

Lehrstuhl fuer Statistik, Oekonometrie und empirische
Wirtschaftsforschung, Universitaet Tuebingen

Diskriminanzanalyse

Dr. S. Prohl

18. Mai 2007/ SoSe 2007

Diskriminanzanalyse

- ▶ Problemstellung der Diskriminanzanalyse
- ▶ Vorgehensweise bei der Diskriminanzanalyse:
 - ▶ Definition der Gruppen
 - ▶ Formulierung der Diskriminanzfunktion
 - ▶ Schätzung der Diskriminanzfunktion
 - ▶ Prüfung der Diskriminanzfunktion
 - ▶ Prüfung der Merkmalsvariablen
 - ▶ Klassifikation neuer Elemente
 - ▶ Klassifizierungsfunktionen
 - ▶ Distanzkonzept
 - ▶ Wahrscheinlichkeitskonzept
 - ▶ Berechnung der Klassifizierungswahrscheinlichkeiten
 - ▶ Überprüfung der Klassifizierung

Diskriminanzanalyse

Literatur:

- 1 Backhaus, K., Erichson, B., Plinke, W., und Weiber, R.:
Multivariate Analysemethoden, 2006, Kapitel 3.

Ergaenzende Literatur:

- 2 Handl, A.: Multivariate Verfahren: Theorie und Praxis
multivariater Verfahren unter besonderer
Beruecksichtigung von S-Plus, 2002, Kapitel 12.3

Diskriminanzanalyse

Ziele der Diskriminanzanalyse

Diskriminanzanalyse ist ein multivariates Verfahren:

- ▶ zur Analyse von Gruppenunterschieden hinsichtlich einer Mehrzahl von Variablen.
- ▶ zur Bestimmung oder Prognose der Gruppenzugehörigkeit von Elementen.

Beziehung der Diskriminanzanalyse zu anderen gruppierenden Verfahren (z.B. Clusteranalyse):

- ▶ Clusteranalyse: Ausgangspunkt sind ungruppierte Daten. Ziel: Bildung von Gruppen.
- ▶ Diskriminanzanalyse: Ausgangspunkt: vorgegebene Gruppen. Ziel: Einordnung von Elementen in vorgegebenen Gruppen.

Beispiel: Kreditwürdigkeitsprüfung.

Definition der Gruppen

Ausgangspunkt:

- ▶ zwei Gruppen sollen moeglichst gut voneinander getrennt werden.
- ▶ Kriterien fuer eine Trennung / Diskriminanz:
 - ▶ z.B.: zwei normalverteilte Gruppen mit unterschiedlichen Mittelwerten und gleicher Varianz (dieselbe Verteilung).
 - ▶ Varianz innerhalb der Gruppen soll moeglichst gering sein.

Beispiel.

Formulierung einer Diskriminanzfunktion

Ziel:

- ▶ optimale Trennung zwischen den Gruppen, und
- ▶ Prüfung der diskriminatorischen Bedeutung der Merkmalsvariablen.

Ausgangsmodell:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_j X_j.$$

wobei:

- ▶ Y ist Diskriminanzvariable,
- ▶ X_j ist Merkmalvariable j ($j=1,2,\dots,J$),
- ▶ b_j ist Diskriminanzkoeffizient fuer Merkmalsvariable j ,
- ▶ b_0 ist konstantes Glied.

Beispiel.

Formulierung einer Diskriminanzfunktion

- ▶ Jede Gruppe lässt sich durch den Centroid (mittleren Diskriminanzwert) unterscheiden:

$$\bar{Y}_g = \frac{1}{I_g} \sum_{i=1}^{I_g} Y_{gi}$$

- ▶ Unterschiedlichkeit zweier Gruppen durch Differenz:

$$\frac{|\bar{Y}_A - \bar{Y}_B|}{s}$$

- ▶ s ist die Standardabweichung der Diskriminanzwerte als Maß für die Streuung der Gruppe.

Schaetzung der Diskriminanzfunktion

- Die unbekanntes Koeffizienten $\beta_j, j = 1, \dots, J$ werden durch Maximierung des Diskriminanzkriteriums bestimmt:

$$\Gamma = \frac{\sum_{g=1}^G l_g (\bar{Y}_g - \bar{Y})^2}{\sum_{g=1}^G \sum_{i=1}^{l_g} (Y_{gi} - \bar{Y}_g)^2}.$$

- Gesamtstreuung = erklarte Streuung + nicht-erklarte Streuung.

$$SS_{total} = SS_{between} + SS_{within}$$

$$\sum_{g=1}^G \sum_{i=1}^{l_g} (\bar{Y}_{gi} - \bar{Y})^2 =$$

$$\sum_{g=1}^G (\bar{Y}_g - \bar{Y})^2 + \sum_{g=1}^G \sum_{i=1}^{l_g} (Y_{gi} - \bar{Y}_g)^2.$$

Pruefung des Diskriminanzkriteriums

Maximalwert des Diskriminanzkriteriums:

- ▶ γ = erklarte Streuung / nicht-erklarte Streuung.
 - ▶ wobei γ liegt in $[0,1]$.

Kanonischer Korrelationskoeffizient:

$$c = \sqrt{\frac{\gamma}{1+\gamma}}$$

Wilk's Lambda:

$$\Lambda = \frac{1}{1+\gamma}$$

- ▶ wobei: $c^2 + \Lambda = 1$.

Signifikanzpruefung der Diskriminanzfunktion:

$$\chi_{emp}^2 = -\left(N - \frac{J+G}{2} - 1\right) \ln \Lambda.$$

- ▶ H_0 : es gibt keinen Unterschied zwischen beiden Gruppen
 - ▶ H_1 : beide Gruppen unterscheiden sich.
- ▶ **Wichtig:** falls $\chi_{emp}^2 > \chi_{theor}^2 \Rightarrow H_0$ wird verworfen.

Diskriminanzanalyse: Kontrollfragen

- ▶ Formulieren Sie die Diskriminanzfunktion.
- ▶ Geben Sie die Kriterien der Unterschiedlichkeit zweier Gruppen an.
- ▶ Definieren Sie das Diskriminanzkriterium. Wie ist die Streuung zwischen den Gruppen gegeben? Geben Sie die Definition fuer die Streuung in den Gruppen an. Diskutieren Sie die Unterschiede zur Varianzanalyse.
- ▶ Laesst sich die Diskriminanzanalyse mit der Regressionsanalyse vergleichen?
- ▶ Geben Sie ein Mass fuer die Guete der Diskriminanzfunktion.