OCR
Oxford Cambridge and RSA

Friday 14 June 2024 – Afternoon

A Level Further Mathematics B (MEI)

Y432/01 Statistics Minor

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

141685057O

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 [M/508/5596]
DC (LK/CGW) 338108/3

OCR is an exempt Charity

stemup.co.uk

1 When a footballer takes a penalty kick the result is that either a goal is scored or a goal is not scored.

It is known that, on average, a certain footballer scores a goal on 85% of penalty kicks. During one practice session, the footballer decides to take penalty kicks until a goal is **not** scored. You may assume that the outcome of each penalty kick that the footballer takes is independent of the outcome of each other penalty kick.

The random variable representing the number of penalty kicks up to and including the first penalty kick that does **not** result in a goal is denoted by X.

(a) State **one** further assumption that is necessary for X to be modelled by a Geometric distribution. **[1]**

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

(b) Find each of the following.

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

(c) Find the probability that the footballer takes exactly 3 penalty kicks. **[2]**

(d) Find the probability that the footballer takes at least 5 penalty kicks. **[2]**

2 The sides of a fair 12-sided spinner are labelled 1, 2, ..., 12.

The spinner is spun and X is the random variable denoting the number on the side of the spinner that it lands on.

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

(b) Find each of the following.

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

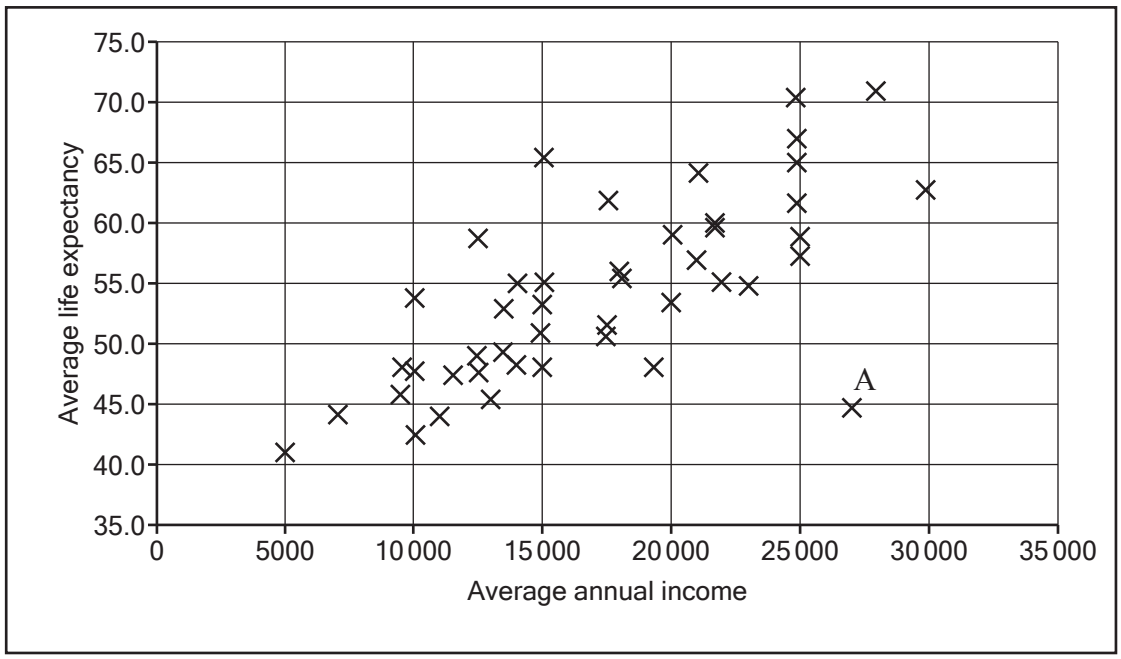
You are given that $\mathrm{E}(X)$ is denoted by μ and $\mathrm{Var}(X)$ is denoted by σ^2.

(c) Determine $\mathrm{P}\left(\left|\frac{2(X-\mu)}{\sigma}\right| > 1\right)$. **[3]**

© OCR 2024

stemup.co.uk

3 The scatter diagram below illustrates data concerning average annual income per person, $\$x$, and average life expectancy, y years, for 45 randomly selected cities.



(a) State whether neither variable, one variable or both variables can be considered to be random in this situation. **[1]**

A student is researching possible positive association between average annual income and average life expectancy. The student decides that the data point labelled A on the scatter diagram is an outlier.

(b) Describe the apparent relationship between average annual income and average life expectancy for this data point relative to the rest of the data. **[1]**

The data for point A is removed. The student now wishes to carry out a hypothesis test using the product moment correlation coefficient for the remaining 44 data points to investigate whether there is positive correlation between average annual income and average life expectancy.

