuierliche Kriteriumsvariable in der Diskriminanzanalyse durch eine nominalskalierte Gruppierungsvariable ersetzt wird. Daher führen im Zwei-Gruppen-Fall Diskriminanz- und Regressionsanalyse zu identischen Resultaten, wenn für die Gruppierungsvariable eine geeignete Dummy-Kodierung verwendet wird (vgl. z.B. Hattemer, 1974).

Analog lässt sich eine Diskriminanzanalyse mit mehr als zwei Gruppen – wie die multivariate Varianz- und Regressionsanalyse – in formaler Hinsicht als Spezialfall der *kanonischen Korrelationsanalyse* auffassen (vgl. Tatsuoka, 1953). Die Diskriminanzfunktionen repräsentieren dann kanonische Variablen; die kanonischen Korrelationskoeffizienten geben den Zusammenhang zwischen Linearkombinationen der Merkmalsvariablen auf der einen und Linearkombinationen der dummykodierten Gruppierungsvariablen auf der anderen Seite wieder. Demnach lässt sich auch die Diskriminanzanalyse unter das Allgemeine Lineare Modell subsumieren (vgl. z.B. Moosbrugger, 1983).

Die rechnerische Lösung des Diskriminationsproblems weist (wie die kanonische Korrelationsanalyse mit ausschließlich intervallskalierten Variablen) starke Bezüge zum Vorgehen bei der *Hauptkomponentenanalyse* auf (vgl. z.B. Tatsuoka, 1971). Als Resultat erhalten wir hier wie dort wechselseitig unabhängige Dimensionen, die gemeinsam einen Unterraum des Variablenraums konstituieren. Anders als die Faktoren der Hauptkomponentenanalyse werden die Diskriminanzfunktionen allerdings nach dem Kriterium extrahiert, eine sukzessiv maximale Trennung zwischen den Gruppen zu erreichen. Die Diskriminanzanalyse kann damit als variablenreduzierendes Verfahren verstanden werden, das unter Ausschöpfung des Diskriminationspotentials der ursprünglichen Merkmalsvariablen eine ökonomische, übersichtliche und gegebenenfalls interpretierbare Darstellung multivariater Gruppenunterschiede ermöglicht.

1.3 Überblick

Ausführliche Darstellungen der linearen Diskriminanzanalyse einschließlich ihrer formalen Bezüge zu anderen multivariaten Analysetechniken finden sich unter anderem bei Tatsuoka (1971), Cooley und Lohnes (1971) sowie Bortz (1993). Wegen der Beschränkung auf die lineare Diskriminanzanalyse können diskriminanzanalytische Verfahren für kategoriale und gemischtskalierte Datensätze (vgl. hierzu Goldstein & Dillon, 1978; Fahrmeir & Hamerle, 1984) und diskriminanzanalytische Verfahren für den Fall nicht-normalverteilter Gruppen (vgl. hierzu Trampisch, 1975) nicht behandelt werden. Die Monographie von Lachenbruch (1975) bietet einen Abriss des gesamten Bereichs der Diskriminanzanalyse. Einen Überblick über diskriminanzanalytische Techniken mit einem Schwerpunkt auf Klassifikationsproblemen geben Krauth (1983) sowie Deichsel und Trampisch (1985). Über die Grundzüge der Entwicklung verschiedener diskriminanzanalytischer Ansätze informiert die systematische Zusammenstellung von Das Gupta (1973).

Diskriminations- und Klassifikationsprobleme lassen sich prinzipiell unabhängig voneinander behandeln. Daher werden die beiden grundlegenden Anwendungsmöglichkeiten der Diskriminanzanalyse in zwei getrennten Abschnitten dargestellt (Abschnitt 2 und Abschnitt 3). Dabei wird die Analyse von Gruppenunterschieden als der logisch vorgängige Verfahrensschritt zuerst erörtert; eine an einer geometrischen Veranschaulichung orientierte Einführung in die Grundprinzipien des Verfahrens (Abschnitt 2.1) wird den mathematischen Grundlagen und der rechnerischen Durchführung der Diskriminanzanalyse (Abschnitt 2.2) vorangestellt. Alle wesentlichen Rechenschritte werden zudem exemplarisch an einem Datensatz aus der Persönlichkeitsforschung (Brednich, 1993) zur Anwendung gebracht (Abschnitt 2.3 und Abschnitt 3.4).

2 Diskrimination zwischen Gruppen

2.1 Grundprinzip der Diskriminanzanalyse

Das Grundprinzip der Diskriminanzanalyse soll zunächst anhand des einfachsten Falles von zwei Gruppen erläutert werden, bei dem *eine* Diskriminanzfunktion zur Gruppentrennung genügt (Abschnitt 2.1.1). Im nächsten Schritt wird dann gezeigt, wie sich das für den Zwei-Gruppen-Fall anschaulich begründete Diskriminanzkriterium auf den Mehr-Gruppen-Mehr-Variablen-Fall verallgemeinern lässt (Abschnitt 2.1.2). Als Resultat dieses allgemeinen Anwendungsfalls ergibt sich der durch eine Folge von Diskriminanzfunktionen konstituierte Diskriminanzraum, dessen wesentliche Eigenschaften im Anschluss dargestellt werden (Abschnitt 2.1.3).

2.1.1 Das Diskriminanzkriterium für den Zwei-Gruppen-Fall

Gegeben seien zwei Untersuchungsgruppen G_1 und G_2 mit jeweils n_1 und n_2 Untersuchungsobjekten, für die – im Minimalfall – Merkmalswerte auf zwei Variablen X_1 und X_2 erhoben wurden. Gesucht wird nun eine neue Achse Y in der von X_1 und X_2 aufgespannten Ebene, die eine optimale Trennung zwischen den Gruppen ermöglicht. Anders ausgedrückt: Es soll eine Diskriminanzvariable Y eingeführt werden, welche die Unterschiedlichkeit der Gruppen hinsichtlich der beiden Merkmalsvariablen simultan und unter Erfüllung eines Optimierungskriteriums (s.u.) beschreibt. Die bivariate Information wird also auf eine einzige Dimension verdichtet. Folgerichtig wird die gesuchte *Diskriminanzachse* oder *Diskriminanzfunktion* als Linearkombination

$$Y = v_1 X_1 + v_2 X_2 \tag{1}$$

der Merkmalsvariablen dargestellt; v_1 und v_2 repräsentieren die Gewichtungs- oder Diskriminanzkoeffizienten, welche die Lage der Diskriminanzachse in der Ebene der Merkmalsvariablen bestimmen. Für ein Untersuchungsobjekt i erhalten wir den *Diskriminanzwert* y_i durch Einsetzen der individuellen Merkmalswerte x_{1i} und x_{2i} in Gleichung 1:

$$y_i = v_1 x_{1i} + v_2 x_{2i}\,. \tag{2}$$

Das Problem besteht nun darin, die Gewichtungskoeffizienten v_1 und v_2 so zu bestimmen, dass durch die resultierende Diskriminanzachse die Gruppen G_1 und G_2 möglichst gut voneinander getrennt werden.

Wie lässt sich ein solches Zielkriterium näher präzisieren? Als Maß für die Güte der Gruppentrennung bieten sich zwei Indikatoren an. Zum einen ist anzustreben, den quadrierten Abstand D^2 der Gruppenmittelwerte $\bar{y}_1$ und $\bar{y}_2$ auf der Diskriminanzvariablen zu maximieren:

$$\begin{aligned} D^2 &= (\bar{y}_1 - \bar{y}_2)^2 \\ &= [v_1(\bar{x}_{11} - \bar{x}_{21}) + v_2(\bar{x}_{12} - \bar{x}_{22})]^2 \end{aligned} \tag{3}$$

Zum anderen können wir davon ausgehen, dass die Gruppen durch eine Diskriminanzfunktion umso besser getrennt sind, je kleiner der Überschneidungsbereich der gruppenspezifischen Streuungen der Diskriminanzwerte ist. Daraus folgt unmittelbar die Forderung, die Gesamtstreuung innerhalb (*within*) der Gruppen G_1 und G_2 (hier ausgedrückt als Quadratsumme $Q_w(Y)$ der Diskriminanzwerte innerhalb der Gruppen) zu minimieren:

$$Q_w(Y) = \sum_{i'=1}^{n_1} (y_{1i'} - \bar{y}_1)^2 + \sum_{i'=1}^{n2} (y_{2i'} - \bar{y}_2)^2 . \tag{4}$$

Wie man sich anhand eines graphischen Beispiels leicht veranschaulichen kann, sind die Forderung nach einer Maximierung der Mittelwertedifferenz der Gruppen auf der Diskriminanzvariablen und die Forderung nach einer Minimierung der Innergruppenstreuung der Diskriminanzvariablen nicht äquivalent.[5] Beide Maße werden von unterschiedlichen Diskriminanzfunktionen optimiert. Wir benötigen daher ein Kriterium, das beide Forderungen zusammenfasst. Dieses gewinnen wir, indem wir das Verhältnis aus der quadrierten Mittelwertedifferenz D^2 und der Innergruppenstreuung $Q_w(Y)$ bilden. Das *Diskriminanzkriterium* λ für den Zwei-Gruppen-Fall lautet demnach:

$$\lambda = \frac{D^2}{Q_w(Y)} \longrightarrow \max . \tag{5}$$

Für den Zwei-Gruppen-Fall gilt also gemäß der in Gleichung 5 ausgedrückten Zielfunktion der Diskriminanzanalyse, dass die Gewichtungskoeffizienten der Diskriminanzfunktion so zu schätzen sind, dass sich einerseits die Mittelwerte der Gruppen auf der Diskriminanzvariablen möglichst deutlich unterscheiden, andererseits die Streuungen der Diskriminanzwerte innerhalb der Gruppen möglichst klein sind.

---

[5] Dies liegt daran, dass sich auch die Gesamtstreuung der Diskriminanzwerte mit unterschiedlicher Lage der Diskriminanzachse im Merkmalsraum ändert.

2.1.2 Verallgemeinerung des Diskriminanzkriteriums auf den Mehr-Gruppen-Fall

Wenn wir uns – wie es in vielen Anwendungen sinnvoll sein mag – für die Diskrimination von mehr als zwei Gruppen interessieren, sind die einfachen Differenzen zwischen Gruppenmittelwerten als Maß für die Unterschiedlichkeit der Gruppen nicht mehr hinreichend, da eine einzige Diskriminanzfunktion die Unterschiede zwischen den Gruppen dann in der Regel nicht vollständig erfassen kann.[6] Wie sich anhand der mathematischen Lösung der Maximierung des Diskriminanzkriteriums zeigen lässt (s. Abschnitt 2.2.1), können bei k Gruppen und p Merkmalsvariablen k–1 Diskriminanzfunktionen extrahiert werden, sofern die Anzahl p der Merkmalsvariablen größer oder gleich der Anzahl k der Gruppen ist; ist die Anzahl k der Gruppen hingegen größer als die Anzahl p der Merkmalsvariablen, ergeben sich p Diskriminanzfunktionen.[7] Die Schätzung der einzelnen Diskriminanzfunktionen erfolgt dann nach dem Kriterium der *sukzessiv maximalen Trennung* der Gruppen: Durch die erste, über die Maximierung des Diskriminanzkriteriums λ zu schätzende Diskriminanzfunktion wird im Mehr-Gruppen-Mehr-Variablen-Fall nur ein Teil der Zwischengruppenvarianz der Merkmalsvariablen aufgeklärt. Wir suchen daher nach einer zweiten Diskriminanzfunktion, die mit der ersten Diskriminanzfunktion unkorreliert ist und eine optimale Trennung der Gruppen unter Ausnutzung der noch verbleibenden Merkmalsvarianz ermöglicht, und unter Umständen nach weiteren maximal diskriminierenden, aber mit den bereits extrahierten Diskriminanzfunktionen unkorrelierten Diskriminanzfunktionen. In dieser Weise wird verfahren, bis das gesamte Diskriminationspotential der Merkmalsvariablen ausgeschöpft ist, in der Regel bei einer Anzahl von $r = \min(p, k-1)$ Diskriminanzfunktionen. Diese bilden eine geordnete Folge,

---

[6] Eine einzige Diskriminanzfunktion wäre bei $k > 2$ Gruppen nur dann hinreichend, wenn – geometrisch gesprochen – sich die Gruppenunterschiede vollständig durch Projektion auf eine Gerade darstellen lassen. Dies ist etwa dann der Fall, wenn lediglich zwei Merkmalsvariablen erhoben wurden bzw. wenn nur zwei der p Merkmalsvariablen überhaupt zwischen den Gruppen diskriminieren (die Gruppenmittelwerte der übrigen p–2 Merkmalsvariablen hingegen identisch sind), da dann die Gruppencentroide auf *einer* Geraden zu liegen kommen. Analog sind auch bei $k > 3$ Gruppen zwei Diskriminanzfunktionen hinreichend, wenn sich die Gruppenunterschiede durch Projektion auf eine Ebene darstellen lassen etc.

[7] Um eine zufriedenstellende Trennung zwischen den Gruppen zu erreichen, empfiehlt es sich normalerweise, $p \geq k$ Merkmalsvariablen in die Analyse einzubeziehen (vgl. Backhaus, Erichson, Plinke & Weiber, 1996).

insofern sich der Beitrag, den eine Diskriminanzfunktion zur Aufklärung der Zwischengruppenvarianz leisten kann, mit jedem Extraktionsschritt verringert.

Für den allgemeinen Fall von $k \geq 2$ Gruppen ist es daher zweckmäßig, als Verallgemeinerung von Gleichung 5 – unter Ausnutzung der aus der Varianzanalyse bekannten Quadratsummenzerlegung der totalen Quadratsumme Q_t in Zwischengruppen-Quadratsumme (*between*-Quadratsumme) Q_b und Innergruppen-Quadratsumme (*within*-Quadratsumme) Q_w – anstelle der einfachen quadrierten Mittelwertdifferenzen die Zwischengruppen-Quadratsumme $Q_b(Y_s)$ der jeweils gesuchten s-ten Diskriminanzfunktion über alle k Gruppen hinweg zu bestimmen:

$$Q_b(Y_s) = \sum_{g=1}^{k} n_g(\bar{y}_{gs} - \bar{\bar{y}}_s)^2 . \tag{6}$$

Die Zwischengruppen-Quadratsumme $Q_b\,(Y_s)$ gibt – als die mit der Gruppengröße n_g gewichtete Summe der quadrierten Abweichungen der Gruppenmittelwerte $\bar{y}_{gs}$ vom Gesamtmittelwert $\bar{\bar{y}}_s$ – den Anteil an der Gesamtstreuung $Q_t\,(Y_s)$ der Diskriminanzwerte auf der s-ten Diskriminanzfunktion an, der durch die Gruppeneinteilung "erklärt" wird. Umgekehrt gibt die Innergruppen-Quadratsumme $Q_w\,(Y_s)$

$$Q_w(Y_s) = \sum_{g=1}^{k} \sum_{i'=1}^{ng} \left(y_{i's} - \bar{y}_{gs}\right)^2 . \tag{7}$$

den durch die Gruppeneinteilung "nicht erklärten" Streuungsanteil der Diskriminanzwerte auf der s-ten Diskriminanzfunktion an, der sich im allgemeinen Fall aus der Summe der quadrierten Abweichungen der einzelnen Diskriminanzwerte $y_{i's}$ von ihrem jeweiligen Gruppenmittel $\bar{y}_{gs}$ bemisst.
Damit können wir das Diskriminanzkriterium λ_s für den allgemeinen Fall von $k \geq 2$ Gruppen als das Verhältnis von Zwischengruppen-Quadratsumme $Q_b(Y_s)$ und Innergruppen-Quadratsumme $Q_w\,(Y_s)$

$$\lambda_s = \frac{Q_b\,(Y_s)}{Q_w\,(Y_s)} \longrightarrow \max \tag{8}$$

definieren, wobei als Nebenbedingung die sukzessive Unkorreliertheit der Diskriminanzfunktionen zu berücksichtigen ist. Folglich sind für den allgemeinen Fall von $k \geq 2$ Gruppen die Gewichtungskoeffizienten der r Diskriminanzfunktionen nacheinander so zu bestimmen, dass die auf die Gruppeneinteilung zurückführbare Streuung der jeweiligen Diskriminanzwerte maximal, die Streuung innerhalb der Gruppen hingegen minimal ist; zusätzlich soll die im t-ten Extraktionsschritt bestimmte Diskriminanzfunktion mit den t–1 bereits extrahierten Diskriminanzfunktionen unkorreliert sein. Man beachte, dass das Diskriminanzkriterium λ_s für den allgemeinen Fall dem F-Bruch der Varianzanalyse entspricht, wobei der über alle möglichen Diskriminanzfunktionen konstante und daher für die Schätzung der Diskriminanzfunktionen irrelevante Term, der die Hypothesen- und Fehlerfreiheitsgrade ausdrückt, weggelassen ist. Diese Entsprechung verdeutlicht nochmals die bereits für den Zwei-Gruppen-Fall angemerkte Grundidee des diskriminanzanalytischen Verfahrens, multivariate Unterschiede zwischen Gruppen durch die Bildung von Linearkombinationen der Merkmalsvariablen analysierbar zu machen, indem eine Reduktion des multivariaten Problems auf ein oder mehrere univariate Probleme vorgenommen wird.

2.1.3 Eigenschaften des Diskriminanzraums

Alle Diskriminanzfunktionen oder – wie wir aufgrund der Ähnlichkeit des Verfahrens zur Hauptkomponentenanalyse auch sagen können – Diskriminanzfaktoren gemeinsam konstituieren den sogenannten *Diskriminanzraum*, der das gesamte Diskriminanzpotential der Merkmalsvariablen enthält, allerdings (im Regelfall von $p \geq k$) von geringerer Dimensionalität ist als der Variablenraum. Darüber hinaus lässt sich die Anzahl diskriminatorisch bedeutsamer Diskriminanzfunktionen meist weiter reduzieren, da zur Entscheidung über die Anzahl von Diskriminanzfunktionen, die in eine endgültige Lösung sinnvollerweise aufzunehmen sind, deskriptive Maße und Möglichkeiten einer inferenzstatistischen Beurteilung zur Verfügung stehen (s. Abschnitt 2.2.3). Da die Diskriminanzfunktionen – im Unterschied zu den ursprünglichen Merkmalsvariablen – wechselseitig unkorreliert sind, ergeben sich keine Überschneidungen hinsichtlich der "diskriminatorischen Information", die durch die Dimensionen des Diskriminanzraums repräsentiert wird. Wird eine inhaltliche Interpretation der Diskriminanzfunktionen angestrebt, können – wie in der Hauptkomponentenanalyse – die Ladungen der Merkmalsvariablen auf den Diskriminanzfaktoren herangezogen werden. Die standardisierten Gewichtungskoeffizienten ermöglichen außerdem eine Beurteilung der diskriminatorischen Bedeutsamkeit der Merkmalsvariablen für einzelne Diskriminanzfunktionen sowie für die Gruppentrennung insgesamt (vgl. Abschnitt 2.2.4).

2.2 Rechnerische Durchführung der Diskriminanzanalyse

2.2.1 Schätzung der Diskriminanzfunktionen

Zunächst muss die Notation um einige Symbole erweitert werden: Für eine Stichprobe von N Untersuchungsobjekten sollen Werte auf p Merkmalsvariablen vorliegen, die in der $(N \times p)$-Datenmatrix $\mathbf{X}$ zusammengefasst sind. Die N Untersuchungsobjekte lassen sich einer von k Gruppen $G_1, ..., G_g, ..., G_k$ der Größe n_g zuordnen. Entsprechend ist die Datenmatrix $\mathbf{X}$ partitionierbar in k Teilmatrizen $\mathbf{X}_1, ..., \mathbf{X}_g, ..., \mathbf{X}_k$ vom Typ $(n_g \times p)$, die sich wiederum zeilenweise aus den n_g Merkmalsvektoren $\mathbf{x}'_i = (x_{i1}, ..., x_{ij}, ..., x_{ip})$ zusammensetzen, in denen die dem i-ten Untersuchungsobjekt aus Gruppe g zugehörigen Werte auf den p Merkmalsvariablen enthalten sind. Die Diskriminanzwerte y_{is} der N Untersuchungsobjekte auf den r Diskriminanzvariablen sollen in den Vektoren $\mathbf{y}_1, ..., \mathbf{y}_s, ..., \mathbf{y}_r$ zusammengefasst werden.

