

Dipl.-Kfm. Sascha Steinmann
Universität Siegen
Lehrstuhl für Marketing
steinmann@marketing.uni-siegen.de**Sommersemester 2010****Marktforschung**

Die erste der beiden Übungsaufgaben zur Diskriminanzanalyse soll Ihnen im wesentlichen eine Hilfestellung sein, um die komplexen mathematischen Schritte für die Ermittlung der Diskriminanzfunktionen aus der letzten Veranstaltung nachvollziehen zu können. Die zweite Übungsaufgabe sollten Sie sich schon genau angucken. Hier sollten Sie nicht nur die einzelnen Schritte nachvollziehen können, sondern auch eine derartige Aufgabe eigenständig lösen können.

Übungsaufgabe zur Diskriminanzanalyse:**Aufgabe 1:**

Für sechs Pralinsorten P_1, \dots, P_6 wurden Daten zu den Merkmalen Qualität (x_1) und Geschmack (x_2) mittels siebenstufiger Ratingskala erhoben. Die Daten sind in der nachstehenden Tabelle zusammengefasst.

	P_1	P_2	P_3	P_4	P_5	P_6
x_1	-1	-1	2	3	2	1
x_2	1	2	-2	-1	-2	-1

Bei einem entsprechenden Produkttest wurden die Alternativen P_1 und P_2 von den bewertenden Probanden akzeptiert. Die Alternativen P_3, \dots, P_6 hingegen wurden abgelehnt. Mit Hilfe der Diskriminanzanalyse soll nun der Einfluss der Merkmale x_1 und x_2 auf die Akzeptanz (bzw. das resultierende Kaufverhalten) untersucht werden.

Aufgabenstellung:

- Erstellen Sie einen Scatterplot für das vorliegende Datenmaterial und interpretieren diesen kurz.
- Bestimmen Sie die Diskriminanzfunktion. Achten Sie hierbei auf einen strukturierten Lösungsweg.
- Normieren Sie die unter a) gefundene Diskriminanzfunktion.

Hinweise:

$$\Gamma = \frac{SS_B(y)}{SS_W(y)} = \frac{\sum_{g=1}^G I_g (\bar{y}_g - \bar{y})^2}{\sum_{g=1}^G \sum_{i=1}^{I_g} (y_{gi} - \bar{y}_g)^2}$$

$$B_{jr} = \sum I_g (\bar{x}_{jg} - \bar{x}_j)(\bar{x}_{rg} - \bar{x}_r)$$

$$W_{jr} = \sum \sum (x_{jgi} - \bar{x}_{jg})(x_{rgi} - \bar{x}_{rg})$$

$$A = W^{-1}B$$

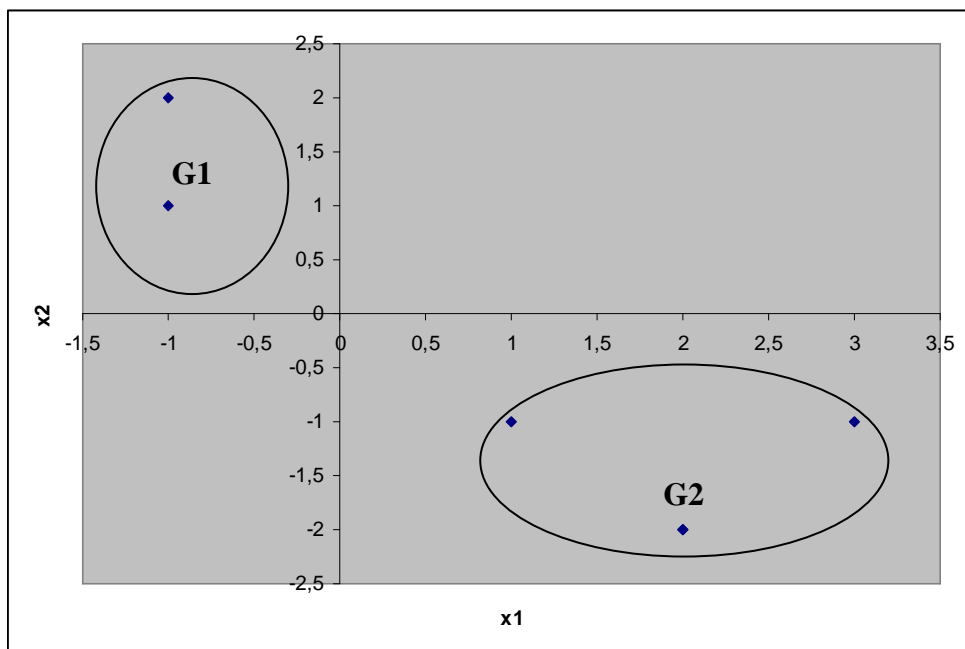
$$b = v \left(\frac{1}{I - G} v' W v \right)^{-1/2}$$

