Data provided: Neave's Tables, Graph Paper

MAS6011

The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2012–2013**

Dependent Data **3 hours**

Marks will be awarded for your best ***five*** *answers.*

RESTRICTED OPEN BOOK EXAMINATION
Candidates may bring to the examination lecture notes and associated lecture material (but no textbooks) plus a calculator that conforms to University regulations.
There are 100 marks available on the paper.

Please leave this exam paper on your desk
Do not remove it from the hall

Registration number from U-Card (9 digits)
to be completed by student

Blank

1 p dimensional observations are available on items which may come from one of k distinct Normal populations $N_p(\mu, \Sigma)$, $i = 1, \ldots, k$, where the μ_i and Σ are known and Σ is a positive definite covariance matrix.

(i) Show that the maximum likelihood rule classifies a new observation x into that population whose Mahalanobis distance from x is the smallest. ***(1 mark)***

(ii) Consider the case of three known Normal populations with common variance matrix Σ and means μ_1, μ_2, μ_3, where $\mu_3 = \frac{\mu_1 + \mu_2}{2}$. Let Δ_{ij} be the Mahalanobis distance between populations i and j. Show that $\Delta_{13} = \frac{\Delta_{12}}{2}$. ***(4 marks)***

(iii) If $P_{ij} = P(\text{classify a type } j \text{ as type } i)$, then it is a standard result that, in the case of $k = 2$, the two misclassification probabilities P_{12} and P_{21} are each equal to $\Phi(-\Delta/2)$, where Δ is the Mahalanobis distance between the two populations and Φ denotes the cumulative distribution function of the standard Normal distribution $N(0, 1)$. Deduce that in the case of part (ii) above

$$P_{13} = P_{23} = \Phi(-\Delta_{12}/4).$$

(2 marks)

(iv) Using μ_1, μ_2, μ_3 defined in part (ii), show that the other four probabilities of misclassification are given by

$$P_{21} = P_{12} = \Phi(-3\Delta_{12}/4) \quad \text{and} \quad P_{31} = P_{32} = \Phi(-\Delta_{12}/4) - \Phi(-3\Delta_{12}/4).$$

[HINT: you may find it useful to sketch μ_1, μ_2, μ_3.] ***(5 marks)***

(v) Bivariate measurements are made on the setting times and bonding strengths from two samples of adhesive, both of size 47, those in the first sample having been previously identified as acceptable and those in the second as being of unacceptable standard. The sample means and pooled variance matrix of the two samples were

$$\mu_1 = \begin{pmatrix} 11 \\ 16 \end{pmatrix}, \quad \mu_2 = \begin{pmatrix} 18 \\ 12 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 10 & 2 \\ 2 & 4 \end{pmatrix}.$$

The data will be used to construct a sample discriminant rule for identifying substandard batches on the basis of such bivariate measurements. The first procedure proposed is to have just two categories 'acceptable' and 'unacceptable'. What proportion of unacceptable batches is likely to be erroneously passed as acceptable when using this rule?
(8 marks)

2 Measurements (in Newtons per square centimetre) were made on each of 11 samples of timber of x_1 (stiffness) and x_2 (bending strength) both before and after a new resin treatment. The sample mean vectors and the variance matrices of the sample of the sample means were as follows

$$\text{Before treatment:} \quad \bar{x}_B = \begin{pmatrix} \bar{x}_{1B} \\ \bar{x}_{2B} \end{pmatrix} = \begin{pmatrix} 1420 \\ 9175 \end{pmatrix}, \quad S_B = \begin{pmatrix} 265 & 350 \\ 350 & 2525 \end{pmatrix}$$

$$\text{After treatment:} \quad \bar{x}_A = \begin{pmatrix} \bar{x}_{1A} \\ \bar{x}_{2A} \end{pmatrix} = \begin{pmatrix} 1440 \\ 9295 \end{pmatrix}, \quad S_A = \begin{pmatrix} 240 & 375 \\ 375 & 2610 \end{pmatrix}.$$

The covariance between before and after treatment measurements of x_1 was 105 and that before and after treatment measurements of x_2 was 375. The covariances between non-corresponding measurements before and after treatment were negligible.