(c) Explain why this type of hypothesis test is appropriate in this situation. Justify your answer. **[2]**

The summary statistics for these 44 data points are as follows.

$\sum x = 751\,120 \quad \sum y = 2397.1 \quad \sum x^2 = 14\,363\,849\,200 \quad \sum y^2 = 133\,014.63 \quad \sum xy = 42\,465\,962$

(d) Determine the value of the product moment correlation coefficient. **[4]**

(e) Carry out the test at the 1% significance level. **[5]**

stemup.co.uk

4 A genetics researcher is investigating whether there is any association between natural hair colour and natural eye colour. A random sample of 800 adults is selected. Each adult can categorise their natural hair colour as blonde, brown, black or red and their natural eye colour as brown, blue or green.

(a) Explain the benefit of using a **random** sample in this investigation. **[1]**

The data collected from the sample are summarised in **Table 4.1**.

Table 4.1

Observed frequency		Hair Colour				
		Blonde	Brown	Black	Red	Total
Eye Colour	Brown	47	153	196	36	432
	Blue	61	78	115	26	280
	Green	19	22	31	16	88
	Total	127	253	342	78	800

The researcher decides to carry out a chi-squared test.

(b) Determine the expected frequencies for each eye colour in the blonde hair category. **[2]**

You are given that the test statistic is 28.62 to 2 decimal places.

(c) Carry out the chi-squared test at the 10% significance level. **[5]**

Table 4.2 shows the chi-squared contributions for some of the categories. The contributions for the categories relating to green eye colour have been deliberately omitted.

Table 4.2

Chi-squared contributions		Hair Colour			
		Blonde	Brown	Black	Red
Eye Colour	Brown	6.791	1.964	0.694	0.889
	Blue	6.162	1.257	0.185	0.062
	Green				

(d) Calculate the chi-squared contribution for the green eye and blonde hair category. **[1]**

(e) With reference to the values in **Table 4.2**, discuss what the data suggest about brown eye colour and blue eye colour for people with blonde hair. **[2]**

(f) A different researcher, carrying out the same investigation, independently takes a different random sample of size 800 and performs the same hypothesis test, but at the 1% significance level, reaching the same conclusion as the original test.

By comparing only the significance level of the two tests, specify which test, the one at the 10% significance level or the one at the 1% significance level, provides stronger evidence for the conclusion. Justify your answer. **[1]**

© OCR 2024

stemup.co.uk

5 Over a long period of time, it is found that the mean number of mistakes made by a certain player when playing a particular piece of music is 5.

The number of mistakes that the player makes when playing the piece is denoted by the random variable Y.

(a) State **two** assumptions necessary for Y to be modelled by a Poisson distribution. **[2]**

For the remainder of this question you may assume that Y **can** be modelled by a Poisson distribution.

(b) (i) Find the probability that the player makes exactly 3 mistakes when playing the piece. **[1]**

(ii) Find the probability that the player makes fewer than 3 mistakes when playing the piece. **[1]**

(iii) Find the probability that the player makes fewer than 6 mistakes in total when playing the piece twice, assuming that the performances are independent. **[2]**

In a recording studio, the player plays the piece once in the morning and once in the afternoon each day for one week (7 days). It can be assumed that all the performances are independent of each other.

The performances are recorded onto two CDs, one for each of two critics, A and B, to review. The critics are interested in the total number of mistakes made by the player **per day**. Unfortunately, there is a recording error in one of the CDs; on this CD, every piece that is supposed to be an afternoon recording is in fact just a repeat of that morning's recording.

The random variables M_1 and M_2 represent the total number of mistakes per day for the correctly recorded CD and for the wrongly recorded CD respectively.

(c) By considering the values of $\mathrm{E}(M_1)$ and $\mathrm{E}(M_2)$ explain why it is not possible to use the mean number of mistakes per day on the CDs to determine which critic received the wrongly recorded CD. **[2]**

Each critic counts the total number of mistakes made **per day**, for each of the 7 days of recordings on their CD. Summary data for this is given below.

Critic A: $n = 7, \quad \Sigma x_{\mathrm{A}} = 70, \quad \Sigma x_{\mathrm{A}}^2 = 812$

Critic B: $n = 7, \quad \Sigma x_{\mathrm{B}} = 72, \quad \Sigma x_{\mathrm{B}}^2 = 800$

(d) By considering the values of $\mathrm{Var}(M_1)$ and $\mathrm{Var}(M_2)$ determine which critic is likely to have received the wrongly recorded CD. **[4]**

stemup.co.uk

6 The probability distribution of a discrete random variable, X, is shown in the table below.

x	0	1	2
$P(X = x)$	$1 - a - b$	a	b

(a) Find $E(X)$ in terms of a and b. **[1]**

(b) (i) In the case where $E(X) = a + 0.4$, find an expression for $\text{Var}(X)$ in terms of a. **[3]**

(ii) In this case, show that the greatest possible value of $\text{Var}(X)$ is 0.65. You must state the associated value of a. **[2]**

(c) You are now given instead that $E(X)$ is not known.

