Data provided: Neave's Tables, Graph Paper

MAS361

The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Autumn Semester 2012–13**

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 Obesity is a well-known risk factor in many diseases and a clinician is hoping to obtain a clear indication of its role in hypertension (high blood pressure) for males in the 'over 40 years' age range.

(i) A quick literature search finds 3 studies in which findings are presented for males in the relevant age range. The results are summarized below. The clinician is keen to quote the studies, but concerned that the small study sizes might mean their results are unreliable.

Study I

	Blood Pressure raised	not	
Obese	28	72	100
Not obese	18	82	100
	46	154	200

Study II

	Blood Pressure raised	not	
Obese	20	36	56
Not obese	23	101	124
	43	137	180

Study III

	Blood Pressure raised	not	
Obese	24	54	78
Not obese	30	140	170
	54	194	248

(a) The authors of Study I used a standard χ^2 test with a 5% significance level to assess whether raised blood pressure is associated with obesity. Assume that 20% of non-obese males over 40 years of age suffer from hypertension. If 30% of obese males over 40 suffer from hypertension, show that the power to detect this difference is actually less than 50%. ***(5 marks)***

(b) The clinician decides to combine all the studies using a meta-analysis. Explain under what conditions this is a sensible approach. Assuming these are satisfactory, carry out the analysis for him. ***(6 marks)***

(ii) In fact, he suspects that numerous factors are important in determining blood pressure, so is pleased to discover a larger and more sophisticated study in which an analysis based on logistic regression was conducted. The following data (from Altman, 1995) show the results. Note that Obesity, Smoking and Snoring are indicators with 1 indicating that a subject exhibits the factor.

1 (continued)

factor	regression coefficient b	s.e.(b)	z	p
Obesity	0.695	0.285	2.44	0.015
Constant	-2.378	0.380		
Smoking	-0.068	0.278	0.24	0.81
Snoring	0.872	0.398	2.199	0.028

(a) Explain briefly what is meant by 'logistic regression', ensuring your terminology and/or notation is clear. ***(4 marks)***

(b) Explain, in terms that might be understood by a non-statistician, how the risk of exhibiting hypertension differs for males in this age group who are obese and non-obese. How does the risk differ for smokers and non-smokers? ***(5 marks)***

2 The following tables (adapted from Matthews 1989) show data and derived statistics from a trial investigating the effects of two treatments (A and B) on asthma (values are FEV1, a measure of lung function, in litres; high values are good). The design is an AB/BA crossover.

Subject	Group	Period1	Period2	sum	diff
1	1	1.28	1.33	2.61	-0.05
2	1	1.60	2.21	3.81	-0.61
3	1	2.46	2.43	4.89	0.03
4	1	1.41	1.81	3.22	-0.40
5	1	1.40	0.85	2.25	0.55
6	1	1.12	1.20	2.32	-0.08
7	1	0.90	0.90	1.80	0.00
8	1	2.41	2.79	5.20	-0.38
9	2	2.68	2.10	4.78	0.58
10	2	2.60	2.32	4.92	0.28
11	2	1.48	1.30	2.78	0.18
12	2	2.08	2.34	4.42	-0.26
13	2	2.72	2.48	5.20	0.24
14	2	1.94	1.11	3.05	0.83
15	2	3.35	3.23	6.58	0.12
16	2	1.16	1.25	2.41	-0.09

2 (continued)

	Group 1: A then B ($n_1 = 8$)			
	Period 1	Period 2	Sum (1+2)	Difference (1-2)
mean	1.5725	1.69	3.2625	-0.1175
s.d.	0.5717829	0.7311244	1.263914	0.3542295
	Group 2: B then A ($n_2 = 8$)			
	Period 1	Period 2	Sum (1+2)	Difference (1-2)
mean	2.25125	2.01625	4.2675	0.235
s.d.	0.7215348	0.7381238	1.417954	0.3468223

(i) Plot the treatment means for each period and make a preliminary graphical assessment of the trial's findings. ***(5 marks)***

(ii) Assess whether there is any evidence of a carryover effect from Period 1 to Period 2. ***(4 marks)***

(iii) Assess whether there is any evidence of a difference in mean response between Periods 1 and 2. ***(3 marks)***

(iv) Assess whether there is any evidence of a Treatment effect, taking into account the results of your analyses in (ii) and (iii). ***(3 marks)***

(v) Suppose it was later discovered that Subjects 13 and 14 had not in fact completed their Period 2 treatment (i.e. Treatment A) correctly, because they found it had unpleasant side effects.

