OCR
Oxford Cambridge and RSA

Wednesday 14 June 2023 – Afternoon

A Level Further Mathematics A

Y542/01 Statistics

Time allowed: 1 hour 30 minutes

9976854652

You must have:
- the Printed Answer Booklet
- the Formulae Booklet for A Level Further Mathematics A
- a scientific or graphical calculator

QP

INSTRUCTIONS
- Use black ink. You can use an HB pencil, but only for graphs and diagrams.
- Write your answer to each question in the space provided in the **Printed Answer Booklet**. If you need extra space use the lined pages at the end of the Printed Answer Booklet. The question numbers must be clearly shown.
- Fill in the boxes on the front of the Printed Answer Booklet.
- Answer **all** the questions.
- Where appropriate, your answer should be supported with working. Marks might be given for using a correct method, even if your answer is wrong.
- Give non-exact numerical answers correct to **3** significant figures unless a different degree of accuracy is specified in the question.
- The acceleration due to gravity is denoted by $g\,\text{m}\,\text{s}^{-2}$. When a numerical value is needed use $g = 9.8$ unless a different value is specified in the question.
- Do **not** send this Question Paper for marking. Keep it in the centre or recycle it.

INFORMATION
- The total mark for this paper is **75**.
- The marks for each question are shown in brackets **[ ]**.
- This document has **8** pages.

ADVICE
- Read each question carefully before you start your answer.

© OCR 2023 [J/508/5510]
DC (LK) 322893/4

OCR is an exempt Charity

1 A certain section of a library contains several thousand books. A lecturer is looking for a book that refers to a particular topic. The lecturer believes that one-twentieth of the books in that section of the library contain a reference to that topic. However, the lecturer does not know which books they might be, so the lecturer looks in each book in turn for a reference to the topic. The first book the lecturer finds that refers to the topic is the Xth book in which the lecturer looks.

(a) A student says, "There is a maximum value of X as there is only a finite number of books. So a geometric distribution cannot be a good model for X."

Explain whether you agree with the student. **[1]**

(b) (i) State **one** modelling assumption (not involving the total number of books) needed for X to be modelled by a geometric distribution in this context. **[1]**

(ii) Suggest a reason why this assumption may **not** be valid in this context. **[1]**

Assume now that X can be well modelled by the distribution Geo(0.05).

(c) The probability that the lecturer needs to look in no more than n books is greater than 0.9.

Find the smallest possible value of n. **[3]**

(d) The lecturer needs to find four different books that refer to the topic.

Find the probability that the lecturer wants to look in exactly 40 books. **[2]**

2 The director of a concert hall wishes to investigate if the price of the most expensive concert tickets affects attendance. The director collects data about the price, £P, of the most expensive tickets and the number of people in the audience, H hundred (rounded to the nearest hundred), for 20 concerts. For each price there are several different concerts. The results are shown in the table.

P (£)	75	65	55	45	35
H (hundred)	27	27	27	26	15
	27	27	20	21	12
		22	18	16	9
		19	18	13	
		12	16	9	

$n = 20 \quad \Sigma p = 1080 \quad \Sigma h = 381 \quad \Sigma p^2 = 61300 \quad \Sigma h^2 = 8011 \quad \Sigma ph = 21535$

(a) Calculate the equation of the regression line of h on p. **[2]**

(b) State what change, if any, there would be to your answer to part **(a)** if H had been measured in thousands (to 1 decimal place) rather than in hundreds. **[1]**

For a special charity concert, the most expensive tickets cost £50.

(c) Use your answer to part **(b)** to estimate the expected size of the audience for this concert. Give your answer correct to **1** decimal place. **[1]**

(d) Comment on the reliability of your answer to part **(c)**. You should refer to

- the value of the product-moment correlation coefficient for the data, which is 0.642
- the value of £50
- any **one** other relevant factor that should be taken into account. **[4]**

3 The discrete random variable W has the distribution U(11). The independent discrete random variable V has the distribution U(5).

(a) It is given that, for constants m and n, with $m > 0$,
$\mathrm{E}(mW+nV) = 0$ and $\mathrm{Var}(mW+nV) = 1$.

Determine the exact values of m and n. **[5]**

The random variable T is the mean of three independent observations of W.

(b) Explain whether the Central Limit Theorem can be used to say that the distribution of T is approximately normal. **[1]**

4 Two magazines give numerical ratings to hi-fi systems. Li wishes to test whether there is agreement between the opinions of the magazines. Li chooses a random sample of 5 hi-fi systems and looks up the ratings given by the two magazines. The results are shown in the table.

