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 MAS461

Autumn Semester 2010-2011 **2 Hours**

Medical Statistics

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 **THREE** answers.*
Questions 3 and 4 both use the data from a study on chemotherapy but otherwise are independent: either or both an be attempted.All questions carry equal marks. Total marks 75.

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

(This page is left blank)

1 In a clinical trial to compare two devices designed to inhibit snoring, 794 men aged between 40 and 65 (all notorious snorers) were allocated either to device A or to device B. Of the 396 who used device A, 233 reported that they had had no episodes of snoring during the week they used the device whereas 203 of those using B reported no snoring.

(a) Do these data provide sufficient evidence that switching men in this age group from device B to device A would reduce the overall incidence of snoring?

(6 marks)

(b) Further investigation revealed that 403 men in the trial were aged under 50. In this age group 28 of the 103 who used device A reported no snoring episodes, whereas 121 of those using device B were relieved of snoring during the week of the study. Do these data suggest that switching men aged under 50 from device B to A would increase or decrease the incidence of snoring in this age group?

(6 marks)

(c) Do the data provided in part (b) suggest that switching men aged between 50 and 65 from device B to A would increase or decrease the incidence of snoring in this age group?

(6 marks)

(d) Using a Mantel-Haenszel test to allow for the differences in snoring rates and in allocation rates to devices between the two age groups, assess the evidence for a difference in effectiveness between devices A and B. Which device should be recommended for use by any notorious male snorer aged between 40 and 65?

(7 marks)

2 Given below is a record (edited in places) of an **R** analysis of the results of a two period crossover trial to investigate the effects of two treatments A (standard) and B (new) for an allergic coughing reaction. The figures represent the numbers of coughing incidents during a three-day period. Because of doubts about normality of the data, nonparametric methods of analysis have been used. The new treatment will be introduced provided that there is good evidence that there is a reduction of at least six from the level of the standard treatment in the median number of coughing incidents in this period. Patients were randomly allocated to two groups: group 1 received treatment A in period 1 and B in period 2. Group 2 received the treatments in the opposite order.

(a) Plot the treatment medians for each group for each period.

(3 marks)

(b) Assess all of the evidence that there is a carryover effect from period 1 to period 2.

(5 marks)

(c) Do the data provide evidence that there is a difference in average response between periods 1 and 2?

(6 marks)

(d) Assess whether the treatments differ in effect, taking into account the results of your assessments of carryover and period effects.

(6 marks)

(e) Describe what further analysis, if any, would be required to assess whether there is evidence to support the introduction of the new treatment.

(5 marks)

Question 2 continued on next page

Question 2 (continued)

```
     ***  Summary Statistics for data in:  coughs ***
group:1
          period1 period2
    Mean:   13.83    7.00
  Median:   15.00    6.00
 Total N:   12.00   12.00
Std Dev.:    5.13    3.59
--------------------------
group:2
          period1 period2
    Mean:    6.73   10.40
  Median:    6.00    8.00
 Total N:   15.00   15.00
Std Dev.:    3.37    7.39
> attach(coughs)
> totalresponse <- period1 + period2
> perioddiffs <- period1 - period2
> treatorder <- 3 - 2 * group
> treatdiffs <- perioddiffs * treatorder
> wilcox.test
+ (totalresponse [group==1], totalresponse [group==2])

     Wilcoxon rank-sum test

data:  totalresponse[group == 1] and totalresp[group ==
2]
rank-sum normal statistic with correction Z = 1.615,
p-value = 0.1062
alternative hypothesis:  mu is not equal to 0

>wilcox.test(perioddiffs[group==1],perioddiffs[group==2])

     Wilcoxon rank-sum test

data:  perioddiffs[group==1] and perioddiffs[group==2]
rank-sum normal statistic with correction Z = 4.091,
p-value = 0
alternative hypothesis:  mu is not equal to 0

> wilcox.test(treatdiffs[group==1],treatdiffs[group==2])

     Wilcoxon rank-sum test

data:  treatdiffs[group == 1] and treatdiffs[group == 2]
rank-sum normal statistic with correction Z = 1.645,
p-value = 0.0999
alternative hypothesis:  mu is not equal to 0

>
```

