



The
University
Of
Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS

**Autumn Semester
2014–15**

Medical Statistics

2 hours

RESTRICTED OPEN BOOK EXAMINATION

*Candidates may bring to the examination lecture notes and associated lectures material (but no textbooks) plus a calculator that conforms to University regulations. All questions will be marked, but credit will be given for only the best **THREE** answers. All questions carry equal marks. Total marks 60.*

**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 A doctor wishes to compare a new drug B with the current standard drug A. Performance is measured by patient ratings on a continuous 0-20 scale, where high scores are good. Ratings typically have a standard deviation of around 2 and if mean ratings could be improved by 3, this would indicate sufficient promise in B for future development. He plans to compare the drugs with a small (10 subjects) crossover trial using a simple AB/BA design. A summary of the results, and some derived statistics, is given below.

Group 1: A then B ($n_1 = 5$)				
	Period 1	Period 2	Sum (1+2)	Difference (1-2)
mean	10.5	9.6	20.1	0.9
s.d.	2.1	2.3	3.1	2.6
Group 2: B then A ($n_2 = 5$)				
	Period 1	Period 2	Sum (1+2)	Difference (1-2)
mean	12.3	13.5	25.8	-1.2
s.d.	2.0	2.1	2.9	2.4

- (i) Demonstrate why it is not appropriate to assess whether drug A or drug B is better using a standard crossover analysis. **(6 marks)**
 - (ii) Perform an alternative test for a treatment effect. **(6 marks)**
 - (iii) Calculate the power of the test performed in (ii) to detect an improvement of 3 in mean ratings. How does knowledge of this power affect the conclusions you drew in (ii)? **(6 marks)**
 - (iv) Suggest how you might modify the trial design to avoid the problems encountered in (i). **(2 marks)**
- 2
- (i) Compare and contrast the statistical advantages and disadvantages of prospective and retrospective epidemiological studies. **(4 marks)**
 - (ii) As part of a much larger study of occupational health risks, data were collected on employment history from 50 people suffering from the lung disease emphysema and a similar number of controls (matched for age, sex and smoking status).

	Case	Control	
High risk occupation	16	8	24
Not high risk occupation	34	40	74
	50	48	98

where ‘high risk occupation’ refers to those employed for 10 years or more in the construction, chemical or automotive industries.

2 (continued)

- (a) Is there any evidence that these occupations affect the odds of developing emphysema? *(6 marks)*
- (b) Could a study of this type provide evidence that these occupations cause emphysema? Explain your answer. *(2 marks)*
- (c) The larger study involves subjects from four different countries. Here similar data have been collected, but the analysis has been via logistic regression in R. The response is the presence of emphysema, coded as 1 if the subject has the disease and 0 if they do not.

The possible prognostic factors are given in the table below, together with their coding and their R format (i.e. whether they are regarded as a variable (V) or factor (F)).

Variable	R format	coding
Sex	F	0=female; 1=male
Smoking	F	0=never smoked; 1=current or former smoker
Age	V	age in years
Country	F	0=UK; 1=Finland; 2=Spain; 3=South Korea
Occupation	F	0=not high risk; 1=high risk

The model fitting has been summarized in the following (partially completed) Analysis of Deviance table (with terms added sequentially, first to last).

	Df	Deviance	Resid. Df	Resid. Deviance
NULL			719	6855.69
Sex	1	936.96	718	5918.73
Smoking	1	2524.56	717	3394.17
Age	1	542.28	716	2851.89
Country	3	68.14	713	2783.75
Occupation	1	604.23	712	2179.52

What conclusions do you draw? Explain your reasoning. *(6 marks)*

- (d) What other information would you need from R to make useful statements about relative odds of developing emphysema for members of the general population? *(2 marks)*

- 3 23 patients with pancreatic cancer were randomised to one of two forms of treatment and followed up until remission. The data below show the times until remission in weeks. 7 patients were lost to follow up before remission, these censored observations are denoted by asterisks in the table.

