

MAS5052



The
University
Of
Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS

**Spring Semester
2011–2012**

Basic Statistics

2 hours

RESTRICTED OPEN BOOK EXAMINATION.

Candidates may bring to the examination lecture notes and associated lecture material (including set textbooks) plus a calculator that conforms to University regulations.

*Candidates should attempt **ALL** questions.*

The maximum marks for the various parts of the questions are indicated.

The paper will be marked out of 80.

**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 publisher is interested in the readership of three of its popular magazines, in particular the age of buyers. A survey is carried out in a local newsagent to record the ages of 50 buyers of each magazine. The summary statistics can be found below:

Magazine Topic	Min	Lowest quartile	Median	Upper quartile	Max
Pop Music	16	21	27	36	41
Current Affairs	18	34	54	76	91
Classic Design	30	40	65	69	81

- (i) Provide a suitable display to illustrate the differences in readership between the three magazines. *(4 marks)*
- (ii) Comment briefly on these differences. *(4 marks)*

- 2 X_1, \dots, X_n are a random sample from the $Be(3, \theta)$ distribution. We wish to test

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta = \theta_1$$

where $\theta_0 < \theta_1$. Show that the most powerful test is to reject H_0 if

$$T = \prod_{i=1}^n (1 - X_i) > k^*$$

for some k^* . *(8 marks)*

- 3 An $Exp(\theta)$ random variables has density

$$f(x; \theta) = \frac{1}{\theta} e^{-x/\theta} \quad x \geq 0 \quad (\text{Note parameterization})$$

Suppose that X_1 and X_2 are taken from a $Exp(\alpha\lambda)$ and $Exp(\beta\lambda)$ where α and β are known and positive.

- (i) Find the maximum likelihood estimator for λ given X_1 and X_2 . *(5 marks)*
- (ii) By calculating bias and variances compare it with the alternative estimators

$$T_1 = \frac{X_1 - X_2}{a - b}$$

$$T_2 = \frac{X_1 + X_2}{a + b}$$

[Hint: Note that $(a - b)^2 \geq 0$ and so $a^2 + b^2 \geq 2ab$ for all a and b .] *(11 marks)*

- 4 In a survey on politics, voters from two regions were asked which of the three main political leaders they supported. The replies are shown in the following table:

	Area		
	A	B	
David Cameron	63	24	87
Ed Milliband	42	37	79
Nick Clegg	16	18	34
	121	79	200

Do the data provide evidence of dependence between area and political leader?
(8 marks)

- 5 In bowel cancer, there are two different type of polyps (tumors) that grow. Six polyps of each type have their size measured (mm) and recorded below:

Polyp Type A	10.1	9.3	4.5	7.6	11.2	11.6
Polyp Type B	9.9	13.2	8.2	13.7	19.1	17.7

- (i) Find the sample mean and variance for the size of each type of polyp.
(4 marks)
- (ii) Test whether polyps of the two types have equal variances. *(5 marks)*
- (iii) In light of this result perform a suitable test to determine whether the mean size of the two types of polyp are equal. *(5 marks)*
- (iv) What assumptions have you made in performing these tests? *(2 marks)*

- 6 Three new drugs are being considered to help reduce the cholesterol level in adults. To study this a trial was designed where each drug was randomly allocated to four male and four female subjects. After taking the drug for a period of time, the following data were collected on the subjects' cholesterol levels.

