RESTRICTED OPEN BOOK EXAMINATION (Not to be removed from the examination hall)
Data provided: "Statistics Tables" by H.R. Neave

MAS5052

The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2017–2018**

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 caluclator 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 July 2017, the British Broadcasting Company (BBC) released their Annual Report. In the Annex, found at:
`http://downloads.bbc.co.uk/aboutthebbc/insidethebbc/reports/pdf/annex_annual_report_201617.pdf`,
they published the earnings of the top 96 salary earners who appeared on the television (defined as "on air talent and contributors") who earned £150,000 or more in the financial year 2016/17. Exact salary figures were not given, however a summarised table of the data, extracted from the report, can be seen below.

Wage (£)	150,000–199,999	200,000–299,999	300,000–499,999	500,000–2,249,999
Frequency	39	34	16	7

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

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

2 The publication of the Annual Report referred to in Question 1 created headlines, as there were claims that (in general) men earn more than women at the BBC. The data below shows the number of women and total number of employees at each pay grade at the BBC as of March 2017 and is taken from:
`http://downloads.bbc.co.uk/diversity/pdf/equality-information-report-2017.pdf`.

Pay Band	**Number of Women**	**Total Number**
Worldwide Non-Senior Manager Equivalent	504	867
Bands 1-4	1,362	2,433
Bands 5-7	4,707	9,395
Bands 8/9	2,398	5,556
Bands 10/11	1,131	2,629
Bands SM2/1	128	359
Grand Total	10,230	21,239

(i) Using this information, carry out a suitable test to investigate the claims regarding the difference in pay between genders. ***(8 marks)***

(ii) Comment on the suitability of the data given for exploring the gender pay issue. ***(2 marks)***

3 The annual survey of the cost of attending an English professional football match was carried out at the beginning of the 2017/18 football season and published in November 2017 (see `http://www.bbc.co.uk/sport/football/41482931`). Each of the 92 clubs in the four divisions were asked (amongst other things) how much the most expensive ticket was for attending a match (the four divisions comprise 20, 24, 24 and 24 teams). In order to investigate differences in price by division, a one-way ANOVA analysis was carried out. The resulting ANOVA table is shown below, with some values omitted.

Source	**Df**	**Sum of Squares**	**Mean Square**	**F**
Division	a	c	e	59.076
Error	b	d	76.728	

(i) Determine values for a, ..., e to complete the table. ***(8 marks)***

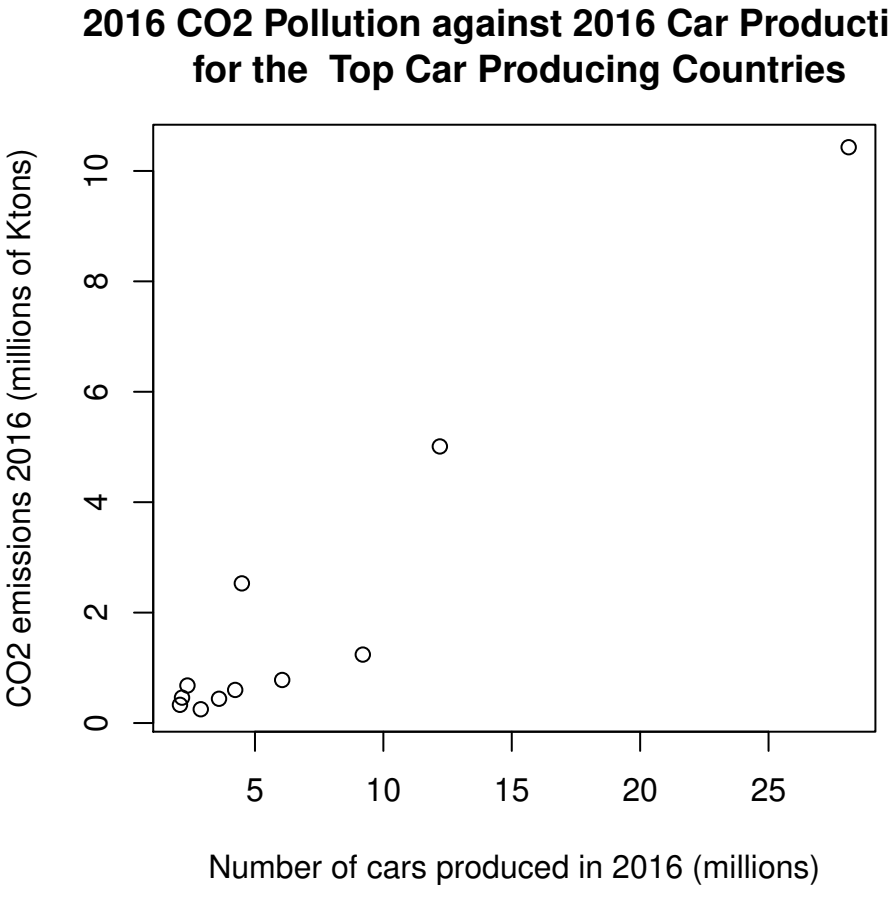
(ii) Is there evidence that the cost of the most expensive ticket differs between the divisions? ***(3 marks)***