Treatment	Remission/Censored Times (weeks)												Total Time in Study
	1	3*	5	9*	10	12	20	26*	43	46*	52	64*	
Standard	1	3*	5	9*	10	12	20	26*	43	46*	52	64*	291
New	1*	1	4	5	7	13*	18	20	28	32	34		163

- (i) Considering time to remission as “survival” time, for each group separately, estimate the survivor function at 20 weeks after follow up using Kaplan-Meier. *(6 marks)*
 - (ii) Now assume that the remission times in each group are exponentially distributed with rates $\lambda_j, j = 1, 2$ respectively. Estimate λ_1 and λ_2 and their 95% confidence intervals. *(4 marks)*
 - (iii) How would you assess the assumption that time to remission follows an exponential distribution? *(2 marks)*
 - (iv) Assuming that the assumption in (ii) is appropriate, perform a LRT test to assess the evidence for a difference in the mean survival times between the two treatments. *(5 marks)*
 - (v) If it was known that the censored patients had in fact died from the cancer how might this affect the suitability of the analysis suggested above? *(3 marks)*
- 4 (i) In a study by Freireich et al.(1963), 42 children with acute leukemia responded to a primary treatment whereby they entered into partial or complete remission where signs of the disease disappeared. The children were then randomized to remission maintenance therapy with the drug 6-mercaptopurine (6-MP) or placebo and time to relapse was studied as the survival time of interest. Patients still in remission at the conclusion of the study who were considered right-censored. The data are stored in `leuk` and coding for the different variables is shown below.

Coding:

treat: treatment (0 = placebo; 1 = 6-MP)

time: remission length in weeks

status: indicator of relapse (1) or censoring (0)

4 (continued)

Some R analysis was performed with the output shown below:

```
> fit <- survreg(Surv(time, status) ~ treat, data=leuk,
+ dist="exponential")
> summary(fit)
```

Call:

```
survreg(formula = Surv(time, status) ~ treat, data = leuk,
+ dist = "exponential")
```

	Value	Std. Error	z	p
(Intercept)	2.16	0.218	9.90	4.33e-23
treat	1.53	0.398	3.83	1.27e-04

Scale fixed at 1

Exponential distribution

Loglik(model)= -108.5 Loglik(intercept only)= -116.8

Chisq= 16.49 on 1 degrees of freedom, p= 4.9e-05

Number of Newton-Raphson Iterations: 4

n= 42

- (a) Describe the analysis performed and the final model for T , the time to relapse, for both the placebo and treatment group. *(4 marks)*
- (b) Assess the evidence that the remission maintenance therapy makes a difference to relapse time. Is it an improvement or not? *(2 marks)*
- (c) Estimate the mean relapse time for a patient in both the placebo and 6-MP group. *(2 marks)*

4(continued)

- (ii) A cohort study was carried out on a group of individuals. Before entering the study, individuals were asked for their smoking habits and whether they performed regular exercise. Their weights were also recorded (kg) and centred around 70 kg (i.e. 80 would be coded as 10 and 55 as -15). The outcome of interest was all-cause mortality. A Cox proportional hazards model was fitted to the data and the output is shown below:

Variable	Coefficient	Standard Error
Sex		
Male	Reference	—
Female	-0.15	0.05
Smoking Status		
Current Smoker	Reference	—
Ex-Smoker	-0.35	0.39
Never Smoker	-0.43	0.17
Weight (centred on 70 kgs as baseline)	-0.0003	0.0007
Regular Exercise		
No	Reference	—
Yes	-0.28	0.14

Table: Log hazard ratio of mortality with standard errors

- (a) Why is a Cox proportional hazards model called a semi-parametric model? What is the general form of the model? **(3 marks)**
- (b) Describe in detail the effects of the variables in the table on survival. **(5 marks)**
- (c) Using the model, calculate the estimate of the hazard ratio comparing
- A male weighing 87kg who denotes themselves as a current smoker and who **does not** take regular exercise
 - A male weighing 68kg who denotes themselves as an ex-smoker who **does** take regular exercise
- (4 marks)**

End of Question Paper