The University Of Sheffield.

PLEASE LEAVE THIS EXAM PAPER ON YOUR DESK.
DO NOT REMOVE IT FROM THE HALL.

Data Provided:
Neaves Tables
Graph Paper

SCHOOL OF MATHEMATICS AND STATISTICS MAS6061

Session 2013-2014 **3 Hours**

Epidemiology and Time Series

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.
All answers will be marked but credit will be given for only the best ***FIVE*** *answers.*
All questions carry equal marks. Total marks 100.

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

(This page is left blank)

1. The data in the table below describe an unmatched case-control study to determine the effect of helmets on the risk of head injuries in skiers and snowboarders. The 693 cases (skiers or snowboarders with head injuries as reported by the ski patrol) and 3294 controls (skiers or snowboarders with non-head injuries who were reported by the ski patrols at the same ski area as the cases) were cross-classified according to helmet use and gender.

Males		
	Helmet Use	
	Yes	No
Cases with head injury	158	263
Controls	655	802
Females		
	Helmet Use	
	Yes	No
Cases with head injury	111	161
Controls	788	1049

We are interested in determining the effect of helmets on the risk of head injuries in skiers and snowboarders.

(i) Explain why we cannot estimate the relative risk directly in this study.

(2 marks)

(ii) Ignoring gender, calculate the odds ratio and 95% confidence interval for the occurrence of head injuries for helmet users, relative to non-helmet users. Comment on the results.

(6 marks)

(iii) Use an appropriate method to calculate an estimate of the odds ratio for occurrence of a head injury in skier or snowboarders due to helmet use, allowing for gender. Calculate a 95% confidence interval for this odds ratio.

(8 marks)

(iv) Compare the results from (ii) and (iii). Is there evidence of a significant association between helmet use and head injuries? Justify your answer.

(4 marks)

2. A novel test based on the characteristics (a count of specific mutations present) of circulating DNA is evaluated for its ability to identify melanoma patients who will relapse (cancer re-occurs) within 12 months. A cohort of 350 melanoma patients were assessed for this test, and 75 of these relapsed. A random sample of 12 of those who relapsed and 12 who did not relapse are listed below.

Score Test:

Patients who relapsed: 1, 2, 4, 5, 7, 9, 9, 10, 13, 13, 14, 14

Patients who did not relapse: 1, 1, 1, 2, 3, 3, 3, 3, 4, 5, 6, 6

i) Find the odds ratio of relapsing with a score over 5. Can we use this to estimate the relative risk of relapsing given a score over 5?

(2 marks)

ii) Find the sensitivity of the predictor for a specificity of 75%.

(2 marks)

iii) Plot the ROC using the three quartiles of the distribution of the score test as cut-off points.

(6 marks)

iv) Find the area under the ROC curve found in (iii) and interpret it.

(5 marks)

v) What statistics would be useful to an individual if this test were to be used in a clinical setting and can you derive them from this table? If not, why not?

(3 marks)

vi) What other factors have to be in place for the score test to be a useful prognostic tool?

(2 marks)

3. A large population based study is conducted to look at the possible effects of early childhood vaccination on risk of autism. Access to medical care may be associated with higher rates of vaccination and preventative medical care. In the study access to medical care is categorised as good or poor access.

	Good Access to Medical Care		**Poor Access to Medical Care**	
	Autism	No autism	Autism	No autism
Vaccination	224	5376	42	2058
No vaccination	56	1344	18	882
	280	6720	60	2940

i) According to the classic definition of confounding, is there evidence of confounding in the relationship between vaccination and being diagnosed with autism?

(5 marks)

ii) Derive the table that would be obtained if no data on access to medical care were available**.**

(1 mark)

iii) Calculate the Risk Difference, the Relative Risk and the Odds Ratio for the situations where access to medical care is good, access to medical care is poor and for the collapsed table derived in part b).

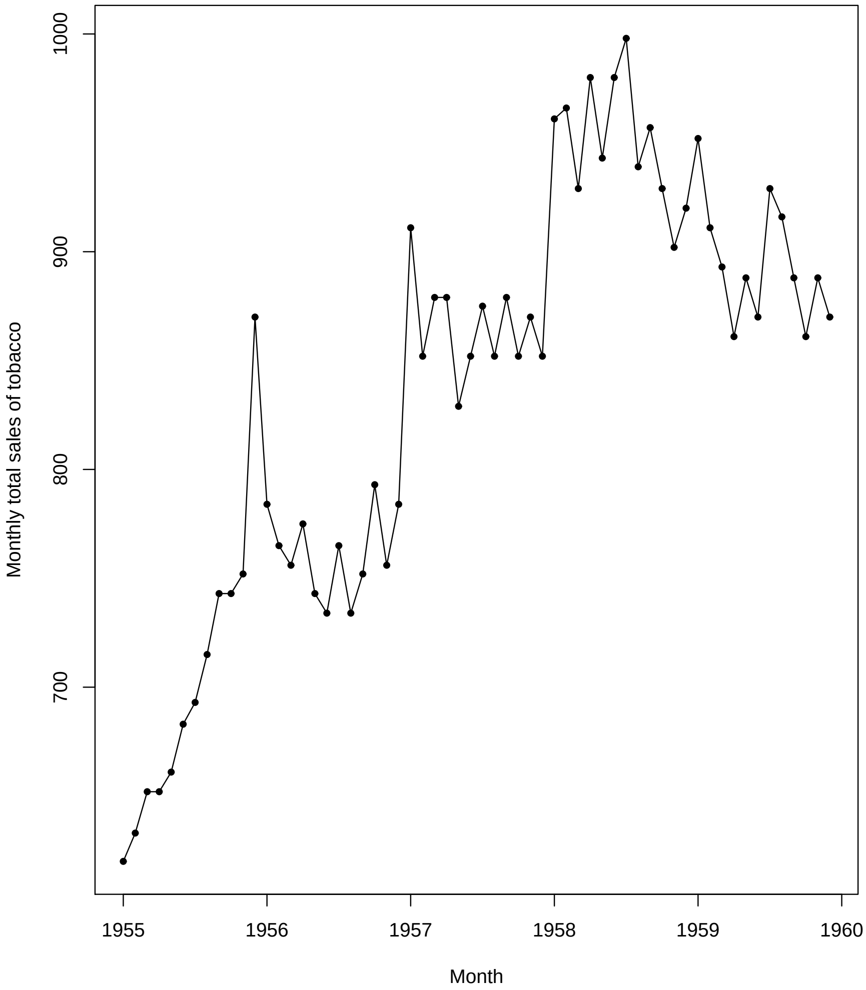
(6 marks)

