



Edexcel AS Further Mathematics

8FM0/24 - Further Statistics 2

Tuesday 24 May 2022

Questions and Fully Worked Model Solutions

This booklet places the **question and the solution together**. The mathematics is set out using proper fractions, powers, roots and exact notation, with extra written explanation so that a student can see *why* each step is being taken.

Solutions have been checked against the Pearson Edexcel Summer 2022 mark scheme. The aim is a clear **full-mark model response**, not merely a list of final answers.

Contents

Question 1	2
Question 2	3
Question 3	4
Question 4	6
Question 5	8

www.mathswithmatt.co.uk

Question 1

Question 1

Abena and Meghan rank the same 10 films from most favourite to least favourite. For the differences d between their ranks, $\sum d^2 = 84$.

(a) Calculate Spearman's rank correlation coefficient. (1)

At the 5% significance level, test for agreement using $H_0 : \rho_S = 0$ and $H_1 : \rho_S > 0$.

(b) (i) Find the critical region. (ii) State the conclusion. (2)

(c) An 11th film is added and both rank it least favourite. Carry out the same 5% test, stating the test statistic and critical value used. (4)

Model solution

(a) Spearman's coefficient is appropriate because the data are ranks. There are no tied ranks, so use the standard rank-correlation formula with $n = 10$.

$$r_S = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} = 1 - \frac{6(84)}{10(10^2 - 1)} = 0.490909 \dots$$

$$r_S = 0.491 \text{ (to 3 decimal places).}$$

(b)(i) The alternative is one-sided and positive, so only sufficiently large positive values of r_S support agreement. From the $n = 10$, 5% critical-values table,

$$\text{critical region: } r_S > 0.5636.$$

$$r_S > 0.5636.$$

(b)(ii) The observed value 0.491 is not in this critical region, so it is not large enough to reject H_0 .

There is insufficient evidence, at the 5% level, of agreement between Abena's and Meghan's ranks.

(c) The extra film has rank 11 for both people, so its rank difference is 0 and $\sum d^2$ remains 84. Recalculate the test statistic with the new number of films, $n = 11$.

$$r_S = 1 - \frac{6(84)}{11(11^2 - 1)} = 0.618181 \dots = 0.618.$$

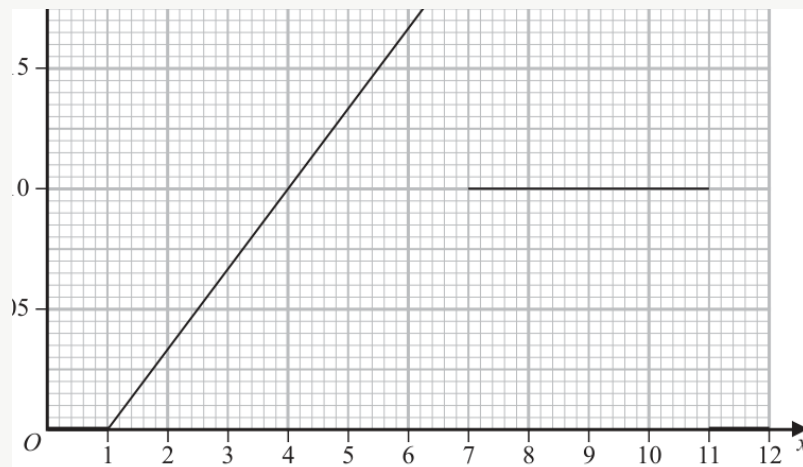
For a one-tailed 5% test with $n = 11$, the critical value is 0.5364. Since $0.618 > 0.5364$, the statistic is in the critical region.

$r_S = 0.618$, critical value 0.5364; reject H_0 . There is sufficient evidence of agreement between their film ranks.

Question 2

Question 2

The graph is the probability density function of a continuous random variable X .



(a) Find $P(X < 4)$.

(2)

(b) Specify the cumulative distribution function of X for $7 \leq x \leq 11$.

(3)

(a) Find $P(X < 4)$.

(2)

(b) Specify the cumulative distribution function of X for $7 \leq x \leq 11$.

