

LEHRSTUHL FÜR STATISTIK UND ÖKONOMETRIE
ÜBUNG ZUR DATENANALYSE

Aufgabenserie 6: Diskriminanzanalyse

Aufgabe 14

Ein Kreditinstitut möchte seine Kreditnehmer in Kunden mit guter Bonität (Gruppe 1) und in Kunden mit schlechter Bonität (Gruppe 2) einteilen. Dabei wird ein Kunde als gute Bonität eingestuft, wenn er seinen Kredit zurückzahlt, als schlechte Bonität, wenn er seinen Zahlungsverpflichtungen nicht nachkommt und somit „ausfällt“.

Die Bank betrachtet folgende Lernstichprobe, in der sie die Merkmale Alter in Jahren (X_1) und Bruttoeinkommen in Tausend Euro (X_2) erhoben hat:

Kunden mit guter Bonität			Kunden mit schlechter Bonität		
Person i	Alter $X_{1i,1}$	Einkommen $X_{2i,1}$	Person j	Alter $X_{1j,2}$	Einkommen $X_{2j,2}$
1	26	2.6	1	54	2.2
2	36	2.3	2	21	0.9
3	60	3.1	3	37	2.2
4	41	4.9	4	32	3.3
5	39	3.6	5	26	1.4
6	47	3.9	6	31	2.0

Gegeben sind die Mittelwertsvektoren $\bar{\mathbf{x}}_1$ und $\bar{\mathbf{x}}_2$ sowie die Stichproben-Varianz-Kovarianz-Matrix \mathbf{S}_1 der Gruppe 1 :

$$\bar{\mathbf{x}}_1 = \begin{pmatrix} 41.5 \\ 3.4 \end{pmatrix}, \quad \bar{\mathbf{x}}_2 = \begin{pmatrix} 33.5 \\ 2.0 \end{pmatrix}, \quad \mathbf{S}_1 = \begin{pmatrix} 129.900 & 2.880 \\ 2.880 & 0.896 \end{pmatrix}.$$

1. Berechnen Sie die gepoolte Stichproben-Varianz-Kovarianz-Matrix \mathbf{S}_{pooled} .
2. Berechnen Sie die lineare Stichproben-Diskriminanzfunktion nach Fisher.
3. Es wird eine Kreditanfrage über 20000 Euro von einer 30-jährigen Person mit einem Bruttoeinkommen von 1800 Euro gestellt. Welcher Gruppe würden Sie diese Person zuordnen, wenn Sie
 - (a) mithilfe der linearen Diskriminanzfunktion nach Fisher klassifizieren;
 - (b) wissen, dass 10% aller Kredite ausfallen und \mathbf{x} bivariat normalverteilt ist? Vernachlässigen Sie zunächst die Kosten. Berücksichtigen Sie dann, dass die Bank 10% der Kreditsumme an Zinsen einnimmt und bei einem Kreditausfall durchschnittlich die Hälfte der vergebenen Kreditsumme nicht zurückgezahlt wird (inkl. Zinsen).
4. Bestimmen Sie die Güte der Stichproben-Diskriminanzfunktion und testen Sie diese bei einer Irrtumswahrscheinlichkeit von 5% auf statistische Signifikanz.

Aufgabe 15

Zur Verbesserung der Auswahl zukünftiger Bewerber untersucht die Zulassungsstelle einer Business School eine Zufallsstichprobe 85 ehemaliger Master-Studenten. Diese können in drei Gruppen eingeteilt werden: „Abschluss innerhalb der Regelstudienzeit“ (Gruppe 1), „verspäteter Abschluss“ (Gruppe 2) und „Abbruch“ (Gruppe 3). Basierend auf der Bachelor-Abschlussnote (X_1) und der im GMAT erzielten Punktzahl (X_2) soll nun eine bestmögliche Separation der drei Gruppen sowie eine Klassifikation neuer Bewerber vorgenommen werden.¹

1. Stellen Sie in einem X_1 - X_2 -Koordinatensystem die drei Gruppen in unterschiedlichen Farben dar.
2. Bilden Sie den Stichproben-Mittelwertsvektor $\bar{\mathbf{x}}_i$ für jede Gruppe und die gepoolte Stichproben-Varianz-Kovarianz-Matrix \mathbf{S}_{pooled} . Berechnen Sie auf deren Basis die Stichproben-Diskriminanzgewichte.
3. Testen Sie die Güte der beiden Stichproben-Diskriminanzfunktionen.
4. Ordnen Sie einen neuen Bewerber mit $X_1 = 2.7$ und $X_2 = 530$ mithilfe der Klassifikationsfunktionen einer der Gruppen zu.
5. Verwenden Sie nun eine zufällig gezogene Lernstichprobe vom Umfang 45, und evaluieren Sie die Klassifikationsgüte an der aus den 40 restlichen Studenten bestehenden Teststichprobe. Verwenden Sie hierbei den Befehl `lda` aus der R-Bibliothek `MASS`.

¹Ein ähnliches Beispiel findet sich bei Johnson, R.A.; Wichern, D.W. (2007). *Applied multivariate statistical analysis*. Die Daten stammen von <http://www.statistik.uni-goettingen.de/veranstaltungen/Multivariate/Daten/GPA.frame>.