(i) Calculate the variance matrix of the change in strength measurements as a result of the new resin treatment. ***(6 marks)***

(ii) Do this data provide evidence that the new resin treatment changes the overall strength of the timber on average? ***(10 marks)***

(iii) What linear combination of the two measurements of strength shows the greatest change as a result of the treatment? ***(4 marks)***

3 A study was made of samples of ellipsoidal pebbles taken from two tributaries of a river, one tributary flowing into the upper part of the river and the other into the lower. On each pebble, the maximum and minimum diameter was measured. The 21 pebbles from the upper tributary gave mean maximum and minimum diameters of 10.3mm and 8.5mm with sample variances 1.69mm^2 and 1.30mm^2 and covariance 0.1mm^2. The measurements of the 31 pebbles from the lower tributary gave mean results of 9.1mm and 7.9mm with sample variances 1.33mm^2 and 1.21mm^2 and sample covariance 0.14mm^2.

(i) Calculate Fisher's linear discriminant function for classifying a pebble with maximum and minimum diameters (x_1, x_2) as deriving from the upper or lower tributary. ***(10 marks)***

(ii) Assuming that these measurements are adequately modelled by bivariate Normal distributions with a common variance matrix and that the classification of pebbles uses Fisher's discriminant function, estimate the probability of misclassifying a randomly selected lower tributary pebble as deriving from the upper tributary. ***(6 marks)***

(iii) Two further pebbles whose labels were lost during the study had diameters $(9.7, 8.2)$ and $(9.9, 8.0)$. What are the best assessments of the sources of these two pebbles? ***(4 marks)***

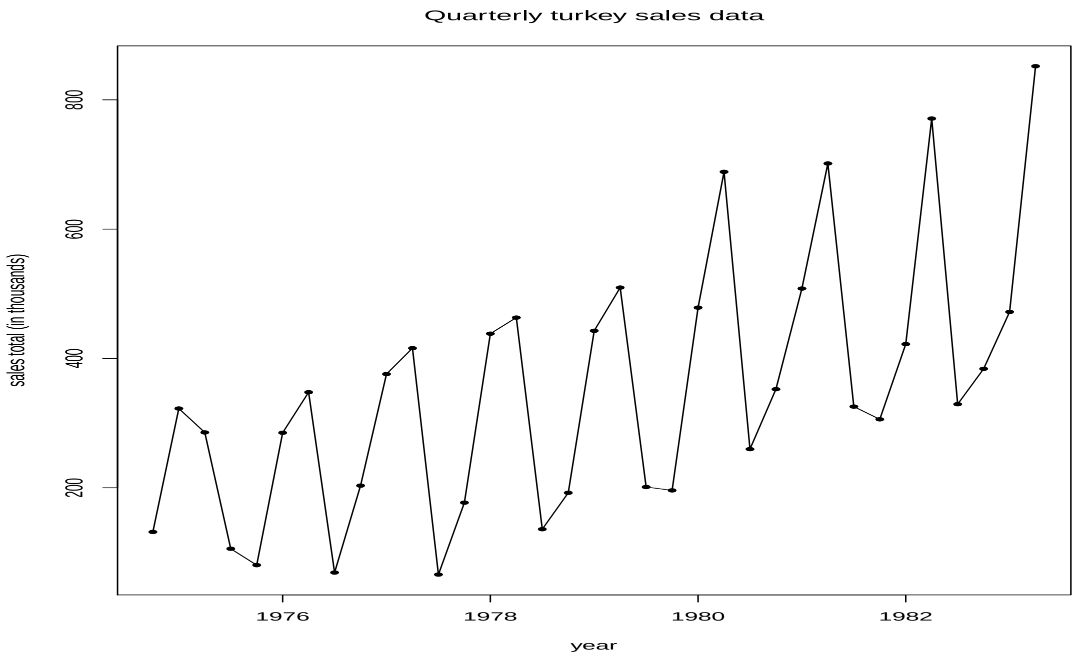

Figure 1: Quarterly sales turkey chicks data

4 The plot above shows quarterly total sales (in thousands) of one-day-old turkey chicks from hatcheries in Eire (source Pole, A., West, M., and Harrison, P.J., 1994, Applied Bayesian Forecasting and Time Series Analysis, Chapman-Hall).