iv) Comment on whether there is evidence of confounding according to the collapsibility definition.

(3 marks)

v) What is the counterfactual definition of confounding? Calculate the crude and standardised expected counts and compare these values. What is the evidence that access to medical care is a confounder in the relationship between vaccination and autism?

(5 marks)



4 (i) The plot above shows data consisting of monthly total sales (in some standard scale) of UK tobacco and related products in the period 1955 to 1959[1]. Briefly describe the features of the data. ***(2 marks)***

[1]Source: West, M. and Harrison, P.J. (1997) Bayesian Forecasting and Dynamic Models, Springer

 Question 4 continued on next page

4 (continued)

(ii) For a new time series z_t with length 20, the sample ACF and the sample PACF are tabulated below:

Lag	1	2	3	4	5
ACF	0.62	0.57	0.30	0.10	0.05

and

Lag	1	2	3	4	5
PACF	$a_1^{(1)}$	$a_2^{(2)}$	0.28	0.15	0.01

(a) Determine whether z_t is stationary or not and give a brief explanation. ***(1 mark)***

(b) Find the values of $a_1^{(1)}$ and $a_2^{(2)}$. ***(4 marks)***

(c) Test whether z_t is a white noise. ***(3 marks)***

(d) Test whether z_t is consistent with autoregressive models. ***(3 marks)***

(e) Test whether z_t is consistent with moving average models. ***(5 marks)***

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

5 Suppose that a model is set up for a seasonal time series y_t so that the transformed series x_t, defined by

$$x_t = (1 - B^4)^3 y_t,$$

where B is the backward shift operator, follows the AR(1) model

$$x_t = 0.7x_{t-1} + \epsilon_t,$$

where ϵ_t is white noise with variance 1.

(i) Write down the abbreviated form of the model of y_t: SARIMA$(p, d, q) \times (P, D, Q)_s$, i.e. identify p, d, q, P, D, Q and s. ***(1 mark)***

(ii) Show that

$$y_t = x_t + 3y_{t-4} - 3y_{t-8} + y_{t-12},$$

for $t > 12$. ***(5 marks)***

(iii) If the first 13 observations of y_t were

t	1	2	3	4	5	6	7	8	9	10	11	12	13
y_t	10	12	15	16	9	13	15	17	8	12	14	15	9

find the one-step, two-step and three-step forecast means of y_{14}, y_{15} and y_{16} respectively. ***(10 marks)***

(iv) If y_{14}, y_{15} and y_{16} were respectively 13, 16 and 18, then calculate the forecast error in each of the forecasts in part (iii) above and briefly comment on the forecast performance, based on these 3 predictions. ***(4 marks)***

6 Consider the time series model

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + z_t \gamma_t + \eta_t,$$

where α_1, α_2 are some parameters, z_t is a known time-varying covariate, γ_t is a time-varying regression parameter and η_t is a white noise with variance 10. The modeller postulates that $\gamma_t \approx \gamma_{t-1}$ (γ_t has a slow evolution), hence she suggests using the evolution model

$$\gamma_t = \gamma_{t-1} + \nu_t,$$

where ν_t is a white noise with variance 1. Assume that η_t and ν_t are independent for all t, each of them following a normal distribution.

(i) Define the state vector

$$\beta_t = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \gamma_t \end{bmatrix}.$$

Write the model of y_t in state-space form, i.e.

$$y_t = x_t^\top \beta_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2)$$
$$\beta_t = F\beta_{t-1} + \zeta_t, \quad \zeta_t \sim N(0, Z)$$

and determine the values of x_t, F, σ^2 and Z. ***(4 marks)***

(ii) If $z_3 = 2$, $y_1 = 5$, $y_2 = 7$, $y_3 = 4$, the posterior mean vector and the posterior covariance matrix of β_2 are

$$\hat{\beta}_{2|2} = \begin{bmatrix} 1 \\ 1/5 \\ 1 \end{bmatrix} \quad \text{and} \quad P_{2|2} = \begin{bmatrix} 20 & 0 & 0 \\ 0 & 20 & 0 \\ 0 & 0 & 20 \end{bmatrix},$$

then calculate the posterior mean vector and the posterior covariance matrix of β_3 at time $t = 3$. ***(12 marks)***

(iii) If $z_4 = 3$, find a 95% prediction interval for y_4. Based on this interval alone, comment on the forecast accuracy for the future observation y_4. ***(4 marks)***

End of Question Paper