(i) State the least possible value of $\text{Var}(X)$. **[1]**

(ii) Give all possible pairs of values of a and b which give the least possible value of $\text{Var}(X)$ stated in part **(c)(i)**. **[2]**

END OF QUESTION PAPER

© OCR 2024

stemup.co.uk

BLANK PAGE

© OCR 2024

stemup.co.uk

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

stemup.co.uk

Surname	StemUp Model Solution
Answer booklet	Write the question number in the left-hand margin at the start of each answer.

Question number

1 (a) The probability that a penalty kick does not result in a goal is constant (0.15) for every kick.

1 (b) $X \sim Geo(p)$ with $p = P(\text{goal not scored}) = 1 - 0.85 = 0.15$

$E(X) = \frac{1}{p} = \frac{1}{0.15} = \frac{20}{3} = 6.67$ (3 s.f.)

$Var(X) = \frac{1-p}{p^2} = \frac{0.85}{0.15^2} = \frac{340}{9} = 37.8$ (3 s.f.)

1 (c) Exactly 3 kicks: goal, goal, then no goal

$P(X = 3) = 0.85^2 \times 0.15 = 0.108375 = 0.108$ (3 s.f.)

1 (d) At least 5 kicks means the first 4 kicks all score

$P(X \geqslant 5) = P(X > 4) = 0.85^4 = 0.52200... = 0.522$ (3 s.f.)

2 (a) Discrete uniform distribution on $\{1, 2, ..., 12\}$

$X \sim U(12)$, with $P(X = r) = \frac{1}{12}$ for $r = 1, 2, ..., 12$

2 (b) $E(X) = \frac{n+1}{2} = \frac{12+1}{2} = 6.5$

$Var(X) = \frac{n^2-1}{12} = \frac{12^2-1}{12} = \frac{143}{12} = 11.9$ (3 s.f.)

Question number

2 (c) $\mu = 6.5, \quad \sigma = \sqrt{\frac{143}{12}} = 3.4520...$

$\left|\frac{2(X-\mu)}{\sigma}\right| > 1 \Rightarrow X - \mu > \frac{\sigma}{2}$ or $X - \mu < -\frac{\sigma}{2}$

$X > 6.5 + \frac{3.4520...}{2} = 8.226...$ or $X < 6.5 - \frac{3.4520...}{2} = 4.773...$

So $X = 1, 2, 3, 4$ or $X = 9, 10, 11, 12$: 8 values

$P\left(\left|\frac{2(X-\mu)}{\sigma}\right| > 1\right) = \frac{8}{12} = \frac{2}{3}$

3 (a) Both variables can be considered to be random.

3 (b) Compared with the rest of the data, city A has a lower average life expectancy than would be expected for its average annual income (high income, low life expectancy).

3 (c) Neither variable is controlled: the data are random on random, and the scatter diagram is roughly elliptical in shape, so the data appear consistent with a bivariate Normal distribution. So a test using the pmcc is appropriate.

3 (d) $S_{xx} = 14\ 363\ 849\ 200 - \frac{751\ 120^2}{44} = 1\ 541\ 547\ 964\ (1.54 \times 10^9)$

$S_{yy} = 133\ 014.63 - \frac{2397.1^2}{44} = 2421.7116...$

$S_{xy} = 42\ 465\ 962 - \frac{751\ 120 \times 2397.1}{44} = 1\ 545\ 285.82\ (1.55 \times 10^6)$

Question number

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{1\,545\,285.82}{\sqrt{1\,541\,547\,964 \times 2421.7116}}$$

$r = 0.7998... = 0.800$ (3 s.f.)

3 (e) H_0: $\rho = 0$, H_1: $\rho > 0$

where ρ is the population product moment correlation coefficient between average annual income and average life expectancy.

$n = 44$, 1% level, one-tailed: critical value = 0.3496

$0.800 > 0.3496$, so the result is significant: reject H_0.

There is sufficient evidence at the 1% level to suggest that there is positive correlation between average annual income and average life expectancy in the population.

4 (a) A random sample is likely to be representative of the population, so valid (unbiased) inferences can be drawn from it.