(i) Briefly describe the features of the data. *(2 marks)*

(ii) Suggest a transformation of the time series data y_t, likely to result in a stationary time series x_t and write down x_t as a function of y_t using appropriate differencing notation. *(2 marks)*

(iii) For a time series x_t (of length 31) derived from y_t by a suitable transformation, the sample ACF and the sample PACF are tabulated below:

Lag	1	2	3	4	5	6	7	8
ACF	r_1	r_2	-0.468	0.700	-0.397	0.134	0.051	0.002

and

Lag	1	2	3	4	5	6	7	8
PACF	-0.650	-0.488	-0.832	-0.565	-0.300	0.124	0.090	0.032

(a) Find the values of r_1 and r_2. *(4 marks)*

(b) Test whether x_t is a white noise. *(2 marks)*

(c) Test whether x_t is consistent with autoregressive models. *(3 marks)*

(d) Test whether x_t is consistent with moving average models. *(5 marks)*

(e) Based on your analysis above, suggest a time series model for x_t that is likely to perform well when fitted to the data. *(2 marks)*

5 Consider the time series model

$$y_t = 19 - \frac{1}{3}y_{t-1} - \frac{1}{4}y_{t-2} + \epsilon_t - \frac{1}{2}\epsilon_{t-1},$$

where ϵ_t is white noise with variance 8.

(i) Write down this model using the Backward shift operator B. ***(2 marks)***

(ii) Show that this model is causal and invertible. ***(5 marks)***

(iii) Find the mean of y_t. ***(3 marks)***

(iv) Find the variance of y_t. ***(10 marks)***

6 The plot below relates y_t the quarterly change in a company's sales to x_t the quarterly change of a market sales indicator variable.

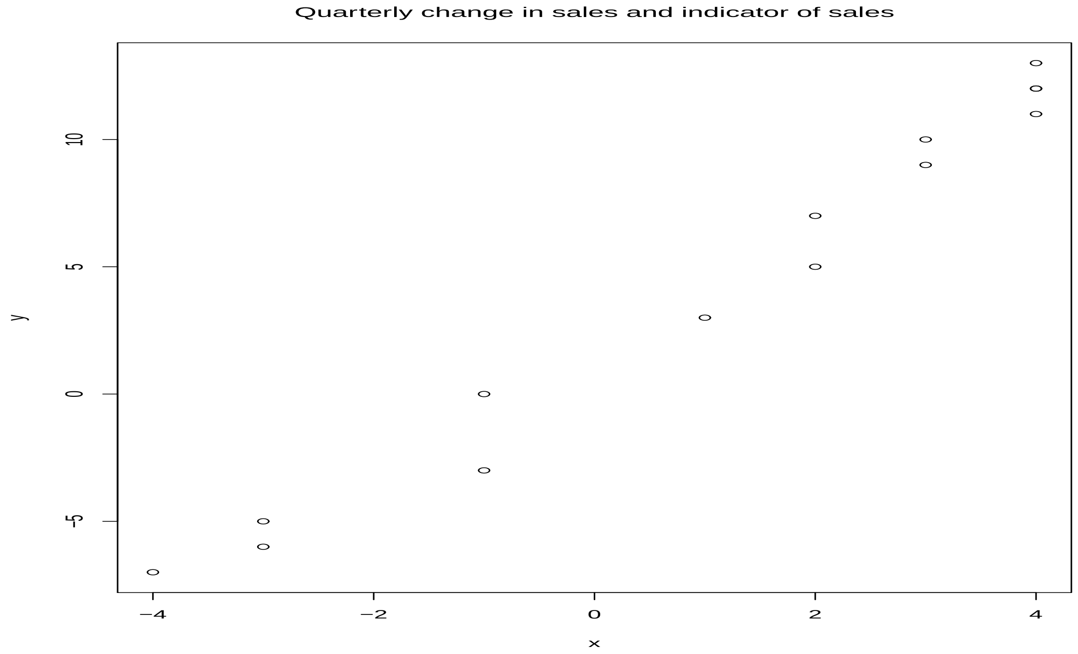

Figure 2: Relation of change of sales y_t and indicator of sales x_t

As a first model it is suggested to regress y_t on x_t, or

$$y_t = x_t\beta + \epsilon_t,$$

where β is a regression coefficient and ϵ_t is a white noise with variance 1.