$$b_0 = -\sum_{j=1}^2 b_j \bar{x}_j$$

$$(A - \gamma^* E) v = 0$$

Lösungsskizze:

a) Scatterplot:



Interpretation selbst?

b) $G=2; v=2$

$$I_1=2; I_2=4; I=2+4=6$$

Gesucht: allgemeine Form der Diskriminanzfunktion:

$$y = vx_1 + vx_2 = x'v$$

Ermitteln durch Maximierung des Diskriminanzkriteriums:

$$\Gamma = \frac{SS_B(y)}{SS_W(y)} = \frac{\sum_{g=1}^G I_g (\bar{y}_g - \bar{y})^2}{\sum_{g=1}^G \sum_{i=1}^{I_g} (y_{gi} - \bar{y}_g)^2} = \frac{\text{erklärte Streuung}}{\text{nicht-erklärte Streuung}} \rightarrow \max!_{v_j}$$

$$\text{mit: } \bar{y}_g = \frac{1}{I_g} \sum_{i=1}^{I_g} y_{gi} \quad \bar{y} = \frac{1}{I} \sum_{g=1}^G \sum_{i=1}^{I_g} y_{gi}$$

Mittelwerte der x_j :

$$\bar{x}_1 = 1$$

$$\bar{x}_2 = -0,5$$

Mittelwerte der x_j in den Gruppen:

$$\bar{x}_{11} = -1$$

$$\bar{x}_{12} = 2$$

$$\bar{x}_{21} = 1,5$$

$$\bar{x}_{22} = -1,5$$

Matrix für die Between-Streuung:

$$B = (B_{jr})_{2 \times 2} = \begin{pmatrix} 12 & -12 \\ -12 & 12 \end{pmatrix}$$

$$B_{jr} = \sum I_g (\bar{x}_{jg} - \bar{x}_j)(\bar{x}_{rg} - \bar{x}_r)$$

$$B_{12} = B_{21} = 2 * (-1 - 1)(1,5 + 0,5) + 4(2 - 1)(-1,5 + 0,5) = -12$$

$$B_{11} = 2(-1 - 1)^2 + 4(2 - 1)^2 = 12$$

$$B_{22} = 2(1,5 + 0,5)^2 + 4(-1,5 + 0,5)^2 = 12$$

Matrix für die Within-Streuung:

$$W = (W_{jr})_{2 \times 2} = \begin{pmatrix} 2 & 0 \\ 0 & 3/2 \end{pmatrix}$$

$$W_{jr} = \sum \sum (x_{jgi} - \bar{x}_{jg})(x_{rgi} - \bar{x}_{rg})$$

Beispielrechnung:

$$\begin{aligned} W_{12} = W_{21} &= (x_{111} - \bar{x}_{11})(x_{211} - \bar{x}_{21}) + (x_{112} - \bar{x}_{11})(x_{212} - \bar{x}_{21}) \\ &+ (x_{121} - \bar{x}_{12})(x_{221} - \bar{x}_{22}) + (x_{122} - \bar{x}_{12})(x_{222} - \bar{x}_{22}) \\ &+ \dots + (x_{124} - \bar{x}_{12})(x_{224} - \bar{x}_{22}) \\ &= 0 + 0 + 0 + 0,5 + 0 + (-0,5) = 0 \end{aligned}$$

$$W^{-1} = \frac{1}{\det W} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \frac{1}{a \cdot d - b \cdot c} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

$$W^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 2/3 \end{pmatrix}$$

Darauf aufbauend können die weiteren Rechenschritte zur Ermittlung der Diskriminanzfunktion durchgeführt werden.