4 (b) $\text{Expected} = \frac{\text{row total} \times \text{column total}}{800}$

Brown: $\frac{432 \times 127}{800} = 68.58$

Blue: $\frac{280 \times 127}{800} = 44.45$

Green: $\frac{88 \times 127}{800} = 13.97$

4 (c) H_0: there is no association between hair colour and eye colour

H_1: there is some association between hair colour and eye colour

Question number	
	Degrees of freedom $\nu = (4 - 1)(3 - 1) = 6$
	Critical value χ^2_6 at 10% = 10.64
	28.62 > 10.64, so the result is significant: reject H_0.
	There is sufficient evidence at the 10% level to suggest that there is some association between natural hair colour and natural eye colour.
4 (d)	Contribution $= \frac{(19 - 13.97)^2}{13.97} = 1.811$ (4 s.f.)
4 (e)	Brown eyes/blonde (6.791) and blue eyes/blonde (6.162) are the two largest contributions, so these cells differ most from what would be expected if there were no association.
	Observed 61 > expected 44.45: more people with blonde hair have blue eyes than would be expected.
	Observed 47 < expected 68.58: fewer people with blonde hair have brown eyes than would be expected.
4 (f)	The test at the 1% level: the probability of rejecting H_0 when it is true is lower, so a test statistic exceeding the critical value is less likely to be due to chance, giving stronger evidence.
5 (a)	Mistakes occur independently of each other.
	Mistakes occur singly and at a constant average rate.
5(b)(i)	$Y \sim Po(5)$
	$P(Y = 3) = e^{-5} \times \frac{5^3}{3!} = 0.14037... = 0.140$ (3 s.f.)

Question number	
5(b)(ii)	$P(Y < 3) = P(Y \leq 2) = e^{-5}(1 + 5 + \frac{5^2}{2}) = 0.12465\ldots = 0.125$ (3 s.f.)
5(b)(iii)	Total for two independent performances ~ $Po(5 + 5) = Po(10)$
	$P(T < 6) = P(T \leq 5) = 0.06708\ldots = 0.0671$ (3 s.f.)
5 (c)	Let Y_M and Y_A be the morning and afternoon mistakes.
	$M_1 = Y_M + Y_A$: $E(M_1) = E(Y_M) + E(Y_A) = 5 + 5 = 10$
	$M_2 = 2Y_M$: $E(M_2) = 2E(Y_M) = 2 \times 5 = 10$
	$E(M_1) = E(M_2) = 10$, so both CDs have the same expected mean
	per day, so the means cannot tell us which CD is wrong.
5 (d)	$Var(M_1) = Var(Y_M) + Var(Y_A) = 5 + 5 = 10$ (independent)
	$Var(M_2) = Var(2Y_M) = 2^2\ Var(Y_M) = 4 \times 5 = 20$
	Critic A: $s^2 = \frac{1}{6}(812 - \frac{70^2}{7}) = \frac{112}{6} = 18.7$ (3 s.f.)
	Critic B: $s^2 = \frac{1}{6}(800 - \frac{72^2}{7}) = \frac{416}{42} = 9.90$ (3 s.f.)
	A: 18.7 is close to 20 = $Var(M_2)$; B: 9.90 is close to 10 = $Var(M_1)$
	So critic A is likely to have received the wrongly recorded CD.
6 (a)	$E(X) = 0(1 - a - b) + 1(a) + 2(b) = a + 2b$
6(b)(i)	$a + 2b = a + 0.4 \Rightarrow b = 0.2$
	$E(X^2) = 1^2(a) + 2^2(0.2) = a + 0.8$
	$Var(X) = E(X^2) - [E(X)]^2 = a + 0.8 - (a + 0.4)^2$
	$= a + 0.8 - a^2 - 0.8a - 0.16$

Question number	
	$\text{Var}(X) = -a^2 + 0.2a + 0.64$
6(b)(ii)	$\text{Var}(X) = -(a^2 - 0.2a) + 0.64$
	$= -[(a - 0.1)^2 - 0.01] + 0.64$
	$= 0.65 - (a - 0.1)^2$
	$(a - 0.1)^2 \geq 0$, so the greatest value is when $a = 0.1$
	Greatest possible $\text{Var}(X) = 0.65$, when $a = 0.1$, as required
6(c)(i)	Least possible $\text{Var}(X) = 0$
6(c)(ii)	$\text{Var}(X) = 0$ only when X takes a single value with probability 1,
	so the probabilities $(1 - a - b, a, b)$ are $(1, 0, 0)$, $(0, 1, 0)$ or $(0, 0, 1)$
	$(a, b) = (0, 0)$, $(1, 0)$ or $(0, 1)$