However, the statistician of the company argues this model is not appropriate to model the data set. To back her argument she has provided the autocorrelation functions of the time series x_t and y_t, given in the plot overleaf (Figure 3).

(i) Explain why the statistician believes the model above is inappropriate.

(1 mark)

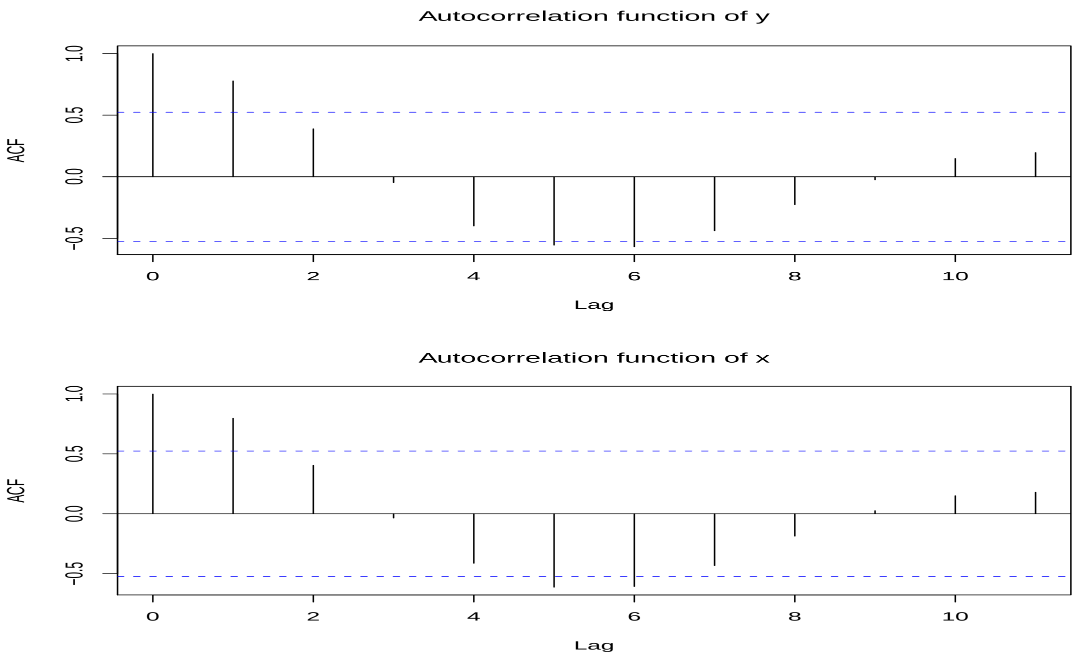

Figure 3: ACF of the time series x_t and y_t

6 (continued)

(ii) The statistician suggested a second model as alternative, given by

$$y_t = x_t\beta_t + \epsilon_t \quad \text{and} \quad \beta_t = \beta_{t-1} + \zeta_t, \tag{1}$$

where ϵ_t is as before and ζ_t is a white noise with variance 10.

(a) Give the name of model (1). ***(1 mark)***

(b) For model (1) show that $P_{t|t}$ the posterior variance of β_t satisfies

$$\frac{1}{P_{t|t}} = \frac{1}{P_{t-1|t-1} + 10} + x_t^2.$$

(7 marks)

(c) If $x_1 = 4$, $x_2 = 4$, $y_1 = 12$, $y_2 = 11$ and the prior of β_0 is $\beta_0 \sim N(2, 0.81)$, then use the result in (b) to calculate the posterior means $\hat{\beta}_{1|1}$, $\hat{\beta}_{2|2}$ and the posterior variances $P_{1|1}$, $P_{2|2}$. ***(5 marks)***

(d) If $x_3 = 3$, use (c) to obtain the one-step forecast mean of $y_3 = 9$ and the associated residual. Comment on the quality of this forecast. ***(2 marks)***

 Question 6 continued on next page

6 (continued)

(e) If instead of $\beta_0 \sim N(2, 0.81)$ the prior distribution of β_0 is set to either of the following:

(α) $\beta_0 \sim N(10, 0.81)$ or

(β) $\beta_0 \sim N(2, 100)$,

comment on whether you expect an improvement on forecasting and the general model performance for all y_t. ***(3 marks)***

(f) Suggest how the model performance can be improved. ***(1 mark)***

End of Question Paper