(a) How would this affect your confidence in the conclusions reached above? ***(3 marks)***

(b) Would you propose any revised action or analysis? Explain your answer. [NB You need not actually carry out any new analysis proposed.] ***(2 marks)***

3 The following table shows some data from a two-part motion sickness experiment adapted from Altman (1995). In the experiment, volunteers were subjected to simulated sea travel for a period of up to 120 minutes. Time until vomiting ('failure') was recorded. In Experiment 1, 13 subjects were subject to 'gentle' motion. Subjects 8 to 13 survived the entire 120 minutes without vomiting; Subjects 3 and 5 left the trial without vomiting after 50 and 66 minutes respectively. In Experiment 2, 20 subjects experienced a more 'violent' motion. Here Subjects 15 to 20 survived the entire 120 minutes without vomiting; Subject 2 left the trial without vomiting after only 6 minutes.

Experiment I

group	time	censor
1	30	1
1	50	1
1	50	0
1	51	1
1	66	0
1	82	1
1	92	1
1	120	0
1	120	0
1	120	0
1	120	0
1	120	0
1	120	0

Experiment II

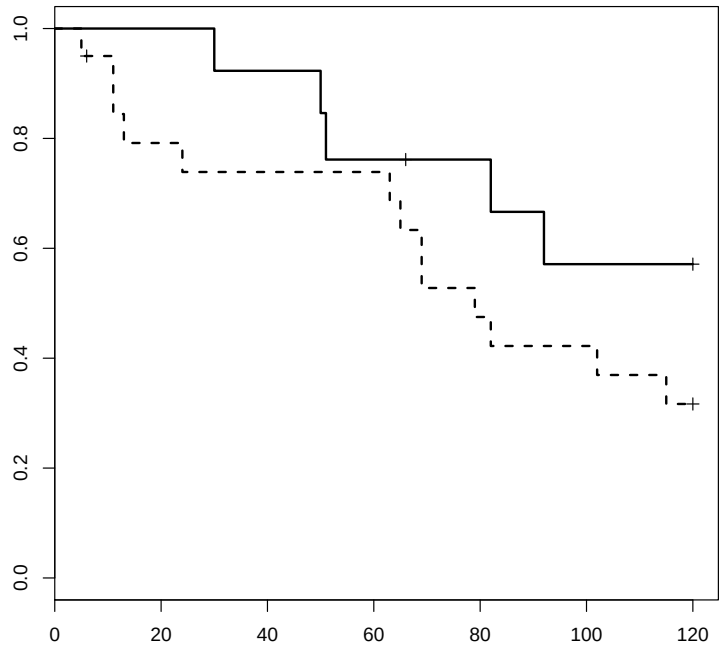
group	time	censor
2	5	1
2	6	0
2	11	1
2	11	1
2	13	1
2	24	1
2	63	1
2	65	1
2	69	1
2	69	1
2	79	1
2	82	1
2	102	1
2	115	1
2	120	0
2	120	0
2	120	0
2	120	0
2	120	0
2	120	0

The figures below present an R analysis of the data, edited in places.

```
> library(survival)
Loading required package: splines
> time<-c(30,50,50,51,66,82,92,120,120,120,120,120,120,5,6,11,11,13,24,63,65,
+69,69,79,82,102,115,120,120,120,120,120,120)
> censor<-c(1,1,0,1,0,1,1,0,0,0,0,0,0,1,0,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0)
> group<-c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2)
> Q3.data<-data.frame(group,time,censor)
> Q3.sv<-Surv(time, censor, type="right")
> plot(survfit(Q3.sv~group), lty=c(1,2),lwd=2)
```

 Question 3 continued on next page

3 (continued)




```
> survdiff(Q3.sv~group)
Call:
survdiff(formula = Q3.sv ~ group)

         N Observed Expected (O-E)^2/E (O-E)^2/V
group=1 13        5     7.85     1.033      1.87
group=2 20       13    10.15     0.799      1.87

 Chisq= 1.9  on 1 degrees of freedom, p= 0.172
>
> sum(time[group==1]); sum(time[group==2])
[1] 1141
[1] 1434
> sum(censor[group==1]); sum(censor[group==2])
[1] 5
[1] 13
>
```