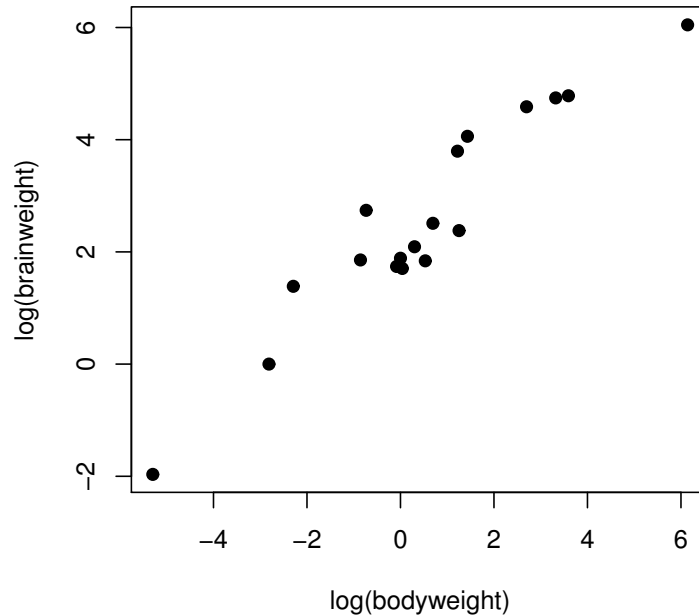
	Drug											
	1				2				3			
Male	1.32	4.38	1.84	1.88	1.60	3.33	3.28	2.97	4.09	4.63	3.42	2.91
Female	2.02	2.81	2.65	2.10	4.79	3.94	4.11	3.31	4.42	6.21	5.21	3.90

Interpret the following R output, making sure you give the model initially fitted and the most appropriate model for the data. Explain your reasoning and make sure you give a conclusion in simple, clear language.

```
> CholFit <- aov(formula = Chol ~ Gender * Drug, data = Chol.Data)
> summary(CholFit)
          Df  Sum Sq Mean Sq F value    Pr(>F)
Gender      1   4.0047   4.0047   5.1684 0.035483 *
Drug        2  15.6064   7.8032  10.0708 0.001161 **
Gender:Drug  2   1.8194   0.9097   1.1741 0.331694
Residuals  18  13.9470   0.7748
---
```

(8 marks)

- 7 The average brain and body weight for 18 mammals was recorded. This data was entered into R as the variables `brain` (measured in g) and `body` (measured in kg) respectively.



The following analysis was performed:

```
> lm1 <- lm(log(brain) ~ log(body))
> summary(lm1)
```

Call:

```
lm(formula = log(brain) ~ log(body))
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	2.20884	0.14152	15.61	4.21e-11	***
log(body)	0.70342	0.05541	12.69	9.07e-10	***

Residual standard error: 0.5884 on 16 degrees of freedom

Multiple R-squared: 0.9097, Adjusted R-squared: 0.904

F-statistic: 161.1 on 1 and 16 DF, p-value: 9.07e-10

- (i) Using standard notation, where y is the brain weight (g) and x is the body weight (kg), what is the model selected by this analysis? *(2 marks)*

- (ii) Give a prediction for the brain weight of a mammal with a body weight of 1.3kg. *(3 marks)*

7 (continued)

- (iii) Give a 95% confidence interval for the expected increase in $\log(\text{brainweight})$ for an increase of one in the value of $\log(\text{bodyweight})$. Hence test a biologist's claim that

$$\text{brainweight} \propto (\text{bodyweight})^{0.8}.$$

(3 marks)

8 A company employs 600 people, of these 200 are women and 400 are men. They wish to carry out a survey of their employees to determine job satisfaction. An overall sampling fraction of 5% has been decided on and they have decided to stratify by gender.

- (i) Why might the company have decided to stratify by gender? (2 marks)
- (ii) Suppose that the company wishes to use a large sampling fraction of female employees. If they wish to sample 9% of the women while still keeping the overall sampling fraction the same at 5%, how many male and female employees will be sampled. (2 marks)
- (iii) They propose to sample people at the Christmas party as they think that it will be an easy opportunity to find interviewees. Give two drawbacks of this proposal. (2 marks)
- (iv) After consultation they decide to instead conduct the survey over email. After the survey was sent out they had quite a large number of non-responses from people who said they were too busy. Should the company be worried about these non-responses or can they just ignore them? Explain your answer very briefly (2 marks)

End of Question Paper