Testen, Diskriminanzanalyse

Aufgabe 1:
Lesen Sie den Datensatz *Atemwegserkrankungen von Schulkindern* aus dem Datenarchiv (http://www.stat.uni-muenchen.de/service/datenarchiv) in R ein (fehlende Werte sind mit -1 kodiert, dies muss R mitgeteilt werden, z.B. direkt beim Einlesen über Option `na.strings` bei Funktion `read.table`).

a) Einstichprobenfall: Test
Logarithmieren Sie die Variablen `fef50` und `fef75` und testen Sie zum Signifikanzniveau $\alpha = 0.01$ die transformierten Variablen multivariat auf den Mittelwertsvektor

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

Die Kovarianzmatrix $\mathbf{\Sigma}$ is hierbei

i) unbekannt

ii) $\mathbf{\Sigma} = \begin{pmatrix} 0.09 & 0.07 \\ 0.07 & 0.12 \end{pmatrix}$.

Vergleichen Sie die Ergebnisse auch mit den beiden univariaten t-Tests.

b) Einstichprobenfall: Konfidenzbereich
Zeichnen Sie die 99%-Konfidenzellipse für den Mittelwertsvektor aus a) und die entsprechenden simultanen Bonferroni-Konfidenzintervalle.

c) Einstichprobenfall: Symmetrie Test
Testen sie ob die Messungen `pef, fef50, fef75` zum selben durchschnittlichen Ergebnis führen.

d) Zweistichprobenfall: Test
Testen Sie die Mittelwerte der Variablen `fef50` und `fef75` multivariat auf Gruppenunterschiede bzgl. des Geschlechts (`sex`). Vergleichen Sie die Ergebnisse mit den univariaten Pedants.

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.

b) Ermitteln Sie in Abhängigkeit vom Parameter 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 Herzinfarktrisiko hat und zur Klasse 2, falls er ein erhöhtes Risiko hat. Das dichotome Mermal 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} ϵ_{21} und ϵ.