MAS5052

The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2018–2019**

Basic Statistics **2 hours and 30 minutes**

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 In England and Wales there have been a lot of reports in the media about the rise in crime and whether this is linked to the fall in police numbers.

In the table below, police numbers (in thousands), as of 31 March for 6 years from 2013 until 2018, are shown alongside the number of recorded crimes (in millions) for the year ending 31 March. These data come from the Office of National Statistics and the Home Office, see

`https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/crimeinenglandandwales/yearendingmarch2018`

`https://www.gov.uk/government/statistics/police-workforce-england-and-wales-31-march-2018`

Year	2013	2014	2015	2016	2017	2018
Police Numbers as of 31 March (thousands)	129.6	127.9	127.2	124.1	123.1	122.4
Total recorded crime for year ending 31 March (millions)	4.1	4.0	4.2	4.5	5.0	5.5

(i) Represent the data in a suitable graphical format. ***(5 marks)***

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

(iii) A model was fitted and the following output was obtained (from `R`).

```
Call:
lm(formula = Crime ~ Police)

Residuals:
       1        2        3        4        5        6
 0.25480 -0.15374 -0.08078 -0.34342 -0.02491  0.34805

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 27.36680    5.57956   4.905  0.00802 **
Police      -0.18149    0.04437  -4.090  0.01497 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.2893 on 4 degrees of freedom
Multiple R-squared:  0.807,Adjusted R-squared:  0.7588
F-statistic: 16.73 on 1 and 4 DF,  p-value: 0.01497
```

Explain carefully what model has been fitted and what assumptions have been made. Give the estimates of the model parameters. ***(6 marks)***

(iv) Include a graphical representation of the model results into the figure produced in (i) ***(2 marks)***

(v) Does this model provide any evidence of a link between crime figures and police numbers? ***(5 marks)***

2 The publication "Equality in higher education: statistical report 2018", see `https://www.ecu.ac.uk/publications/equality-higher-education-statistical-report-2018/`, was published in September 2018. The table below shows extracted data from the publication.

Ethnic group	**Professors**	**Non-Professors**
White	17,130	143,700
Black	115	3,390
Chinese	455	6,085
Other	1,250	17,270

Using this information, carry out a suitable test to investigate differences in the numbers of professors by ethnic group. Once obtaining a test statistic write a suitable R command to assess the significance. Given the values of your test statistic what conclusion do you think is likely? ***(9 marks)***

3 (continued)

3 One of the topics that is debated regarding Brexit (i.e. Britain leaving the European Union (EU)) is that of trade. Will Britain do more or less trade with the EU after Brexit and did the effect of the referendum result already have an impact before Brexit? To assess the latter of these points, the table below gives trade figures between Britain and 6 EU countries for the year before the referendum (2015) and the year after the referendum (2017). The data come from:
`http://www.intracen.org/itc/market-info-tools/trade-statistics/`
Using the definition, "Balance of Trade is the difference between the value of exports and imports", assess, stating assumptions made, whether there has been a difference in the Balance of Trade between Britain and the rest of the EU from 2015 to 2017. Figures are in millions of GB pounds.

Country	Exported value (2015)	Exported value (2017)	Imported value (2015)	Imported value (2017)
Germany	30.5	36.2	61.7	69.6
France	17.8	23.6	25.3	28.3
Netherlands	17.3	21.4	31.1	39.9
Ireland	16.7	19.5	12.5	14.6
Belgium	11.6	13.8	20.5	25.0
Spain	8.9	10.5	14.1	15.7

Some of the following R output may be of use:

```
> qt(0.99,5)        > qt(0.99,6)
[1] 3.36493         [1] 3.142668
> qnorm(0.99,6)     > qnorm(0.99,5)
[1] 8.326348        [1] 7.326348
> qt(0.95,6)        > qnorm(0.95,6)
[1] 1.94318         [1] 7.644854
> qnorm(0.95,5)     > qt(0.95,5)
[1] 6.644854        [1] 2.015048
> qt(0.90,6)        > qnorm(0.90,6)
[1] 1.439756        [1] 7.281552
> qnorm(0.90,5)     > qt(0.90,5)
[1] 6.281552        [1] 1.475884
> qt(0.80,6)        > qnorm(0.80,6)
[1] 0.9057033       [1] 6.841621
> qt(0.80,5)        > qnorm(0.80,5)
[1] 0.9195438       [1] 5.841621
> qnorm(0.75,5)     > qt(0.75,5)
[1] 5.67449         [1] 0.7266868
> qt(0.75,6)        > qnorm(0.75,6)
[1] 0.7175582       [1] 6.67449
```

(14 marks)

4 Suppose that a standard treatment of athletes foot is to be compared with a new one. The measure of success is the proportion of patients who are symptom free for a year. Patients are allocated at random either to the new treatment or to the old one. Of the 54 patients receiving the new treatment 41 are symptom free for a year, while 18 of the 31 that received the old treatment are symptom free for a year. The company is interested in whether there is evidence that the new treatment is better then the old one.

(i) Formulate the problem statistically, including specifying the appropriate hypotheses. ***(4 marks)***

(ii) Do the data provide evidence that the new treatment is better than the old one? Please provide R code to assess the significance of the test statistic that you use. ***(5 marks)***

(iii) If the test statistic gave $p = 0.046$, what would your conclusion be? ***(2 marks)***

5 The positive random variables $X_1, X_2, \ldots X_n$ are independent observations having the Gamma distribution $Ga(3, 1/\eta)$, with density function

$$\frac{x^2}{2\eta^3} e^{-x/\eta} \quad (x > 0)$$

which depends on the unknown parameter $\eta \in (0, \infty)$.

(i) Find the maximum likelihood estimator of η. ***(6 marks)***

(ii) Show that the maximum likelihood estimator of η is unbiased and find its variance. ***(4 marks)***

(iii) Give the maximum likelihood estimator of η^2. Find the bias of this estimator. Propose an unbiased estimator of η^2. ***(6 marks)***

6 A survey of a university's students is to be carried out to estimate both the proportion, P, who own a bicycle and the average weekly spend on junk food, $\bar{X}$. It is desired to estimate P with standard error of no more than 0.02 and $\bar{X}$ with a 90% confidence interval of length no greater than £20. If P is thought to lie in the range 0.05 to 0.25 and the population variance in junk food spending is thought not to exceed 450 pounds^2, how large a simple random sample would be necessary to meet both these requirements simultaneously? ***(10 marks)***

End of Question Paper