Zur Bestimmung der s-ten Diskriminanzfunktion (vgl. Abschnitt 2.1) wurde der Vektor der Gewichtungskoeffizienten $\mathbf{v}_s = (v_{s1}, ..., v_{sj}, ..., v_{sp})'$ in der Linearkombination

$$\mathbf{y}_s = \mathbf{X}\mathbf{v}_s \tag{9}$$

so gewählt, dass er zu einer Maximierung des Diskriminanzkriteriums

$$\lambda_s = \frac{Q_b(Y_s)}{Q_w(Y_s)} \longrightarrow \max \tag{10}$$

führt. Der Vektor $\mathbf{v}_1$ der Gewichtungskoeffizienten v_{1j}, die der ersten Diskriminanzfunktion zuzuordnen sind, soll demnach eine Maximierung von λ_1 erzielen; sofern sich weitere Diskriminanzfunktionen bilden lassen, müssen deren Gewichtungsvektoren $\mathbf{v}_2, ..., \mathbf{v}_s, ..., \mathbf{v}_r$ so gewählt werden, dass eine Maximierung von $\lambda_1, ..., \lambda_s, ..., \lambda_r$ hinsichtlich der von den bereits extrahierten Diskriminanzfunktionen noch nicht aufgeklärten Varianz erzielt wird.

Zur Lösung dieses Maximierungsproblems werden zuerst die Quadratsummen Q_b und Q_w der noch unbekannten Diskriminanzvariablen $Y_1, ..., Y_s, ..., Y_r$ als Ausdrücke der in der Datenmatrix $\mathbf{X}$ enthaltenen Merkmalsvariablen und der jeweils gesuchten, zugehörigen Gewichtungsvektoren $\mathbf{v}_1, ..., \mathbf{v}_s, ..., \mathbf{v}_r$ notiert, wobei die in Gleichung 9 ausgedrückte Beziehung zu verwenden ist. Die totale Quadratsumme Q_t einer durch

Linearkombination der Merkmalsvariablen X_1, ..., X_p gebildeten Variablen Y_s ergibt sich demnach als

$$Q_t(Y_s) = \mathbf{v}_s{}'\mathbf{T}\mathbf{v}_s, \tag{11}$$

wobei die (p×p)-Streuungsmatrix **T** in der Hauptdiagonalen die Quadratsummen der p Merkmalsvariablen, außerhalb der Hauptdiagonalen deren Kreuzproduktsummen enthält (vgl. z.B. Tatsuoka, 1971; Bortz, 1993). Nach der Beziehung

$$\mathbf{T} = \mathbf{B} + \mathbf{W} \tag{12}$$

lässt sich die Gesamtstreuungsmatrix **T** in die (p×p)-Matrix **B** der Zwischengruppen-Quadratsummen und –Kreuzprodukte und die (p×p)-Matrix **W** der Innergruppen-Quadratsummen und –Kreuzprodukte zerlegen. Mit Hilfe von **B** und **W** lassen sich nun auch die Zwischenquadratsumme Q_b (Y_s) nach Gleichung 6 und die Innergruppen-Quadratsumme Q_w (Y_s) nach Gleichung 7 allein unter Rückgriff auf die gesuchten Gewichtungsvektoren $\mathbf{v}_s$ und die Streuungen der ursprünglichen Merkmalsvariablen ausdrücken:

$$Q_b(Y_s) = \mathbf{v}'_s\,\mathbf{B}\mathbf{v}_s, \tag{13}$$

$$Q_w(Y_s) = \mathbf{v}'_s\,\mathbf{W}\mathbf{v}_s. \tag{14}$$

Matrix **W** wird auch als Matrix der gepoolten Innergruppen-Streuung bezeichnet, da sie sich als Summe der gruppenspezifischen Streuungsmatrizen $\mathbf{W}_g$ darstellen lässt:

$$\mathbf{W} = \sum_{g=1}^{k} \mathbf{W}_g \tag{15}$$

mit

$$\mathbf{W}_g = \sum_{i'=1}^{n_g} (\mathbf{x}_{i'} - \bar{\mathbf{x}}_g)(\mathbf{x}_{i'} - \bar{\mathbf{x}}_g)', \tag{16}$$

wobei $\bar{\mathbf{x}}_g$ der Vektor der gruppenspezifischen Mittelwerte der p Merkmalsvariablen in Gruppe g ist. Definiert man zusätzlich einen Vektor $\bar{\bar{\mathbf{x}}}$, der die Gesamtmittelwerte der p Merkmalsvariablen enthält, kann Matrix **B** anschaulich hergeleitet werden als

$$\mathbf{B}=\sum_{g=1}^{k} n_g\left(\bar{\mathbf{x}}_g-\bar{\bar{\mathbf{x}}}\right)\left(\bar{\mathbf{x}}_g-\bar{\bar{\mathbf{x}}}\right)' \\ =\left(\bar{\mathbf{X}}-\bar{\bar{\mathbf{X}}}\right)'\left(\bar{\mathbf{X}}-\bar{\bar{\mathbf{X}}}\right). \tag{17}$$

Dabei sind in der ($N\times p$)-Matrix $\bar{\mathbf{X}}$ die Gruppenmittelwerte der Merkmalsvariablen so aufgelistet, dass die ersten n_1 Zeilen jeweils $\bar{\mathbf{x}}'_1$, also dem transponierten Vektor der Gruppenmittelwerte in G_1, die folgenden n_2 Zeilen hingegen jeweils $\bar{\mathbf{x}}'_2$, also dem transponierten Vektor der Gruppenmittelwerte in G_2, entsprechen usw., bis zu den letzten n_k Zeilen, die jeweils dem Vektor $\bar{\bar{\mathbf{x}}}'_k$, also dem transponierten Vektor der Gruppenmittelwerte in G_k, entsprechen. Alle Zeilen der ($N\times p$)-Matrix $\bar{\bar{\mathbf{X}}}$ enthalten hingegen die Gesamtmittelwerte der p Merkmalsvariablen, entsprechen also $\bar{\bar{\mathbf{x}}}'$, dem transponierten Vektor der Gesamtmittelwerte der p Merkmalsvariablen.

Allgemein, d.h. ohne Festlegung, um die wievielte Diskriminanzfunktion es sich handelt, können wir nun das Diskriminanzkriterium λ gemäß der in Gleichung 13 und Gleichung 14 definierten Beziehungen als Funktion der bekannten Matrizen **B**, **W** und des gesuchten Gewichtungsvektors **v** ausdrücken:

$$\lambda=\frac{\mathbf{v}'\mathbf{B}\mathbf{v}}{\mathbf{v}'\mathbf{W}\mathbf{v}}\longrightarrow \max . \tag{18}$$

Wir lösen das Maximierungsproblem, indem wir den Gradienten von λ berechnen und dem Nullvektor gleichsetzen:

$$\frac{\partial\lambda}{\partial\mathbf{v}}=\frac{2\left[(\mathbf{B}\mathbf{v})(\mathbf{v}'\mathbf{W}\mathbf{v})-(\mathbf{v}'\mathbf{B}\mathbf{v})(\mathbf{W}\mathbf{v})\right]}{(\mathbf{v}'\mathbf{W}\mathbf{v})^2}=\mathbf{0}. \tag{19}$$

Durch die Division von Zähler und Nenner durch **v'Wv** und Verwendung der in Gleichung 18 gegebenen Definition von λ für das zweite Glied der Differenz im Zähler vereinfacht sich die Gleichung zu

$$\frac{2\left[\mathbf{B}\mathbf{v}-\lambda\mathbf{W}\mathbf{v}\right]}{\mathbf{v}'\mathbf{W}\mathbf{v}}=\mathbf{0}. \tag{20}$$

Durch Multiplikation mit **v'Wv**, Division durch 2 und Ausklammern von **v** erhalten wir

$$(\mathbf{B}-\lambda\mathbf{W})\mathbf{v}=\mathbf{0}. \tag{21}$$

Sofern Matrix **W** nicht singulär ist, können wir in Gleichung 21 durch Multiplikation mit der Inversen $\mathbf{W}^{-1}$ das Diskriminanzkriterium λ isolieren

$$(\mathbf{W}^{-1}\mathbf{B} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0} \tag{22}$$

und erhalten die Bestimmungsgleichung für die gesuchten Vektoren **v**.[8] Die Lösung von Gleichung 22 stellt ein klassisches Eigenwertproblem dar; es gilt, die Eigenwerte λ und die dazugehörigen Eigenvektoren **v** der quadratischen Matrix $\mathbf{W}^{-1}\mathbf{B}$ zu finden. Die interessierende nichttriviale Lösung $\mathbf{v} \neq \mathbf{0}$ kann nur dann gefunden werden, wenn die Matrix $(\mathbf{W}^{-1}\mathbf{B} - \lambda\mathbf{I})$ singulär ist, d.h. eine Determinante von null hat (vgl. z.B. Moosbrugger, 1997). Aus dieser Überlegung erhalten wir mit

$$\left|\mathbf{W}^{-1}\mathbf{B} - \lambda\mathbf{I}\right| = 0 \tag{23}$$

die charakteristische Gleichung der Matrix $\mathbf{W}^{-1}\mathbf{B}$. Die Entwicklung der Determinante führt zu einem Polynom r-ter Ordnung mit $r = \min(p, k-1)$ Lösungen $\lambda_1, ..., \lambda_s, ..., \lambda_r$ für λ, die – eingesetzt in die Bestimmungsgleichung 22 – zu den dazugehörigen Eigenvektoren $\mathbf{v}_1, ..., \mathbf{v}_s, ..., \mathbf{v}_r$ führen. Setzt man den Eigenvektor $\mathbf{v}_1$, der dem größten Diskriminanzkriterium (Eigenwert) λ_1 zugeordnet ist, als Gewichtungsvektor in Gleichung 9 ein, ergibt sich die Diskriminanzfunktion, die von allen möglichen Diskriminanzfunktionen die höchste Diskriminationsfähigkeit besitzt. Die analog zu bildende Diskriminanzfunktion mit dem Gewichtungsvektor $\mathbf{v}_2$ erzielt unter allen möglichen Linearkombinationen, die nicht mit der ersten Diskriminanzfunktion korreliert sind, das höchste Diskriminanzkriterium λ_2. Ebenso erzielt die Diskriminanzfunktion mit dem Gewichtungsvektor $\mathbf{v}_3$ das höchste Diskriminanzkriterium λ_3 unter allen möglichen Linearkombinationen, die nicht mit den beiden ersten Diskriminanzfunktionen korreliert sind, usw. Wie man sieht, führt die Lösung des zunächst allgemein für einen Gewichtungsvektor **v** gestellten Maximierungsproblems zur simultanen Auffindung aller extrahierbaren Diskriminanzfunktionen.

---

[8] Nach Tatsuoka (1971) sind auch dann verschiedene Lösungsstrategien möglich, wenn **W** singulär ist, was z.B. dann der Fall ist, wenn einige der Merkmalsvariablen ipsative Messungen darstellen. Entweder man rechnet mit den betreffenden Variablen eine Hauptkomponentenanalyse und fügt nach Eliminierung der betreffenden Variablenwerte die erhaltenen Faktorwerte in die Datenmatrix ein, oder man löst Gleichung 21, ohne sie zuvor zu Gleichung 22 zu vereinfachen. Im letzteren Fall handelt es sich um ein sogenanntes verallgemeinertes Eigenwertproblem (*generalized eigenvalue problem*) mit der charakteristischen Gleichung $|\mathbf{B} - \lambda\mathbf{W}| = 0$.

Die Anzahl von Eigenwerten der quadratischen Matrix $\mathbf{W}^{-1}\mathbf{B}$, die ungleich null sind, entspricht dem Rang der Matrix **B**, da dieser höchstens gleich, in der Regel sogar kleiner ist als der Rang der Matrix $\mathbf{W}^{-1}$, die vollen Rang haben muss, da Matrix **W** als nonsingulär angenommen wurde. In dieser Eigenschaft der Matrix $\mathbf{W}^{-1}\mathbf{B}$ ist die Begründung dafür zu finden, dass sich allgemein $r = \min(p, k{-}1)$ Diskriminanzfunktionen extrahieren lassen (vgl. Tatsuoka, 1971). An der Herleitung von **B** aus dem Matrixprodukt in Gleichung 17 ist ersichtlich, dass der Rang von **B** gleich dem Rang der Matrix $\overline{\mathbf{X}} - \overline{\overline{\mathbf{X}}}$ ist. Diese Matrix hat wie Matrix $\overline{\mathbf{X}}$ höchstens k unterschiedliche Zeilen, die zudem einer linearen Restriktion unterworfen sind, da gilt:

$$\sum_{g=1}^{k} n_g (\bar{x}_{gj} - \bar{\bar{x}}_j) = 0 . \tag{24}$$

Daher existieren in der Matrix $\overline{\mathbf{X}} - \overline{\overline{\mathbf{X}}}$ und damit in Matrix **B** nicht mehr als k–1 linear unabhängige Zeilen, so dass der Rang der Matrix **B** im Höchstfall entweder k–1 oder p beträgt, je nach dem, welcher der beiden Werte kleiner ist.

2.2.2 Normierung der Diskriminanzfunktionen

Die Maximierung des Diskriminanzkriteriums führt zwar zu einer eindeutigen Bestimmung der Lage der Diskriminanzfunktionen im Variablenraum, d.h. zu einer Festlegung des Verhältnisses der Diskriminanzkoeffizienten einer Diskriminanzfunktion untereinander, ohne dass damit jedoch die Skaleneinheit der Diskriminanzwerte festgelegt wäre. Um die Diskriminanzwerte mehrerer Diskriminanzfunktionen untereinander vergleichen zu können, muss eine Skalierung der Diskriminanzfunktionen vorgenommen werden. Nach Backhaus et al. (1996) hat sich dafür die Normierungsvorschrift durchgesetzt, die normierten Gewichtungskoeffizienten (enthalten in den Vektoren $\mathbf{b}_1, \ldots, \mathbf{b}_s, \ldots, \mathbf{b}_r$) so zu bestimmen, dass die gepoolte (d.h. über die k Gruppen vereinte) Innergruppenvarianz der Diskriminanzwerte den Wert eins ergibt:[9]

$$\frac{1}{N-k} \mathbf{b}'_s \mathbf{W} \mathbf{b}_s = 1 . \tag{25}$$

[9] Anstelle der Matrix **W** kann zur Normierung auch auf die Matrix **T** zurückgegriffen werden, welche die Gesamtstreuungen der Merkmalsvariablen enthält, woraus für die Diskriminanzwerte eine Gesamtvarianz von eins resultiert (vgl. Cooley & Lohnes, 1971).

Demnach errechnet sich der Vektor $\mathbf{b}_s$ der normierten Diskriminanzkoeffizienten nach

$$\mathbf{b}_s = \mathbf{v}_s \frac{1}{s_{(w)ys}} . \tag{26}$$

Dabei ist $s_{(w)ys}$ die gepoolte Innergruppen-Standardabweichung der auf Basis der nicht normierten Diskriminanzkoeffizienten der s-ten Diskriminanzfunktion bestimmten Diskriminanzwerte, also

$$s_{(w)ys} = \sqrt{\frac{Q_w(Y_s)}{N-k}} = \sqrt{\frac{1}{N-k}\mathbf{v}'_s \mathbf{W} \mathbf{v}_s} . \tag{27}$$

Anschließend bestimmt man für jede Diskriminanzfunktion eine additive Konstante b_{s0}

$$b_{s0} = -\mathbf{b}'_s \overline{\overline{\mathbf{x}}} , \tag{28}$$

welche so gewählt wird, dass der Gesamtmittelwert der Diskriminanzwerte der s-ten Diskriminanzfunktion null ergibt:

$$\overline{\overline{y}}_s = \frac{1}{N}\sum_{i=1}^{N} y_{is} = \sum_{j=1}^{p} b_{sj}\overline{\overline{x}}_j = 0 . \tag{29}$$

Den Vektor $\mathbf{y}_s$ der normierten Diskriminanzwerte erhält man nun über die normierte Diskriminanzfunktion

$$\mathbf{y}_s = b_{s0}\mathbf{x}_0 + b_{s1}\mathbf{x}_1 + \ldots + b_{sj}\mathbf{x}_j + \ldots + b_{sp}\mathbf{x}_p , \tag{30}$$

wobei $\mathbf{x}_0$ ein N-dimensionaler Einsenvektor ist und $\mathbf{x}_j$ jeweils ein N-dimensionaler Vektor, der die Merkmalswerte der N Untersuchungsobjekte auf Merkmalsvariable j enthält. Nach der Normierung aller r Diskriminanzfunktionen liegen also ausnahmslos normierte Diskriminanzwerte mit einem Mittelwert von null und einer Innergruppenvarianz von eins vor.

2.2.3 Beurteilung der Diskriminanzfunktionen

Neben einer Reihe deskriptiver Maße zur Beurteilung der Diskriminationsfähigkeit (Abschnitt 2.2.3.1) stehen Möglichkeiten einer inferenzstatistischen Prüfung der Diskriminanzfunktionen zur Verfügung (Abschnitt 2.2.3.2). Bei der Diskriminanzanalyse handelt es sich daher auch um ein konfirmatorisches Verfahren. Es lässt sich nicht nur prüfen, ob zwischen den Gruppen ein Gesamtunterschied hinsichtlich des untersuchten Variablenkomplexes besteht, sondern auch die detailliertere Fragestellung beantworten, wie viele Diskriminanzfunktionen einen signifikanten Beitrag zur Gruppentrennung leisten. Als Folge kann im Mehr-Gruppen-Mehr-Variablen-Fall die Anzahl diskriminatorisch bedeutsamer Diskriminanzfunktionen in der Regel reduziert werden.

2.2.3.1 Eigenwertanteil und kanonischer Korrelationskoeffizient

Die relative Bedeutung einer Diskriminanzfunktion s für die Trennung der Gruppen, also ihr Anteil an der durch alle Gruppen insgesamt geleisteten Varianzaufklärung, wird durch ihren Eigenwertanteil

$$\frac{\lambda_s}{\lambda_1 + \ldots + \lambda_s + \ldots + \lambda_r} \tag{31}$$

ausgedrückt. Da das Diskriminanzkriterium λ_s zwar per definitionem ein Maß für die Diskriminationsfähigkeit von Diskriminanzfunktionen darstellt, aber wie die ursprünglichen Gewichtungsvektoren nicht normiert ist, zieht man zur Beurteilung der Diskriminationsfähigkeit einer Diskriminanzfunktion den kanonischen Korrelationskoeffizienten ρ oder das sogenannte Wilks Λ heran (zu letzterem s. Abschnitt 2.2.3.2).

