

Testen, Diskriminanzanalyse

Aufgabe 1:

Der Datensatz `atemwege.rda` (Download von der Veranstaltungshomepage) enthält Daten zu Atemwegserkrankungen von Schulkindern. Enthalten sind folgende Variablen:

Variable	Beschreibung
pef	Maximaler Ausatemstrom
fef50	Ausatemstrom nach 50%iger Ausatmung
fef75	Ausatemstrom nach 75%iger Ausatmung
sex	Geschlecht

Lesen Sie den Datensatz ein und logarithmieren Sie die Variablen `pef`, `fef50` und `fef75`. Verwenden Sie in den nachfolgenden Aufgaben jeweils die transformierten Variablen.

- a) Testen Sie zum Signifikanzniveau $\alpha = 0.01$ die Variablen `fef50` und `fef75` multivariat auf den Mittelwertsvektor

$$\mu_0 = \begin{pmatrix} \log(3) \\ \log(1.4) \end{pmatrix}.$$

Die Kovarianzmatrix Σ ist hierbei unbekannt.

Zeichnen Sie das dem Test entsprechende 99%-Konfidenzellipsoid und vergleichen Sie dieses mit den simultanen Bonferroni-Konfidenzintervallen.

Hinweis: Die univariaten 99.5%-Konfidenzintervalle der beiden Variablen ergeben sich durch $[1.061, 1.108]$ für `fef50` und $[0.315, 0.369]$ für `fef75`.

- b) Testen Sie zum Signifikanzniveau $\alpha = 0.01$, ob die Messungen `pef`, `fef50`, `fef75` zum selben durchschnittlichen Ergebnis führen.
- c) Testen Sie zum Signifikanzniveau $\alpha = 0.01$ die Mittelwerte der Variablen `fef50` und `fef75` multivariat auf Gruppenunterschiede bzgl. des Geschlechts. Vergleichen Sie ihr Ergebnis mit den univariaten Pendants.

Aufgabe 2:

Die bedingte Verteilung $f(x|y)$ eines dichotomen Merkmals in zwei Klassen und die a-priori-Wahrscheinlichkeiten für die jeweilige Klassenzugehörigkeit seien durch die folgende Tabelle bestimmt:

	A	B	a-priori-Wahrscheinlichkeiten
Klasse 1	p	$1 - p$	0.1
Klasse 2	0.05	0.95	0.9

- a) Bestimmen Sie die individuellen Fehlerraten ϵ_{12} und ϵ_{21} der Bayes-Zuordnung in Abhängigkeit vom Parameter p .
- b) Ermitteln Sie in Abhängigkeit von p die Gesamtfehlerrate ϵ . Welchen Wert nimmt ϵ höchstens an?

Aufgabe 3:

Die bedingte Verteilung $f(x|y)$ eines dichotomen Merkmals $X \in \{G, S\}$ in zwei Klassen und die a priori-Wahrscheinlichkeiten für die jeweiligen Klassenzugehörigkeiten $Y \in \{1, 2\}$ seien durch die folgende Tabelle bestimmt. Die statistischen Objekte sind Patienten einer Kardiologenpraxis. Ein Patient gehört zur Klasse 1, falls er kein erhöhtes Herzinfarkttrisiko hat und zur Klasse 2, falls er ein erhöhtes Risiko hat. Das dichotome Merkmal sagt aus, ob das Elektrokardiogramm gut oder schlecht ist.

	Elektrokardiogramm gut G	Elektrokardiogramm schlecht S	a priori- Wahrscheinlichkeiten
Klasse1	0.95	0.05	p
Klasse2	0.10	0.90	$1 - p$

- Bestimmen Sie die Bayes-Zuordnung in Abhängigkeit vom Parameter p . Ist keine eindeutige Zuordnung möglich, so erfolgt eine Zuordnung in Klasse 1.
- Ermitteln Sie für $p = 0.2$ die Fehlerraten ϵ_{12} und ϵ_{21} sowie ϵ .