OCR
Oxford Cambridge and RSA

AS Level Further Mathematics A

Y532/01 Statistics

Thursday 17 May 2018 – Afternoon

Time allowed: 1 hour 15 minutes

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

You may use:
- a scientific or graphical calculator

MODEL SOLUTIONS

INSTRUCTIONS
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Complete the boxes provided on the Printed Answer Booklet with your name, centre number and candidate number.
- Answer **all** the questions.
- **Write your answer to each question in the space provided in the Printed Answer Booklet.** If additional space is required, use the lined page(s) at the end of this booklet. The question number(s) must be clearly shown.
- Do **not** write in the barcodes.
- You are permitted to use a scientific or graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.
- 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 **60**.
- 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 **12** pages. The Question Paper consists of **4** pages.

1 A book reviewer estimates that the probability that he receives a delivery of books to review on any one weekday is 0.1. The first weekday in September on which he receives a delivery of books to review is the Xth weekday of September.

(i) State an assumption needed for X to be well modelled by a geometric distribution. [1]

• Receipt of books is independent.

(ii) Find $P(X = 11)$. [2]

$P(X=11) = 0.9^{10} \times 0.1 = 0.03486$ (4sf)

(iii) Find $P(X \leqslant 8)$. [2]

$P(X \leqslant 8) = 1 - P(X \geqslant 9)$

$= 1 - 0.9^8$

$= 0.5695$ (4sf)

(iv) Find $\text{Var}(X)$. [2]

$\text{Var}(X) = \dfrac{0.9}{0.1^2} = 90$

(v) Give a reason why a geometric distribution might not be an appropriate model for the first weekday in a calendar year on which the reviewer receives a delivery of books to review. [1]

Probabilities different for this period, so not independent.

2 The probability distribution for the discrete random variable W is given in the table.

w	1	2	3	4
$P(W = w)$	0.25	0.36	x	x^2

(i) Show that $\text{Var}(W) = 0.8571$. [7]

Tot Probability = 1

0.2 $.36 + x + x^2 = 1$

$x^2 + x - 0.39 = 0$

$x = 0.3$ $x = -1.3$
(reject as $x > 0$)

$E(W) = (1 \times 0.25) + (2 \times 0.36) + (3 \times 0.3) + (4 \times 0.3^2)$

$= 0.25 + 0.72 + 0.9 + 0.36$

$= 2.23$

$E(W^2) = (1^2 \times 0.25) + (2^2 \times 0.36) + (3^2 \times 0.3) + (4^2 \times 0.3^2)$
$= 0.25 + 1.44 + 2.7 + 1.44$
$= 5.83$

$Var(W) = E(W^2) - (E(W))^2$
$= 5.83 - 2.23^2$
$= \underline{0.8571 \text{ (4sf)}}$

(ii) Find Var$(3W+6)$. **[1]**

$Var(3W+6) = 3^2 \times Var(W)$
$= 9 \times 0.8571$
$= \underline{7.7139}$

3 In the manufacture of fibre optical cable (FOC), flaws occur randomly. Whether any point on a cable is flawed is independent of whether any other point is flawed. The number of flaws in 100 m of FOC of standard diameter is denoted by X.

(i) State a further assumption needed for X to be well modelled by a Poisson distribution. **[1]**

Flaws occur at constant average rate

Assume now that X can be well modelled by the distribution Po(0.7).

(ii) Find the probability that in 300 m of FOC of standard diameter there are exactly 3 flaws. **[2]**

In 300m, $\lambda = 0.7 \times 3 = 2.1$

Assume Y can be modelled by distribution $P_o(2.1)$

$P(Y=3) = \frac{e^{-\lambda} \times \lambda^x}{x!} = \frac{e^{-2.1} \times 2.1^3}{3!} = \underline{0.1890 \text{ (4sf)}}$

The number of flaws in 100 m of FOC of a larger diameter has the distribution Po(1.6).

