The University Of Sheffield.

MAS5052

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2014–2015**

Basic Statistics **2 hours**

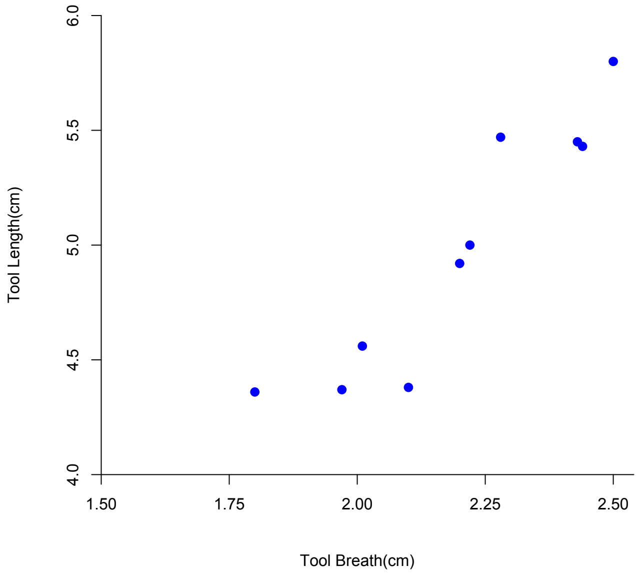
RESTRICTED OPEN BOOK EXAMINATION.
Candidates may bring to the examination lecture notes and associated lecture material (including set textbooks) plus a calculator that conforms to University regulations.
Candidates should attempt ***ALL*** *questions.*
The maximum marks for the various parts of the 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 The following table summarises data provided by an archaeologist. They relate to the length (at the longest point) and breadth (at the widest point) in centimetres of each of ten flint tools. Alongside it is a plot of the same data.



Breadth (x_i)	Length (y_i)
1.97	4.37
2.44	5.43
2.22	5.00
2.10	4.38
2.43	5.45
2.50	5.80
1.80	4.36
2.28	5.47
2.20	4.92
2.01	4.56

(i) Comment briefly on the main features of the plot. *(2 marks)*

(ii) Use the following information to obtain the equation of the least squares regression line for predicting tool length from tool breadth on the basis of these data. *(4 marks)*

$$\sum x_i = 21.95 \quad \sum x_i^2 = 48.646$$
$$\sum y_i = 49.79 \quad \sum y_i^2 = 250.039 \quad \sum x_y y_i = 110.209$$

(iii) Suppose that the archaeologist has two other tools of the same type which have had the points knocked off (and so are shorter, but not narrower, than they originally were). One has a breadth of 2.4cm and the other 3.2cm. Use your regression line to obtain estimates of the length of these tools and comment on the reliability of your estimates. *(2 marks)*

2 The weight (in kg) of 70 individuals undertaking an exercise class were recorded and can be found below

Weight (kg)	50–60	60–70	70–75	75–80	80–90	90–110	110–150
Frequency	3	16	4	9	6	18	14

(i) Represent the data in a suitable graphical format. *(6 marks)*

(ii) Provide a (very) brief interpretation of the data. *(2 marks)*

3 Independent observations X, Y, Z are made, with gamma distributions as follows:

$$X \sim \text{Ga}(1, 1/\mu), \quad Y \sim \text{Ga}(2, 2/\mu), \quad Z \sim \text{Ga}(3, 3/\mu)$$

[The pdf of a r.v. $T \sim \text{Ga}(a,b) = b^a/\Gamma(a)t^{a-1}e^{-bt}$ and $E[T] = a/b, \text{Var}[T] = a/b^2$.]

Three estimators of μ are considered:

$$t_1(X,Y,Z) = (X+Y+Z)/3\,, \qquad t_2(X,Y,Z) = (X+2Y+3Z)/6\,,$$
$$t_3(X,Y,Z) = (X+2Y+3Z)/7\ .$$

(i) Which of these estimators are unbiased? ***(3 marks)***

(ii) Prove that $\text{Var}[t_2(X,Y,Z)] = \mu^2/6$ and that $\text{Var}[t_1(X,Y,Z)]$ is larger than this for all μ. ***(3 marks)***

(iii) Prove that $t_3(X,Y,Z)$ has mean squared error $\mu^2/7$. Which estimator would you prefer (give reasons)? ***(4 marks)***

(iv) Find the maximum likelihood estimator of μ given the three observations X, Y and Z. ***(6 marks)***

4 Sam is testing out a new method for statistics teaching. In order to see if it makes a difference to student understanding he signs up 16 people onto his course and randomly assigns them to one of two groups. In one group he tries out his old teaching method and in the other group he tries his new method. At the end of the year he sets them all an exam and records their marks

Old Style Teaching	41	73	58	54	85	60	81	50
New Style Teaching	70	71	85	93	87	71	88	70

(i) Test whether the marks of the students taught by the two learning approaches have equal variances. ***(7 marks)***

(ii) In light of this result, perform a suitable test to determine if the mean exam marks of the two types of teaching are equal. ***(7 marks)***

(iii) What assumptions have you made in performing these tests? ***(2 marks)***

5 Suppose that a Science Faculty at a UK university has 810 male students and 680 female students in the final year. The Dean of the Faculty commissions a study to learn about the experiences of final year students. Assume that a decision has already been made to sample 10% of the final year students as a whole and to stratify the study by gender.

(i) Why might the organisers of the study have decided to stratify by gender? ***(2 marks)***

(ii) Suggest a suitable proportional sampling scheme for the Science Faculty study outlined above. ***(2 marks)***

(iii) Suppose that there is particular concern about the experiences of female students. If the sampling fraction for female students is to be raised to 15%, but the overall sample fraction is to remain the same, how many male and female students will be sampled? ***(2 marks)***

(iv) One of the questions the Dean proposes to ask is "How many hours did you spend studying last week?" Comment on this choice of question. ***(2 marks)***

6 A new drug is being trialled and doctors are interested in the potential side effects. A series of subjects are either given the drug or a placebo and the outcome recorded was whether or not the subject suffered a stroke in the next 5 years. The data can be seen below:

	Outcome		
	Stroke	No Stroke	
Drug	18	957	975
Placebo	7	967	974
	25	1924	1949

Test if there is any association between taking the drug and having a stroke. ***(8 marks)***

7 Each of the following statements (taken from Campbell & Machin, 1999) concerning p-values is either true or false. For each statement, identify whether it is true or false and if it is false, give an explanation for your decision.

(i) The p-value is the probability that the null hypothesis is true.

(ii) The p-value is the probability of the observed result, or one more extreme, if the null hypothesis were true.

(iii) The p-value is one minus the Type II error probability.

(iv) The p-value can only take a limited number of values such as 0.1, 0.05, 0.01, etc. ***(8 marks)***

8 Let $X_1, \ldots, X_n \sim f(\theta) = \theta x^{\theta-1}$ for $x \in (0,1)$ and $\theta > 0$. Show that the uniformly most powerful test for

$$H_0 \,:\, \theta = 1 \quad \text{vs} \quad H_1 \,:\, \theta = \theta_1 > 1.$$

rejects the null hypothesis if

$$T(\mathbf{X}) = -\sum_{i=1}^{n} \log X_i$$

is small enough. ***(8 marks)***

End of Question Paper