Der quadrierte kanonische Korrelationskoeffizient ρ_s^2 steht in der folgenden Beziehung zu dem Diskriminanzkriterium λ_s(vgl. Tatsuoka, 1953):

$$\rho_s^2 = \frac{\lambda_s}{1 + \lambda_s}. \tag{32}$$

Damit repräsentiert ρ_s^2 den Quotienten von erklärter Varianz und Gesamtvarianz der Diskriminanzwerte,[10] gibt also den Anteil der durch die s-te Diskriminanzfunktion erklärten Varianz an. Bei Durchführung einer kanonischen Korrelationsanalyse mit k-1 in geeigneter Weise dummykodierten Gruppierungsvariablen ergeben sich direkt r Werte für ρ_s^2 als Lösung eines Eigenwertproblems (vgl. z.B. Tatsuoka, 1971). Der kanonische Korrelationskoeffizient $\rho_s = \sqrt{\rho_s^2}$ gibt demnach an, wie stark der Zusammenhang zwischen den Diskriminanzfunktionen – den kanonischen Variablen – und der Gruppeneinteilung ist.[11]

2.2.3.2 Inferenzstatistische Prüfung anhand von Wilks Λ

Zur inferenzstatistischen Überprüfung von Diskriminanzfunktionen greift man üblicherweise auf Wilks Λ zurück. Wilks Λ ist als Quotient der durch die Gruppeneinteilung nicht erklärten Streuung und der Gesamtstreuung definiert, bestimmt sich also im Falle des sogenannten univariaten Wilks Λ_s für die s-te Diskriminanzfunktion als

$$\Lambda_s = \frac{1}{1+\lambda_s}. \tag{33}$$

Bei Wilks Λ, das auf einen Wertebereich zwischen eins und null normiert ist, handelt es sich um ein inverses Gütemaß, so dass sehr kleine Werte von Λ_s im Sinne einer hohen Diskrimination der Gruppen durch die s-te Diskriminanzfunktion zu interpretieren sind. Als Gütemaß für die Diskriminationsfähigkeit aller Diskriminanzfunktionen gemeinsam wird das sogenannte multivariate Wilks Λ berechnet:

---

[10] Diese Beziehung gilt unabhängig davon, ob bereits normierte Diskriminanzfunktionen vorliegen oder nicht, da die Ausprägung des Diskriminanzkriteriums von einer Normierung nicht beeinflusst wird. Zur Veranschaulichung kann aber folgender Gedankengang hilfreich sein: Nach einer Normierung unter Rückgriff auf die Matrix **W** ist die nicht erklärte Varianz auf den Wert eins normiert; da das Diskriminanzkriterium aber nach Gleichung 5 als Quotient der Zwischengruppen-Quadratsumme $Q_b(Y_s)$ und der Innergruppen-Quadratsumme $Q_w(Y_s)$ definiert ist, muss in Folge einer Normierung unter Rückgriff auf **W** die erklärte Varianz λ_s, die Gesamtvarianz $1+\lambda_s$ betragen.

[11] Im Zwei-Gruppen-Fall, in dem nur eine Diskriminanzfunktion extrahiert werden kann, entspricht der nach Gleichung 32 ermittelbare Wert dem multiplen Korrelationskoeffizienten der Regressionsanalyse; das diskriminanzanalytische Verfahren ist dann äquivalent zu dem einer Regressionsanalyse mit dummykodierter Gruppierungsvariablen.

$$\Lambda = \prod_{s=1}^{r} \Lambda_s \,. \tag{34}$$

Unter den Voraussetzungen, dass die p Merkmalsvariablen in den k Populationen, aus denen die untersuchten Stichproben entnommen sind, multivariat normalverteilt sind und die Varianz-Kovarianzmatrizen in den Gruppen als homogen angesehen werden können, lässt sich Wilks Λ in einen sogenannten V-Wert (Bartletts V-Statistik; vgl. Tatsuoka, 1971) überführen, der näherungsweise χ^2-verteilt ist. Über den dem multivariaten Wilks Λ zugehörigen V-Wert

$$\begin{aligned} V &= -[N-1-(p^*+k)/2]\ln\Lambda \\ &= [N-1-(p^*+k)/2]\ln[(1+\lambda_1)(1+\lambda_2)\ldots(1+\lambda_r)] \\ &= [N-1-(p^*+k)/2]\sum_{s=1}^{r}\ln[1+\lambda_s] \end{aligned} \tag{35}$$

kann die statistische Bedeutsamkeit aller r Diskriminanzfunktionen gemeinsam überprüft werden, wobei für p^* die Anzahl der Merkmalsvariablen p einzusetzen ist;[12] der V-Wert für das multivariate Wilks Λ hat $p^*(k-1)$ Freiheitsgrade. Als Alternativhypothese H_1 fungiert hier die globale Unterschiedshypothese, dass in der Grundgesamtheit bezüglich der Mittelwertevektoren $\mu_1, \ldots, \mu_g \ldots, \mu_k$ Unterschiede bestehen; die dazugehörige Nullhypothese H_0 lautet formal: $\mu_1 = \ldots = \mu_g = \ldots = \mu_k$.

Erweist sich nach Gleichung 35 das gesamte Diskriminanzpotential als signifikant, kann über die V-Statistik für das residuelle Wilks Λ weiterhin geprüft werden, ob das nach der Extraktion und "Herauspartialisierung" der ersten, varianzstärksten Diskriminanzfunktion verbleibende Diskriminanzpotential noch signifikant zwischen den Gruppen unterscheidet.[13] Wie die einzelnen Diskriminanzfunktionen sind die

---

12 Bei inferenzstatistischen Überprüfungen, die sich auf das vollständige Modell mit allen p Merkmalsvariablen beziehen, gilt immer $p^* = p$. Wenn hingegen das Inkrement einer Untermenge von Merkmalsvariablen geprüft werden soll, ist für p^* ein Wert $p' < p$ einzusetzen (s. Abschnitt 0).

13 Formal ausgedrückt bezieht sich die inferenzstatistische Prüfung über das residuelle Wilks Λ also auf die Alternativhypothese H_1, dass in der Population mindestens eines der Elemente in den Gewichtungsvektoren $\beta_2, \ldots, \beta_s, \ldots, \beta_r$ ungleich null ist, bzw. auf die Nullhypothese $\beta_2 = \ldots = \beta_s = \ldots = \beta_r = \mathbf{0}$. Im allgemeinen Fall der Prüfung der nach Extraktion der t-ten Diskriminanzfunktion verbleibenden residuellen Diskriminanz lautet die Alternativhypothese H_1, dass in der Population mindestens eines der Elemente in den Gewichtungsvektoren $\beta_{t+1}, \ldots, \beta_s, \ldots, \beta_r$ ungleich null ist, die Nullhypothese H_0 hingegen $\beta_{t+1} = \ldots = \beta_s = \ldots = \beta_r = \mathbf{0}$.

einzelnen Terme des logarithmierten Produkts in Gleichung 35 statistisch unabhängig. Daraus folgt, dass auch der für die s-te Diskriminanzfunktion ermittelbare V-Wert

$$\begin{aligned} V_s &= -[N-1-(p^*+k)/2]\ln \Lambda_s \\ &= [N-1-(p^*+k)/2]\ln(1+\lambda_s) \end{aligned} \tag{36}$$

approximativ χ^2-verteilt ist, und zwar mit p^*+k-2s Freiheitsgraden (vgl. z.B. Tatsuoka, 1971), wobei p^* wiederum gleich p ist. Die Summe der univariaten V-Werte für die einzelnen Diskriminanzfunktionen ist dem V-Wert für das multivariate Wilks Λ äquivalent. Die residuelle Diskriminanz nach Extraktion der ersten t Diskriminanzfunktionen lässt sich daher über den V-Wert für das residuelle Wilks Λ prüfen, den wir erhalten, indem wir die V-Werte der ersten, bereits extrahierten t Diskriminanzfunktionen vom V-Wert für das multivariate Wilks Λ abziehen:

$$\begin{aligned} V_{t+1,\ldots,r} &= V_1 - \ldots - V_t \\ &= [N-1-(p^*+k)/2]\cdot \sum_{s=t+1}^{r} \ln(1+\lambda_s). \end{aligned} \tag{37}$$

Die Freiheitsgrade für die Signifikanzprüfung anhand der χ^2-Verteilung lauten in diesem Fall $(p^*-t)(k-t-1)$ (vgl. z.B. Tatsuoka, 1971), wobei für p^* die Anzahl der Merkmalsvariablen p einzusetzen ist. Sobald der V-Wert für das residuelle Wilks Λ den der festgelegten Irrtumswahrscheinlichkeit zugeordneten χ^2-Wert *unter*schreitet, ist die Nullhypothese beizubehalten und folglich der Schluss zu ziehen, dass nur die ersten t Diskriminanzfunktionen signifikant zur Gruppentrennung beitragen.

Wie bei allen inferenzstatistischen Verfahren erweisen sich bei Anwendung der V-Statistik auch kleine Unterschiede zwischen Gruppen (bzw. Diskriminanzfunktionen mit kleinem Eigenwertanteil) als signifikant, wenn nur ein hinreichend großer Stichprobenumfang N untersucht wird. Daher ist es für die Beurteilung von Diskriminanzfunktionen sinnvoll, neben der Anwendung von Signifikanztests eine genaue Inspektion der deskriptiven Maße für die Diskriminationsfähigkeit der einzelnen Diskriminanzfunktionen (Eigenwertanteile und kanonische Korrelationskoeffizienten) vorzunehmen. Auch in empirischen Anwendungen mit hohen Gruppen- und Variablenzahlen erweisen sich unabhängig von den Ergebnissen der inferenzstatistischen Prüfung meist nur die ersten beiden Diskriminanzfunktionen als praktisch relevant, insofern diese bereits einen großen Teil der insgesamt erreichten Varianzaufklärung – abzulesen an ihrem Eigenwertanteil – zu leisten vermögen (vgl. Cooley & Lohnes, 1971).[14]

2.2.4 Beurteilung der Merkmalsvariablen

Neben einer Beurteilung der diskriminatorischen Bedeutsamkeit der Diskriminanzfunktionen können wir daran interessiert sein, den simultanen Beitrag der p Merkmalsvariablen zur Diskrimination der Gruppen im multivariaten Kontext zu ermitteln. Daraus ergeben sich zwei miteinander verwandte Fragestellungen: Zum einen kann ein theoretisches Anliegen darin bestehen, multivariate Gesamtunterschiede zwischen den Gruppen einer detaillierten Analyse zu unterziehen und gegebenenfalls die Diskriminanzfunktionen inhaltlich zu interpretieren (Abschnitte 2.2.4.1 und 2.2.4.2). Zum anderen kann es vor einer praktischen Anwendung der ermittelten Diskriminanzfunktionen zu Zwecken der Klassifikation "neuer" Objekte (vgl. Abschnitt 3.2) sinnvoll sein, jene Merkmalsvariablen zu identifizieren, die nur eine geringe Trennfähigkeit aufweisen, um sie aus der Diskriminanzfunktion zu eliminieren (Abschnitt 2.2.4.3).

2.2.4.1 Standardisierte Diskriminanzkoeffizienten

Die normierten Diskriminanzkoeffizienten b_{sj} in den normierten Diskriminanzfunktionen

$$\mathbf{y}_s = b_{s0}\mathbf{x}_0 + b_{s1}\mathbf{x}_1 + \ldots + b_{sj}\mathbf{x}_j + \ldots + b_{sp}\mathbf{x}_p \tag{30}$$

geben zwar den absoluten Beitrag der Merkmalsvariablen zur Variabilität der Diskriminanzwerte der s-ten Diskriminanzfunktion an, sind aber untereinander nicht vergleichbar, da ihre Größe, sofern nicht schon vor Durchführung der Diskriminanz-

---

[14] Bei kleinen Stichproben ist die χ^2-Approximation der V-Werte ungenau, insbesondere wenn p eine ungerade und k eine gerade Zahl ist; für diesen Fall hat Schatzoff (1966, zitiert nach Tatsuoka, 1971) Korrekturfaktoren ermittelt. Alternativ lässt sich Wilks Λ auch in eine approximativ F-verteilte Prüfgröße transformieren (Raos R-Statistik; s. z.B. Moosbrugger, 1997), deren Genauigkeit nicht von der Stichprobengröße abhängt (vgl. z.B. Rao, 1973; Tatsuoka, 1971).

analyse eine Variablenstandardisierung vorgenommen wurde, von der Skalierung der Merkmalsvariablen abhängt. Deshalb formen wir die unstandardisierten, normierten Koeffizienten b_{s1}, ..., b_{sj}, ..., b_{sp} in standardisierte Koeffizienten b^*_{s1}, ..., b^*_{sj},..., b^*_{sp} um, indem wir sie mit der gepoolten Innergruppen-Standardabweichung $s_{(w)j}$ der zugehörigen Merkmalsvariablen multiplizieren:

$$b^*_{sj} = s_{(w)j} b_{sj} . \tag{38}$$

Fassen wir die Diagonalenelemente der Matrix $(N-k)^{-1}\,\mathbf{W}$, nämlich die Innergruppen-Varianzen der einzelnen Merkmalsvariablen, in der $(p \times p)$-Diagonalmatrix $\mathbf{D}$ zusammen, lässt sich der Vektor $\mathbf{b}^*_s$ der standardisierten Diskriminanzkoeffizienten b^*_{sj} auch ermitteln als

$$\mathbf{b}^*_s = \mathbf{D}^{1/2} \mathbf{b}_s . \tag{39}$$

Je größer ein standardisierter Diskriminanzkoeffizient b^*_{sj} ist, desto größer ist der Beitrag der Merkmalsvariablen j zur s-ten Diskriminanzfunktion. Folglich kann der relative Beitrag einer Merkmalsvariablen zur Diskrimination einer Diskriminanzfunktion über das Verhältnis des Betrags des betreffenden standardisierten Diskriminanzkoeffizienten zur Summe der Beträge aller Diskriminanzkoeffizienten bestimmt werden:

$$\frac{|b^*_{sj}|}{\sum_{j=1}^{p} |b^*_{sj}|} . \tag{40}$$

Schließlich erlauben die standardisierten Diskriminanzkoeffizienten auch die Ermittlung eines Maßes für den Beitrag jeder einzelnen Merkmalsvariablen zur insgesamt erreichten Gruppentrennung. Der über die r Diskriminanzfunktionen gemittelte Diskriminanzkoeffizient $\bar{b}^*_j$ errechnet sich als die Summe der mit den Eigenwertanteilen gewichteten Beträge der Diskriminanzkoeffizienten in allen Diskriminanzfunktionen:

$$\bar{b}^*_j = \sum_{s=1}^{r} |b^*_{sj}| \cdot \frac{\lambda_s}{\sum_{s=1}^{r} \lambda_s} . \tag{41}$$

Die standardisierten Diskriminanzkoeffizienten erklären multivariate Gesamtunterschiede, insofern sie die Bedeutung einer Merkmalsvariablen für die Gruppentrennung im Rahmen des vollständigen Modells aller in die Analyse miteinbezogenen Merkmalsvariablen wiedergeben. Will man mit Hilfe der standardisierten Diskriminanzkoeffizienten Diskriminanzwerte ermitteln, sind anhand der gepoolten Innergruppen-Standardabweichung $s_{(w)j}$ standardisierte Merkmalswerte zu verwenden. Liegt für ein Untersuchungsobjekt etwa auf der Merkmalsvariablen j ein um eine Standardabweichung höherer Wert als bei einem anderen Untersuchungsobjekt vor, während alle anderen Merkmalswerte gleich sind, so ergibt sich für dieses Untersuchungsobjekt auf der s-ten Diskriminanzfunktion ein um den b_{sj}-ten Teil einer Standardabweichung höherer Diskriminanzwert.

2.2.4.2 Strukturkoeffizienten

In dem Maße, in dem die Merkmalsvariablen untereinander korreliert sind, ist damit zu rechnen, dass sich die diskriminatorische Information der Merkmalsvariablen – im Unterschied zu den Diskriminanzfunktionen – überschneidet. Daher sind die standardisierten Diskriminanzkoeffizienten zu einer vor allem im Mehr-Gruppen-Mehr-Variablen-Fall interessanten inhaltlichen Interpretation der Diskriminanzfunktionen nur bedingt tauglich. Ein dafür geeigneteres, den Faktorladungen der Hauptkomponentenanalyse analoges Maß gewinnen wir, indem wir die Korrelationen zwischen den Diskriminanzwerten und den Merkmalswerten über alle N Untersuchungsobjekte berechnen. Die Korrelationskoeffizienten der Korrelationen zwischen Merkmals- und Diskriminanzwerten werden *Strukturkoeffizienten* genannt, da sie sich geometrisch als Cosinus der Winkel zwischen Diskriminanz- und Merkmalsvektoren im Personenraum ergeben und somit die vektorielle Struktur des Personenraums widerspiegeln. Der Strukturkoeffizientenvektor $\mathbf{t}_s$ bezüglich der s-ten Diskriminanzfunktion lässt sich mit Hilfe der Interkorrelationsmatrix $\mathbf{R}$ der p Merkmalsvariablen berechnen. Wir ermitteln zunächst Matrix $\mathbf{R}$ als

$$\mathbf{R} = \mathbf{S}^{-1}\mathbf{C}\mathbf{S}^{-1}. \tag{42}$$

Hierin ist die Matrix $\mathbf{C}$ die Gesamt-Varianz-Kovarianzmatrix, die sich als $N^{-1}\mathbf{T}$ ergibt, während $\mathbf{S}$ eine Diagonalmatrix ist, auf deren Hauptdiagonalen die Standardabweichungen der p Merkmalsvariablen eingetragen sind. Nun erhalten wir den Vektor $\mathbf{t}_s$ als

$$\mathbf{t}_s = \mathbf{RSv}_s \cdot \frac{1}{s_{(t)ys}} . \tag{43}$$

mit $s_{(t)ys}$ als Gesamt-Standardabweichung der Diskriminanzwerte der s-ten Diskriminanzfunktion, welche sich nach der Beziehung

$$s_{(t)ys} = \sqrt{\mathbf{v}'_s \mathbf{C} \mathbf{v}_s} \tag{44}$$

berechnen lässt.

Resultieren für die einzelnen Diskriminanzfunktionen inhaltlich plausible Muster von Strukturkoeffizienten, können diese wie bei der Hauptkomponentenanalyse zur Interpretation der Diskriminanzfunktionen herangezogen werden. Um die Interpretierbarkeit zu verbessern, ist auch eine Rotation der Matrix der Strukturkoeffizienten bzw. der Diskriminanzachsen z.B. nach dem Varimax-Kriterium möglich, wodurch die Diskriminationsfähigkeit des Modells nicht beeinträchtigt wird (vgl. Dillon & Goldstein, 1984).

2.2.4.3 Selektion von Merkmalsvariablen

In der Literatur zur Diskriminanzanalyse wird teilweise vorgeschlagen, die Strukturkoeffizienten auch zur Identifikation derjenigen Merkmalsvariablen zu verwenden, welche in einem optimal diskriminierenden Modell beibehalten werden sollen. Seligiert werden demnach z.B. im Zwei-Gruppen-Fall diejenigen Variablen, welche die höchsten Strukturkoeffizienten bezüglich der Diskriminanzfunktion aufweisen (vgl. z.B. Hartung & Elpelt, 1984). Dass dieses Vorgehen zumindest im Mehr-Gruppen-Fall nicht unproblematisch ist, wird daran augenfällig, dass die Struktur des Diskriminanzraumes durch die nachträgliche Eliminierung der nicht selektierten Variablen unter Umständen erhebliche Unterschiede aufweist gegenüber der Struktur vor der Eliminierung, welche zur Variablenauswahl herangezogen worden war. Zudem handelt es sich bei dieser Verwendung von Strukturkoeffizienten um ein – dem univariaten F-Test äquivalentes – univariates Vorgehen, mit dem der diskriminatorische Beitrag einer Variablen im Kontext der anderen Variablen nicht ermittelt werden kann (vgl. Rencher, 1988). Verwendet man hingegen ein multivariates Kriterium in Verbindung mit einer geeigneten Selektionsprozedur, so ist eine Variablenauswahl zumindest bei umfangreichen Datensätzen aus mehreren Gründen anzuraten:

- Sollen praktische Anwendungen folgen, bei denen die Gruppenzugehörigkeiten zu prognostizieren sind, kann der Aufwand für die Erhebung und Auswertung der notwendigen Informationen erheblich reduziert werden, wenn Variablen von geringer diskriminatorischer Bedeutung nicht berücksichtigt werden müssen.
- Die Diskriminanzanalyse zeitigt bei hohen Variablenzahlen instabile Ergebnisse, insofern die Fehlerrate bei der Klassifikation "neuer" Objekte bei gleichbleibender Größe der untersuchten Teilstichproben und gleichbleibender Gruppentrennung mit steigender Variablenzahl zunimmt (vgl. z.B. Läuter, 1992).
- Eine ungünstige Relation von Variablen- und Probandenzahlen kann bereits die Qualität der Lösung von Diskriminanzproblemen beeinträchtigen. Im nachteiligsten Fall wird die für die Schätzung von Diskriminanzfunktionen zentrale Matrix **W** singulär und ist folglich nicht invertierbar.
- Liegen interkorrelierte Variablen vor (was außer bei Verwendung von orthogonalen Faktorwerten die Regel sein dürfte), bieten diese bei gemeinsamer Aufnahme in das Modell nur geringe inkrementelle Beiträge, auch wenn sie an sich für die Diskrimination geeignet wären. Trifft man eine methodisch begründete Auswahl unter derartigen Variablen, kann die Effizienz der Schätzung von Diskriminanzfunktionen erhöht werden.