System	A	B	C	D	E
Magazine I	68	75	77	83	92
Magazine II	30	25	40	35	45

(a) Give a reason why Li might choose to use a test based on Spearman's rank correlation coefficient rather than on Pearson's product-moment correlation coefficient. **[1]**

(b) Calculate the value of Spearman's rank correlation coefficient for the data. **[4]**

(c) Use your answer to part **(b)** to carry out a hypothesis test at the 5% significance level. **[4]**

(d) The value of Spearman's rank correlation coefficient between the ratings given by magazine I and by a third magazine, magazine III, has the same numerical value as the answer to part **(b)** but with the sign changed.

In the Printed Answer Booklet, complete the table showing the rankings given by magazine III. **[1]**

5 An historian has reason to believe that the average age at which men got married in the seventeenth century was higher in urban areas compared to rural areas. The historian collected data from a random sample of 8 men in an urban area and a random sample of 6 men in a rural area, all of whom were married in the seventeenth century. The results were as follows, given in the form years/months.

Urban: 18/3 18/5 19/9 20/7 25/6 34/6 41/8 46/3
Rural: 18/0 18/1 18/4 19/11 22/2 28/11

(a) Use an appropriate non-parametric method to test at the 5% significance level whether the average age at marriage of men is higher in urban areas than in rural areas. **[7]**

(b) When checking the data, the historian found that the age of one of the men, Mr X, which had been recorded as 28/11, had been wrongly recorded. When corrected, the result of the test in part **(a)** was unchanged.

Determine the youngest age that Mr X could have been, given that it was not the same, in years and months, as that of any of the other men in the sample. **[3]**

© OCR 2023

6 The continuous random variable X has a uniform distribution on the interval $[-\pi, \pi]$.

The random variable Y is defined by $Y = \sin X$.

Determine the cumulative distribution function of Y. **[7]**

7 A club secretary collects data about the time, T minutes, needed to process the details of a new member. The mean of T is denoted by μ. The variance of T is denoted by σ^2. The results of a random sample of 40 observations of T are summarised as follows.

$n = 40 \quad \Sigma t = 396.0 \quad \Sigma t^2 = 4271.40$

(a) Determine a 99% confidence interval for μ. **[7]**

(b) The secretary discovers that over a long period the value of σ^2 is in fact 10.0. The secretary collects an independent random sample of 50 observations of T and constructs a new 99% confidence interval for μ based on this sample of size 50, but using $\sigma^2 = 10.0$.

Find the probability that this new confidence interval contains the value $\mu + 1.6$. **[3]**

TURN OVER FOR QUESTION 8

8 A team of researchers have reason to believe that the number of calls received in randomly chosen 10-minute intervals to a call centre can be well modelled by a Poisson distribution. To test this belief the researchers record the number of telephone calls received in 60 randomly chosen 10-minute intervals. The results, together with relevant calculations, are shown in the following table.

							Total
Number of calls, r	0	1	2	3	4	$\geqslant 5$	
Observed frequency, f	18	13	12	9	8	0	60
rf	0	13	24	27	32	0	96
r^2f	0	13	48	81	128	0	270
Expected frequency	12.114	19.382	15.506	8.270	3.308	1.421	60
Contribution to test statistic	2.860	2.101	0.793	1.232			6.99

(a) Calculate the mean of the observed number of calls received. **[1]**

(b) Calculate the variance of the observed number of calls received. **[1]**

(c) Comment on what your answers to parts **(a)** and **(b)** suggest about the proposed model. **[1]**

(d) Explain why it is necessary to combine some cells in the table. **[1]**

(e) Show how the values 15.506 and 0.793 in the table were obtained. **[4]**

(f) Carry out the test, at the 5% significance level. **[5]**

In the light of the result of the test, the team consider that a different model is appropriate. They propose the following improved model:

$$P(R=r)=\begin{cases}\frac{1}{60}(a+(2-r)b) & r=0,1,2,3,4\\ 0 & \text{otherwise,}\end{cases}$$

where a and b are integers.

(g) Use **at least three** of the observed frequencies to suggest appropriate values for a and b. You should consider more than one possible pair of values, and explain which pair of values you consider best. (Do not carry out a goodness-of-fit test.) **[3]**

END OF QUESTION PAPER

© OCR 2023

BLANK PAGE

© OCR 2023

OCR
Oxford Cambridge and RSA

Copyright Information

OCR is committed to seeking permission to reproduce all third-party content that it uses in its assessment materials. OCR has attempted to identify and contact all copyright holders whose work is used in this paper. To avoid the issue of disclosure of answer-related information to candidates, all copyright acknowledgements are reproduced in the OCR Copyright Acknowledgements Booklet. This is produced for each series of examinations and is freely available to download from our public website (www.ocr.org.uk) after the live examination series.

If OCR has unwittingly failed to correctly acknowledge or clear any third-party content in this assessment material, OCR will be happy to correct its mistake at the earliest possible opportunity.

For queries or further information please contact The OCR Copyright Team, The Triangle Building, Shaftesbury Road, Cambridge CB2 8EA.

OCR is part of Cambridge University Press & Assessment, which is itself a department of the University of Cambridge.

© OCR 2023