



Edexcel AS Further Mathematics

8FM0/24 - Further Statistics 2

Friday 17 May 2024

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 2024 mark scheme. The aim is a clear **full-mark model response**, not merely a list of final answers.

Contents

Question 1	2
Question 2	4
Question 3	5
Question 4	7
Question 5	9

www.mathswithmatt.co.uk

Question 1

Question 1

A continuous random variable X has cumulative distribution function

$$F(x) = \begin{cases} 0, & x < -1, \\ \frac{1}{5}(x+1)^2, & -1 \leq x \leq 0, \\ 1 - \frac{1}{20}(4-x)^2, & 0 < x \leq 4, \\ 1, & x > 4. \end{cases}$$

(a) Find the probability density function, $f(x)$. (2)

(b) (i) Sketch $f(x)$. (2)

(ii) Hence describe the skewness of the distribution. (1)

(c) Find, to 3 significant figures, the value of c such that

$$P(1 < X < c) = P(c < X < 2).$$

(3)

Model solution

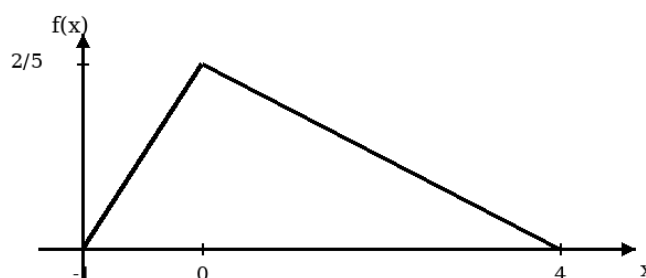
(a) A density function is the derivative of the cumulative distribution function on each interval where the cdf changes.

$$f(x) = \begin{cases} \frac{d}{dx} \left[\frac{1}{5}(x+1)^2 \right] = \frac{2}{5}(x+1), & -1 \leq x \leq 0, \\ \frac{d}{dx} \left[1 - \frac{1}{20}(4-x)^2 \right] = \frac{1}{10}(4-x), & 0 < x \leq 4, \\ 0, & \text{otherwise.} \end{cases}$$

$$f(x) = \frac{2}{5}(x+1) \text{ for } -1 \leq x \leq 0, f(x) = \frac{1}{10}(4-x) \text{ for } 0 < x \leq 4, \text{ and } 0 \text{ otherwise.}$$

(2)

(b)(i) Each piece of the density is linear. The first rises from $(-1, 0)$ to $(0, 2/5)$, and the second falls from $(0, 2/5)$ to $(4, 0)$. Thus the required sketch is a triangle with these three vertices.



A triangular density with horizontal intercepts $x = -1, 4$ and maximum height $2/5$ at $x = 0$.

(2)

(b)(ii) The triangle has a longer tail to the right of its peak than to the left, which is the visual clue for positive skew.

The distribution is positively skewed.

(1)

(c) The two probabilities split the interval $(1, 2)$ into equal probability, so $F(c)$ is halfway between $F(1)$ and $F(2)$.

$$F(1) = 1 - \frac{1}{20}(4-1)^2 = \frac{11}{20}, \quad F(2) = 1 - \frac{1}{20}(4-2)^2 = \frac{4}{5}.$$

$$F(c) = \frac{F(1) + F(2)}{2} = \frac{1}{2} \left(\frac{11}{20} + \frac{16}{20} \right) = \frac{27}{40} = 0.675.$$

Since c lies between 1 and 2, use the $0 < x \leq 4$ part of the cdf.

$$1 - \frac{1}{20}(4-c)^2 = \frac{27}{40} \Rightarrow (4-c)^2 = \frac{13}{2}.$$

$$c = 4 \pm \sqrt{\frac{13}{2}}.$$

The plus root is greater than 4 and cannot lie between 1 and 2, so it is rejected.

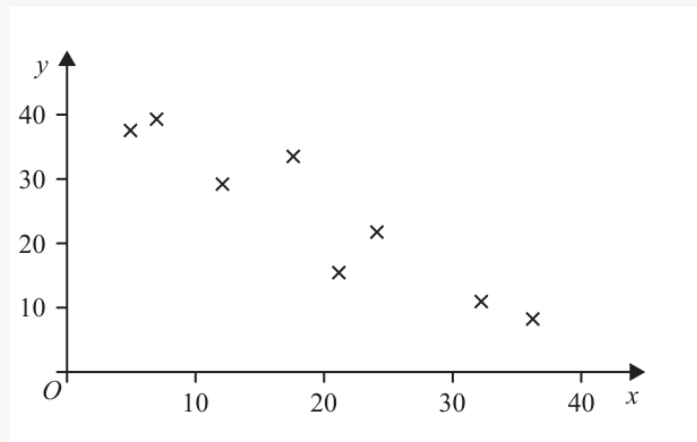
$$c = 4 - \sqrt{\frac{13}{2}} = 1.45 \text{ (3 s.f.)}$$

(3)

Question 2

Question 2

A random sample of size $n = 8$ of paired data is shown in the scatterplot.