$$A = W^{-1}B = \begin{pmatrix} 1/2 & 0 \\ 0 & 2/3 \end{pmatrix} \begin{pmatrix} 12 & -12 \\ -12 & 12 \end{pmatrix} = \begin{pmatrix} 6 & -6 \\ -8 & 8 \end{pmatrix}$$

→ charakteristisches Polynom von A:

$$P_A(\gamma) = \det \begin{pmatrix} 6 - \gamma & -6 \\ -8 & 8 - \gamma \end{pmatrix} = (6 - \gamma)(8 - \gamma) - 48$$

$$= 48 - 6\gamma - 8\gamma + \gamma^2 - 48 = \gamma^2 - 14\gamma$$

$$\text{Nullstellen: } \gamma_{1,2} = 0; 14$$

$\gamma^* = 14$ ist ein maximaler Eigenwert.

Berechnung des Eigenvektors v durch $(A - \gamma * E)v = 0$

$$(6 - 14)v_1 - 6v_2 = 0 \wedge -8v_1 + (8 - 14)v_2 = 0$$

$$\begin{pmatrix} -8 & -6 \\ -8 & -6 \end{pmatrix} v = 0$$

$$\Rightarrow v_2 = -\frac{4}{3}v_1$$

$$\Rightarrow y = v_1x_1 - \frac{4}{3}v_1x_2$$

c) Normierung der Diskriminanzfunktion:

$$\begin{aligned} b &= v \left(\frac{1}{I - G} v' W v \right)^{-\frac{1}{2}} \\ &= \begin{pmatrix} v_1 \\ -\frac{4}{3}v_1 \end{pmatrix} \left(\frac{1}{4} (v_1 - \frac{4}{3}v_1) \begin{pmatrix} 2 & 0 \\ 0 & \frac{3}{2} \end{pmatrix} \begin{pmatrix} v_1 \\ -\frac{4}{3}v_1 \end{pmatrix} \right)^{-\frac{1}{2}} \\ &= \begin{pmatrix} v_1 \\ -\frac{4}{3}v_1 \end{pmatrix} \frac{1}{v_1} \sqrt{\frac{6}{7}} \\ &= \begin{pmatrix} -\sqrt{\frac{6}{7}} \\ \frac{4}{3}\sqrt{\frac{6}{7}} \end{pmatrix} \\ \Rightarrow b_0 &= -\sum_{j=1}^2 b_j \bar{x}_j = \frac{5}{3} \sqrt{\frac{6}{7}} \end{aligned}$$

somit erhalten wir nachstehende normierte Diskriminanzfunktion:

$$y^{(n)}(x_1, x_2) = b_0 + b_1x_1 + b_2x_2$$

$$y^{(n)}(x_1, x_2) = 1,543 - 0,926x_1 + 1,234x_2$$

Aufgabe 2:

Die Marketingabteilung eines Handyherstellers ist daran interessiert inwieweit sich die Käufer der zwei am Markt führenden Handytypen A und B im Hinblick auf die getroffene bzw. zu treffende Wahlentscheidung unterscheiden. Im Rahmen einer entsprechenden Marktstudie wurden zu diesem Zweck bei sechs als repräsentativ erachteten Käufern die Merkmale Differenz zum bundesweiten Durchschnittseinkommen (in Tsd. Euro, x_1) und Differenz zur durchschnittlichen Haushaltsgröße (Anzahl Personen, x_2) erhoben.

