Model Solutions

OCR
Oxford Cambridge and RSA

Thursday 13 June 2019 – Afternoon

A Level Further Mathematics A

Y542/01 Statistics

Time allowed: 1 hour 30 minutes

You must have:
- Printed Answer Booklet
- Formulae A Level Further Mathematics A

You may use:
- a scientific or graphical calculator

* 7 7 7 1 7 4 9 7 5 7 *

INSTRUCTIONS
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Answer **all** the questions.
- **Write your answer to each question in the space provided in the Printed Answer Booklet.** If additional space is required, you should use the lined page(s) at the end of the Printed Answer Booklet. The question number(s) must be clearly shown.
- You are permitted to use a scientific or graphical calculator in this paper.
- 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 $g\,\text{m}\,\text{s}^{-2}$. Unless otherwise instructed, when a numerical value is needed, use $g = 9.8$.

INFORMATION
- The total mark for this paper is **75**.
- The marks for each question are shown in brackets **[ ]**.
- **You are reminded of the need for clear presentation in your answers.**
- The Printed Answer Booklet consists of **16** pages. The Question Paper consists of **4** pages.

Answer **all** the questions.

1 A set of bivariate data (X, Y) is summarised as follows.

$n = 25, \sum x = 9.975, \sum y = 11.175, \sum x^2 = 5.725, \sum y^2 = 46.200, \sum xy = 11.575$

(a) Calculate the value of Pearson's product-moment correlation coefficient. **[1]**

(b) Calculate the equation of the regression line of y on x. **[2]**

It is desired to know whether the regression line of y on x will provide a reliable estimate of y when $x = 0.75$.

(c) State one reason for believing that the estimate will be reliable. **[1]**

(d) State what further information is needed in order to determine whether the estimate is reliable. **[1]**

a) using $S_{xx} = \Sigma x^2 = \frac{(\Sigma x)^2}{n}$

$S_{yy} = \Sigma y^2 - \frac{(\Sigma y)^2}{n}$

$S_{xy} = \Sigma xy - \frac{\Sigma x \Sigma y}{n}$

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

$\therefore$ correlation coefficient = 0.8392 (4dp)

b) $(y - \bar{y}) = \frac{S_{xy}}{S_{xx}}(x - \bar{x})$

$\therefore$ $y = -1.180 + 4.0781x$

c) r is high so points lie to line.

d) If $x = 0.75$ is within the data range.

2 The average numbers of cars, lorries and buses passing a point on a busy road in a period of 30 minutes are 400, 80 and 17 respectively.

(a) Assuming that the numbers of each type of vehicle passing the point in a period of 30 minutes have independent Poisson distributions, calculate the probability that the total number of vehicles passing the point in a randomly chosen period of 30 minutes is at least 520. **[3]**

(b) Buses are known to run in approximate accordance with a fixed timetable.

Explain why this casts doubt on the use of a Poisson distribution to model the number of buses passing the point in a fixed time interval. **[1]**

a) Let X be average number of vehicles passing a point in 30 minutes. $(X = 400 + 80 + 17 = 497)$

$X \sim Po(497)$

$P(X \geq 520) = 1 - P(X \leq 519)$

$= 0.1564$ (4dp)

b) Occurance of a bus is not a random event if it runs on schedule.

3 Six red counters and four blue counters are arranged in a straight line in a random order.

Find the probability that

(a) no blue counter has fewer than two red counters between it and the nearest other blue counter, [3]

(b) no two blue counters are next to one another. [3]

a) B R R B R R B R R B

$$\frac{4! \times 6!}{10!} = \frac{24 \times 720}{3628800} = \frac{1}{210}$$

b) B R B R B R B R B R B R B

7 possible positions for 4 red counters.

$$\frac{{}^7C_4 \times 6! \times 4!}{10!} = \frac{35 \times 720 \times 24}{3628800} = \frac{1}{6}$$

4 The greatest weight W N that can be supported by a shelving bracket of traditional design is a normally distributed random variable with mean 500 and standard deviation 80.

A sample of 40 shelving brackets of a new design are tested and it is found that the mean of the greatest weights that the brackets in the sample can support is 473.0 N.