Aus der Äquivalenz von Diskriminanzanalyse und kanonischer Korrelationsanalyse ist bereits erkennbar, dass die Prüfung des diskriminatorischen Beitrags einzelner Merkmalsvariablen dem Vorgehen bei der schrittweisen Regressionsanalyse nahe verwandt ist. Alle dort üblichen Kriterien (z.B. R^2 oder der inkrementelle F-Wert) sind ebenso für den Zwei-Gruppen-Fall der Diskriminanzanalyse geeignet. Für den Fall der Diskriminanzanalyse mit mehr als zwei Gruppen können verschiedene Prüfgrößen berechnet werden, neben dem inkrementellen Wilks Λ etwa Hotellings Spur-Statistik oder auch der inkrementelle F-Wert (vgl. Krzanowski & Marriott, 1995; Fujikoshi, 1989; Rencher, 1993).

2.2.4.3.1 Prüfung des inkrementellen Beitrags einzelner Merkmalsvariablen

Die Logik der Prüfung des inkrementellen Beitrags einzelner Variablen zur Diskrimination soll exemplarisch anhand des inkrementellen Wilks Λ für folgende Problemsituation vorgestellt werden: Das ursprüngliche Modell umfasst die Variablen $X_1, \ldots, X_p$, die in den k verschiedenen Gruppen durch ihre $(p\times 1)$-Mittelwertevektoren

$\bar{\mathbf{x}}_1, \ldots, \bar{\mathbf{x}}_g, \ldots, \bar{\mathbf{x}}_k$ (mit $\bar{\mathbf{x}}_g = (\bar{x}_{g1}, \ldots, \bar{x}_{gp})$) repräsentiert sind; geprüft werden soll der inkrementelle Beitrag einer (p+1)-ten Variablen Z,[15] wodurch die ((p+1)×1)-Mittelwertevektoren $\bar{\mathbf{x}}_1^{(Z)}, \ldots, \bar{\mathbf{x}}_g^{(Z)}, \ldots, \bar{\mathbf{x}}_k^{(Z)}$ (mit $\bar{\mathbf{x}}_g^{(Z)} = (\bar{x}_{g1}, \ldots, \bar{x}_{gp}, \bar{z}_g)$) als zusätzliches Element den jeweiligen Gruppenmittelwert der Variablen Z enthalten. Die Nullhypothese H_0 lautet, dass sich die hinter den $\bar{\mathbf{x}}_g$ liegenden Populationsmittelwerte $\boldsymbol{\mu}_1^{(Z)}, \ldots, \boldsymbol{\mu}_g^{(Z)}, \ldots, \boldsymbol{\mu}_k^{(Z)}$ unter Berücksichtigung der Variablen X_1 , ..., X_p und Z von den ursprünglichen Populationsmittelwerten $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_g, \ldots, \boldsymbol{\mu}_k$, welche die Variable Z nicht berücksichtigen, nicht unterscheiden.[16] Berechnet man gemäß Gleichung 34 sowohl für das ursprüngliche als auch für das um Z erweiterte Modell je einen Wert für das multivariate Wilks Λ, so erhält man das inkrementelle Wilks $\Lambda_{Z|X}$ als den Quotienten aus $\Lambda_{X,Z}$ (für das erweiterte Modell) und Λ_X (für das ursprüngliche Modell) (vgl. Rencher, 1993):

$$\Lambda_{Z|X} = \frac{\Lambda_{X,Z}}{\Lambda_X}. \tag{45}$$

Nach Gleichung 35 kann $\Lambda_{Z|X}$ in einen approximativ χ^2-verteilten V-Wert umgeformt werden, wobei für p^* die Anzahl jener Variablen einzusetzen ist, deren Inkrement geprüft werden soll. Bezeichnen wir die Anzahl der Variablen im ursprünglichen Modell mit p' und die Anzahl der Variablen im erweiterten Modell mit p, ist folglich für p^* die Differenz p–p' einzusetzen, im Falle einer einzigen hinzugefügten Variablen also eins. Der inkrementelle Beitrag der Variablen lässt sich unter den Voraussetzungen der multivariaten Normalverteilung und der Gleichheit der Varianz-Kovarianzmatrizen in den Gruppen wiederum anhand der χ^2-Verteilung mit $p^*(k-1)$ Freiheitsgraden auf Signifikanz prüfen.

2.2.4.3.2 Schrittweise Diskriminanzanalyse

Vor der Durchführung einer *schrittweisen Diskriminanzanalyse* ist neben der Festlegung eines Prüfkriteriums die Spezifikation eines Algorithmus nötig, der eine Selektionsstrategie und eine Terminationsregel beinhaltet, welche die Bedingungen für die Beendigung des Auswahlprozesses angibt. Wegen ihres vergleichsweise geringen

---

[15] Nach demselben Verfahren kann auch der inkrementelle Beitrag mehrerer, gleichzeitig hinzugefügter Variablen geprüft werden.

[16] Zur Logik der Überprüfung von Inkrementen vgl. auch Moosbrugger (1994), S. 50ff.

Rechenaufwandes sind nach wie vor einfache Algorithmen wie die aufsteigende und die absteigende Variablenauswahl (*forward* bzw. *backward selection*) verbreitet. Bei der aufsteigenden Strategie wird mit einer leeren Variablenmenge begonnen und zunächst diejenige Variable in das Modell aufgenommen, welche die höchste Trennkraft (d.h. die deutlichsten univariaten Unterschiede) aufweist, um dann in jedem weiteren Schritt diejenigen Variablen aufzunehmen, die das gewählte Prüfkriterium optimieren und einen festgelegten Zuwachs des Prüfkriteriums erreichen, oder Variablen wieder zu eliminieren, bei denen ein festgelegter Zuwachs des Prüfkriteriums nicht erreicht wird. Die absteigende Strategie nimmt ihren Ausgang hingegen von einem vollständigen Modell mit allen zu berücksichtigenden Variablen, wobei schrittweise Variablen eliminiert werden. Der Auswahlprozess endet normalerweise, wenn die für Variablenaufnahme oder -elimination bestimmten Werte des Prüfkriteriums über- bzw. unterschritten werden. Daran wird deutlich, dass das Ergebnis einer schrittweisen Diskriminanzanalyse immer von der Reihenfolge der Selektionsschritte abhängt. Ein optimales Endresultat ist daher im Grunde nur dann garantiert, wenn die Diskriminationsfähigkeit aller möglichen Untermengen der zur Verfügung stehenden Variablenmenge geprüft wird. Werden bei der schrittweisen Diskriminanzanalyse als Prüfkriterien Signifikanzgrenzen festgelegt, bei deren Über- oder Unterschreiten die Variablen in das Modell aufgenommen werden, so ist zu beachten, dass die ermittelbaren Irrtumswahrscheinlichkeiten aufgrund der Vielzahl der bezogen auf eine Zusammenstellung von Variablen durchgeführten Signifikanztests nicht die "wahren" α-Wahrscheinlichkeiten widerspiegeln. Empirisch fundierte Hinweise für die Gestaltung schrittweiser Prozeduren im Hinblick auf eine möglichst präzise Klassifikation von Objekten geben z.B. Nakache und Dussere (1975) oder Moran (1975).

2.3 Rechenbeispiel: Diskrimination zwischen "Okkultistisch Aktiven" und "Okkultistisch nicht Aktiven"

Für ein illustrierendes Rechenbeispiel sollen Daten einer Fragebogenstudie aus der Okkultismusforschung (Brednich, 1993; vgl. auch Mischo, 1996) herangezogen werden. An einer Stichprobe von 401 Erwachsenen wurden insgesamt 15 intervallskalierte Merkmalsvariablen erhoben, die in vier inhaltliche Bereiche gliederbar sind:

1. Der Bereich *Magisch-irrationales Denken* umfasst die Merkmale "Psi", "Hexen", "Astrologie", "Okkultes Wissen" und "Okkultes Tun" (für eine Beschreibung der Skalen s. Mischo, 1991).
2. *Religiosität* wurde anhand einer Skala "Religiosität" (s. Mischo, 1991) erhoben.
3. Die Skalen "Schizotypie" (Schizotypal Personality Scale, Claridge & Broks, 1984), "Magische Vorstellungen" (Magical Ideation Scale, Eckbald & Chapman, 1983) und "Wahrnehmungsabweichung" (Perceptual Abberation Scale, Chapman, Chapman & Raulin, 1978) beziehen sich auf definierende Symptome der *Schizotypischen Persönlichkeitsstörung*.
4. Zusätzlich liegen Werte auf den Skalen "Depression", "Neurotizismus", "Extraversion" und "Körperliche Beschwerden" des Freiburger Persönlichkeits-Inventars (FPI-R, Fahrenberg, Selg & Hampel, 1984) vor, die der Schizotypischen Persönlichkeitsstörung *Zugehörige Symptome* und Merkmale erfassen.

Mit einer diskriminanzanalytischen Auswertung dieses Datensatzes sollen Aufschlüsse darüber gewonnen werden, ob und in welchen Hinsichten sich die mit okkulten Praktiken vertrauten Personen von solchen Personen unterscheiden, die über keine oder nur sehr geringe Erfahrungen mit okkulten Handlungen verfügen. Eine präzisere Formulierung der Fragestellung lautet, dass untersucht werden soll, wie gut die den *einzelnen* Merkmalsbereichen zugeordneten Variablen *im Gesamtkontext* der Merkmale, mit denen die Aspekte von Magisch-irrationalem Denken, Religiosität, Schizotypie und zugehörigen Persönlichkeitsmerkmalen erfasst werden, zwischen den Gruppen der "Okkultistisch Aktiven" und der "Okkultistisch nicht Aktiven" zu trennen vermögen. Gegenüber getrennten univariaten Analysen bietet der multivariate Zugang der Diskriminanzanalyse dabei einen wesentlichen Informationsgewinn, da – unbeschadet der Ergebnisse univariater Analysen – beispielsweise folgende inhaltliche Probleme bearbeitet werden können:

- Diskriminieren die klinischen Merkmale, die sich auf die definierenden Symptome der Schizotypischen Persönlichkeitsstörung beziehen, auch dann zwischen den beiden Gruppen, wenn gleichzeitig Merkmalswerte für Magisch-irrationales Denken in die Analyse einbezogen werden, oder dupliziert sich die diskriminatorische Information?
- Kommt denjenigen mit dem FPI erhobenen Merkmalen, welche sich auf kontingente, Zugehörige Symptome der Schizotypischen Persönlichkeitsstörung beziehen, eine von den definierenden Symptomen der Schizotypischen Persönlichkeitsstörung unabhängige diskriminatorische Bedeutung zu?

- Stellt Religiosität ein gegenüber den Aspekten des Magisch-irrationalen Denkens eigenständiges Merkmal dar, hinsichtlich dessen sich "Okkultistisch Aktive" von "Okkultistisch nicht Aktiven" unterscheiden?

Darüber hinaus kann die Identifikation derjenigen Variablen, die gemeinsam die bestmögliche Gruppentrennung erzielen, auch für praktische Anwendungen von Nutzen sein. Diese Variablen können nämlich zur Beurteilung der okkultistischen Aktivität von solchen Personen herangezogen werden, bei denen das Ausmaß okkulten Tuns nicht direkt erhoben werden kann.

Zunächst sind die zu diskriminierenden Teilstichproben der "Okkultistisch Aktiven" und der "Okkultistisch nicht Aktiven" zu definieren. Dazu werden die Extremgruppen der intervallskalierten Variablen "Okkultes Tun" gebildet[17]. Dies bereitet bei dem gewählten Datensatz von Brednich (1993) keine Schwierigkeiten, da die Variable "Okkultes Tun" – wie alle anderen Variablen – bereits in Stanine-normierter Form vorliegt. Bestimmen wir als "Okkultistisch nicht Aktive" (G_1) die Probanden mit den Standardwerten 1 und 2 und als "Okkultistisch Aktive" (G_2) die Probanden mit den Standardwerten 8 und 9, ergeben sich zwei ungefähr gleich große Stichproben ($n_1 = 44$ und $n_2 = 45$), die jeweils etwa 11% der Gesamtstichprobe der Untersuchung enthalten.

Aus Gründen der Übersichtlichkeit soll eine Vorauswahl der für die Diskrimination zu berücksichtigenden Merkmalsvariablen getroffen werden.[18] Da sich die Variablen bestimmten inhaltlichen Bereichen zuordnen lassen, ist vorab zu vermuten, dass entsprechende Interkorrelationsmuster bestehen. Aus diesem Grund könnte die Situation eintreten, dass sich für Variablen aus ein- und demselben inhaltlichen Bereich nur geringe Diskriminationskoeffizienten ergeben, weil sich die diskriminatorischen

[17] Unter versuchsplanerischen Gesichtspunkten liegt also ein korrelatives Design vor, da die Gruppenzuordnung post hoc aufgrund von erst in der Untersuchungssituation erhobenen Daten erfolgt. Damit unterscheidet sich das Rechenbeispiel vom eingangs skizzierten, typischen Anwendungsfall der Diskriminanzanalyse, in dem Zufallsstichproben aus wohldefinierten Populationen gezogen werden.

[18] Die Vorauswahl geschieht in erster Linie aus darstellungstechnischen Gründen. Bei Verfügbarkeit eines entsprechenden Rechenprogramms lässt sich eine Variablenselektion auch allein mit diskriminanzanalytischen Mitteln, etwa einer schrittweisen Diskriminanzanalyse, vornehmen (vgl. Abschnitt 2.2.4.3). Das der schrittweisen Diskriminanzanalyse zugrundeliegende Prinzip, die Bestimmung und inferenzstatistische Prüfung des Inkrements einzelner Merkmalsvariablen, wird im Rahmen des Rechenbeispiels illustriert.

Beiträge überschneiden, obwohl die Merkmale an sich zur Diskrimination zwischen den beiden Gruppen geeignet sind. Die Vermutung, dass Variablen aus denselben Bereichen hoch korreliert sind, lässt sich empirisch anhand der Ergebnisse einer Hauptkomponentenanalyse (unter Einbeziehung der Gesamtstichprobe von $N=401$) im Wesentlichen erhärten: Die nach dem Varimax-Kriterium rotierte Faktorlösung enthält drei Faktoren (Abbruchkriterium: Kaiser-Kriterium), wobei die den Bereichen *Magisch-irrationales Denken* (mit Ausnahme der Variablen "Okkultes Wissen") und *Religiosität* zugeordneten Variablen sowie die Variable "Magische Vorstellungen" die höchsten Ladungen auf dem ersten Faktor aufweisen (alle genannten Faktorladungen $>.55$; Varianzaufklärung 38.0%). Der zweite Faktor fasst hingegen die beiden übrigen Variablen aus dem Bereich *Schizotypische Persönlichkeitsstörung* und die Variablen aus dem Bereich der *Zugehörigen Symptome* (mit Ausnahme der Variablen "Extraversion") zusammen (alle genannten Faktorladungen $>.55$; Varianzaufklärung 15.9%). Auf dem dritten Faktor haben lediglich die Variablen "Extraversion" und Okkultes Wissen" ihre höchsten Ladungen (beide Faktorladungen $>.6$; Varianzaufklärung 9.0%). Unter Berücksichtigung der Faktorladungen, des Varianzanteils der einzelnen Faktoren und unserer Fragestellung, gemäß der alle inhaltlichen Bereiche zu berücksichtigen sind, scheint für das Rechenbeispiel die folgende Variablenauswahl gerechtfertigt:

1. "Hexen" (X_1)
2. "Okkultes Wissen" (X_2)
3. "Religiosität" (X_3)
4. "Magische Vorstellungen" (X_4)
5. "Extraversion" (X_5)
6. "Körperliche Beschwerden" (X_6)

Die $n_1+n_2=89$ Messwerte sind in der (89×6)-Datenmatrix $\mathbf{X}$ (s. Tabelle 1) zusammengefasst. Neben der Partitionierung der Datenmatrix $\mathbf{X}$ in die gruppenspezifischen Submatrizen $\mathbf{X}_1$ und $\mathbf{X}_2$ lassen sich der Tabelle 1 auch zeilenweise die (1x6)-Vektoren $\mathbf{x}'_i$ der einzelnen Probanden sowie spaltenweise die gruppenspezifischen (89x1)-Vektoren $\mathbf{x}_{gj}$ der einzelnen Merkmalsvariablen entnehmen.

Tabelle 1. Submatrix X_1 der Datenmatrix X

Gruppe g	Pb. i	Merkmalswerte x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}	x_{i6}	
1	1	6	1	8	4	4	5	$\mathbf{x}'_1$
	2	4	1	3	3	4	4	$\mathbf{x}'_2$
	3	1	5	2	2	5	6	$\mathbf{x}'_3$
	4	2	1	4	2	6	3	$\mathbf{x}'_4$
	5	1	4	2	3	5	7	$\mathbf{x}'_5$
	6	2	6	1	3	7	2	$\mathbf{x}'_6$
	7	2	7	2	2	6	8	$\mathbf{x}'_7$
	8	1	3	6	3	5	2	$\mathbf{x}'_8$
	9	1	4	6	3	7	4	$\mathbf{x}'_9$
	10	4	5	3	3	1	7	$\mathbf{x}'_{10}$
	11	7	7	2	6	5	7	$\mathbf{x}'_{11}$
	12	1	4	3	2	3	5	$\mathbf{x}'_{12}$
	13	1	1	4	2	7	6	$\mathbf{x}'_{13}$
	14	1	6	4	1	7	1	$\mathbf{x}'_{14}$
	15	2	1	3	3	7	6	$\mathbf{x}'_{15}$
	16	5	4	9	6	6	6	$\mathbf{x}'_{16}$
	17	3	7	5	3	2	5	$\mathbf{x}'_{17}$
	18	5	4	9	3	4	4	$\mathbf{x}'_{18}$
	19	3	7	1	4	3	3	$\mathbf{x}'_{19}$
	20	3	3	3	1	3	3	$\mathbf{x}'_{20}$
	21	3	3	5	6	4	9	$\mathbf{x}'_{21}$
	22	5	2	4	4	4	7	$\mathbf{x}'_{22}$
	23	4	5	6	4	3	6	$\mathbf{x}'_{23}$
	24	1	2	1	4	5	5	$\mathbf{x}'_{24}$
	25	1	6	2	2	4	3	$\mathbf{x}'_{25}$
	26	4	2	4	4	4	8	$\mathbf{x}'_{26}$
	27	2	5	5	2	6	3	$\mathbf{x}'_{27}$
	28	2	2	2	2	4	9	$\mathbf{x}'_{28}$
	29	5	2	6	5	6	4	$\mathbf{x}'_{29}$
	30	3	2	4	4	4	5	$\mathbf{x}'_{30}$
	31	2	4	5	3	4	1	$\mathbf{x}'_{31}$
	32	3	2	4	3	9	3	$\mathbf{x}'_{32}$
	33	3	3	3	3	5	2	$\mathbf{x}'_{33}$
	34	6	4	6	7	3	7	$\mathbf{x}'_{34}$
	35	8	6	9	5	2	4	$\mathbf{x}'_{35}$
	36	3	4	3	5	8	7	$\mathbf{x}'_{36}$
	37	1	2	3	2	6	5	$\mathbf{x}'_{37}$
	38	1	2	4	2	7	7	$\mathbf{x}'_{38}$
	39	5	4	5	3	5	5	$\mathbf{x}'_{39}$
	40	3	6	6	3	7	3	$\mathbf{x}'_{40}$
	41	3	3	3	3	5	5	$\mathbf{x}'_{41}$
	42	2	4	2	3	2	6	$\mathbf{x}'_{42}$
	43	1	3	5	2	6	5	$\mathbf{x}'_{43}$
	44	3	3	3	4	9	9	$\mathbf{x}'_{44}$
		$\mathbf{x}_{1\,1}$	$\mathbf{x}_{1\,2}$	$\mathbf{x}_{1\,3}$	$\mathbf{x}_{1\,4}$	$\mathbf{x}_{1\,5}$	$\mathbf{x}_{1\,6}$	
		Submatrix $\mathbf{X}_1$						