Kalibrierungsdaten			
Käufer	Gerätetyp	x_1	x_2
1	A	-1	1
2	A	-1	2
3	B	2	-3
4	B	1	-1

Validierungsdaten			
Käufer	Gerätetyp	x_1	x_2
5	A	2	-2
6	B	3	-1

- Erstellen Sie einen Scatterplot für die vorliegenden Kalibrierungsdaten! Bestimmen Sie die Gruppenzentroide für die Kalibrierungsdaten und fügen Sie diese dem Scatterplot hinzu.
- Welche Kalibrierungsdaten würden Sie welcher der beiden Gruppen unter Verwendung der euklidischen Distanz im Merkmalsraum zuordnen? Zu welcher Trefferquote gelangen Sie mittels dieser Zuordnungsprozedur? Zur Lösung der Aufgabenstellung wählen Sie bitte eine geeignete tabellarische Darstellung.
- Die Anwendung der Diskriminanzanalyse hat zu folgenden Klassifikationsfunktionen geführt:

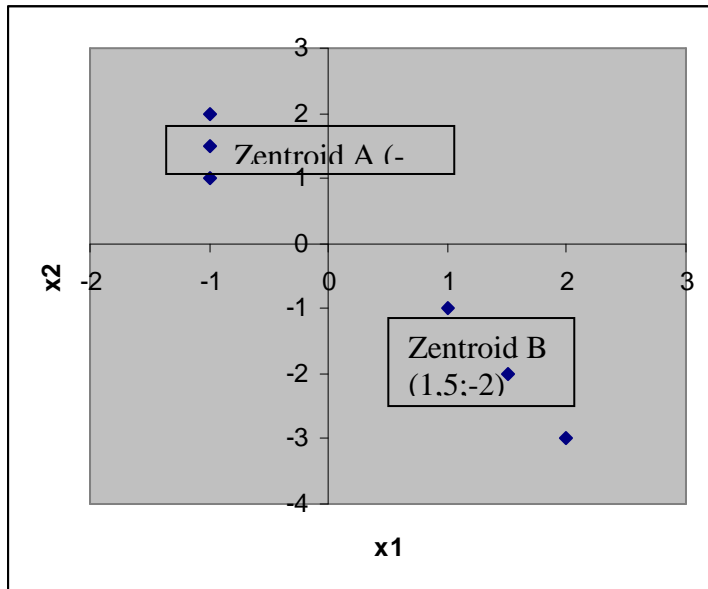
$$L_A = -2,5 - 8x_1 - 2x_2$$

$$L_B = -6,5 + 14x_1 + 4x_2$$

Bestimmen Sie die Funktionswerte der Klassifikationsfunktionen für die Validierungsdaten und führen Sie anhand dieser eine Zuordnung der Beobachtungen zu den Gruppen durch! Zu welcher Trefferquote gelangen Sie nun? Wie schätzen Sie infolgedessen die Güte der gefundenen Klassifikationsfunktion ein? Stellen Sie Ihre Resultate wieder in einer geeigneten tabellarischen Darstellung dar.

Musterlösung (Skizze):

a) Scatterplot:



Interpretation

b) geeignete tabellarische Darstellung:

Käufer	tatsächl. Gerätetyp	geschät. Gerätetyp	Treffer	Fehler
1	A	A	Ja	0
2	A	A	Ja	0
3	B	B	Ja	0
4	B	B	ja	0

Euklidische Distanz (Beispiel für Käufer 4):

$$d_{4(B)} = \sqrt{(1-1,5)^2 + ((-1) - (-2))^2} = \sqrt{1,25}$$

$$d_{4(A)} = \sqrt{(1 - (-1))^2 + ((-1) - 1,5)^2} = \sqrt{10,25}$$

d.h., Käufer 4 gehört in Gruppe B, da die Distanz kleiner ist als zu Gruppe A. (für die anderen Käufer analog)

 Trefferquote: $TQ=100\%$; $TQ_A=100\%$; $TQ_B=100\%$

c) einsetzen in die gegebenen Diskriminanz- bzw. Klassifikationsfunktionen führt zu:

$$L_A^{(5)} = -2,5 - 8 * 2 - 2 * (-2) = -14,5$$

$$L_B^{(5)} = -6,5 + 14 * 2 + 4 * (-2) = 13,5$$

$$L_A^{(6)} = -2,5 - 8 * 3 - 2 * (-1) = -24,5$$

$$L_B^{(6)} = -6,5 + 14 * 3 + 4 * (-1) = 31,5$$

5 → B (tatsächlicher Gerätetyp ist aber A: Fehlzuordnung)

6 → B (tatsächlicher Gerätetyp ist B)

d.h. TQ=50%

tabellarische Darstellung selbst (analog zu oben)!!