(3)

Model solution

(a) A probability from a density graph is an area. From $x = 1$ to $x = 4$ the region is a triangle with base $4 - 1 = 3$ and height $f(4) = 0.10$.

$$P(X < 4) = \frac{1}{2}(3)(0.10) = 0.15.$$

$$P(X < 4) = 0.15.$$

(b) For $7 \leq x \leq 11$, the cumulative probability consists of all the area up to 7, followed by the rectangular area from 7 to x . The area up to 7 is the triangle of base 6 and height 0.20.

$$F(7) = \frac{1}{2}(6)(0.20) = 0.60.$$

$$F(x) = F(7) + \int_7^x 0.10 \, dt = 0.60 + 0.10(x - 7) = 0.10x - 0.10.$$

$$F(x) = 0.1x - 0.1 \quad (7 \leq x \leq 11).$$

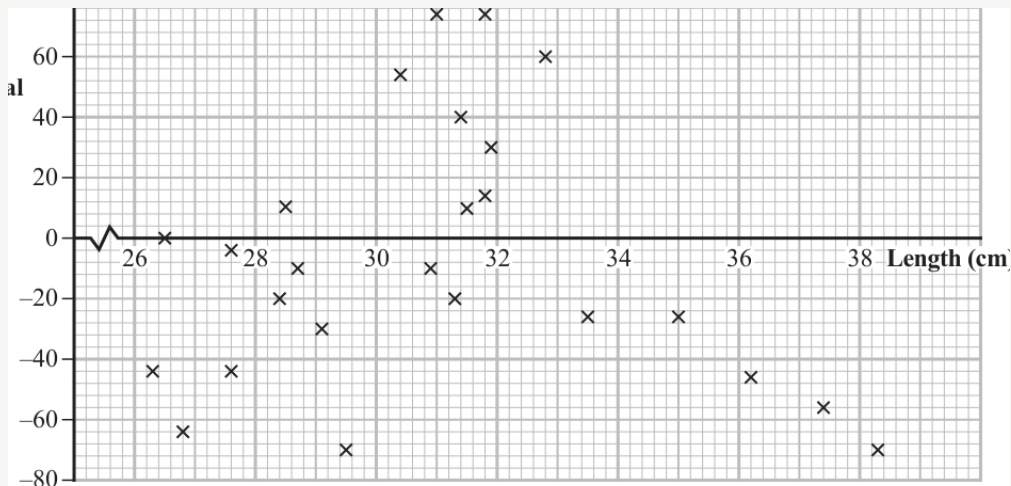
Question 3

Question 3

Gabriela models bream weight w grams from length x cm. For $n = 27$ bream,

$$\bar{x} = 31.07, \quad \bar{w} = 628.59, \quad \sum w^2 = 11\,386\,134, \quad S_{xw} = 13\,082.3, \quad S_{xx} = 260.8.$$

- Find the product moment correlation coefficient between x and w . (3)
 - Explain whether this is consistent with a linear model. (1)
 - Find the regression line of w on x in the form $w = a + bx$. (3)
- The residual plot is shown below. One bream has length 32 cm.



One of the bream in the sample has a length of 32 cm.

Find its weight.

(2)

- Find its weight. (2)
- With reference to the residual plot, comment on the model for bream longer than 33 cm. (1)

Model solution

(a) The product moment correlation coefficient uses the centred sums S_{xx} , S_{ww} and S_{xw} . First calculate the missing centred sum for weight.

$$S_{ww} = \sum w^2 - n\bar{w}^2 = 11\,386\,134 - 27(628.59)^2 = 717\,748.5213.$$

$$r = \frac{S_{xw}}{\sqrt{S_{xx}S_{ww}}} = \frac{13\,082.3}{\sqrt{260.8(717\,748.5213)}} = 0.95618\dots$$

$$r = 0.956 \text{ (to 3 decimal places).}$$

(b) A correlation coefficient close to +1 indicates a strong positive linear association.

Yes. Since $r = 0.956$ is close to 1, the data are consistent with a linear model.

(c) For the regression of weight on length, the gradient is $b = S_{xw}/S_{xx}$; this predicts w from a given x . The line must also pass through $(\bar{x}, \bar{w})$.

$$b = \frac{13\,082.3}{260.8} = 50.162\dots$$

$$a = \bar{w} - b\bar{x} = 628.59 - 50.162\dots(31.07) = -930.0\dots$$

$$w = -930 + 50.2x.$$

(d) The residual is actual weight – predicted weight. At $x = 32$, the plot gives residual $+80$, so add it to the model prediction.

$$\hat{w} = -930 + 50.2(32) = 676.4, \quad w = 676.4 + 80 = 756.4.$$

$$756.4 \text{ g.}$$

(e) For lengths above 33 cm, all five plotted residuals are negative. Thus actual weights are consistently below the model predictions.

The model systematically overestimates the weights of the longest bream, so is not reliable for lengths above 33 cm.

Question 4

Question 4

The random variable X has probability density function

$$f(x) = \begin{cases} 0.8 - 6.4x^{-3}, & 2 \leq x \leq 4, \\ 0, & \text{otherwise.} \end{cases}$$

The median of X is m .

