MATH4424 Homework3

1. We want to find out whether students from HKU, CU, and HKUST are up to the same standard in the subject of mathematics and literature. For this purpose, 100 students are randomly selected from each university and are assigned same exams of mathematics and literature. The sample mean and sample variances are

$$\text{HKU}: \quad \bar{X}_1 = \begin{pmatrix} 99 \\ 94 \end{pmatrix}, \quad S_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\text{CU}: \quad \bar{X}_2 = \begin{pmatrix} 90 \\ 96 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$$

$$\text{HKUST}: \quad \bar{X}_3 = \begin{pmatrix} 96 \\ 95 \end{pmatrix}, \quad S_3 = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

Construct the MANOVA table and test the hypothesis whether the three university students are at about same level in terms of mathematics and literature. (Use the χ^2 approximation for the distribution of the test statistic, and use significance level 5%).

5.20. A wildlife ecologist measured x_1 = tail length (in millimeters) and x_2 = wing length (in millimeters) for a sample of $n = 45$ female hook-billed kites. These data are displayed in Table 5.12. Using the data in the table,

Table 5.12 Bird Data

x_1 (Tail length)	x_2 (Wing length)	x_1 (Tail length)	x_2 (Wing length)	x_1 (Tail length)	x_2 (Wing length)
191	284	186	266	173	271
197	285	197	285	194	280
208	288	201	295	198	300
180	273	190	282	180	272
180	275	209	305	190	292
188	280	187	285	191	286
210	283	207	297	196	285
196	288	178	268	207	286
191	271	202	271	209	303
179	257	205	285	179	261
208	289	190	280	186	262
202	285	189	277	174	245
200	272	211	310	181	250
192	282	216	305	189	262
199	280	189	274	188	258

Source: Data courtesy of S. Temple.

(a) Find and sketch the 95% confidence ellipse for the population means μ_1 and μ_2. Suppose it is known that $\mu_1 = 190$ mm and $\mu_2 = 275$ mm for *male* hook-billed kites. Are these plausible values for the mean tail length and mean wing length for the female birds? Explain.

(b) Construct the simultaneous 95% T^2-intervals for μ_1 and μ_2 and the 95% Bonferroni intervals for μ_1 and μ_2. Compare the two sets of intervals. What advantage, if any, do the T^2-intervals have over the Bonferroni intervals?

(c) Is the bivariate normal distribution a viable population model? Explain with reference to Q–Q plots and a scatter diagram.

6.22. Researchers interested in assessing pulmonary function in nonpathological populations asked subjects to run on a treadmill until exhaustion. Samples of air were collected at definite intervals and the gas contents analyzed. The results on 4 measures of oxygen consumption for 25 males and 25 females are given in Table 6.12 on page 348. The variables were

$$X_1 = \text{resting volume } O_2 \text{ (L/min)}$$
$$X_2 = \text{resting volume } O_2 \text{ (mL/kg/min)}$$
$$X_3 = \text{maximum volume } O_2 \text{ (L/min)}$$
$$X_4 = \text{maximum volume } O_2 \text{ (mL/kg/min)}$$

(a) Look for gender differences by testing for equality of group means. Use $\alpha = .05$. If you reject H_0: $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$, find the linear combination most responsible.

(b) Construct the 95% simultaneous confidence intervals for each $\mu_{1i} - \mu_{2i}, i = 1, 2, 3, 4$. Compare with the corresponding Bonferroni intervals.

(c) The data in Table 6.12 were collected from graduate-student volunteers, and thus they do not represent a random sample. Comment on the possible implications of this information.

7.25. Amitriptyline is prescribed by some physicians as an antidepressant. However, there are also conjectured side effects that seem to be related to the use of the drug: irregular heartbeat, abnormal blood pressures, and irregular waves on the electrocardiogram, among other things. Data gathered on 17 patients who were admitted to the hospital after an amitriptyline overdose are given in Table 7.6. The two response variables are

$$Y_1 = \text{Total TCAD plasma level (TOT)}$$
$$Y_2 = \text{Amount of amitriptyline present in TCAD plasma level (AMI)}$$

The five predictor variables are

$$Z_1 = \text{Gender: 1 if female, 0 if male (GEN)}$$
$$Z_2 = \text{Amount of antidepressants taken at time of overdose (AMT)}$$
$$Z_3 = \text{PR wave measurement (PR)}$$
$$Z_4 = \text{Diastolic blood pressure (DIAP)}$$
$$Z_5 = \text{QRS wave measurement (QRS)}$$

(a) Perform a regression analysis using only the first response Y_1.

(i) Suggest and fit appropriate linear regression models.

(ii) Analyze the residuals.

(iii) Construct a 95% prediction interval for Total TCAD for $z_1 = 1$, $z_2 = 12$ $z_3 = 140$, $z_4 = 70$, and $z_5 = 85$.

(b) Repeat Part a using the second response Y_2.

(c) Perform a multivariate multiple regression analysis using both responses Y_1 and Y_2.

(i) Suggest and fit appropriate linear regression models.

(ii) Analyze the residuals.

(iii) Construct a 95% prediction ellipse for both Total TCAD and Amount of amitriptyline for $z_1 = 1$, $z_2 = 1200$, $z_3 = 140$, $z_4 = 70$, and $z_5 = 85$. Compare this ellipse with the prediction intervals in Parts a and b. Comment.

In 7.25 (a) (iii) z_2=1200

Table 7.6 Amitriptyline Data

y_1 TOT	y_2 AMI	z_1 GEN	z_2 AMT	z_3 PR	z_4 DIAP	z_5 QRS
3389	3149	1	7500	220	0	140
1101	653	1	1975	200	0	100
1131	810	0	3600	205	60	111
596	448	1	675	160	60	120
896	844	1	750	185	70	83
1767	1450	1	2500	180	60	80
807	493	1	350	154	80	98
1111	941	0	1500	200	70	93
645	547	1	375	137	60	105
628	392	1	1050	167	60	74
1360	1283	1	3000	180	60	80
652	458	1	450	160	64	60
860	722	1	1750	135	90	79
500	384	0	2000	160	60	80
781	501	0	4500	180	0	100
1070	405	0	1500	170	90	120
1754	1520	1	3000	180	0	129

Source: See [24].