OCR
Oxford Cambridge and RSA

Friday 16 June 2017 – Afternoon

AS GCE MEI STATISTICS

G243/01   Statistics 3 (Z3)

QUESTION PAPER

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

These instructions are the same on the Printed Answer Book and the Question Paper.
*   The Question Paper will be found inside the Printed Answer Book.
*   Write your name, centre number and candidate number in the spaces provided on the Printed Answer Book. Please write clearly and in capital letters.
*   **Write your answer to each question in the space provided in the Printed Answer Book.** If additional space is required, you should use the lined page(s) at the end of the Printed Answer Book. The question number(s) must be clearly shown.
*   Use black ink. HB pencil may be used for graphs and diagrams only.
*   Read each question carefully. Make sure you know what you have to do before starting your answer.
*   Answer **all** the questions.
*   Do **not** write in the barcodes.
*   You are permitted to use a scientific or graphical calculator in this paper.
*   Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

This information is the same on the Printed Answer Book and the Question Paper.
*   The number of marks is given in brackets **[ ]** at the end of each question or part question on the Question Paper.
*   You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.
*   The total number of marks for this paper is **72**.
*   The Printed Answer Book consists of **12** pages. The Question Paper consists of **4** pages. Any blank pages are indicated.

INSTRUCTION TO EXAMS OFFICER/INVIGILATOR

*   Do not send this Question Paper for marking; it should be retained in the centre or recycled. Please contact OCR Copyright should you wish to re-use this document.

OCR is an exempt Charity

Turn over

Section A (46 marks)

1   A television viewer who is hearing impaired uses subtitles to help him understand programmes. Many programmes such as news programmes are 'live subtitled' which means that the subtitles appear on the screen a few seconds after the words are spoken. The time taken (in seconds) between the words being spoken and the subtitles appearing on screen is known as the 'latency'. The viewer asks a friend to help record the latency in samples of both national and local television news programmes. The data collected by the friend are given below.

National news	6.5	5.9	5.1	5.8	4.6	5.3	5.5	7.1	4.4	7.8	4.1
Local news	4.3	5.2	3.4	3.9	3.7	4.8	4.9	5.4			

   (i) The viewer wishes to carry out a hypothesis test to investigate whether latencies in the two types of programme are different. What property of the two samples is required in order to carry out such a hypothesis test? **[1]**

   (ii) You are given that the distributions of the populations are not known, and that the property of the samples in part **(i)** does hold. Use a suitable test to examine, at the 5% significance level, whether it appears that, on the whole, there is any difference in latency in the two types of programme. **[10]**

   (iii) The viewer's friend, who is studying statistics, suggests that it would have been better to use the same news stories for both national and local programmes, so that the data are paired.

   (*A*) Name a test for paired data which requires no distributional assumptions. **[2]**

   (*B*) Suppose that there is a difference in latency for the two types of programme, and that the assumptions required for a paired sample *t* test hold. Explain whether the test you have named in part (*A*) would be more likely, equally likely or less likely to detect this difference than a paired sample *t* test. **[2]**

2   A medical student is investigating whether there is any difference in blood pressure when people are standing and when they are seated. She randomly selects 9 students from her university and measures their blood pressures, both seated and standing. The systolic blood pressures, in suitable units, are as follows.

Student	A	B	C	D	E	F	G	H	I
Standing	119	126	116	124	120	151	116	120	126
Seated	118	130	128	134	123	142	121	116	134

   (i) Briefly explain why it is better to use the same students for both positions, standing and seated. **[2]**

   (ii) State the distributional assumption necessary for the use of a paired sample *t* test. **[2]**

   (iii) Use a paired sample *t* test to examine, at the 5% significance level, whether it appears that there is any difference between mean systolic blood pressure when people are standing or seated. **[11]**

3 A student wishes to know if heavier species of mammals tend to have heavier brains. He selects a random sample of 7 species and obtains their average body weights and their average brain weights from a reputable source, as shown in the table below.

Item removed due to third party copyright restrictions.

   (i) Draw a scatter diagram to illustrate these data. **[3]**

  (ii) Explain why the shape of the scatter diagram suggests that the distribution may not be bivariate Normal, and so the assumptions required to carry out a test based on the product moment correlation coefficient may not be satisfied. **[1]**

 (iii) Calculate the value of Spearman's rank correlation coefficient. **[5]**

 (iv) Using your answer to part **(iii)**, carry out a test at the 5% level of significance to investigate whether it is reasonable to assume that average body weight and average brain weight are positively associated. **[7]**

Section B (26 marks)

4 At a small bakery company, the baker is investigating the heights of the loaves which she bakes. She wishes to see if there is any difference in the heights when using two new bread recipes, A and B, and will use this information to choose which recipe to use in future.

 (i) Initially she considers baking 3 loaves with each recipe, all in one batch, to compare the heights. Give 2 reasons why this would not be a sensible approach. **[2]**

After getting advice from a statistician, the baker adopts a better approach. She bakes 5 batches, each containing 60 loaves of each recipe. She uses the same total volume of ingredients for each loaf baked with each recipe.

 (ii) Explain how the baker can choose a stratified sample of 50 loaves of recipe A from the 300 which she bakes. **[4]**

The baker decides instead to use simple random samples to select 50 loaves of each recipe. The heights of the loaves, measured in cm, are denoted by x and y for recipes A and B respectively. You are given that $\Sigma x = 709.5$, $\Sigma x^2 = 10092$ and the sample mean and sample standard deviation of y are 14.51 and 0.718 respectively.

 (iii) Calculate the sample mean and sample standard deviation of x. **[3]**

 (iv) Carry out a test at the 5% significance level to investigate whether there is any difference in the mean heights of loaves produced using the two recipes. **[11]**

 (v) Explain briefly why it is not necessary to make any assumptions about the distributions of the parent populations in order to carry out the test. **[2]**

 (vi) If the two samples had each been of size 10 rather than 50, a Wilcoxon rank sum test could have been used. Name a second test which could have been used under these circumstances. State the assumptions necessary for this second test. **[4]**

END OF QUESTION PAPER