(a) Test at the 1% significance level whether the mean of the greatest weight that a bracket of the new design can support is less than the mean of the greatest weight that a bracket of the traditional design can support. [7]

(b) State an assumption needed in carrying out the test in part (a). [1]

(c) Explain whether it is necessary to use the central limit theorem in carrying out the test. [1]

a) $H_0 : \mu = 500$
$H_1 : \mu < 500$

let μ be the mean of the greatest weight that the new design can support

$\bar{X} \sim N\left(500, \frac{80^2}{40}\right)$

$P(\bar{X} < 473) = 0.0164$ (4dp)

As $0.0164 > 0.01$

Do not reject H_0. Insufficient evidence that the greatest weight that the new design can support is less than the greatest weight that the original design can support.

b) Standard deviation remains unchanged.

c) It is, as we do not know that the distribution of weights for the new design is 'normal'.

OR

It is not, as the population distribution is known to be 'normal'.

5 Five runners, A, B, C, D and E, take part in two different races.

Spearman's rank correlation coefficient for the orders in which the runners finish is calculated and a test for positive agreement is carried out at the 5% significance level.

(a) State suitable hypotheses for the test. **[1]**

(b) Find the largest possible value of $\sum d^2$ for which the result of the test is to reject the null hypothesis. **[3]**

(c) In the first race, the order in which the five runners finished was: A, B, C, D, E. In the second race, three of the runners finished in the same positions as in the first race. The result of the test is to reject the null hypothesis.

Find a possible order for the runners to finish in the second race. **[3]**

a) H_0: No association between orders in races of this type.
H_1: Association between orders in races of this type.

b) $1 - \dfrac{6\sum d^2}{5 \times 24} \geqslant 0.9$ $\left(r_s = 1 - \dfrac{6\sum d^2}{n(n^2+1)} \right)$

$0.1 \geqslant \dfrac{6\sum d^2}{5 \times 24}$

$2 \geqslant \sum d^2$ $\Rightarrow$ Largest possible value is 2.

c) As $\sum d^2 = 2$
$\Rightarrow d = 0, 0, 0, 1, -1$

$\therefore$ a possible order for the runners to finish is: ABCED
(Any condition where 2 neighbouring letters are swapped).

6 Yusha investigates the proportion of left-handed people living in two cities, A and B. He obtains data from random samples from the two cities. His results are shown in the table, in which L denotes "left-handed".

	L	L'	Total
A	14	9	23
B	26	51	77
Total	40	60	100

(a) Test at the 10% significance level whether there is association between being left-handed and living in a particular city. **[7]**

A person is chosen at random from one of the cities A and B.
Let A denote "the person lives in city A".

(b) State the relationship between $P(L)$ and $P(L|A)$ according to the model implied by the null hypothesis of your test. **[1]**

(c) Use the data in the table to suggest a value for $P(L|A)$ given by an improved model. **[2]**

H_0: No association between city and description of handedness.
H_1: Association between city and description of handedness.

O_i	14	9	26	51
E_i	$\frac{(23\times 40)}{100} = 9.2$	$\frac{(23\times 60)}{100} = 13.8$	$\frac{(77\times 40)}{100} = 30.8$	$\frac{(77\times 60)}{100} = 46.2$
$\frac{(\lvert O_i-E_i\rvert-0.5)^2}{E_i}$	2.0097	1.3398	0.6003	0.4002

$$\sum_i \frac{(\lvert O_i-E_i\rvert-0.5)^2}{E_i} = 4.35 \ (3sf)$$

Degrees of freedom = $(2-1)(2-1) = 1$

$\chi^2_1(0.1) = 2.705$

As $4.35 > 2.705$
Reject H_0. Sufficient evidence to suggest association between city and description of handedness.

b) $P(L|A) = P(L)$

c) $P(L\cap A) = \frac{14}{100}$

$P(A) = \frac{23}{100}$

$$\therefore P(L|A) = \frac{P(L\cap A)}{P(A)} = \frac{\frac{14}{100}}{\frac{23}{100}} = \frac{14}{23}$$