Tabelle 1. (Fortsetzung): Submatrix $\mathbf{X}_2$ der Datenmatrix **X**

Gruppe g	Pb. i	Merkmalswerte x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}	x_{i6}	
2	45	7	4	7	6	6	7	$\mathbf{x}'_{45}$
	46	8	5	7	8	4	9	$\mathbf{x}'_{46}$
	47	2	3	1	3	5	3	$\mathbf{x}'_{47}$
	48	5	7	7	8	4	5	$\mathbf{x}'_{48}$
	49	8	5	7	5	6	4	$\mathbf{x}'_{49}$
	50	5	9	5	5	9	2	$\mathbf{x}'_{50}$
	51	6	6	3	5	6	4	$\mathbf{x}'_{51}$
	52	4	7	4	3	3	3	$\mathbf{x}'_{52}$
	53	7	6	7	7	8	5	$\mathbf{x}'_{53}$
	54	6	9	5	6	7	8	$\mathbf{x}'_{54}$
	55	7	6	3	5	7	8	$\mathbf{x}'_{55}$
	56	4	8	1	5	7	5	$\mathbf{x}'_{56}$
	57	7	4	7	6	5	6	$\mathbf{x}'_{57}$
	58	9	7	9	7	4	8	$\mathbf{x}'_{58}$
	59	5	6	6	4	6	5	$\mathbf{x}'_{59}$
	60	6	7	5	5	5	6	$\mathbf{x}'_{60}$
	61	8	6	8	6	3	6	$\mathbf{x}'_{61}$
	62	8	7	8	8	5	9	$\mathbf{x}'_{62}$
	63	7	5	1	5	8	5	$\mathbf{x}'_{63}$
	64	4	6	4	5	6	5	$\mathbf{x}'_{64}$
	65	6	3	5	7	6	4	$\mathbf{x}'_{65}$
	66	8	5	6	9	9	5	$\mathbf{x}'_{66}$
	67	4	8	6	7	8	4	$\mathbf{x}'_{67}$
	68	5	3	6	5	6	5	$\mathbf{x}'_{68}$
	69	4	9	2	2	8	1	$\mathbf{x}'_{69}$
	70	6	5	6	9	4	4	$\mathbf{x}'_{70}$
	71	4	7	4	5	1	3	$\mathbf{x}'_{71}$
	72	5	8	5	6	5	7	$\mathbf{x}'_{72}$
	73	7	6	5	8	5	4	$\mathbf{x}'_{73}$
	74	9	8	7	5	4	4	$\mathbf{x}'_{74}$
	75	5	8	4	5	7	3	$\mathbf{x}'_{75}$
	76	6	2	5	6	2	6	$\mathbf{x}'_{76}$
	77	6	7	5	8	7	4	$\mathbf{x}'_{77}$
	78	7	5	6	6	5	7	$\mathbf{x}'_{78}$
	79	5	5	6	4	2	7	$\mathbf{x}'_{79}$
	80	8	6	5	8	3	7	$\mathbf{x}'_{80}$
	81	8	6	6	8	6	5	$\mathbf{x}'_{81}$
	82	8	4	4	7	6	3	$\mathbf{x}'_{82}$
	83	6	6	7	7	5	8	$\mathbf{x}'_{83}$
	84	6	5	6	8	5	4	$\mathbf{x}'_{84}$
	85	9	5	9	8	6	5	$\mathbf{x}'_{85}$
	86	7	7	5	9	6	5	$\mathbf{x}'_{86}$
	87	6	6	6	7	7	3	$\mathbf{x}'_{87}$
	88	6	5	2	5	5	4	$\mathbf{x}'_{88}$
	89	6	5	2	6	8	5	$\mathbf{x}'_{89}$
		$\mathbf{x}_{21}$	$\mathbf{x}_{22}$	$\mathbf{x}_{23}$	$\mathbf{x}_{24}$	$\mathbf{x}_{25}$	$\mathbf{x}_{26}$	
		Submatrix $\mathbf{X}_2$						

Anmerkungen. Die Submatrizen $\mathbf{X}_1$ und $\mathbf{X}_2$ enthalten spaltenweise die gruppenspezifischen Merkmalsvektoren $\mathbf{x}_{1j}$ bzw. $\mathbf{x}_{2j}$ der $p = 6$ Merkmalsvariablen und zeilenweise die Merkmalsvektoren $\mathbf{x}'_i$ der $N = 89$ Untersuchungsobjekte. Beide Submatrizen gemeinsam bilden die Datenmatrix **X**.

Die Submatrizen $\mathbf{X}_1$ und $\mathbf{X}_2$ bilden gemeinsam die Datenmatrix $\mathbf{X}$ nach der folgenden Anordnung:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}.$$

Zur Berechnung der Elemente der Matrizen $\mathbf{B}$ und $\mathbf{W}$ benötigen wir zunächst die in den Vektoren $\bar{\mathbf{x}}_1$ und $\bar{\mathbf{x}}_2$ enthaltenen gruppenspezifischen Mittelwerte der $p=6$ Merkmalsvariablen:[19, 20]

$$\begin{aligned}\bar{\mathbf{x}}'_1 &= \frac{\mathbf{u}'\mathbf{X}_1}{n_1} = \frac{\mathbf{u}'\mathbf{X}_1}{44} = \begin{bmatrix} \bar{x}_{11} & \bar{x}_{12} & \bar{x}_{13} & \bar{x}_{14} & \bar{x}_{15} & \bar{x}_{16} \end{bmatrix} \\ &= \begin{bmatrix} 2.93 & 3.86 & 4.09 & 3.27 & 4.98 & 5.05 \end{bmatrix},\end{aligned}$$

$$\begin{aligned}\bar{\mathbf{x}}'_2 &= \frac{\mathbf{u}'\mathbf{X}_2}{n_2} = \frac{\mathbf{u}'\mathbf{X}_2}{45} = \begin{bmatrix} \bar{x}_{21} & \bar{x}_{22} & \bar{x}_{23} & \bar{x}_{24} & \bar{x}_{25} & \bar{x}_{26} \end{bmatrix} \\ &= \begin{bmatrix} 6.22 & 5.93 & 5.22 & 6.16 & 5.56 & 5.11 \end{bmatrix}.\end{aligned}$$

Weiterhin berechnen wir die im Vektor $\bar{\bar{\mathbf{x}}}$ enthaltenen Gesamtmittelwerte als

$$\begin{aligned}\bar{\bar{\mathbf{x}}}' &= \frac{\mathbf{u}'\mathbf{X}}{N} = \frac{\mathbf{u}'\mathbf{X}}{89} = \begin{bmatrix} \bar{\bar{x}}_1 & \bar{\bar{x}}_2 & \bar{\bar{x}}_3 & \bar{\bar{x}}_4 & \bar{\bar{x}}_5 & \bar{\bar{x}}_6 \end{bmatrix} \\ &= \begin{bmatrix} 4.60 & 4.82 & 4.66 & 4.73 & 5.27 & 5.08 \end{bmatrix}.\end{aligned}$$

Unter Verwendung der Gleichungen 15 und 16 ergibt sich nun die Matrix $\mathbf{W}$ der Innergruppen-Quadratsummen und -Kreuzprodukte als

[19] Zugunsten eines übersichtlichen Druckbildes wurden bei der Darstellung der Rechenbeispiele (Abschnitt 2.3 und Abschnitt 3.4) Zahlenrundungen vorgenommen. Da die Berechnungen aber unter höherer Genauigkeit erfolgt sind, können beim Nachrechnen "per Hand" unter Verwendung der angegebenen Zahlenwerte marginale Abweichungen gegenüber den berichteten Ergebnissen auftreten.

[20] Der Vektor $\mathbf{u}'$ ist in den folgenden drei Gleichungen jeweils der transponierte Einheitsvektor der Länge n_g bzw. der Länge N.

$$\mathbf{W} = \mathbf{W}_1 + \mathbf{W}_2$$

$$= \begin{bmatrix} 138.80 & 15.05 & 81.27 & 71.82 & -48.07 & 26.14 \\ 15.05 & 143.55 & -10.73 & 5.82 & -32.32 & -25.36 \\ 81.27 & -10.73 & 185.64 & 39.91 & -12.91 & -25.18 \\ 71.82 & 5.82 & 39.91 & 78.73 & -14.73 & 43.46 \\ -48.07 & -32.32 & -12.91 & -14.73 & 148.98 & -9.96 \\ 26.14 & -25.36 & -25.18 & 43.46 & -9.96 & 191.91 \end{bmatrix}$$

$$+ \begin{bmatrix} 111.78 & -16.33 & 80.78 & 65.44 & -4.56 & 54.89 \\ -16.33 & 120.80 & -5.33 & -13.53 & 30.67 & -13.67 \\ 80.78 & -5.33 & 177.78 & 75.44 & -41.56 & 67.89 \\ 65.44 & -13.53 & 75.44 & 123.91 & 3.11 & 41.22 \\ -4.56 & 30.67 & -41.56 & 3.11 & 149.11 & -38.78 \\ 54.89 & -13.67 & 67.89 & 41.22 & -38.78 & 150.44 \end{bmatrix}$$

$$= \begin{bmatrix} 250.57 & -1.29 & 162.05 & 137.26 & -52.62 & 81.03 \\ -1.29 & 264.35 & -16.06 & -7.72 & -1.65 & -39.03 \\ 162.05 & -16.06 & 363.41 & 115.35 & -54.47 & 42.71 \\ 137.26 & -7.72 & 115.35 & 202.64 & -11.62 & 84.68 \\ -52.62 & -1.65 & -54.47 & -11.62 & 298.09 & -48.73 \\ 81.03 & -39.03 & 42.71 & 84.68 & -48.73 & 342.35 \end{bmatrix}.$$

Die Matrix **B** der Zwischengruppenstreuungen erhalten wir nach Gleichung 17 als

$$\mathbf{B} = \begin{bmatrix} 240.87 & 164.82 & 82.82 & 211.04 & 42.33 & 4.81 \\ 164.82 & 112.78 & 56.66 & 144.4 & 28.97 & 3.28 \\ 82.82 & 56.66 & 28.48 & 72.56 & 14.56 & 1.65 \\ 211.04 & 144.4 & 72.56 & 184.89 & 37.09 & 4.21 \\ 42.33 & 28.97 & 14.56 & 37.09 & 7.44 & 0.84 \\ 4.81 & 3.28 & 1.65 & 4.21 & 0.84 & 0.1 \end{bmatrix}.$$

Für die Inverse $\mathbf{W}^{-1}$ ermitteln wir

$$\mathbf{W^{-1}} = \begin{bmatrix} 7.76\cdot10^{-3} & -2.78\cdot10^{-4} & -2.09\cdot10^{-3} & -3.80\cdot10^{-3} & 7.47\cdot10^{-4} & -5.61\cdot10^{-4} \\ -2.78\cdot10^{-4} & 3.87\cdot10^{-3} & 2.55\cdot10^{-4} & -9.13\cdot10^{-6} & 9.89\cdot10^{-5} & 4.92\cdot10^{-4} \\ -2.09\cdot10^{-3} & 2.55\cdot10^{-4} & -9.13\cdot10^{-6} & -9.77\cdot10^{-4} & 3.82\cdot10^{-4} & 3.17\cdot10^{-4} \\ -3.80\cdot10^{-3} & -9.13\cdot10^{-6} & -9.77\cdot10^{-4} & 8.52\cdot10^{-3} & -7.12\cdot10^{-4} & -1.19\cdot10^{-3} \\ 7.47\cdot10^{-4} & 9.89\cdot10^{-5} & 3.82\cdot10^{-4} & -7.12\cdot10^{-4} & 3.61\cdot10^{-3} & 4.76\cdot10^{-4} \\ -5.61\cdot10^{-4} & 4.92\cdot10^{-4} & 3.17\cdot10^{-4} & -1.19\cdot10^{-3} & 4.76\cdot10^{-4} & 3.43\cdot10^{-3} \end{bmatrix}.$$

Nun können wir das Matrixprodukt $\mathbf{W}^{-1}\mathbf{B}$ berechnen:

$$\mathbf{W}^{-1}\mathbf{B} = \begin{bmatrix} 0.877 & 0.600 & 0.302 & 0.769 & 0.154 & 0.018 \\ 0.597 & 0.408 & 0.205 & 0.523 & 0.105 & 0.012 \\ -.316 & -.216 & -.109 & -.277 & -.056 & -.006 \\ 0.764 & 0.523 & 0.263 & 0.670 & 0.134 & -.015 \\ 0.233 & 0.159 & 0.080 & 0.204 & 0.041 & 0.005 \\ -.242 & -.166 & -.083 & -.212 & -.043 & -.005 \end{bmatrix},$$

mit der charakteristischen Gleichung

$$\left|\mathbf{W}^{-1}\mathbf{B}-\lambda\mathbf{I}\right| = \begin{vmatrix} 0.877-\lambda & 0.600 & 0.302 & 0.769 & 0.154 & 0.018 \\ 0.597 & 0.408-\lambda & 0.205 & 0.523 & 0.105 & 0.012 \\ -.316 & -.216 & -.109-\lambda & -.277 & -.056 & -.006 \\ 0.764 & 0.523 & 0.263 & 0.670-\lambda & 0.134 & -.015 \\ 0.233 & 0.159 & 0.080 & 0.204 & 0.041-\lambda & 0.005 \\ -.242 & -.166 & -.083 & -.212 & -.043 & -.005-\lambda \end{vmatrix} = 0.$$

Durch Entwickeln der Determinante erhalten wir folgendes Polynom 6. Ordnung:

$$\lambda^6 - 1.883\lambda^5 - 1.996\cdot10^{-4}\lambda^4 - 5.842\cdot10^{-3}\lambda^3 + 6.942\cdot10^{-6}\lambda^2 - 2.982\cdot10^{-6}\lambda + 4.039\cdot10^{-24} = 0.$$

Da die Anzahl der extrahierbaren Diskriminanzfunktionen $r = \min(p, k-1)$ betragen muss (vgl. Abschnitt 2.2.1), ist bei $k=2$ Gruppen nur ein von null verschiedener Eigenwert zu erwarten. In der Tat sind das dritte bis sechste Glied sowie die additive Konstante des Polynoms bis auf Rundungsungenauigkeiten null, so dass sich über

$$\lambda^6 - 1.883\lambda^5 = 0$$

nur eine einzige Lösung für λ ergibt:

$$\lambda_1 = 1.883 .$$

Zur Berechnung des Vektors $\mathbf{v}_1$ der nicht-normierten Diskriminanzkoeffizienten der einzigen extrahierbaren Diskriminanzfunktion setzen wir den ermittelten Wert für λ_1 in die Bestimmungsgleichung 22 ein:

$$(\mathbf{W}^{-1}\mathbf{B} - \lambda_1 \mathbf{I})\mathbf{v}_1 = \mathbf{0}$$

$$= \begin{bmatrix} 0.877-1.883 & 0.600 & 0.302 & 0.769 & 0.154 & 0.018 \\ 0.597 & 0.408-1.883 & 0.205 & 0.523 & 0.105 & 0.012 \\ -.316 & -.216 & -.109-1.883 & -.277 & -.056 & -.006 \\ 0.764 & 0.523 & 0.263 & 0.670-1.883 & 0.134 & -.015 \\ 0.233 & 0.159 & 0.080 & 0.204 & 0.041-1.883 & 0.005 \\ -.242 & -.166 & -.083 & -.212 & -.043 & -.005-1.883 \end{bmatrix} \cdot \begin{bmatrix} v_{11} \\ v_{12} \\ v_{13} \\ v_{14} \\ v_{15} \\ v_{16} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Die Lösungen des homogenen Gleichungssystem, das nach Ausführen des Matrixprodukts resultiert, sind nur bis auf eine Proportionalitätskonstante bestimmbar, so dass ein Element frei gewählt werden kann (vgl. z.B. Moosbrugger, 1997). Setzen wir $v_{11} = 1$, resultiert der Eigenvektor

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0.680 \\ -.361 \\ 0.871 \\ 0.265 \\ -.276 \end{bmatrix},$$

der als Elemente die noch nicht normierten Diskriminanzkoeffizienten enthält, welche nach Gleichung 27 zu einer gepoolten Innergruppen-Varianz der Diskriminanzwerte von

$$s^2_{(w)y1} = 6.773$$

führen. Der Vektor $\mathbf{b}_1$ der normierten Diskriminanzkoeffizienten errechnet sich sodann nach Gleichung 26 wie folgt:

$$\mathbf{b}_1 = \mathbf{v}_1 \frac{1}{s_{(w)y1}}$$

$$= \begin{bmatrix} 1 \\ 0.680 \\ -.361 \\ 0.871 \\ 0.265 \\ -.276 \end{bmatrix} \cdot \frac{1}{2.602} = \begin{bmatrix} 0.384 \\ 0.261 \\ -.139 \\ 0.335 \\ 0.102 \\ -.106 \end{bmatrix}.$$

Die additive Konstante b_{10} ergibt sich nach Gleichung 29 als

$$b_{10} = -\mathbf{b}'_1 \,\overline{\overline{\mathbf{x}}}$$

$$= -\begin{bmatrix} 0.384 & 0.261 & -.139 & 0.335 & 0.102 & -.106 \end{bmatrix} \cdot \begin{bmatrix} 4.60 \\ 4.82 \\ 4.66 \\ 4.73 \\ 5.27 \\ 5.08 \end{bmatrix} = -3.962.$$

Die normierte Diskriminanzfunktion für unser Beispiel lautet demnach

$$\mathbf{y}_1 = -3.962\mathbf{x}_0 + 0.384\mathbf{x}_1 + 0.261\mathbf{x}_2 - .139\mathbf{x}_3 + 0.335\mathbf{x}_4 + 0.102\mathbf{x}_5 - .106\mathbf{x}_6;$$

ihre Anwendung liefert für die beiden untersuchten Teilstichproben folgende Gruppenmittelwerte:

$$\overline{y}_{11} = b_{10} + \overline{\mathbf{x}}'_1 \,\mathbf{b}_1$$

$$= -3.962 + \begin{bmatrix} 2.93 & 3.86 & 4.09 & 3.27 & 4.98 & 5.05 \end{bmatrix} \cdot \begin{bmatrix} 0.384 \\ 0.261 \\ -.139 \\ 0.335 \\ 0.102 \\ -.106 \end{bmatrix} = -1.325.$$

$$\bar{y}_{21} = b_{10} + \bar{\mathbf{x}}'_2 \mathbf{b_1}$$

$$= -3.962 + \begin{bmatrix} 6.22 & 5.93 & 5.22 & 6.16 & 5.56 & 5.11 \end{bmatrix} \cdot \begin{bmatrix} 0.384 \\ 0.261 \\ -.139 \\ 0.335 \\ 0.102 \\ -.106 \end{bmatrix} = 1.342 .$$

Die Beurteilung der erzielten Diskriminationsfähigkeit kann nun sowohl deskriptiv als auch inferenzstatistisch vorgenommen werden. Da sich nur eine Diskriminanzfunktion ermitteln ließ, erübrigt sich die Berechnung der verschiedenen Eigenwert*anteile*. Die Beurteilung der Diskriminanzfunktion kann daher direkt nach Gleichung 32 durch den – in diesem Fall mit dem multiplen Korrelationskoeffizienten R identischen – kanonischen Korrelationskoeffizienten ρ_1 erfolgen:

$$\rho_1 = \sqrt{\frac{\lambda_1}{1+\lambda_1}} = \sqrt{\frac{1.883}{1+1.883}} = .808 .$$

Der Anteil der durch die einzige extrahierbare Diskriminanzfunktion erklärten Varianz der Gruppierungsvariablen beträgt also $\rho_1^2 = .652$. Das zu ρ_1^2 komplementäre Maß Wilks Λ_1 – im Beispielfall mit dem multivariaten Wilks Λ nach Gleichung 34 identisch – erhalten wir nach Gleichung 33 als

$$\Lambda_1 = \frac{1}{1+\lambda_1} = \frac{1}{1+1.883} = .347 .$$

Die inferenzstatistische Überprüfung, ob die Diskriminanzfunktion auch in der Grundgesamtheit zu einer Unterscheidung der beiden Gruppen führt, erfolgt für die Nullhypothese $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ mit dem nach Gleichung 36 ermittelbaren V-Wert

$$\begin{aligned} V &= -[N - 1 - (p^* + k)/2] \ln \Lambda \\ &= -[89 - 1 - (6+2)/2] \ln .347 \\ &= 88.927 . \end{aligned}$$

Wählt man ein Signifikanzniveau von $\alpha = .01$, so erweist sich der V-Wert nach der χ^2-Verteilung mit $p^*+k-2s=6$ Freiheitsgraden als signifikant (bei $p^*=p=6$). Mit einem Risiko von höchstens 1% verwerfen wir daher die Nullhypothese $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ und nehmen die Alternativhypothese an: Wir können schließen, dass auch in der Grundgesamtheit ein multivariater Gesamtunterschied der beiden Gruppen hinsichtlich der untersuchten Variablen besteht, dass sich also auch in der Grundgesamtheit die Mittelwertevektoren von "Okkultistisch nicht Aktiven" (G_1) und "Okkultistisch Aktiven" (G_2) unterscheiden. Da der Gesamtunterschied nur auf eine einzige Diskriminanzfunktion zurückgeht, erübrigt sich eine gesonderte Prüfung des univariaten bzw. residuellen Wilks Λ: Aus dem signifikanten Ergebnis für den Gesamtunterschied resultiert, dass mindestens einer der Diskriminanzkoeffizienten der erhaltenen Diskriminanzfunktion auch in der Grundgesamtheit von null verschieden ist.