3 The data given below represent survival times in days of 26 patients randomized to one of two forms of chemotherapy following surgery for ovarian cancer, where status records whether the observation is censored (status = 0) or complete (status =1),
(Source: Collett, 2003):

	Treatment A		Treatment B	
	time	status	time	status
	59	1	353	1
	115	1	365	1
	156	1	377	0
	268	1	421	0
	329	1	464	1
	431	1	475	1
	448	0	563	1
	477	0	744	0
	638	1	769	0
	803	0	770	0
	855	0	1129	0
	1040	0	1206	0
	1106	0	1227	0
Totals	***6725***	***7***	***8863***	***5***

Given below is a record (edited in places) of some initial analyses of these data performed in **R**. Careful examination of this record reveals that the data have been entered into **R** incorrectly: the censoring indicators for time 855 for treatment A and 563 for treatment B have been entered incorrectly.

(a) Compute the Kaplan-Meier product limit estimates of the survivor functions for treatments A and B for the corrected data as given in the table above and provide estimates of the median survival times based upon the Kaplan Meier estimates.

(8 marks)

(b) Assuming that the survival times are exponentially distributed with rates λ_j, j=A, B, estimate λ_A and λ_B and hence the median survival times and provide approximate 95% confidence intervals for the median survival times for each group.

(8 marks)

(c) By using a parametric test, assess the evidence for a difference in the mean survival times between the two stages.

(9 marks)

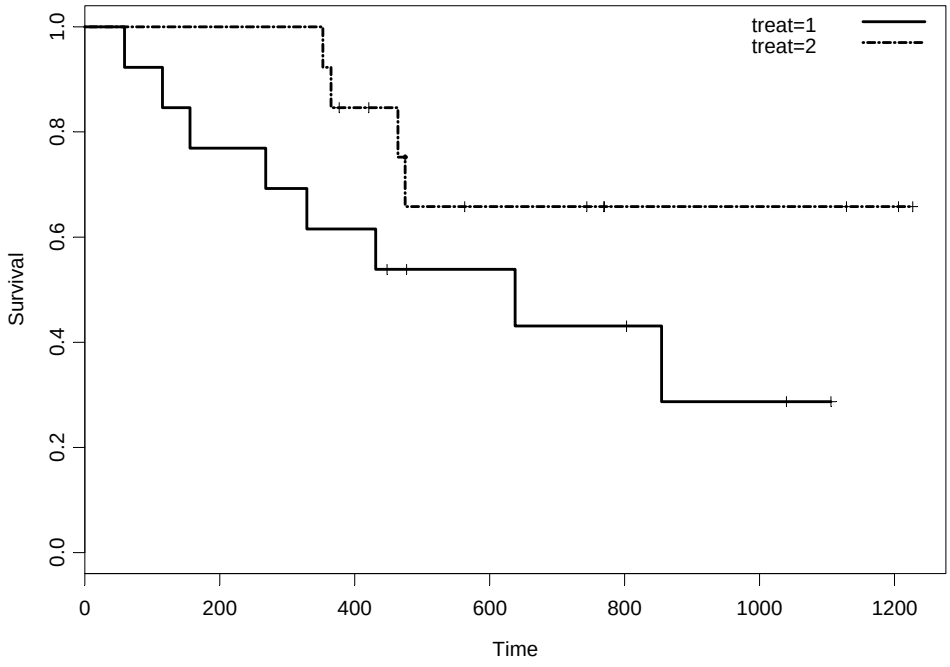
Question 3 continued on next page

Question 3 continued

Analysis of Survival Times of Ovarian Cancer

```
> library(survival)
> attach(ovarian)
> ovarian.sv<-Surv(time, status)
> survfit(ovarian.sv~treat)
Call: survfit(formula = ovarian.sv ~ treat)

        records n.max n.start events median 0.95LCL 0.95UCL
treat=0      13    13      13      7    638     268      NA
treat=1      13    13      13      5     NA     475      NA>
summary(survfit(ovarian.sv~treat))
                treat=1                          treat=2
time n.risk n.event survival    time n.risk n.event survival
  59     13       1    0.923     353     13       1    0.923
 115     12       1    0.846     365     12       1    0.846
 156     11       1    0.769     377     11       0    0.846
 268     10       1    0.692     421     10       0    0.846
 329      9       1    0.615     464      9       1    0.752
 431      8       1    0.538     475      8       1    0.658
 448      7       0    0.538     563      7       0    0.658
 477      6       0    0.538     744      6       0    0.658
 638      5       1    0.431     769      5       0    0.658
 803      4       0    0.431     770      4       0    0.658
 855      3       1    0.287    1129      3       0    0.658
1040      2       0    0.287    1206      2       0    0.658
1106      1       0    0.287    1227      1       0    0.658
```