7 The random variable D has the distribution Geo(p). It is given that $\mathrm{Var}(D) = \frac{40}{9}$.

Determine

(a) $\mathrm{Var}(3D+5)$, **[1]**

(b) $\mathrm{E}(3D+5)$, **[6]**

(c) $\mathrm{P}(D > \mathrm{E}(D))$. **[3]**

a) $\mathrm{Var}(3D+5) = 3^2 \mathrm{Var}(D)$
$= 9 \times \frac{40}{9}$
$= 40$

b) $\frac{1-p}{p^2} = \frac{40}{9}$

$40p^2 = 9 - 9p$
$40p^2 + 9p - 9 = 0$
$p = \frac{3}{8}$ $\quad p = -\frac{3}{5}$ (reject)

$E(D) = \frac{1}{p} = \frac{8}{3}$

$E(3D+5) = 3E(D) + 5$
$= 3 \times \frac{8}{3} + 5$
$= 13$

c) $P(D > E(D)) = P(D \geq 3)$
$= (1-p)^2$
$= \left(1 - \frac{3}{8}\right)^2$
$= \frac{25}{64}$

8 A university course was taught by two different professors. Students could choose whether to attend the lectures given by Professor Q or the lectures given by Professor R. At the end of the course all the students took the same examination.

The examination marks of a random sample of 30 students taught by Professor Q and a random sample of 24 students taught by Professor R were ranked. The sum of the ranks of the students taught by Professor Q was 726.

Test at the 5% significance level whether there is a difference in the ranks of the students taught by the two professors. **[10]**

Let m_Q and m_R represent the medians of rankings given to Q and R.

$H_0: m_Q = m_R$

$H_1: m_Q \neq m_R$

Sum of ranks $= \frac{1}{2} \times 54 \times 55$

$= 1485$

$R_m = 1485 - 726 = 759$

R_m is seen as 'normal' $\left(W \sim N\left(\frac{1}{2}m(m+n+1), \frac{1}{12}mn(m+n+1)\right)\right)$

mean $= \frac{1}{2} \times 24 \times 55$

$= 660$

$R_m \sim N(660, 3300)$

$P(R_m \geq 759) = 0.0432$ (3sf)

two-tailed test:

$\frac{0.05}{2} = 0.025$

As $0.0432 > 0.025$

Do not reject. Insufficient evidence to suggest a difference between the ranks.

9 The continuous random variable T has cumulative distribution function

$$F(t) = \begin{cases} 0 & t < 0, \\ 1 - e^{-0.25t} & t \geqslant 0. \end{cases}$$

(a) Find the cumulative distribution function of $2T$. **[3]**

(b) Show that, for constant k, $E(e^{kt}) = \frac{1}{1-4k}$.

You should state with a reason the range of values of k for which this result is valid. **[7]**

(c) T is the time before a certain event occurs.

Show that the probability that no event occurs between time $T = 0$ and time $T = \theta$ is the same as the probability that the value of a random variable with the distribution $Po(\lambda)$ is 0, for a certain value of λ. You should state this value of λ in terms of θ. **[4]**

END OF QUESTION PAPER

a) Let $H(x)$ be the CDF of $2T$.

$H(x) = P(X \leq x) = P(2T \leq x)$

$= P(T \leq \frac{x}{2})$

$= F(\frac{x}{2})$

$= 1 - e^{-0.25 \times \frac{x}{2}}$

$= 1 - e^{-0.125x}$

b) $f(x) = 0.25e^{-0.25t}$

$E(e^{kt}) = \int_0^\infty 0.25e^{kt}e^{-0.25t}\,dt$

$\int_0^N 0.25e^{-(0.25-k)t}\,dt = \left[-\frac{0.25e^{-(0.25-k)t}}{0.25-k}\right]_0^N$

$= \frac{0.25e^{-(0.25-k)N}}{0.25-K} + \frac{0.25}{0.25-K}$

As first term will converge for $K < 0.25$

Then $\lim_{N \to 0} \frac{0.25e^{-(0.25-k)N}}{0.25-K} = 0$

b continued)

$$\int_0^\infty 0.25e^{-(0.25-K)t}\,dt$$

$$= \lim_{N\to\infty}\left\{-\frac{0.25e^{-(0.25-K)N}}{0.25-K} + \frac{0.25}{0.25-K}\right\}$$

$$= \frac{0.25}{0.25-K} = \frac{1}{1-4K}$$

c) P(no event between 0 and θ)

$= P(T > \theta)$
$= e^{-0.25\theta}$

$P(0)$ from $Po(\lambda) = e^{-\lambda}$

Hence same expression
$\lambda = 0.25\theta$