Test, at the 1% level of significance, whether there is evidence of a negative rank correlation between the two variables. State hypotheses and the critical value, and show your working clearly. (7)

Model solution

The question asks about rank correlation, so Spearman's rank correlation coefficient is appropriate. A negative alternative gives a one-tailed lower-tail test.

$$H_0 : \rho_s = 0, \quad H_1 : \rho_s < 0.$$

Rank each variable from smallest to largest. The plotted values give the following ranks; d is the difference between the two ranks for a pair.

rank of x	1	2	3	4	5	6	7	8
rank of y	7	8	5	6	3	4	2	1
d^2	36	36	4	4	4	4	25	49

$$\sum d^2 = 162.$$

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} = 1 - \frac{6(162)}{8(8^2 - 1)} = -0.928571 \dots$$

For $n = 8$ and a one-tailed 1% test, the critical value is -0.8333 . Values more negative than this are in the rejection region.

$$-0.928571 \dots < -0.8333.$$

Reject H_0 . At the 1% level, there is sufficient evidence of negative rank correlation between the variables.

(7)

Question 3

Question 3

The continuous random variable Y has probability density function

$$f(y) = \begin{cases} \frac{1}{24}(y+2)(4-y), & 0 \leq y \leq 3, \\ 0, & \text{otherwise.} \end{cases}$$

- (a) Show that the mode of Y is 1, justifying your reasoning. (2)
- (b) Given that $P(Y < 1) = \frac{13}{36}$, determine whether the median of Y is less than, equal to, or greater than 2. Give a reason. (2)
- (c) Given that $E(Y^2) = \frac{213}{80}$, find, using algebraic integration, $\text{Var}(2Y)$. (5)

Model solution

(a) The mode is the value of y where the density is greatest. Differentiate the density and find its stationary point.

$$f(y) = \frac{1}{24}(-y^2 + 2y + 8), \quad f'(y) = \frac{1}{24}(2 - 2y).$$

$$f'(y) = 0 \Rightarrow 2 - 2y = 0 \Rightarrow y = 1.$$

The coefficient of y^2 in $f(y)$ is negative, so the parabola opens downwards: this stationary point is a maximum. Also, 1 is in the permitted interval $[0, 3]$.

The mode is 1.

(2)

(b) The quadratic density has equal values at points equally far from $y = 1$. Hence the areas from 0 to 1 and from 1 to 2 are equal.

$$P(Y < 2) = 2P(Y < 1) = 2\left(\frac{13}{36}\right) = \frac{13}{18} > \frac{1}{2}.$$

By $y = 2$, more than half of the total probability has accumulated. Therefore the point at which the cdf reaches $1/2$ must occur before 2.

The median is less than 2.

(2)

(c) Use algebraic integration to obtain $E(Y)$, then use $\text{Var}(Y) = E(Y^2) - [E(Y)]^2$. Finally,

scaling a random variable by 2 multiplies its variance by 2^2 .

$$\begin{aligned} E(Y) &= \int_0^3 yf(y) \, dy \\ &= \frac{1}{24} \int_0^3 y(y+2)(4-y) \, dy \\ &= \frac{1}{24} \int_0^3 (-y^3 + 2y^2 + 8y) \, dy \\ &= \frac{1}{24} \left[-\frac{y^4}{4} + \frac{2y^3}{3} + 4y^2 \right]_0^3 = \frac{45}{32}. \end{aligned}$$

$$\text{Var}(Y) = \frac{213}{80} - \left(\frac{45}{32} \right)^2 = \frac{3507}{5120},$$

$$\text{Var}(2Y) = 2^2 \text{Var}(Y) = 4 \left(\frac{3507}{5120} \right).$$

$$\text{Var}(2Y) = \frac{3507}{1280} = 2.74 \text{ (3 s.f.)}$$

(5)

Question 4

Question 4

The continuous random variable X is uniformly distributed over $[2, 7]$.