Vor einer detaillierten Beurteilung der einzelnen Merkmalsvariablen muss zunächst eine Standardisierung der Diskriminanzkoeffizienten vorgenommen werden. Nach Ermittlung der Diagonalmatrix $\mathbf{D}$, die die Varianzen der Merkmalsvariablen enthält, lässt sich der Vektor $\mathbf{b}^*_1$ der standardisierten Diskriminanzkoeffizienten nach Gleichung 39 berechnen:

$$\mathbf{b}^*_1 = \mathbf{D}^{1/2}\mathbf{b}_1$$

$$= \begin{bmatrix} 2.880 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3.038 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4.177 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2.329 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3.426 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3.935 \end{bmatrix}^{1/2} \cdot \begin{bmatrix} 0.384 \\ 0.261 \\ -.139 \\ 0.335 \\ 0.102 \\ -.106 \end{bmatrix} = \begin{bmatrix} 0.652 \\ 0.456 \\ -.283 \\ 0.511 \\ 0.189 \\ -.210 \end{bmatrix}.$$

Drei der Merkmalsvariablen scheinen eine relativ große Bedeutung für die Gruppentrennung zu besitzen, da ihre standardisierten Diskriminanzkoeffizienten um das Zwei- bis Dreifache höher sind als die der übrigen drei Variablen. Es handelt sich hierbei um die Variablen "Hexen" (X_1) mit $b^*_{11}=0.652$, "Magische Vorstellungen" (X_4) mit $b^*_{14}=0.511$ und "Okkultes Wissen" (X_2) mit $b^*_{12}=0.455$. Nach Gleichung 40 können wir den relativen Beitrag, den diese drei Variablen zur Gruppentrennung liefern, bestimmen:

$$\frac{|b^*_{11}|+|b^*_{12}|+|b^*_{14}|}{\sum\limits_{j=1}^{p}|b^*_{1j}|}=\frac{1.61865}{2.300}=.704\,.$$

Wir sehen also, dass der relative Beitrag, den diese drei Variablen zur Diskrimination leisten, ca. 70% beträgt. Ein ähnliches Bild ergibt sich, wenn wir den Vektor $\mathbf{t}_1$ der Strukturkoeffizienten – die "Ladungen" der Merkmalsvariablen auf der Diskriminanzvariablen – über die Korrelationen der Merkmalswerte mit den Diskriminanzwerten – ermitteln. Haben wir die Interkorrelationsmatrix $\mathbf{R}$ nach Gleichung 42 und die Gesamtstreuung $s_{(t)y1}$ der Diskriminanzvariablen nach Gleichung 44 berechnet, ergibt sich $\mathbf{t}_1$ als

$$\mathbf{t}_1=\mathbf{RSv}_1\cdot\frac{1}{s_{(t)y1}}$$

$$=\begin{bmatrix}1 & .380 & .558 & .798 & -.027 & .209\\ .380 & 1 & .106 & .358 & .081 & -.100\\ .558 & .106 & 1 & .482 & -.115 & .121\\ .798 & .358 & .482 & 1 & .074 & .244\\ -.027 & .081 & -.115 & .074 & 1 & -.148\\ .209 & -.100 & .121 & .244 & -.148 & 1\end{bmatrix}\cdot\begin{bmatrix}2.36 & 0 & 0 & 0 & 0 & 0\\ 0 & 2.07 & 0 & 0 & 0 & 0\\ 0 & 0 & 2.11 & 0 & 0 & 0\\ 0 & 0 & 0 & 2.10 & 0 & 0\\ 0 & 0 & 0 & 0 & 1.86 & 0\\ 0 & 0 & 0 & 0 & 0 & 1.97\end{bmatrix}$$

$$\cdot\begin{bmatrix}1\\ 0.680\\ -.361\\ 0.871\\ 0.265\\ -.276\end{bmatrix}\cdot\frac{1}{4.393}=\begin{bmatrix}.866\\ .677\\ .334\\ .855\\ .193\\ .021\end{bmatrix}.$$

Die Variablen "Hexen" (X_1) mit $t_{11}=.866$, "Okkultes Wissen" (X_2) mit $t_{12}=.677$ aus dem inhaltlichen Bereich *Magisch-irrationales Denken* und "Magische Vorstellungen" (X_4) mit $t_{14}=.855$ aus dem Bereich *Schizotypische Persönlichkeitsstörung* weisen die mit Abstand höchsten Strukturkoeffizienten auf. Die Variable "Religiosität" (X_3) erreicht immerhin noch einen Strukturkoeffizienten von $t_{13}=.334$, während die beiden Variablen aus dem Bereich *Zugehörige Symptome*, "Extraversion" (X_5) mit $t_{15}=.193$ und "Körperliche Beschwerden" (X_6) mit $t_{16}=.021$, nur niedrige Korrelationen mit den Diskriminanzwerten aufweisen.

Aufgrund der niedrigen standardisierten Diskriminanzkoeffizienten und der niedrigen Strukturkoeffizienten der beiden letztgenannten Variablen lässt sich vermuten, dass deren inkrementeller Beitrag zur Diskriminationsfähigkeit sich nicht als signifikant erweist. Zur Überprüfung des Inkrements der Variablen "Extraversion" (X_5) und "Körperliche Beschwerden" (X_6), bei simultaner Aufnahme in das Modell, benötigt man zunächst das multivariate Wilks Λ für das Modell mit den Variablen X_1, ..., X_4, welches analog dem bereits demonstrierten Rechengang mit einer auf $p' = 4$ eingeschränkten Variablenzahl erzielt wird:

$$\Lambda_{X_1,\ldots,X_4} = .366\,.$$

Das multivariate Wilks Λ des gegenüber diesem Modell um X_5 und X_6 erweiterten Modells wurde mit der uneingeschränkten Variablenzahl $p = 6$ bereits als

$$\Lambda_{X_1,\ldots,X_6} = .347$$

ermittelt. Das inkrementelle Wilks $\Lambda_{X_5,X_6|X_1,\ldots,X_4}$ errechnet sich nach Gleichung 45 als Quotient der beiden Werte für Wilks Λ:

$$\Lambda_{X_5,X_6|X_1,\ldots,X_4} = \frac{\Lambda_{X_1,\ldots,X_6}}{\Lambda_{X_1,\ldots,X_4}} = \frac{.347}{.366} = .947\,.$$

Den V-Wert des inkrementellen Wilks Λ für die $p^* = p'-p = 2$ interessierenden Variablen X_5 und X_6 ermittelt man nach der zur Überprüfung von Inkrementen modifizierten Gleichung 35:

$$\begin{aligned} V &= -[N - 1 - (p^* + k)/2]\ln \Lambda_{X_6|X_1,\ldots,X_4} \\ &= -[89 - 1 - (6 - 4 + 2)/2]\ln .947 \\ &= 4.683\,. \end{aligned}$$

Wählt man ein Signifikanzniveau von $\alpha = .05$, erweist sich der inkrementelle Beitrag der Variablen "Extraversion" (X_5) und "Körperliche Beschwerden" (X_6) bei simultaner Betrachtung nach der χ^2-Verteilung mit $p^*(k-1) = 2$ Freiheitsgraden als nicht signifikant ($p = .094$).

Für die diskriminanzanalytische Lösung, die lediglich die Merkmalsvariablen X_1, ..., X_4 enthält, lässt sich – wiederum analog dem bereits demonstrierten Rechengang – ein kanonischer Korrelationskoeffizient von

$$\rho_{1(X_1,\ldots,X_4)} = \left(\frac{\lambda_1}{1+\lambda_1}\right)^{\frac{1}{2}} = \left(\frac{1.729}{1+1.729}\right)^{\frac{1}{2}} = .796$$

berechnen; man erkennt, dass die mit den vier Merkmalsvariablen X_1, ..., X_4 gebildete Diskriminanzfunktion mit $\rho^2_{1(X_1,\ldots,X_4)} = .633$ einen erklärten Varianzanteil erzielt, der nur geringfügig niedriger ist als die erklärte Varianz, die durch die mit allen sechs Merkmalsvariablen X_1, ..., X_6 gebildete Diskriminanzfunktion erzielt wird ($\rho^2_{1(X_1,\ldots,X_6)} = .652$). Die beiden Variablen "Extraversion" (X_5) und "Körperliche Beschwerden" (X_6) aus dem Bereich *Zugehörige Symptome* bieten demnach im Gesamtkontext der einbezogenen Merkmalsvariablen einen mit ca. 2% nur minimalen Gewinn an diskriminatorischer Information, der sich zudem nicht zufallskritisch absichern lässt. Demgegenüber führt die allein mit den Merkmalsvariablen X_1, ..., X_4 gebildete Diskriminanzfunktion auch in der Grundgesamtheit zu einer Unterscheidung zwischen den beiden Gruppen; der nach Gleichung 36 ermittelbare *V*-Wert mit $p^* = p'$

$$\begin{aligned} V &= -\left[N - 1 - (p^* + k)/2\right] \ln \Lambda_{X_1,\ldots,X_4} \\ &= -\left[89 - 1 - (4+2)/2\right] \ln .366 \\ &= 85.435 \end{aligned}$$

erweist sich bei einem Signifikanzniveau von $\alpha = .01$ nach der χ^2-Verteilung mit $p^*+k-2s = 4$ Freiheitsgraden als signifikant. Die Berücksichtigung der Variablen "Hexen" (X_1) und "Okkultes Wissen" (X_2) aus dem Bereich *Magisch-irrationales Denken*, der Variablen "Religiosität" (X_3) sowie der Variablen "Magische Vorstellungen" (X_4) aus dem Bereich *Schizotypische Persönlichkeitsstörung* erweist sich zur Diskrimination zwischen den Gruppen der "Okkultistisch nicht Aktiven" (G_1) und der "Okkultistisch Aktiven" (G_2) als hinreichend: Mit den Variablen "Extraversion" (X_5) und "Körperliche Beschwerden" (X_6) aus dem Bereich *Zugehörige Symptome* geht kein bedeutsamer Informationsgewinn einher. Die Effizienz der Schätzung der Diskriminanzfunktion könnte sich ggf. erhöhen, wenn die beiden letztgenannten Variablen gleich von Beginn an unberücksichtigt bleiben.

3 Klassifikation von Objekten

In den Termini des erstmals von Welch (1939) in die diskriminanzanalytische Literatur eingeführten entscheidungstheoretischen Ansatzes stellt sich das Klassifikationsproblem folgendermaßen dar (vgl. Krzanowski & Marriott, 1995): Ein zufällig gezogenes Untersuchungsobjekt i mit den Ausprägungen x_{i1}, ..., x_{ip} in den p Merkmalsvariablen soll einer von k Populationen zugeordnet werden, wobei die einzelnen Populationen durch ihre (px1)-Mittelwertevektoren μ_1, ..., μ_g, ..., μ_k sowie in Form der Dichtefunktionen f_g (X_1, ..., X_p) der p Merkmale charakterisiert sind. Gesucht wird nun eine Klassifikationsregel, mit deren Hilfe für jedes Untersuchungsobjekt i ein Schätzwert für die Populationszugehörigkeit zu *einer* der k Populationen ermittelt werden kann. Dazu wird der p-dimensionale Merkmalsraum in k disjunkte, den jeweiligen Populationen zugeordnete Regionen unterteilt. Von *optimalen* Klassifikationsregeln kann dann gesprochen werden, wenn sie die tatsächliche Fehlerrate bzw. den Anteil der insgesamt falsch zugeordneten Untersuchungsobjekte minimieren und zugleich den Anteil der insgesamt richtig zugeordneten Untersuchungsobjekte maximieren.[21]

3.1 Klassifikationsregeln

Die Auswahl der hier besprochenen Verfahren erfolgt unter dem Aspekt, dass sie mit den Anwendungsvoraussetzungen der linearen Diskriminanzanalyse vereinbar und somit auch auf Diskriminanzwerte anwendbar sind (Abschnitt 3.2). Zunächst werden Maximum-Likelihood-Regeln auf der Basis des verallgemeinerten Distanzmaßes von Mahalanobis (1936) vorgestellt (Abschnitt 3.1.1), darauf aufbauend dann Regeln, die eine Maximierung der a posteriori-Wahrscheinlichkeit der Gruppenzugehörigkeit zur Grundlage haben (Abschnitt 3.1.2). Der letzte Abschnitt beschäftigt sich mit verschiedenen Methoden der Schätzung der tatsächlichen Fehlerrate, die zur Beurteilung der Klassifikationsfähigkeit einer diskriminanzanalytischen Lösung von Bedeutung sind (Abschnitt 3.3). Die wichtigsten der vorgestellten Verfahren werden wiederum

[21] Wir berücksichtigen hier nur Klassifikationsregeln, die in erster Linie die tatsächliche Fehlerrate minimieren. In bestimmten Kontexten ist die Minimierung anderer Fehlerraten sinnvoll (vgl. z.B. die Ansätze in Krauth, 1983). Eine Unterscheidung verschiedener Fehlerraten findet sich bei Trampisch (1977).

in einem konkreten Rechenbeispiel anhand des Datensatzes von Bednich (1993) verdeutlicht (Abschnitt 3.4).

3.1.1 Maximum-Likelihood-Regeln

Eine erste, auch intuitiv gut nachvollziehbare Klassifikationsregel besteht darin, ein Untersuchungsobjekt i derjenigen Population g zuzuordnen, welcher es hinsichtlich der Ausprägungen auf den p Variablen am "ähnlichsten" ist, d.h., zu deren Populationscentroid $\boldsymbol{\mu}_g = (\mu_{g1}, ..., \mu_{gj}, ..., \mu_{gp})'$ der durch $\mathbf{x}_i = (x_{i1}, ..., x_{ij}, ..., x_{ip})'$ definierte Punkt im Merkmalsraum die geringste Distanz aufweist. Zur Bestimmung dieser Distanz eignet sich das verallgemeinerte Distanzmaß von Mahalanobis (1936), das die Streuungen und Interkorrelationen der p Merkmalsvariablen berücksichtigt. Kann von Gleichheit der $(p\text{x}p)$-Varianz-Kovarianzmatrizen in den k Populationen ausgegangen werden ($\boldsymbol{\Sigma}_1 = ... = \boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}$), lautet die quadrierte Distanz der empirischen Werte x_{ij} zu den jeweiligen Populationsmittelwerten μ_{jg}:

$$D_{ig}^2 = (\mathbf{x}_i - \boldsymbol{\mu}_g)' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_g). \tag{46}$$

Das Untersuchungsobjekt i ist derjenigen Population zuzuordnen, für die D_{ig}^2 minimal ist. Unterstellen wir multivariate Normalverteilung der p Merkmalsvariablen, so folgt die Verteilung des verallgemeinerten Distanzmaßes D_{ig}^2 einer χ^2-Verteilung mit p Freiheitsgraden (vgl. Tatsuoka, 1971). Deshalb wird diese erste Klassifikationsregel auch "*Minimum-Chi-Quadrat-Regel*" genannt. Sind die Populationsparameter $\boldsymbol{\mu}_g$ und $\boldsymbol{\Sigma}$ nicht bekannt, benutzen wir den Stichprobenschätzer $\overline{\mathbf{x}}_g$ bzw. $(N-k)^{-1}\mathbf{W}$.[22] Ist die Annahme der Gleichheit der Varianz-Kovarianzmatrizen $\boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_k$ nicht haltbar, ist $\boldsymbol{\Sigma}$ in Gleichung 46 durch die jeweilige gruppenspezifische Varianz-Kovarianzmatrix $\boldsymbol{\Sigma}_g$ zu ersetzen. Zur Klassifikation ist dann eine modifizierte Minimum-Chi-Quadrat-Regel zu verwenden (vgl. Tatsuoka, 1975), gemäß der bei der Entscheidung über die Gruppenzugehörigkeit der Wert

---

[22] In der Praxis dürfte es für gewöhnlich der Fall sein, dass die für die Anwendung der berichteten Klassifikationsregeln benötigten Populationsparameter aus Stichprobendaten zu schätzen sind. Dies folgt im Grunde aus der Anwendungslogik von Klassifikationsverfahren: Wenn die Populationsparameter bekannt sind, setzt dies eine Vollerhebung aller interessierenden Grundgesamtheiten voraus. Eine nachträgliche Klassifikation eines Untersuchungsobjekts wäre dann nur sinnvoll, wenn die Klassenzugehörigkeit, z.B. aufgrund von verlorengegangenen Daten, nicht mehr bekannt ist.

$$D'^2_{ig} = (\mathbf{x}_i - \boldsymbol{\mu}_g)' \boldsymbol{\Sigma}_g^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_g) + \ln|\boldsymbol{\Sigma}_g| \tag{47}$$

zu minimieren ist. Sind die Populationsparameter $\boldsymbol{\mu}_g$ und $\boldsymbol{\Sigma}_g$ nicht bekannt, sind auch in Gleichung 47 die Stichprobenschätzer $\bar{\mathbf{x}}_g$ bzw. $(n_g-1)^{-1}\mathbf{W}_g$ zu verwenden.