(iii) Apart from independence of the observations, what assumptions have to be met for this analysis to be valid? ***(2 marks)***

4 The table and subsequent plot below show data from the eleven countries that produced the largest number of new cars in 2016, together with the amount of CO_2 emitted by each county in 2016, and is taken respectively from:

`https://en.wikipedia.org/wiki/List_of_countries_by_motor_vehicle_production`
and
`http://edgar.jrc.ec.europa.eu/overview.php?v=CO2andGHG1970-2016&sort=des8`

Country	Car Production 2016 (millions) x	CO_2 Pollution 2016 (millions Ktons) y
China	28.12	10.43
United States	12.20	5.01
Japan	9.20	1.24
Germany	6.06	0.78
India	4.49	2.53
South Korea	4.23	0.60
Mexico	3.60	0.44
Spain	2.89	0.25
Canada	2.37	0.68
Brazil	2.16	0.46
France	2.08	0.33



 Question 4 continued on next page

4 (continued)

Summary statistics are:

$$\sum x = 77.40, \sum y = 22.75,$$
$$\sum x^2 = 1134.91, \sum y^2 = 143.83, \sum xy = 390.04.$$

(i) Describe the relationship between car production and CO_2 pollution and explain why car production is shown on the horizontal axis. ***(2 marks)***

(ii) Calculate the correlation coefficient, does this match with what you found in part (i)? ***(5 marks)***

(iii) Find the least squares estimate of the best-fit straight line for these data. ***(4 marks)***

(iv) Test the hypothesis that there was no relationship between car production and CO_2 pollution in 2016. ***(6 marks)***

5 Suppose that $X_1, X_2, \ldots, X_n$ are independently and identically distributed as $Ge(\theta)$.

(i) Find the maximum likelihood estimator of θ. ***(5 marks)***

(ii) Hence show that the maximum likelihood estimator of $\phi = (1-\theta)/\theta$ is the sample mean $\bar{X}$. ***(2 marks)***

(iii) Prove that the alternative estimator $t(\mathbf{X}) = n\bar{X}/(n+1)$ has smaller mean squared error for estimating ϕ than $\bar{X}$, for all θ. ***(9 marks)***

6 A cardiologist is planning to monitor systolic blood pressure levels amongst those people diagnosed with hypertension in 2017. There has been anecdotal evidence that a new treatment has caused the systolic blood pressure to raise, thus meaning the patients are at risk of going from hypertension to super-hypertension. It is known that the population mean reading of systolic blood pressure for those with hypertension is 140. The cardiologist would be very concerned if there was an increase of 5 or more for those on the new treatment. The cardiologist proposes to take a random sample of 16 people on the new treatment and measure their systolic blood pressure. The cardiologist thinks it is reasonable to assume that the standard deviation of the systolic blood pressure of the sample is the same as for the population, which was 8 and that it is Normally distributed.

(i) Specify the null and alternative hypotheses for the standard one-sided test the cardiologist could use to analyze their results. ***(2 marks)***

6 (continued)

(ii) Assuming that a fixed significance level of 5% is to be used, identify what mean systolic blood pressure for the sample of 16 observations would lead to rejection of the null hypothesis. *(3 marks)*

(iii) Using the rejection criterion identified in (ii), what is the power of the test procedure if the increase in mean systolic blood pressure of 5 has just been achieved? *(3 marks)*

(iv) How many patients would the cardiologist need to measure to guarantee a 90% chance of detecting an increase of 5 in systolic blood pressure? *(5 marks)*

(v) If the cardiologist was not happy with the validity of the assumptions that they made regarding the distribution and variability of systolic blood pressure in the sample above, outline how they might proceed. *(3 marks)*

End of Question Paper