OCR
Oxford Cambridge and RSA

MODEL ANSWERS

Thursday 15 October 2020 – Afternoon

A Level Further Mathematics A

Y542/01 Statistics

Time allowed: 1 hour 30 minutes

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

831934749 5

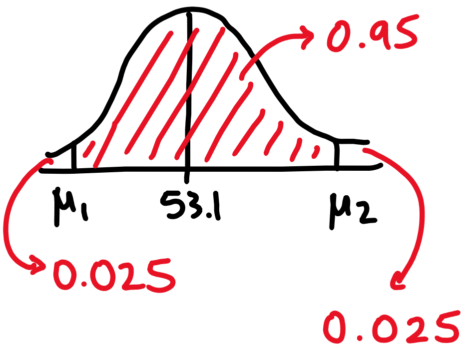
INSTRUCTIONS
- Use black ink. You can use an HB pencil, but only for graphs and diagrams.
- Write your answer to each question in the space provided in the **Printed Answer Booklet**. If you need extra space use the lined pages at the end of the Printed Answer Booklet. The question numbers must be clearly 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.

1 The continuous random variable X has the distribution $N(\mu, 30)$. The mean of a random sample of 8 observations of X is 53.1.

Determine a 95% confidence interval for μ. You should give the end points of the interval correct to 4 significant figures. **[4]**



$\mu = 53.1$

$\sigma = \sqrt{\frac{30}{8}} = \frac{\sqrt{15}}{2}$

$X \sim N\left(53.1, \frac{30}{8}\right)$

For $p = 0.025$, $z = 1.96$.

$\therefore \mu_1 = 53.1 + 1.96\sqrt{\frac{30}{8}} = 56.895\ldots \approx 56.90$ (4sf)

$\therefore \mu_2 = 53.1 - 1.96\sqrt{\frac{30}{8}} = 49.304\ldots \approx 49.30$ (4sf)

$\boxed{\therefore (49.30, 56.90)}$

2 A book collector compared the prices of some books, £x, when new in 1972 and the prices of copies of the same books, £y, on a second-hand website in 2018.
The results are shown in Table 1 and are summarised below the table.

(b.)

Book	A	B	C	D	E	F	G	H	I	~~J~~	K	L
x	0.95	0.65	0.70	0.90	0.55	1.40	1.50	0.50	1.15	0.35	0.20	0.35
y	6.06	7.00	2.00	5.87	4.00	5.36	7.19	2.50	3.00	8.29	1.37	2.00

Table 1

$$n = 12, \sum x = 9.20, \sum y = 54.64, \sum x^2 = 8.9950, \sum y^2 = 310.4572, \sum xy = 46.0545$$

(a) It is given that the value of Pearson's product-moment correlation coefficient for the data is 0.381, correct to 3 significant figures.

pmcc

(i) State what this information tells you about a scatter diagram illustrating the data. **[1]**

(ii) Test at the 5% significance level whether there is evidence of positive correlation between prices in 1972 and prices in 2018. **[5]**

(b) The collector noticed that the second-hand copy of book J was unusually expensive and he decided to ignore the data for book J. → Anomaly

Calculate the value of Pearson's product-moment correlation coefficient for the other 11 books. **[2]**

(a.) (i) The points do not lie very close to a straight line.

(ii) $H_0: \rho = 0$

$H_1: \rho > 0$

ρ is the population pmcc between prices in 1972 and prices in 2018.

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

$$S_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 4.16383\dot{3}$$

$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 1.941\dot{6}$$

$$S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} = 61.663\ldots$$

$\therefore r = 0.381$ (3sf)

pmcc at 5% for $n=12$: 0.4973

$0.381 < 0.4973$ $\therefore$ Accept H_0. Reject H_1.

There is insufficient evidence of positive correlation between prices in the two years.

(b.) Updated stats:

$n = 11$

$\Sigma x = 9.2 - 0.35 = 8.85$

$\Sigma y = 54.64 - 8.29 = 46.35$

$\Sigma x^2 = 8.995 - 0.35^2 = 8.8725$

$\Sigma y^2 = 310.4572 - 8.29^2 = 241.7331$

$\Sigma xy = 46.0545 - (0.35)(8.29) = 43.153$

$S_{xy} = 5.8623\ldots$

$S_{xx} = 1.752\dot{2}\dot{7}$

$S_{yy} = 46.4310\ldots$

$\boxed{\therefore r = 0.6499\ldots \approx 0.650 \text{ (3sf)}}$

3 Jo can use either of two different routes, A or B, for her journey to school. She believes that route A has shorter journey times. She measures how long her journey takes for 17 journeys by route A and 12 journeys by route B. She ranks the 29 journeys in increasing order of time taken, and she finds that the sum of the ranks of the journeys by route B is 219.

(a) Test at the 10% significance level whether route A has shorter journey times than route B. **[8]**

(b) State an assumption about the 29 journeys which is necessary for the conclusion of the test to be valid. **[1]**

(a.) m_A: Median journey time for Route A.
m_B: Median journey time for Route B.