Die "*Maximum-Centour-Regel*" (vgl. Rulon, Tiedemann, Tatsuoka & Langmuir, 1967) beruht unmittelbar auf der Minimum-Chi-Quadrat-Regel und führt daher prinzipiell nicht zu anderen Klassifikationsentscheidungen, hat aber den Vorteil größerer Anschaulichkeit, da sie sich die Verteilungsfunktion der mit p Freiheitsgraden χ^2-verteilten verallgemeinerten Distanzwerte D^2_{ig} zunutze macht. Der Centour-Wert C_{ig} von Untersuchungsobjekt i bezogen auf Population g ist definiert als die Differenz von 100 und dem Prozentrang PR_{ig}, der sich dem jeweiligen D^2_{ig}-Wert anhand der entsprechenden χ^2-Verteilung zuweisen lässt:

$$C_{ig} = 100 - PR_{ig} \,. \tag{48}$$

Der Ausdruck $(100-PR_{ig})/100$ gibt die Überschreitungswahrscheinlichkeit dafür an, dass in Population g die Merkmalsausprägungen $\mathbf{x}_i$ oder – bezogen auf den Gruppencentroid $\boldsymbol{\mu}_g$ – extremere Werte auftreten. Daher lässt sich an dem Centourwert C_{ig} unmittelbar die Plausibilität einer Zuordnung von i zu Population g ablesen: das Untersuchungsobjekt i wird derjenigen Population zugeordnet, für die C_{ig} maximal ist. Ergeben sich durchweg kleine Centour-Werte und erscheint daher keine der k Klassifikationsalternativen plausibel, kann unter Umständen auf eine Zuordnung überhaupt verzichtet werden.

Alle bisher berichteten Klassifikationsregeln können auch als *Maximum-Likelihood-Regeln* aufgefasst werden, da sich die Minimierung von D^2_{ig} und die Maximierung der *Likelihood* $L(\mathbf{x}_i \mid G_g)$ der Zugehörigkeit des Untersuchungsobjekts i mit dem Merkmalsvektor $\mathbf{x}_i$ zu Population g als äquivalent erweisen.[23] Unter der Voraussetzung von multivariater Normalverteilung und Gleichheit der Varianz-Kovarianzmatrizen lautet die zu Gleichung 46 äquivalente Dichtefunktion zur Ermittlung der Likelihood von $\mathbf{x}_i$ bei gegebener Population g:

$$L(\mathbf{x}_i|G_g) = f_g(\mathbf{x}_i) = (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left[(-1/2)(\mathbf{x}_i - \boldsymbol{\mu}_g)' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_g)\right]. \tag{49}$$

[23] Man erhält eine Minimum-Chi-Quadrat-Regel, wenn statt der $\mathbf{x}_i$ zugeordneten Werte der gruppenspezifischen Dichtefunktionen f_g (s. Gleichung 49) die natürlichen Logarithmen der Dichtefunktionen Verwendung finden (vgl. Tatsuoka, 1975).

Liegt die Voraussetzung gleicher Varianz-Kovarianzmatrizen nicht vor, ist die zu Gleichung 47 äquivalente Dichtefunktion zu verwenden:

$$L(\mathbf{x}_i | G_g) = f_g{}'(\mathbf{x}_i) = (2\pi)^{-p/2} |\boldsymbol{\Sigma}_g|^{-1/2} \exp\left[(-1/2)(\mathbf{x}_i - \boldsymbol{\mu}_g)' \boldsymbol{\Sigma}_g{}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_g)\right]. \quad (50)$$

Das Untersuchungsobjekt i wird derjenigen Population zugeordnet, bei der die Likelihood $L(\mathbf{x}_i | G_g)$ maximal wird.[24] Auch für Gleichung 49 und Gleichung 50 gilt wieder, dass die Stichprobenschätzer $\bar{\mathbf{x}}_g$ und $(N-k)^{-1}\mathbf{W}$ bzw. $(n_g-1)^{-1}\mathbf{W}_g$ zu verwenden sind, sofern die Populationsparameter $\boldsymbol{\mu}_g$ und $\boldsymbol{\Sigma}$ bzw. $\boldsymbol{\Sigma}_g$ nicht bekannt sind.

3.1.2 Maximum-Probability-Regel

Die Anwendung von Maximum-Likelihood-Regeln minimiert die tatsächliche Fehlerrate genau dann, wenn keine a-priori-Wahrscheinlichkeiten der Gruppenzugehörigkeit zu berücksichtigen sind. Diese Einschränkung hängt damit zusammen, dass die Likelihood $L(\mathbf{x}_i | G_g)$ unter der hypothetischen Annahme bestimmt wird, das Untersuchungsobjekt i sei der Population g zugehörig. Eine auch bei ungleichen a-priori-Wahrscheinlichkeiten optimale Klassifikationsregel muss sich dagegen auf die *bedingten Wahrscheinlichkeiten* $p(G_g | \mathbf{x}_i)$ stützen, dass die Population g bei Vorliegen des Merkmalsvektors $\mathbf{x}_i$ die zutreffende ist.

Dieser Überlegung folgt die "*Maximum-Probability*"-*Regel*, der zufolge ein Untersuchungsobjekt mit den Merkmalsausprägungen $\mathbf{x}_i$ genau derjenigen Population g zuzuordnen ist, für die sich eine maximale bedingte Wahrscheinlichkeit $p(G_g | \mathbf{x}_i)$ ergibt. Nach dem Bayes-Theorem bestimmt sich die gesuchte bedingte oder a-posteriori-Wahrscheinlichkeit $p(G_g | \mathbf{x}_i)$ wie folgt:

$$p(G_g | \mathbf{x}_i) = \frac{p(G_g) \cdot p(\mathbf{x}_i | G_g)}{\sum_{g=1}^{k} p(G_g) \cdot p(\mathbf{x}_i | G_g)}. \quad (51)$$

Der Ausdruck $p(G_g)$ bezeichnet die a-priori-Wahrscheinlichkeit der Zugehörigkeit zu Gruppe g, die vor allem aufgrund des Anteils der verschiedenen Teilpopulationen an der Gesamtpopulation variieren kann. Sofern multivariate Normalverteilung der p

[24] Dies kann auch so umschrieben werden, dass die Annahme, dass das jeweilige Untersuchungsobjekt i aus Population g stammt, am "plausibelsten" ist.

Merkmalsvariablen angenommen werden kann, lässt sich die bedingte Wahrscheinlichkeit $p(\mathbf{x}_i \mid G_g)$ – wie die Likelihood $L(\mathbf{x}_i \mid G_g)$ – aus den Dichtefunktionen ableiten. Dadurch kann in Gleichung 51 statt der bedingten Wahrscheinlichkeit die Wahrscheinlichkeitsdichte eingesetzt werden, die anhand von Gleichung 49 bzw. 50 ermittelt werden muss.[25] Nach Vereinfachungen ergibt sich für die a-posteriori-Wahrscheinlichkeit $p(G_g \mid \mathbf{x}_i)$ folgende Bestimmungsgleichung, die unter Voraussetzung der Gleichheit der Varianz-Kovarianzmatrizen anzuwenden ist:

$$p(G_g|\mathbf{x}_i) = \frac{p(G_g)\cdot \exp\left[-(1/2)(\mathbf{x}_i - \boldsymbol{\mu}_g)'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_g)\right]}{\sum_{g=1}^{k} p(G_g)\cdot \exp\left[(-1/2)(\mathbf{x}_i - \boldsymbol{\mu}_g)'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_g)\right]}. \tag{52}$$

Kann nicht von Gleichheit der Varianz-Kovarianzmatrizen ausgegangen werden, modifiziert sich die Bestimmungsgleichung zu

$$p(G_g|\mathbf{x}_i) = \frac{p(G_g)\cdot |\boldsymbol{\Sigma}_g|^{-1/2} \exp\left[(-1/2)(\mathbf{x}_i - \boldsymbol{\mu}_g)'\boldsymbol{\Sigma}_g^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_g)\right]}{\sum_{g=1}^{k} p(G_g)\cdot |\boldsymbol{\Sigma}_g|^{-1/2} \exp\left[(-1/2)(\mathbf{x}_i - \boldsymbol{\mu}_g)'\boldsymbol{\Sigma}_g^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_g)\right]}. \tag{53}$$

Für Gleichung 52 und Gleichung 53 ist wiederum anzumerken, dass die Stichprobenschätzer $\bar{\mathbf{x}}_g$ und $(N-k)^{-1}\mathbf{W}$ bzw. $(n_g-1)^{-1}\mathbf{W}_g$ einzusetzen sind, sofern die Populationsparameter $\boldsymbol{\mu}_g$ und $\boldsymbol{\Sigma}$ bzw. $\boldsymbol{\Sigma}_g$ nicht bekannt sind. Die Schätzung der a-priori-Wahrscheinlichkeiten aus den Größen der Teilstichproben als n_g/N ist natürlich nur dann sinnvoll, wenn tatsächlich eine Zufallsstichprobe der interessierenden Gesamtpopulation erhoben wurde.

[25] Wegen der Beziehung $p(\mathbf{x}_i|g) = L(\mathbf{x}_i|g)(\Delta X_1)\cdot\ldots\cdot(\Delta X_j)\cdot\ldots\cdot(\Delta X_p)$ gilt, dass sich die Proportionalitätskonstante $(\Delta X_1)\cdot\ldots\cdot(\Delta X_j)\cdot\ldots\cdot(\Delta X_p)$ – unabhängig von dem für das Intervall ΔX_j gewählten Wert – bei Einsetzen in Gleichung 51 wegkürzen lässt.

3.2 Verwendung von Diskriminanzwerten für die Klassifikation

Sämtliche der referierten Klassifikationsregeln können auch auf Basis von (normierten) Diskriminanzwerten anstelle der ursprünglichen Merkmalsvariablen zur Anwendung gebracht werden. Aufgrund der wechselseitigen Unabhängigkeit der Diskriminanzfunktionen reduziert sich Gleichung 46 bei der Verwendung normierter Diskriminanzwerte zur euklidischen Distanz

$$D_{ig}^2 = \left(\mathbf{y}_i - \overline{\mathbf{y}}_g\right)'\left(\mathbf{y}_i - \overline{\mathbf{y}}_g\right). \tag{55}$$

Werden sämtliche Diskriminanzfunktionen einbezogen, so führen Klassifikationen, die mit den ursprünglichen Merkmalsvariablen und den Diskriminanzwerten vorgenommen werden, zu identischen Resultaten (vgl. Tatsuoka, 1956, zitiert nach Tatsuoka, 1971). Um die Zuverlässigkeit der Zuordnungen zu erhöhen, empfiehlt es sich jedoch, lediglich die statistisch bedeutsamen Diskriminanzfunktionen zu berücksichtigen (vgl. Abschnitt 2.2.3.2), wodurch der Einfluss von zufälligen, stichprobenbedingten Gruppenunterschieden auf die Klassifikation neuer Objekte reduziert werden kann. Weiterhin führt eine diskriminanzanalytische Selektion derjenigen Merkmalsvariablen, die zur Gruppentrennung am besten geeignet sind (vgl. Abschnitt 2.2.4.3), im Allgemeinen auch zu einer größeren Stabilität von Klassifikationen.

3.3 Methoden zur Schätzung von Fehlerraten

Die Güte einer diskriminanzanalytischen Lösung kann (neben den in Abschnitt 2.2.3 dargestellten Möglichkeiten) anhand der tatsächlichen Fehlerraten beurteilt werden, die bei ihrer klassifikatorischen Anwendung entstehen. Die Klassifikationsfähigkeit lässt sich dabei auch inferenzstatistisch prüfen, indem die Klassifikationsmatrix, die zeilenweise die tatsächlichen Gruppenzugehörigkeiten und spaltenweise die prognostizierten Gruppenzugehörigkeiten enthält, kontingenzanalytisch ausgewertet wird. Insbesondere für eine bestimmte Variablenauswahl (vgl. Abschnitt 2.2.4.3) lässt sich mit einer solchen "probeweisen" Klassifikation ein empirisches Kriterium gewinnen, anhand dessen die Zuverlässigkeit der im Einzelfall zu treffenden klassifikatorischen Entscheidung in Abhängigkeit von der jeweiligen Variablenauswahl eingeschätzt werden kann. Im Folgenden sollen drei gängige Methoden der Schätzung der tatsächlichen Fehlerraten kurz vorgestellt werden.

Die *"Resubstitutions"-Methode* (nach Smith, 1947) ist das am wenigsten aufwendige Verfahren, da die Klassifizierung einfach auf denselben Datensatz appliziert wird, der bereits zur Bestimmung der Diskriminanzfunktionen verwendet wurde. Unter der realistischen Annahme, dass der Datensatz Verzerrungen aufweist, da er nicht aus einer Vollerhebung stammt, führt die Resubstitutionsmethode zu einer optimistischen *Unterschätzung* der tatsächlichen Fehlerrate, wenn die diskriminanzanalytische Lösung auch zur Klassifikation von stichprobenfremden Untersuchungsobjekten verwendet werden soll. Die Unterschätzung wird allerdings mit zunehmendem Stichprobenumfang geringer.

Im Unterschied dazu fallen die Fehlerratenschätzungen mit der *"Hold-out sample"-Methode* tendenziell zu pessimistisch aus (vgl. Deichsel & Trampisch, 1985). Bei der "Hold-out sample"-Methode wird die Gesamtstichprobe in eine sogenannte Konstruktionsstichprobe, anhand welcher die Diskriminanzfunktionen ermittelt werden und eine Klassifikationsstichprobe aufgeteilt, mit deren Hilfe man die Fehlerrate schätzt. Dabei kann dieselbe Gesamtstichprobe mehrmals in verschiedene Konstruktions- und Klassifikationsstichproben zerlegt werden, um die Genauigkeit der Fehlerratenschätzung zu erhöhen. Da jedoch für die Ermittlung der Diskriminanzfunktionen jeweils nur ein Teil der zur Verfügung stehenden Information ausgenutzt wird, ergeben sich – zumindest bei kleinen Gesamtstichproben – suboptimale Lösungen mit *Überschätzungen* der Fehlerrate.

Die *"Leave-one-out"-Methode* (Lachenbruch, 1967) ist sozusagen ein Spezialfall der "Hold-out sample"-Methode. Dabei wird jeweils ein Untersuchungsobjekt aus der Gesamtstichprobe anhand der Diskriminanzfunktionen klassifiziert, die auf Basis der Daten der übrigen N-1 Untersuchungsobjekte bestimmt worden sind. Dieses Verfahren wird dann N-mal wiederholt, so dass jede der N Personen (d.h. die jeweils ausgelassene Person) auf Basis einer Konstruktionsstichprobe von $N-1$ Personen klassifiziert werden kann. Die relativ aufwendige "Leave-one-out"-Methode liefert unter vollständiger Ausnutzung der vorhandenen Information eine nahezu unverzerrte Schätzung der tatsächlichen Fehlerraten.

3.4 Rechenbeispiel: Zuordnung von Probanden zur Gruppe der "Okkultistisch Aktiven" oder zur Gruppe der "Okkultistisch nicht Aktiven"

Zur Illustration der Anwendung einiger der genannten Klassifikationsregeln soll wieder der Datensatz aus der Untersuchung von Brednich (1993) herangezogen werden. Die exemplarische Veranschaulichung der Zuordnung eines Untersuchungsobjektes soll anhand des Diskriminanzwerts eines Probanden erfolgen, dessen Gruppenzugehörigkeit bereits bekannt ist. Dies entspricht dem Vorgehen bei der Überprüfung der Klassifikationsfähigkeit einer diskriminanzanalytischen Lösung anhand der Resubstitutionsmethode. Im Anschluss daran soll die Klassifikationsmatrix erläutert werden, die man erhält, wenn alle $N = 89$ Untersuchungsobjekte in der dargelegten Weise klassifiziert werden.

Als Untersuchungsobjekt wird einfachheitshalber der erste Proband aus der Teilstichprobe der "Okkultistisch nicht Aktiven" (G_1) mit dem Merkmalsvektor

$$\mathbf{x}'_1 = \begin{bmatrix} 6 & 1 & 8 & 4 & 4 & 5 \end{bmatrix}$$

gewählt. Durch Einsetzen in die normierte Diskriminanzfunktion erhält man den Diskriminanzwert

$$\begin{aligned} y_{11} &= b_{10} + \mathbf{x}'_1 \mathbf{b}_1 \\ &= -3.962 + \begin{bmatrix} 6 & 1 & 8 & 4 & 4 & 5 \end{bmatrix} \cdot \begin{bmatrix} 0.384 \\ 0.261 \\ -.139 \\ 0.335 \\ 0.102 \\ -.106 \end{bmatrix} = -1.285. \end{aligned}$$

Da mit normierten Diskriminanzwerten gearbeitet wird und Gleichheit der Varianz-Kovarianz-Matrizen in den Gruppen unterstellt werden kann, werden zur Ermittlung der Distanzen von Untersuchungsobjekt 1 zu den Gruppenmittelwerten der Diskriminanzvariablen die euklidischen Distanzen nach Gleichung 55 berechnet:

$$D_{11}^2 = (y_{11} - \bar{y}_{11})^2 = (-1.285 + 1.325)^2 = 1.568 \cdot 10^{-3},$$

$$D_{12}^2 = (y_{11} - \bar{y}_{12})^2 = (-1.285 - 1.342)^2 = 6.898\,.$$

Da die euklidische Distanz zum ersten Gruppenmittelwert mit $D_{11}^2 = 0.0016$ im Vergleich zu $D_{12}^2 = 6.898$ sehr klein ausfällt, wird das Untersuchungsobjekt 1 nach der Minimum-Chi-Quadrat-Regel eindeutig der Gruppe der "Okkultistisch nicht Aktiven" (G_1) zugeordnet, wobei ein Vergleich mit der tatsächlichen Gruppenzugehörigkeit zeigt, dass die Klassifikationsregel zu einer richtigen Zuordnung geführt hat.

Da nur eine Diskriminanzfunktion zu berücksichtigen ist, die euklidischen Distanzen bereits vorliegen und der Stichprobenschätzer $(N\text{–}k)^{-1}\mathbf{W}$ für die Varianz-Kovarianz-Matrix $\mathbf{\Sigma}$ aufgrund der Normierung und wechselseitigen Unabhängigkeit der Diskriminanzfunktionen eine Einheitsmatrix ist, vereinfacht sich die Berechnung der Likelihoods bzw. Dichten nach Gleichung 49:

$$L(y_{11}|G_1) = f_1(y_{11}) = (2\pi)^{-\frac{1}{2}} \exp\left[(-1/2)D_{11}^2\right] = (2\pi)^{-\frac{1}{2}} \exp\left[(-1/2)\cdot 1.568\cdot 10^{-3}\right] = .399\,,$$

$$L(y_{11}|G_2) = f_2(y_{11}) = (2\pi)^{-\frac{1}{2}} \exp\left[(-1/2)D_{12}^2\right] = (2\pi)^{-\frac{1}{2}} \exp\left[(-(1/2)\cdot 6.898\right] = 1.268\cdot 10^{-2}\,.$$

Wie zu erwarten ist, ergibt sich für die Klassifikation von Untersuchungsobjekt 1 in die Gruppe der "Okkultistisch nicht Aktiven" (G_1) auch eine viel größere Likelihood als für die Klassifikation in die Gruppe der "Okkultistisch Aktiven" (G_2).

Die Bestimmung der a-posteriori-Wahrscheinlichkeiten nach Gleichung 52 lässt sich ebenfalls vereinfachen, da – zumindest auf Stichprobenebene – von gleichen a-priori-Wahrscheinlichkeiten ausgegangen wird:

$$\begin{aligned} p(G_1|y_{11}) &= \frac{\exp\left[(-1/2)D_{11}^2\right]}{\exp\left[(-1/2)D_{11}^2\right] + \exp\left[(-1/2)D_{12}^2\right]} \\ &= \frac{\exp\left[(-1/2)\cdot 1.568\cdot 10^{-3}\right]}{\exp\left[(-1/2)\cdot 1.568\cdot 10^{-3}\right] + \exp\left[(-1/2)\cdot 6.898\right]} = .969\,, \end{aligned}$$

$$\begin{aligned} p(G_2|y_{11}) &= \frac{\exp\left[(-1/2)D_{12}^2\right]}{\exp\left[(-1/2)D_{11}^2\right] + \exp\left[(-1/2)D_{12}^2\right]} \\ &= \frac{\exp\left[(-1/2)\cdot 6.898\right]}{\exp\left[(-1/2)\cdot 1.568\cdot 10^{-3}\right] + \exp\left[(-1/2)\cdot 6.898\right]} = 3.082\cdot 10^{-2}\,. \end{aligned}$$

Die a-posteriori-Wahrscheinlichkeit von $p(G_1|y_{11}) = .969$ für die Zugehörigkeit von Untersuchungsobjekt zur Gruppe der "Okkultistisch nicht Aktiven" (G_1) ist wesentlich höher als $p(G_2|y_{11}) = 3.082\cdot 10^{-2}$, so dass erwartungsgemäß (da keine unterschiedlichen a-priori-Wahrscheinlichkeiten zu berücksichtigen waren) auch nach der

Maximum-Probability-Regel das Untersuchungsobjekt 1 der Gruppe G_1 zuzuordnen ist.