4 The study of the effects of chemotherapy on ovarian cancer reported by Collett, (2003) gives the survival times in days of 26 patients randomized to one of two forms of chemotherapy following surgery for ovarian cancer, where status records whether the observation is censored (status = 0) or complete (status =1). In addition to treatment (coded as 0 and 1 for treatments 1 and 2 respectively) and survival times various covariates were measured on each subject. These were AGE (in years), RDISEAE (whether residual disease was present, coded as 1 and 2), and PERF (performance status at the start of the trial, coded as 1 and 2).. Given below is some further analysis in **R** (edited in places) which does not make any assumption of the distribution of survival times with the aim of investigating the effects of the various covariates.

(a) What conclusions can be drawn from the three individual log-rank tests on the binary factors?

(5 marks)

(b) Specify the form of the proportional hazards model used for an analysis in terms of the baseline hazard function $h_0(t)$ and the main effects of all four covariates.

(2 marks)

(c) Describe in detail the effects of the various covariates on the survival time of the subjects, providing approximate 95% confidence intervals for the hazard ratios of the two-level factors TREAT, RDISEASE and PERF.

(9 marks)

(d) What graphical diagnostic plots would you use to investigate the validity of the proportional hazards assumptions in the regression analyses and what features would these have if the assumption is satisfied?

(4 marks)

(e) How would you extend the model to allow for any interaction between the treatment and age? What would be the two algebraic forms of this model for those subjects on treatments 1 and 2?

(5 marks)

Question 4 continued on next page

Question 4 continued

```
> library(survival)
> attach(ovarian)
> ovarian[1:5,]
  patient time status treat age rdisease perf
1       1  156      1     0  66        2    2
2       2 1040      0     0  38        2    2
3       3   59      1     0  72        2    1
4       4  421      0     1  53        2    1
5       5  329      1     0  43        2    1
>
> ovarian.sv<-Surv(time, status)
> survdiff(ovarian.sv~treat)
         N Observed Expected (O-E)^2/E (O-E)^2/V
treat=0 13        7     5.23     0.596      1.06
treat=1 13        5     6.77     0.461      1.06
 Chisq= 1.1  on 1 degrees of freedom, p= 0.303
>
> survdiff(ovarian.sv~rdisease)
            N Observed Expected (O-E)^2/E (O-E)^2/V
rdisease=1 11        3     6.26      1.70      3.62
rdisease=2 15        9     5.74      1.85      3.62
 Chisq= 3.6  on 1 degrees of freedom, p= 0.057
>
> survdiff(ovarian.sv~perf)
        N Observed Expected (O-E)^2/E (O-E)^2/V
perf=1 14        5     6.18     0.226     0.468
perf=2 12        7     5.82     0.240     0.468
 Chisq= 0.5  on 1 degrees of freedom, p= 0.494
>
> ovarian.phtreat<-coxph(ovarian.sv~treat)
> ovarian.phtreat
coxph(formula = ovarian.sv ~ treat)

        coef exp(coef) se(coef)     z    p
treat -0.596     0.551    0.587 -1.02 0.31

Likelihood ratio test=1.05  on 1 df, p=0.305  n= 26
>
> ovarian.ph<-coxph(ovarian.sv~treat+age+rdisease+perf)
>
> ovarian.ph
Call:
coxph(formula = ovarian.sv ~ treat + age + rdisease + perf)

            coef exp(coef) se(coef)     z      p
treat    -0.909     0.403   0.6538 -1.39 0.1600
age       0.124     1.132   0.0467  2.66 0.0078
rdisease  0.834     2.302   0.7903  1.06 0.2900
perf      0.380     1.462   0.6440  0.59 0.5600

Likelihood ratio test=17.0  on 4 df, p=0.00191  n= 26
```

End of Question Paper