Oxford Cambridge and RSA

Friday 7 June 2024 – Afternoon

AS Level Further Mathematics B (MEI)

Y412/01 Statistics a

Time allowed: 1 hour 15 minutes

You must have:
- the Printed Answer Booklet
- the Formulae Booklet for Further Mathematics B (MEI)
- 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 your final answers to a degree of accuracy that is appropriate to the context.
- Do **not** send this Question Paper for marking. Keep it in the centre or recycle it.

INFORMATION
- The total mark for this paper is **60**.
- 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 2024 [F/508/5554]
DC (PQ) 334974/3
OCR is an exempt Charity

1 The probability distribution for a discrete random variable X is given in the table below.

x	0	1	2	3
$\mathrm{P}(X = x)$	$2c$	$3c$	$0.5 - c$	c

(a) Find the value of c. **[2]**

(b) Find the value of each of the following.

- $\mathrm{E}(X)$
- $\mathrm{Var}(X)$ **[3]**

The random variable Y is defined by $Y = 2X - 3$.

(c) Find the value of each of the following.

- $\mathrm{E}(Y)$
- $\mathrm{Var}(Y)$ **[2]**

2 In a game of chance there are 32 slots, numbered 1 to 32, and on each turn a ball lands in one of them. You may assume that the process is completely random.

You are given that X is the random variable denoting the number of the slot that the ball lands in on a given turn.

(a) Suggest a suitable distribution to model X. You should state the value(s) of any parameter(s). **[2]**

(b) Write down $\mathrm{P}(X = 7)$. **[1]**

Players of the game start with a score of 0. On each turn a player may choose to play the game by selecting a number. If the ball lands in the slot with that number then 15 is added to the player's score. Otherwise, the player's score is reduced by 1. A player's score may become negative.

A player decides to play the game, selecting the number 7 on each turn, until the ball lands in the slot numbered 7.

You are given that Y is the random variable denoting the number of turns up to and including the turn in which the ball lands in the slot numbered 7.

(c) Determine $\mathrm{P}(Y \leqslant 15)$. **[3]**

(d) Determine the player's expected final score. **[3]**

3 A glassware factory produces a large number of ornaments each week. Just before they leave the factory, all the ornaments are checked and some may be found to be defective. The Quality Assurance Manager of the factory wishes to model the number of defective ornaments that are found each week using a Poisson distribution.

The numbers of defective ornaments found each week in a period of 40 weeks are shown in **Table 3.1**.

Table 3.1

No. of defective ornaments in a week, r	0	1	2	3	4	5	6	$\geqslant 7$
No. of weeks with r defective ornaments, f	2	14	13	5	3	1	2	0

You are given that summary statistics for the data are $\sum f = 40$, $\sum rf = 84$ and $\sum r^2 f = 256$.

(a) By using the summary statistics to determine estimates for the mean and variance of the number of defective ornaments produced by the factory each week, explain how the data support the suggestion that the number of defective ornaments produced each week can be modelled using a Poisson distribution. **[3]**

The Quality Assurance Manager is asked by the head office to carry out a chi-squared hypothesis test for goodness of fit based on a Po(2) distribution.

(b) **Table 3.2**, which is incomplete, gives observed frequency, probability, expected frequency and chi-squared contribution.

Table 3.2

No. of defective ornaments in a week, r	Observed frequency	Probability	Expected frequency	Chi-squared contribution
0	2	0.13534	5.4134	2.15232
1	14			
2	13	0.27067		0.43620
3	5		7.2179	
$\geqslant 4$	6	0.14288		0.01421

(i) Complete the copy of the table in the Printed Answer Booklet. **[4]**

(ii) Carry out the test at the 10% significance level. **[6]**

(c) On one occasion a fork-lift truck in the factory drops a crate containing eight ornaments and all of them are subsequently found to be defective.

Explain why the Poisson model cannot model defects occurring in this manner. **[1]**

4 A chemist is conducting an experiment in which the concentration of a certain chemical, A, is supposed to be recorded at the start of the experiment and then every 30 seconds after the start. The time after the start is denoted by t s and the concentration by z mg cm^{-3}. The collected data are shown in the table below. Note that the concentration at $t = 90$ was not recorded.

Time, t	0	30	60	120	150
Concentration of A, z	40.0	31.3	27.5	12.8	11.4

The chemist wishes to plot the data on a graph.

(a) Explain why t should be plotted on the horizontal axis. **[1]**

You are given that the summary statistics for the data are as follows.

$n = 5 \quad \sum t = 360 \quad \sum z = 123.0 \quad \sum t^2 = 41400 \quad \sum z^2 = 3629.74 \quad \sum tz = 5835$

The regression line of z on t is given by $z = a + bt$ and is used to model the concentration of chemical A for $t \geqslant 0$.

(b) (i) Use the summary statistics to determine the value of a and the value of b. **[4]**

(ii) Find the value of the residual at each of the following values of t.

- $t = 60$
- $t = 120$ **[2]**

(c) (i) Use the equation of the regression line to estimate the value of the concentration at 90 seconds. **[1]**

(ii) With reference to your answers to part **(b)(ii)**, comment on the reliability of your answer to part **(c)(i)**. **[1]**

Further experiments indicate that the model is reasonably reliable for times greater than 150 seconds up to about 200 seconds.

(d) Show that the model **cannot** be valid beyond a time of about 200 seconds. **[1]**

5 A student is investigating possible association between the amount of coffee that an adult drinks each day and the number of hours that they remain awake each day. In an initial investigation, a random sample of 8 adults is selected. The student obtains the following information from each of these adults: the amount of coffee that they drink each day and the number of hours that they remain awake each day.

The student analyses the data and finds that the associated product moment correlation coefficient is 0.6030.

(a) State **one** assumption that must be made for a hypothesis test based on the product moment correlation coefficient to be carried out. **[1]**

For the remainder of this question you may assume that this assumption is true.

(b) Carry out a test at the 5% significance level to investigate whether there is any correlation between amount of coffee drunk and number of hours awake. **[5]**

The student conducts a second investigation which is similar to the first but this time based on a random sample of 30 adults. The product moment correlation coefficient for the new data is 0.5487. The student carries out an equivalent hypothesis test to the one carried out in part **(b)**, again using a 5% significance level.

(c) Identify any **differences** between the two tests and their results. You do **not** need to restate the hypotheses or explain the conclusion in context. **[2]**

(d) You may assume the following guidelines for considering effect size.

Product moment correlation coefficient	**Effect size**
0.1	Small
0.3	Medium
0.5	Large

Explain briefly why the results of the student's second investigation are likely to be more reliable than the results of the initial investigation. **[2]**

Turn over for question 6

6 A bank monitors the amounts of cash withdrawn from a cash machine. It categorises any withdrawal of an amount of £50 or less as 'small' and any withdrawal of an amount greater than £50 as 'large'.

Over a long period of time the bank finds that the proportion of withdrawals that are small is 0.43.

The bank wishes to model a sample of 10 withdrawals to examine the number of small withdrawals.

(a) **(i)** State a suitable probability distribution for such a model, justifying your answer. **[3]**

(ii) State **one** assumption needed for the model to be valid. **[1]**

(b) **(i)** Find the probability that exactly 4 of the 10 withdrawals are small. **[1]**

(ii) Find the probability that exactly 4 of the 10 withdrawals are large. **[1]**

(iii) Find the probability that no more than 4 of the 10 withdrawals are large. **[2]**

(c) Find the probability that, in the 10 withdrawals, the 7th withdrawal is large and there are exactly 3 that are small. **[2]**

END OF QUESTION PAPER

BLANK PAGE

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 2024