- (a) Show that $m^3 - 3.625m^2 + 4 = 0$. (3)
 (b) (i) Find $f'(x)$. (ii) Explain why the mode of X is 4. (2)
 (c) Given that $E(X^2) = 10.5$ to 3 significant figures, find $\text{Var}(X)$, showing your working clearly. (4)

Model solution

(a) The median divides the total probability into two equal halves, so $P(X \leq m) = 0.5$. Since the density is non-zero from 2 to 4, integrate it from 2 to m .

$$\int_2^m (0.8 - 6.4x^{-3}) \, dx = 0.5.$$

$$[0.8x + 3.2x^{-2}]_2^m = 0.5$$

$$0.8m + \frac{3.2}{m^2} - \left(0.8(2) + \frac{3.2}{2^2}\right) = 0.5.$$

$$0.8m + \frac{3.2}{m^2} - 2.9 = 0.$$

Multiplying by $m^2/0.8$ clears the fraction and gives the required polynomial.

$$m^3 - 3.625m^2 + 4 = 0.$$

$$m^3 - 3.625m^2 + 4 = 0, \text{ as required.}$$

(b)(i) Differentiate the density on its support, using $\frac{d}{dx}(x^{-3}) = -3x^{-4}$.

$$f'(x) = 0 - 6.4(-3x^{-4}) = 19.2x^{-4}.$$

$$f'(x) = 19.2x^{-4} \text{ for } 2 \leq x \leq 4.$$

(b)(ii) On $2 \leq x \leq 4$, $19.2x^{-4} > 0$, so the density is increasing throughout its support. Its largest value is therefore at the right-hand endpoint.

The mode is 4.

(c) Use $\text{Var}(X) = E(X^2) - [E(X)]^2$. The supplied value gives $E(X^2)$, so first find $E(X)$ from the density.

$$\begin{aligned} E(X) &= \int_2^4 xf(x) \, dx = \int_2^4 (0.8x - 6.4x^{-2}) \, dx \\ &= [0.4x^2 + 6.4x^{-1}]_2^4 = \left(0.4(4)^2 + \frac{6.4}{4}\right) - \left(0.4(2)^2 + \frac{6.4}{2}\right) = 3.2. \end{aligned}$$

$$\text{Var}(X) = 10.5 - (3.2)^2 = 0.26.$$

$$\text{Var}(X) = 0.26.$$

www.mathswithmatt.co.uk

Question 5

Question 5

$X \sim U[0.5, 2.5]$. Talia selects T from this distribution.

(a) Find $P(T < 1)$. (1)

Malik calculates $M = \frac{1}{T^2}$. Find the probability that both T and M are less than 2.25. (3)

Raja's score is R , selected from the distribution of X . Greta's score is $G = \frac{2}{R^2}$. Determine, giving a reason, who you expect to have the higher total score when they play many times. (5)

Model solution

(a) For a continuous uniform distribution, probability is the required interval length divided by the total interval length.

$$P(T < 1) = \frac{1 - 0.5}{2.5 - 0.5} = \frac{1}{4}.$$

$$\frac{1}{4}.$$

(b) Both inequalities must hold. Since T is positive, the condition on Malik's number can be rearranged safely:

$$\frac{1}{T^2} < 2.25 \iff T^2 > \frac{1}{2.25} = \frac{4}{9} \iff T > \frac{2}{3}.$$

Together with $T < 2.25$, this gives $\frac{2}{3} < T < 2.25$. Use the uniform distribution to find the probability from the length of this interval.

$$P\left(\frac{2}{3} < T < 2.25\right) = \frac{2.25 - \frac{2}{3}}{2.5 - 0.5} = \frac{19}{24}.$$

$$\frac{19}{24}.$$

(c) Over many games, total scores are compared using expected scores. Raja's score is uniformly distributed on $[0.5, 2.5]$, so its expectation is its midpoint.

$$E(R) = \frac{0.5 + 2.5}{2} = 1.5.$$

For Greta, calculate the expectation of her transformation using the uniform density $1/(2.5 - 0.5) = 1/2$.

$$E(G) = E\left(\frac{2}{R^2}\right) = \int_{0.5}^{2.5} \frac{2}{r^2} \left(\frac{1}{2}\right) dr = \int_{0.5}^{2.5} r^{-2} dr = [-r^{-1}]_{0.5}^{2.5} = 1.6.$$

Greta's expected score per game is greater, so her total is expected to be greater after many games.

Greta, since $E(G) = 1.6 > 1.5 = E(R)$.