(iii) Find the probability that in 200 m of FOC of standard diameter and 100 m of FOC of the larger diameter the total number of flaws is at least 4. **[3]**

$X_2 \sim P_o(3)$

$\lambda = (2 \times 0.7) + 1.6$
$= 1.4 + 1.6$
$= 3$

$P(X_2 \geqslant 4) = 1 - P(X_2 \leqslant 3)$
$= 1 - 0.64723$
$= 0.35277$
$= \underline{0.3528 \text{ (4sf)}}$

4 Judith believes that mathematical ability and chess-playing ability are related. She asks 20 randomly chosen chess players, with known British Chess Federation (BCF) ratings X, to take a mathematics aptitude test, with scores Y. The results are summarised as follows.

$$n = 20,\ \Sigma x = 3600,\ \Sigma x^2 = 660500,\ \Sigma y = 1440,\ \Sigma y^2 = 105280,\ \Sigma xy = 260990$$

(i) Calculate the value of Pearson's product-moment correlation coefficient r. [2]

$$S_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{n} = 660500 - \frac{3600^2}{20} = 12500$$

$$S_{yy} = \Sigma y^2 - \frac{(\Sigma y)^2}{n} = 105280 - \frac{1440^2}{20} = 1600$$

$$S_{xy} = \Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n} = 260990 - \frac{3600 \times 1440}{20}$$

$$= 1790$$

$$\therefore r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \frac{1790}{\sqrt{1600 \times 12500}} = \underline{0.400 \text{ (3sf)}}$$

(ii) State an assumption needed to be able to carry out a significance test on the value of r. [1]

Data needs to have a bivariate normal distribution.

(iii) Assume now that the assumption in part (ii) is valid. Test at the 5% significance level whether there is evidence that chess players with higher BCF ratings are better at mathematics. [4]

H_0: No association between maths test score and BCF rating.

H_1: Positive association between Maths test score and BCF rating.

Critical value = 0.3783

0.400 > 0.3783 so reject H_0.

sufficient evidence to suggest higher maths test scores are associated to high BCF rating.

(iv) There are two different grading systems for chess players, the BCF system and the international ELO system. The two sets of ratings are related by

$$\text{ELO rating} = 8 \times \text{BCF rating} + 650.$$

Magnus says that the experiment should have used ELO ratings instead of BCF ratings. Comment on Magnus's suggestion. [1]

This is a linear transformation so does not affect the conclusion.

5 **(i)** A team of 9 is chosen at random from a class consisting of 8 boys and 12 girls. Find the probability that the team contains no more than 3 girls. **[4]**

There are 3 options you could have:

1 girl and 8 boys

2 girls and 7 boys

3 girls and 6 boys

$$\text{Probability} = \frac{({}^{12}C_1 \times {}^{8}C_8) + ({}^{12}C_2 \times {}^{8}C_7) + ({}^{12}C_3 \times {}^{8}C_6)}{{}^{20}C_9}$$

$$= \frac{12 + 528 + 6160}{167960} = \underline{0.03989 \quad (4sf)}$$

(ii) A group of n people, including Mr and Mrs Laplace, are arranged at random in a line. The probability that Mr and Mrs Laplace are placed next to each other is less than 0.1. Find the smallest possible value of n. **[4]**

Treat Mr and Mrs Laplace as one. They can stand in either order. There are now $n-1$ 'people' to arrange, which can be done in $(n-1)!$ ways.

$$\text{Probability} = \frac{2 \times (n-1)!}{n!} < 0.1$$

$$= \frac{2}{n} < 0.1$$

$$n > 20$$

$$\therefore n_{min} = 21$$

6 **In this question you must show detailed reasoning.**

The random variable T has a binomial distribution. It is known that $E(T) = 5.625$ and the standard deviation of T is 1.875. Find the values of the parameters of the distribution. **[5]**

$T \sim B(n, p)$

$E(T) = np$

$np = 5.625$

$Var(T) = np(1-p)$

$SD(T)^2 = np(1-p)$

$1.875^2 = 5.625(1-p)$

$1-p = 0.625$

$p = 0.375$

$n = \frac{5.625}{p} = \frac{5.625}{0.375}$

$n = 15$

7 An environmentalist measures the mean concentration, c milligrams per litre, of a particular chemical in a group of rivers, and the mean mass, m pounds, of fish of a certain species found in those rivers. The results are given in the table.

c	1.94	1.78	1.62	1.51	1.52	1.4
m	6.5	7.2	7.4	7.6	8.3	9.7

(i) State which, if either, of m and c is an independent variable. **[1]**

Neither are an independent variable.