Wird die Maximum-Probability-Regel zur Überprüfung der Klassifikationsgüte aller $N=89$ Untersuchungsobjekte angewendet, so erhält man die in Tabelle 2 wiedergegebene Klassifikationsmatrix:

Tabelle 2. Klassifikationsmatrix

	Prognostizierte Gruppenzugehörigkeit		
Tatsächliche Gruppenzugehörigkeit	G_1	G_2	Zeilensummen
G_1	40 44.94%	4 4.49%	44
G_2	5 5.62%	40 44.94%	45
Spaltensummen	45	44	N=89

Anmerkungen. G_1: "Okkultistisch nicht Aktive"; G_2: "Okkultistisch Aktive". In der Hauptdiagonalen findet sich der Anteil der richtig klassifizierten Untersuchungsobjekte, außerhalb der Hauptdiagonalen der Anteil der falsch klassifizierten Untersuchungsobjekte.

Wie man an der Klassifikationsmatrix ablesen kann, wird mit einer Klassifikation auf Basis der Diskriminanzwerte eine mit etwa 10% recht geringfügige tatsächliche Fehlerrate erzielt. Eine Klassifikation auf Basis der Merkmalswerte würde zu identischen Resultaten führen, da die einzige extrahierbare Diskriminanzfunktion das gesamte Diskriminanzpotential der Merkmalsvariablen ausschöpft. Zu beachten ist allerdings, dass die Fehlerratenschätzung mit Hilfe der Resubstitutionsmethode die tatsächlichen Fehlerraten optimistisch unterschätzt. Aus einer zusätzlichen Fehlerratenschätzung anhand der "Leave-one-out"-Methode würde wahrscheinlich eine etwas geringere Trefferquote resultieren.

Literatur

Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (1996). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung* (8. Aufl.). Berlin: Springer.

Bortz, J. (1993). *Statistik für Sozialwissenschaftler* (4. Aufl.). Berlin: Springer.

Brednich, A. (1993). *Eine Fragebogenuntersuchung zur Erfassung von magisch-irrationalem Denken und der Schizotypischen Persönlichkeitsstörung bei Erwachsenen*. Unveröff. Diplomarbeit, Albert-Ludwigs-Universität, Freiburg.

Bryan, J. G. (1951). The generalized discriminant function: Mathematical foundation and computational routine. *Harvard Educational Review*, *21*, 90-95.

Chapman, L.J., Chapman, J.P. & Raulin, M.L. (1978). Body-image abberation in schizophrenie. *Journal of Abnormal Psychology*, *87*, 399-407.

Claridge, G. & Broks, P. (1984). Schizotypy and hemisphere function I: Theoretical considerations and the measurement of schizotypy. *Personality and Individual Differences*, *5*, 633-648.

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

Das Gupta, S. (1973). Theories and methods in classification: A review. In T. Cacoullos (Ed.), *Discriminant analysis and applications* (pp. 77-137). New York: Academic Press.

Deichsel, G. & Trampisch, H.J. (1985). *Clusteranalyse und Diskriminanzanalyse*. Stuttgart: Fischer.

Dillon, W.R. & Goldstein, M. (1984). *Multivariate analysis. Methods and applications*. New York: Wiley.

Dixon, W.J. (Ed.) (1992). *BMDP statistical software manual: To accompany the 7.0 software release*. Berkeley, CA: University of California Press.

Eckbald, M. & Chapman, L.J. (1983). Magical ideation as an indicator of schizotypy. *Journal of Consulting and Clinical Psychology*, *51*, 215-255.

Erb, W. (1990). *Anwendungsmöglichkeiten der linearen Diskriminanzanalyse in Geographie und Regionalwissenschaft*. Hamburg: Verlag Weltarchiv.

Fahrenberg, J., Selg, H. & Hampel, R. (1984). *Das Freiburger Persönlichkeits-Inventar FPI. Revidierte Fassung FPI-R und teilweise geänderte Fassung FPI-A1* (4., teilw. geänderte Aufl.). Göttingen: Hogrefe.

Fahrmeir, L. & Hamerle, A. (Hrsg.) (1984). *Multivariate statistische Verfahren*. Berlin: de Gruyter.

Fisher, R. A. (1936). The use of multiple measurement in taxonomic problems. *Annals of Eugenics*, *7*, 179-188.

Fujikoshi, Y. (1989). Tests for redundancy of some variables in multivariate analysis. In Y. Dodge (Ed.), *Statistical data analysis and inference*. Amsterdam: Elsevier Science Publishers.

Goldstein, M. & Dillon, W.R. (1978). *Discrete discriminant analysis*. New York: Wiley.

Hartung, J. & Elpelt, B. (1984). *Multivariate Statistik. Lehr- und Handbuch der angewandten Statistik.* München: Oldenbourg.

Hattemer, H. (1974). Ein Zusammenhang zwischen Diskriminanz- und Regressionsanalyse. *EDV in Medizin und Biologie, 5,* 7-10.

Huberty, C.J. (1975). Discriminant analysis. *Review of Educational Research, 45,* 543-598.

Kendall, M.G. (1966). Discrimination and classification. In P.R. Krishnaiah (Ed.), *Multivariate analysis* (pp. 165-185). New York: Academic Press.

Krauth, J. (1983). Diskriminanzanalyse. In J. Bredenkamp & H. Feger (Hrsg.), *Enzyklopädie der Psychologie, Themengebiet B: Methodologie und Methoden, Serie 1: Forschungsmethoden der Psychologie, Bd. 4: Strukturierung und Reduzierung von Daten* (S. 293-350). Göttingen: Hogrefe.

Krzanowski, W.J. & Marriott, F.H.C. (1995). *Multivariate analysis, part 2. Classification, covariance structures and repeated measurements.* London: Arnold.

Lachenbruch, P.A. (1967). An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. *Biometrics, 23,* 639-645.

Lachenbruch, P. A. (1975). *Discriminant analysis.* New York: Hafner.

Lachenbruch, P. A. & Goldstein, M. (1979). Discriminant analysis. *Biometrics, 35,* 69-85.

Läuter, J. (1992). *Stabile multivariate Verfahren. Diskriminanzanalyse–Regressionsanalyse–Faktoranalyse.* Berlin: Akademie Verlag.

Mahalanobis, P.C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Science (India), 12,* 49-55.

Mischo, J. (1991). *Okkultismus bei Jugendlichen.* Mainz: Grünewald-Verlag.

Mischo, J. (1996). Schizotypische Muster im Denken und Verhalten? *TW Neurologie Psychiatrie, 10,* 266-272.

Moosbrugger, H. (1983). Modelle zur Beschreibung statistischer Zusammenhänge in der psychologischen Forschung. In J. Bredenkamp & H. Feger (Hrsg.), *Enzyklopädie der Psychologie, Themengebiet B: Methodologie und Methoden, Serie 1: Forschungsmethoden der Psychologie, Bd. 4: Strukturierung und Reduzierung von Daten* (S. 1-58). Göttingen: Hogrefe.

Moosbrugger, H. (1994). *Lineare Modelle. Regressions- und Varianzanalyse.* Bern: Huber.

Moosbrugger, H. (1997). *Multivariate statistische Analyseverfahren* (3. Aufl.). Münster: Institut für Sozialwissenschaftliche Forschung.

Moosbrugger, H. & Frank, D. (1992). *Clusteranalytische Methoden in der Persönlichkeitsforschung. Eine anwendungsorientierte Einführung in taxometrische Klassifikationsverfahren.* Göttingen: Huber.

Moran, M.A. (1975). The effects of selecting variables for use in the linear discriminant function. *EDV in Medizin und Biologie, 1/2,* 24-30.

Nakache, J.P. & Dusserre, L. (1975). Practical problems in linear discriminant analysis. *EDV in Medizin und Biologie*, *1/2*, 30-35.

Rao, C. R. (1948). The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society, Series B, 10*, 159-203.

Rao, C.R. (1973). *Linear statistical inference and its applications* (2nd ed.). New York: Wiley.

Rencher, A.C. (1988). On the use of correlations to interpret canonical functions. *Biometrika*, *75*, 363-365.

Rencher, A.C. (1993). The contribution of individual variables to Hotelling's T^2, Wilk's Λ, and R^2. *Biometrics*, *49*, 479-489.

Rulon, P.J., Tiedemann, D.V., Tatsuoka, M.M. & Langmuir, C.R. (1967). *Multivariate statistics for personnel classification*. New York: Wiley.

SAS Institute Inc. (1998). *SAS/STAT user's guide, Version 6* (4th ed.). Cary, NC: SAS Institute Inc.

Schweizer, K. (1996). Klassifikation. In G. Strube (Hrsg.). *Wörterbuch der Kognitionswissenschaft* (S. 299). Stuttgart: Klett-Cotta.

Smith, C.A.B. (1947). Some examples of discrimination. *Annals of Eugenics*, *13*, 272-282.

SPSS Inc. (1996). *SPSS Professional Statistics 7.5*. Chicago: SPSS Inc.

Tatsuoka, M.M. (1953). *The relationship between canonical correlation and discriminant analysis*. Cambridge, MA: Educational Research Corporation.

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

Tatsuoka, M.M. (1975). Classification procedures. In D.J. Amick & H.J. Walberg (Eds.), *Introductory multivariate analysis for educational, psychological and social research* (pp. 257-284). Berkeley, CA: McCutchan.

Trampisch, H.J. (1975). Trennprobleme bei unvollständiger Information – Eine Übersicht. *EDV in Medizin und Biologie*, *6*, 2-8.

Trampisch, H.-J. (1977). Grundbegriffe der Diskriminanzanalyse. *Metamed*, *1*, 365-373.

Welch, B.L. (1939). Note on discriminant functions. *Biometrika*, *31*, 218-222.

Anhang: Index der verwendeten Notation in alphabetischer Reihenfolge

Symbol	Bedeutung
b_{s0}	Additive Konstante in der s-ten, normierten Diskriminanzfunktion
b_{sj}	Normierter Diskriminanzkoeffizient von Merkmalsvariable j in der s-ten Diskriminanzfunktion
b^*_{sj}	Standardisierter Diskriminanzkoeffizient von Merkmalsvariable j in der s-ten Diskriminanzfunktion
$\bar{b}^*_j$	Mittlerer Diskriminanzkoeffizient von Merkmalsvariable j
$\mathbf{b}_s$	$(p\text{x}1)$-Vektor der normierten Diskriminanzkoeffizienten der s-ten Diskriminanzfunktion
$\mathbf{b}^*_s$	$(p\text{x}1)$-Vektor der standardisierten Diskriminanzkoeffizienten der s-ten Diskriminanzfunktion
$\boldsymbol{\beta}_s$	$(p\text{x}1)$-Vektor der Diskriminanzkoeffizienten der s-ten Diskriminanzfunktion in der Grundgesamtheit
$\mathbf{B}$	$(p\text{x}p)$-Matrix der Zwischengruppenstreuungen (Zwischengruppen-Quadratsummen und -Kreuzprodukte)
C_{ig}	Centour-Wert von Untersuchungsobjekt i bezogen auf Gruppe g
$\mathbf{C}$	$(p\text{x}p)$-Gesamt-Varianz-Kovarianzmatrix der p Merkmalsvariablen
D^2_{ig}	Quadrierter Abstand von Untersuchungsobjekt i und dem Gruppencentroid von Gruppe g
$\mathbf{D}$	$(p\text{x}p)$-Diagonalmatrix mit den Innergruppen-Varianzen der p Merkmalsvariablen auf der Hauptdiagonalen
$f_g(\mathbf{x}_i)$	Wert von Untersuchungsobjekt i in der Dichtefunktion der p Merkmalsvariablen in Gruppe g
g	Laufindex für die Gruppen, mit $g = 1, ..., k$
G_g	Gruppe g
i	Laufindex für die Untersuchungsobjekte, mit $i = 1, ..., N$
i'	Laufindex für die Untersuchungsobjekte, mit $i' = 1, ..., n_g$
$\mathbf{I}$	$(p\text{x}p)$-Einheitsmatrix
j	Laufindex für die Merkmalsvariablen, mit $j = 1, ..., p$
k	Anzahl der Gruppen
$L(\mathbf{x}_i\|g)$	Likelihood der Zugehörigkeit von Untersuchungsobjekt i zu Gruppe g
λ_s	Diskriminanzkriterium der s-ten Diskriminanzfunktion
Λ	Multivariates Wilks Lambda
Λ_s	Univariates Wilks Lambda der s-ten Diskriminanzfunktion
$\Lambda_{X,Z}$	Multivariates Wilks Lambda für das um die Variable Z erweiterte Modell

$\Lambda_{Z\mid X}$	Inkrementelles Wilks Lambda für die Erweiterung des ursprünglichen Modells um die Merkmalsvariable Z
μ_{gj}	Mittelwert von Merkmalsvariable j in der g-ten Subpopulation
$\boldsymbol{\mu}_g$	(px1)-Mittelwertevektor der p Merkmalsvariablen in der g-ten Subpopulation
$\boldsymbol{\mu}_g^{(Z)}$	((p+1)x1)-Mittelwertevektor der p Merkmalsvariablen und der zusätzlichen Merkmalsvariablen Z in der g-ten Subpopulation
n_g	Anzahl der Untersuchungsobjekte in Gruppe g
N	Anzahl der Untersuchungsobjekte in allen k Gruppen
p	Anzahl der Merkmalsvariablen
p'	Anzahl der Merkmalsvariablen im eingeschränkten Modell
$p(g)$	A-priori Wahrscheinlichkeit der Zugehörigkeit zu Gruppe g
$p(g\mid\mathbf{x}_i)$	A-posteriori-Wahrscheinlichkeit der Zugehörigkeit von Untersuchungsobjekt i zu Gruppe g
PR_{ig}	Prozentrang (Chi-Quadrat-Verteilung) des quadrierten Abstands von Untersuchungsobjekt i und Populationscentroid von Gruppe g
$Q_b(Y_s)$	Quadratsumme zwischen den Gruppen auf der Diskriminanzvariablen s
$Q_t(Y_s)$	Totale Quadratsumme auf der Diskriminanzvariablen s
$Q_w(Y_s)$	Quadratsumme innerhalb der Gruppen auf der Diskriminanzvariablen s
r	Anzahl der Diskriminanzfunktionen
$\mathbf{R}$	(pxp)-Interkorrelationsmatrix der p Merkmalsvariablen
ρ_s	Kanonischer Korrelationskoeffizient für die s-te Diskriminanzfunktion
s	Laufindex für die Diskriminanzfunktionen, mit $s = 1, ..., r$
$s_{(w)j}$	Gepoolte Innergruppen-Standardabweichung der Merkmalsvariablen j
$s_{(t)ys}$	Gesamt-Standardabweichung der Diskriminanzwerte der s-ten, nicht normierten Diskriminanzfunktion
$s_{(w)ys}$	Gepoolte Innergruppen-Standardabweichung der Diskriminanzwerte der s-ten, nicht normierten Diskriminanzfunktion
$\mathbf{S}$	(pxp)-Diagonalmatrix mit den Standardabweichungen der p Merkmalsvariablen auf der Hauptdiagonalen
$\boldsymbol{\Sigma}$	(pxp)-Varianz-Kovarianzmatrix innerhalb der Gruppen in der Grundgesamtheit
$\boldsymbol{\Sigma}_g$	Varianz-Kovarianzmatrix der p Merkmalsvariablen in der g-ten Subpopulation
t	Anzahl der bereits extrahierten Diskriminanzfunktionen
t_{sj}	Strukturkoeffizient von Merkmalsvariable j bezüglich der s-ten Diskriminanzfunktion

$\mathbf{t}_s$	$(p\mathrm{x}1)$-Vektor der Strukturkoeffizienten bezüglich der s-ten Diskriminanzfunktion
$\mathbf{T}$	$(p\mathrm{x}p)$-Matrix der Gesamtstreuungen (Quadratsummen und Kreuzprodukte)
$\mathbf{u}$	Einheitsvektor der Länge n_g bzw. N
v_{sj}	Nicht normierter, unstandardisierter Diskriminanzkoeffizient von Merkmalsvariable j in der s-ten Diskriminanzfunktion
$\mathbf{v}_s$	$(p\mathrm{x}1)$-Vektor der Diskriminanzkoeffizienten der s-ten Diskriminanzfunktion
V	V-Wert für das multivariate Wilks Lambda
V_s	V-Wert für das univariate Wilks Lambda der s-ten Diskriminanzfunktion
$V_{t+1,\ldots,r}$	V-Wert für das residuelle Wilks Lambda nach Extraktion der ersten t Diskriminanzfunktionen
$\mathbf{W}$	$(p\mathrm{x}p)$-Matrix der gepoolten Innergruppenstreuung (Innergruppen-Quadratsummen und -Kreuzprodukte)
$\mathbf{W}_g$	$(p\mathrm{x}p)$-Matrix der Innergruppenstreuung (Innergruppen-Quadratsummen und -Kreuzprodukte) in Gruppe g
x_{ij}	Merkmalswert von Untersuchungsobjekt i auf Merkmalsvariable j
X_j	Merkmalsvariable j
$\bar{x}_{gj}$	Gruppenmittelwert von Gruppe g auf Merkmalsvariable j
$\bar{\bar{x}}_j$	Gesamtmittelwert der Merkmalsvariablen j
$\mathbf{x}_{gj}$	$(n_g\mathrm{x}1)$-Merkmalsvektor der n_g Untersuchungsobjekte auf Merkmalsvariable j in Gruppe g
$\mathbf{x}'_i$	$(1\mathrm{x}p)$-Merkmalsvektor von Untersuchungsobjekt i
$\mathbf{x}_j$	$(N\mathrm{x}1)$-Merkmalsvektor der N Untersuchungsobjekte auf Merkmalsvariable j
$\bar{\mathbf{x}}_g$	$(p\mathrm{x}1)$-Mittelwertevektor der p Merkmalsvariablen in Gruppe g
$\bar{\bar{\mathbf{x}}}$	$(p\mathrm{x}1)$-Mittelwertevektor der p Merkmalsvariablen in der Gesamtstichprobe
$\mathbf{X}$	$(N\mathrm{x}p)$-Datenmatrix
$\mathbf{X}_g$	$(n_g\mathrm{x}p)$-Teilmatrizen von $\mathbf{X}$ für Gruppe g
$\bar{\mathbf{X}}$	$(N\mathrm{x}p)$-Matrix der Gruppenmittelwerte der p Merkmalsvariablen
$\bar{\bar{\mathbf{X}}}$	$(N\mathrm{x}p)$-Matrix der Gesamtmittelwerte der p Merkmalsvariablen
y_{is}	Diskriminanzwert von Untersuchungsobjekt i auf Diskriminanzvariable s
Y_s	Diskriminanzvariable s
$\bar{y}_{gs}$	Gruppenmittelwert der Diskriminanzwerte von Gruppe g auf Diskriminanzvariable s.
$\bar{\bar{y}}_s$	Gesamtmittelwert der Diskriminanzwerte auf Diskriminanzvariable s
$\mathbf{y}_s$	$(N\mathrm{x}1)$-Vektor der Diskriminanzwerte der s-ten Diskriminanzfunktion
Z	Zum ursprünglichen Modell hinzugefügte Merkmalsvariable

Methoden für die Analyse von Fragebogendaten

Mit Anwendungen aus den Grenzgebieten der Psychologie

herausgegeben von

Karl Schweizer

Hogrefe · Verlag für Psychologie
Göttingen · Bern · Toronto · Seattle