$H_0: m_A = m_B$

$H_1: m_A < m_B$

Wilcoxon Rank-Sum Test: $m = 12, n = 17$

$\mu = \frac{1}{2}m(m+n+1) = 180$

$\sigma^2 = \frac{1}{12}mn(m+n+1) = 510$

$\Rightarrow W \sim N(180, 510)$

$R_m = 219$, $m(m+n+1) - R_m = 141$

$p = \phi\left(\frac{141.5 - 180}{\sqrt{150}}\right) = 0.0441\ldots < 0.1$

$\therefore$ Reject H_0. Accept H_1.

$\therefore$ There is significant evidence that route B takes longer.

(b.) There must be a random sample of all journeys.

OR: Distribution must be same shape.

4 The random variable X is equally likely to take any of the n integer values from $m+1$ to $m+n$ inclusive. It is given that $\mathrm{E}(3X) = 30$ and $\mathrm{Var}(3X) = 36$.

Determine the value of m and the value of n. **[7]**

$E(3X) = 3E(X) = 30 \Rightarrow E(X) = 10$

$3^2 \times \mathrm{Var}(X) = 36 \Rightarrow \mathrm{Var}(X) = 4$

Variation of Uniform Distribution $= \frac{1}{12}(b-a)^2$

$\frac{1}{\cancel{12}}(n^2-1) = 4$

$\times \cancel{12}$ $\quad$ $\times 12$

$n^2 - 1 = 48$

$n^2 = 49$

$\therefore n = \sqrt{49} = 7$

$E(X-m) = \frac{1}{2}(n+1)$

$10 - m = \frac{1}{2}(7+1)$

$10 - m = 4$

$\therefore m = 10 - 4 = 6$

$\boxed{\therefore m = 6,\ n = 7}$

5 26 cards are each labelled with a different letter of the alphabet, A to Z. The letters A, E, I, O and U are vowels.

$\frac{5}{26}$ vowels

(a) Five cards are selected at random without replacement.

Determine the probability that the letters on at least three of the cards are vowels. **[4]**

(b) All 26 cards are arranged in a line, in random order.

(i) Show that the probability that all the vowels are next to one another is $\frac{1}{2990}$. **[3]**

(ii) Determine the probability that three of the vowels are next to each other, and the other two vowels are next to each other, but the five vowels are not all next to each other. **[4]**

(a.) $P(X \geq 3) = P(X=3) + P(X=4) + P(X=5)$

$P(X=3) = \frac{5}{26} \times \frac{4}{25} \times \frac{3}{24} \times \frac{21}{23} \times \frac{20}{22} \times 10 = \frac{105}{3289}$

$P(X=4) = \frac{5}{26} \times \frac{4}{25} \times \frac{3}{24} \times \frac{2}{23} \times \frac{21}{22} \times 5 = \frac{21}{13156}$

$P(X=5) = \frac{5}{26} \times \frac{4}{25} \times \frac{3}{24} \times \frac{2}{23} \times \frac{1}{22} = \frac{1}{65780}$

$\boxed{\therefore P(X \geq 3) = \frac{1103}{32890} \approx 0.0335 \text{ (3sf)}}$

(b.) (i) $\frac{22! \times 5!}{26!} = \boxed{\frac{1}{2990}}$

(ii) 22 fences: 22 for [VVV] x 21 for [VV]

$22 \times 21 = 462$

Consonants arranged in 21! ways.

Vowels arranged in 5! ways.

$\frac{21! \times 5!}{26!} \times 462 = \boxed{\frac{21}{2990}}$

6 The numbers of CD players sold in a shop on three consecutive weekends were 7, 6 and 2. It may be assumed that sales of CD players occur randomly and that nobody buys more than one CD player at a time. The number of CD players sold on a randomly chosen weekend is denoted by X.

(a) How appropriate is the Poisson distribution as a model for X? **[2]**

Now assume that a Poisson distribution with mean 5 is an appropriate model for X.

(b) Find

(i) $\mathrm{P}(X = 6)$, **[2]**

(ii) $\mathrm{P}(X \geqslant 8)$. **[2]**

The number of integrated sound systems sold in a weekend at the same shop can be assumed to have the distribution $\mathrm{Po}(7.2)$.

(c) Find the probability that on a randomly chosen weekend the total number of CD players and integrated sound systems sold is between 10 and 15 inclusive. **[3]**

(d) State an assumption needed for your answer to part **(c)** to be valid. **[1]**

(e) Give a reason why the assumption in part **(d)** may not be valid in practice. **[1]**

(a.) Events occur independently and at constant average rate.