3 (continued)

```
> Q3.regexp<-survreg(Q3.sv ~group, dist="exponential")          ***
> summary(Q3.regexp)
Call:
survreg(formula = Q3.sv ~ group, dist = "exponential")
             Value Std. Error     z        p
(Intercept)  6.157      0.936  6.58 4.86e-11
group       -0.727      0.526 -1.38 1.67e-01

Scale fixed at 1

Exponential distribution
Loglik(model)= -106.3   Loglik(intercept only)= -107.3
        Chisq= 2.09 on 1 degrees of freedom, p= 0.15
Number of Newton-Raphson Iterations: 4
n= 33
```

(i) Estimate the median survival time for Experiment 2 from the plot. ***(2 marks)***

(ii) Assume an Exponential survival model, with constant hazard rate λ, is appropriate for the data from each Experiment.

(a) Does this assumption seem reasonable from your Kaplan-Meier plot? Explain your answer. What would be a better check of this assumption? ***(2 marks)***

(b) Assuming the assumption is reasonable, estimate λ, and hence estimate the mean survival time μ, for Experiment 2. ***(3 marks)***

(c) Compare your estimate of μ and the median survival time estimated in (i). What relationship would you expect to see between these two quantities? ***(3 marks)***

(d) Use a suitable parametric test to assess whether there is a difference in mean survival time under the two sets of experimental conditions. Explain why you have chosen this test. ***(4 marks)***

(iii) In this study, only two sets of experimental conditions have been used. It is likely that the inclination to vomit is dependent on both the frequency and acceleration of the motion, so an extended study might collect data (say 'newdata') from a range of conditions described by their frequency ('freq') and acceleration ('acc') values.

(a) Write down a model which would admit this possibility (again assuming an Exponential model). ***(3 marks)***

(b) Explain how you would modify the R `survreg` call at *** in the output above to consider this possibility. ***(3 marks)***

4 (i) Consider the Cox Proportional Hazards Model.

(a) Write down the model, specifying your notation clearly. ***(2 marks)***

(b) Why is this model described as 'semi-parametric'? ***(1 mark)***

(c) What assumptions does it make and how might these be verified? ***(2 marks)***

(ii) The table below gives the results of fitting a Cox Proportional Hazards model to survival time data (from Altman, 1995) from a trial of Azathioprine as a therapy for Primary Biliary Cirrhosis (a liver disease). The coding of the variables is explained in the lower table.

Variable	regression coefficient b	s.e.(b)	exp(b)
Serum bilirubin	2.0510	0.316	12.31
Age	0.00690	0.00162	1.01
Cirrhosis	0.879	0.216	2.41
Serum albumin	-0.0504	0.0181	0.95
Central cholestasis	0.679	0.275	1.97
Therapy	0.520	0.207	1.68

Variable	scoring
Serum bilirubin	$\log_{10}$ (value in μ mol/l)
Age	exp[(age in years - 20)/10]
Cirrhosis	0=No; 1=Yes
Serum albumin	value in g/l
Central cholestasis	0=No; 1=Yes
Therapy	0=Azathioprine; 1=Placebo

(a) Is there evidence that the therapy is beneficial? Explain your answer. ***(4 marks)***

(b) If the Serum bilirubin value (on original scale) increases by a factor of 10, all other variables remaining constant, how will the hazard change? ***(3 marks)***

(c) What can be said about the probabilities of surviving 3 years for two patients of the same age, same Serum bilirubin value and same Cirrhosis and Central cholestasis status, one of whom has a Serum albumin level of 40 g/l and is given Azathioprine, while the other has Serum albumin of 50.32g/l but is given the Placebo? ***(4 marks)***

4 (continued)

(iii) Suppose a further study of the same Therapy and covariates is to be undertaken. Because of the large number of covariates, a dynamic randomization procedure is to be used to allocate incoming patients to either Azathioprine or Placebo in such a way as maintain as much balance as possible. To simplify handling of continuous covariates, these have all been dichotomized to 'High' or 'Low'. After 40 patients have been enrolled, the marginal counts are as in the table below. If the next patient has high Serum bilirubin, low Age, Cirrhosis, high Serum albumin and no Central cholestasis, should they be allocated to Azathioprine or Placebo?

Variable	Level	Azathioprine	Placebo
Serum bilirubin	High	15	16
	Low	5	4
Age	High	11	12
	Low	9	8
Cirrhosis	0	14	12
	1	6	8
Serum albumin	High	14	13
	Low	6	7
Central cholestasis	0	7	8
	1	13	12

(4 marks)

End of Question Paper