- (a) Write down $E(X)$. (1)
- (b) Find $P(1 < X < 4)$. (1)
- (c) Find $P(2X^2 - 15X + 27 > 0)$. (3)
- (d) Find $E\left(\frac{3}{X^2}\right)$. (3)

Model solution

(a) For a uniform distribution, the mean is the midpoint of its interval.

$$E(X) = \frac{2+7}{2}.$$

$$E(X) = 4.5.$$

(1)

(b) Only values in $[2, 7]$ are possible, so $1 < X < 4$ is equivalent to $2 < X < 4$. For a uniform distribution, probability is favourable interval length divided by total interval length.

$$P(1 < X < 4) = \frac{4-2}{7-2} = \frac{2}{5}.$$

$$P(1 < X < 4) = \frac{2}{5}.$$

(1)

(c) First find where the quadratic is positive. Its positive leading coefficient means it is positive outside its two roots.

$$2x^2 - 15x + 27 = (2x - 9)(x - 3),$$

$$2x^2 - 15x + 27 > 0 \Rightarrow x < 3 \text{ or } x > \frac{9}{2}.$$

Intersecting these regions with the support $[2, 7]$ gives $[2, 3)$ and $(9/2, 7]$.

$$P = \frac{3-2}{7-2} + \frac{7-\frac{9}{2}}{7-2} = \frac{1}{5} + \frac{1}{2} = \frac{7}{10}.$$

$$P(2X^2 - 15X + 27 > 0) = \frac{7}{10}.$$

(3)

(d) For a continuous random variable, $E[g(X)]$ is found by integrating $g(x)$ times the density. Here $f_X(x) = 1/(7-2) = 1/5$ on $[2, 7]$.

$$\begin{aligned} E\left(\frac{3}{X^2}\right) &= \int_2^7 \frac{3}{x^2} \cdot \frac{1}{5} dx \\ &= \frac{3}{5} \int_2^7 x^{-2} dx = \frac{3}{5} [-x^{-1}]_2^7 \\ &= \frac{3}{5} \left(-\frac{1}{7} + \frac{1}{2}\right) = \frac{3}{14}. \end{aligned}$$

$$E\left(\frac{3}{X^2}\right) = \frac{3}{14}.$$

(3)

www.mathswithmatt.co.uk

Question 5

Question 5

A random sample of 24 adults gives the summaries

$$S_{hh} = 0.377, \quad S_{sh} = 0.352, \quad \bar{s} = 1.70, \quad \bar{h} = 1.68.$$

The least-squares regression line of h on s is

$$h = a + 0.919s,$$

where a is a constant.

- Calculate the product moment correlation coefficient. (3)
- A doctor uses the line as a model to predict height from arm span. Use the model to predict the height of an adult whose arm span is 1.79 m. (2)
- Ewan, with arm span 1.70 m and height 1.75 m, is added as the 25th adult. Explain how the gradient of the regression line compares with the gradient for the original 24 adults. Give a reason. (3)

Model solution

(a) For the regression of h on s , the gradient is $b = S_{sh}/S_{ss}$. Find S_{ss} first, then use $r = S_{sh}/\sqrt{S_{ss}S_{hh}}$.

$$0.919 = \frac{0.352}{S_{ss}} \Rightarrow S_{ss} = \frac{0.352}{0.919} = 0.383025 \dots$$

$$r = \frac{0.352}{\sqrt{(0.383025 \dots)(0.377)}} = 0.9263 \dots$$

$$r = 0.926 \text{ (3 d.p.)}$$

(3)

(b) A least-squares regression line passes through $(\bar{s}, \bar{h})$, so write it in centred form. This avoids having to calculate the intercept a separately.

$$h - 1.68 = 0.919(s - 1.70).$$

$$h - 1.68 = 0.919(1.79 - 1.70) = 0.08271, \quad h = 1.76271.$$

$$\text{Predicted height} = 1.76 \text{ m (3 s.f.)}$$

(2)

(c) The new arm span equals the original mean, $s = \bar{s} = 1.70$. Therefore its deviation in s is zero.

$$(s - \bar{s})^2 = 0, \quad (s - \bar{s})(h - \bar{h}) = 0.$$

Consequently the added observation changes neither S_{ss} nor S_{sh} . Since the regression gradient is S_{sh}/S_{ss} , its value does not change.

The gradient is unchanged.

(3)

www.mathswithmatt.co.uk