(b.) $\lambda = 5$

$X \sim Po(5)$

(i) $P(X=6) = \dfrac{e^{-5}(5)^6}{6!} = 0.14622... \approx \boxed{0.146\ (3sf)}$

(ii) $P(X \geq 8) = 1 - P(X \leq 7)$

$= 1 - 0.86662...$

$= 0.1333...$

$\approx \boxed{0.133\ (3sf)}$

(c.) CD Players: $X \sim Po(5)$

Integrated Sound System: $Y \sim Po(7.2)$

$5 + 7.2 = 12.2$

$X + Y \sim Po(12.2)$

$$P(10 \leq X+Y \leq 15) = P(X+Y \leq 15) - P(X+Y \leq 9)$$
$$= 0.82957... - 0.22535...$$
$$= 0.60422...$$
$$\approx \boxed{0.604 \text{ (3sf)}}$$

(d.) Sales of CD players and integrated systems need to be independent.

(e.) If a customer buys a CD player, they probably won't or will buy an integrated system as well.

7 A biased spinner has five sides, numbered 1 to 5. Elmer spins the spinner repeatedly and counts the number of spins, X, up to and including the first time that the number 2 appears. He carries out this experiment 100 times and records the frequency f with which each value of X is obtained. His results are shown in Table 1, together with the values of xf.

x	1	2	3	4	5	6	$\geqslant 7$	Total
Frequency f	20	15	9	13	10	10	23	100
xf	20	30	27	52	50	60	161	400

Table 1

(a) State an appropriate distribution with which to model X, determining the value(s) of any parameter(s). **[3]**

Elmer carries out a goodness-of-fit test, at the 5% level, for the distribution in part **(a)**. Table 2 gives some of his calculations, in which numbers that are not exact have been rounded to 3 decimal places.

x	1	2	3	4	5	6	$\geqslant 7$
Observed frequency O	20	15	9	13	10	10	23
Expected frequency E	25	18.75	14.063	10.547	7.910	5.933	17.798
$(O-E)^2/E$	1	0.75	1.823	0.571	0.552	2.789	1.520

Table 2

(b) Show how the expected frequency corresponding to $x \geqslant 7$ was obtained. **[2]**

(c) Carry out the test. **[5]**

(a.) Geometric Distribution

$$\text{Mean} = \frac{400}{100} = 4 = \frac{1}{p} \Rightarrow \therefore p = \frac{1}{4}$$

$\boxed{\therefore X \sim \text{Geo}(0.25)}$

(b.) $P(X \geq 7) = 1 - P(X \leq 6) = (1-0.25)^6 = 0.1779...$

$\boxed{\text{Expected Frequency} = 0.1779... \times 100 = 17.798}$

(c) H_0: Data consistent with geometric distribution.

H_1: Data not consistent.

$$\chi^2 = \sum \frac{(O - E)^2}{E} = 9.005$$

Degrees of Freedom, $v = 5$

$\chi^2(5\%, v = 5) = 11.07$

$9.005 < 11.07$ $\therefore$ Accept H_0. Reject H_1.

$\therefore$ There is insufficient evidence that a geometric distribution is not a good fit.

8 The continuous random variable X has probability density function

$$f(x) = \begin{cases} \frac{k}{x^n} & x \geqslant 1, \\ 0 & \text{otherwise,} \end{cases}$$

where n and k are constants and n is an integer greater than 1.

(a) Find k in terms of n. **[3]**

(b) **(i)** When $n = 4$, find the cumulative distribution function of X. **[3]**

(ii) Hence determine $P(X > 7 \mid X > 5)$ when $n = 4$. **[4]**

(c) Determine the values of n for which $\text{Var}(X)$ is not defined. **[5]**

(a.) $\int_1^\infty kx^{-n}\,dx = \left[\frac{k}{(1-n)x^{n-1}}\right]_1^\infty$

$= \frac{k}{n-1}$

$\frac{k}{n-1} = 1 \Rightarrow \boxed{k = n-1}$

(b.) i) $\int 3x^{-4}\,dx = \frac{3x^{-3}}{-3} + c = -\frac{1}{x^3} + c$

$x = 1, F(x) = 0 \Rightarrow -\frac{1}{1^3} + c = 0$

$-1 + c = 0$

$\therefore c = 1$

$$\boxed{\therefore F(x) = \begin{cases} 0 & x < 1 \\ 1 - \frac{1}{x^3} & x \geq 1 \end{cases}}$$

(ii) $P(X>7 \mid X>5) = \dfrac{P[(X>7) \cap (X>5)]}{P(X>5)}$

$$= \frac{P(X>7)}{P(X>5)}$$

$$= \frac{1-F(7)}{1-F(5)}$$

$$= \frac{1}{343} \div \frac{1}{125}$$

$$= \boxed{\frac{125}{343}}$$

(c) $E(X^2) = \int_1^\infty k x^{2-n}\, dx = \left[\dfrac{k x^{3-n}}{3-n}\right]_1^\infty \quad (n \neq 3)$

If $n=3$, $E(X^2) = \lim\limits_{x \to \infty} [2\ln(x)]$.

Infinite integral doesn't converge if $3-n \geq 0$.

If $n \geq 4$ then $E(X) = \left[\dfrac{k x^{2-n}}{2-n}\right]$ converges.

$\therefore$ Var(X) is not defined if and only if $n=2$ or $n=3$.