(ii) Calculate the equation of the least squares regression line of c on m. **[3]**

$a = 2.848$

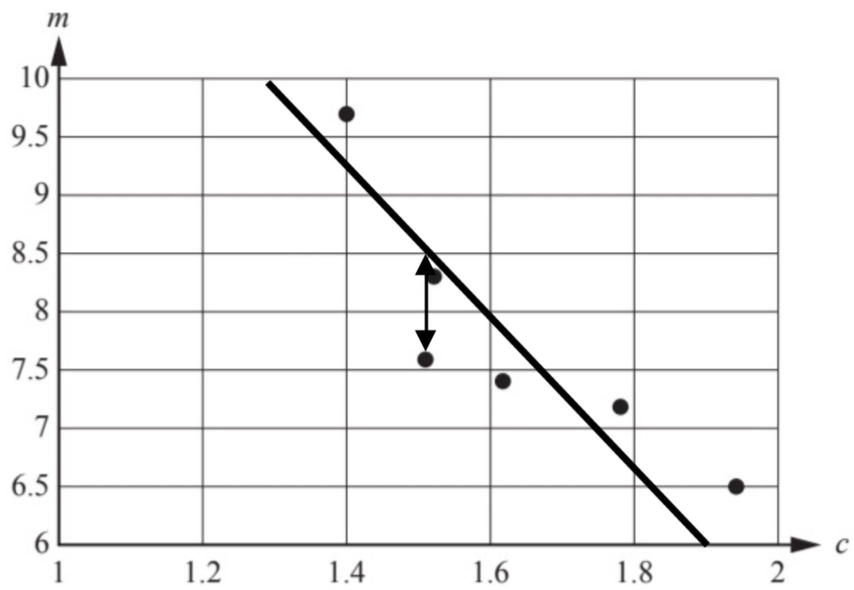
$b = -0.1567$

$\therefore c = 2.848 - 0.1567m$

(iii) State what effect, if any, there would be on your answer to part **(ii)** if the masses of the fish had been recorded in kilograms rather than pounds. (1 kg ≈ 2.2 pounds.) **[1]**

a is unchanged, b could be multiplied by 2.2

(iv) The data is illustrated in the scatter diagram. Explain what is meant by 'least squares', illustrating your answer using the copy of this diagram in the Printed Answer Booklet. **[3]**



The line of best fit minimises the sum of squares of the distances from the line to each point.

8 The table shows the results of a random sample drawn from a population which is thought to have the distribution U(20).

Range	$1 \leqslant x \leqslant 8$	$9 \leqslant x \leqslant 12$	$13 \leqslant x \leqslant 20$
Observed frequency	12	y	$28-y$

Find the range of values of y for which the data are not consistent with the distribution at the 5% significance level. **[9]**

Total people in sample $= 12 + y + 28 - y$
$= 40$

Expected frequencies : $1 \leq x \leq 8$: $\frac{8}{20} \times 40 = 16$

$9 \leq x \leq 12$: $\frac{4}{20} \times 40 = 8$

$13 \leq x \leq 20$: $\frac{8}{20} \times 40 = 16$

Range	$1 \leq x \leq 8$	$9 \leq x \leq 12$	$13 \leq x \leq 20$
Observed frequency	12	y	$28-y$
Expected frequency	16	8	16

$$\sum_i \frac{(O_i - E_i)^2}{E_i} = \frac{(12-16)^2}{16} + \frac{(y-8)^2}{8} + \frac{(28-y-16)^2}{16}$$

$$= 1 + \frac{(y-8)^2}{8} + \frac{(12-y)^2}{16}$$

C.V. = 5.991

For the data to not be consistent with the distribution we need.

$$1 + \frac{(y-8)^2}{8} + \frac{(12-y)^2}{16} > 5.991$$

$$16 + 2(y^2 - 16y + 64) + (144 - 24y + y^2) > 95.856$$

$$3y^2 - 56y + 192.144 > 0$$

Critical values $= \dfrac{56 \pm \sqrt{56^2 - 4(3)(192.144)}}{2(3)}$

$$= \frac{56 \pm \sqrt{830.272}}{6}$$

$= 4.53$ or $= 14.14$

y has to be an integer

$0 \leq y \leq 4$ or $15 \leq y \leq 28$

END OF QUESTION PAPER