



Edexcel AS Further Mathematics

8FM0/24 - Further Statistics 2

Friday 19 May 2023

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 2023 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	6
Question 4	8

www.mathswithmatt.co.uk

Question 1

Question 1

Every applicant completes tasks P, Q and R. For nine applicants the scores on P and Q are

Applicant	A	B	C	D	E	F	G	H	I
P	19	16	16	12	8	17	12	12	5
Q	17	11	14	7	6	18	15	11	10

- (a) Calculate Spearman's rank correlation coefficient for the scores on P and Q. (4)
 (b) At the 1% level of significance, test whether there is evidence of a positive correlation between the ranks of scores on P and Q. State hypotheses and critical value clearly. (4)
 The Spearman's rank correlation coefficients for P and R, and Q and R, are respectively 0.290 and 0.795.
 (c) Giving a reason, state which two tasks should be retained. (2)

Model solution

(a) *Spearman's coefficient measures correlation of ranks, so first rank each row. Rank 1 is assigned to the highest score. Equal scores receive the mean of the ranks they would have occupied.*

Applicant	A	B	C	D	E	F	G	H	I
p	1	3.5	3.5	6	8	2	6	6	9
q	2	5.5	4	8	9	1	3	5.5	7

The tied scores 16 on P have ranks 3 and 4, hence rank 3.5; the three scores of 12 have mean rank 6. Similarly the two scores of 11 on Q have rank 5.5. We use the product-moment correlation coefficient of these ranks.

$$\sum p = 45, \quad \sum q = 45, \quad \sum p^2 = 282.5, \quad \sum q^2 = 284.5, \quad \sum pq = 271.25.$$

$$r_s = \frac{\sum pq - \frac{(\sum p)(\sum q)}{9}}{\sqrt{\left(\sum p^2 - \frac{(\sum p)^2}{9}\right) \left(\sum q^2 - \frac{(\sum q)^2}{9}\right)}} = \frac{271.25 - \frac{45 \times 45}{9}}{\sqrt{\left(282.5 - \frac{45^2}{9}\right) \left(284.5 - \frac{45^2}{9}\right)}} = 0.7907 \dots$$

$$r_s = 0.791(\text{to 3d.p.}).$$

(4)

(b) *The wording "positive correlation" makes this a one-tailed Spearman rank test. Let ρ denote the population rank correlation coefficient.*

$$H_0 : \rho = 0, \quad H_1 : \rho > 0.$$

For $n = 9$ at the 1% one-tailed level, the critical value is 0.7833. The critical region is therefore $r_s > 0.7833$.

$$0.7907 > 0.7833.$$

The test statistic is in the critical region, so the null hypothesis is rejected.

At the 1% level, there is evidence of a positive correlation between the ranks of these

(4)

(c) To obtain two tasks that give the manager different information about applicants, choose the pair with the weakest agreement. The correlation for P and R is only 0.290, whereas the pairs involving Q have much stronger positive correlations.

Use tasks P and R.

(2)

www.mathswithmatt.co.uk

Question 2

Question 2

The continuous random variable X has probability density function

$$f(x) = \begin{cases} \frac{x}{16}(9 - x^2), & 1 \leq x \leq 3, \\ 0, & \text{otherwise.} \end{cases}$$

- (a) Find the cumulative distribution function of X . (3)
- (b) Calculate $P(X > 1.8)$. (2)
- (c) Use calculus to find $E\left(\frac{3}{X} + 2\right)$. (3)
- (d) Show that the mode of X is $\sqrt{3}$. (3)

Model solution

(a) A cumulative distribution function is the area under the density to the left of x . On the support, integrate the density from its lower endpoint 1 to x .

$$F(x) = \int_1^x \frac{t}{16}(9 - t^2) dt = \left[\frac{9t^2}{32} - \frac{t^4}{64} \right]_1^x = \frac{18x^2 - x^4 - 17}{64}.$$

Below 1 no area has accumulated; above 3 all the probability has accumulated. Hence

$$F(x) = \begin{cases} 0, & x < 1, \\ \frac{18x^2 - x^4 - 17}{64}, & 1 \leq x \leq 3, \\ 1, & x > 3. \end{cases}$$

$F(x)$ is the piecewise function above.

(3)

(b) A right-tail probability is most directly found from the CDF using $P(X > c) = 1 - F(c)$.

$$P(X > 1.8) = 1 - F(1.8) = 1 - \frac{18(1.8)^2 - (1.8)^4 - 17}{64} = 0.5184.$$

$$P(X > 1.8) = 0.5184.$$

(2)

(c) For a function of a continuous random variable, use $E[g(X)] = \int g(x)f(x) dx$ over the

support. Here the factor x in the density cancels the $1/x$ term.

$$\begin{aligned} E\left(\frac{3}{X} + 2\right) &= \int_1^3 \left(\frac{3}{x} + 2\right) \frac{x}{16}(9 - x^2) dx \\ &= 3 \int_1^3 \frac{9 - x^2}{16} dx + 2 \\ &= 3 \left[\frac{9x - x^3/3}{16} \right]_1^3 + 2 \\ &= 3 \left(\frac{18 - 26/3}{16} \right) + 2 = \frac{15}{4}. \end{aligned}$$

$$E\left(\frac{3}{X} + 2\right) = \frac{15}{4} = 3.75.$$

(3)

(d) The mode is where the density is greatest. Since the density is differentiable inside its support, find stationary points by differentiating f .

$$f(x) = \frac{9x - x^3}{16}, \quad f'(x) = \frac{9}{16} - \frac{3x^2}{16}.$$

$$f'(x) = 0 \implies 9 - 3x^2 = 0 \implies x^2 = 3.$$

Only the positive root lies in $[1, 3]$, so the stationary point is $x = \sqrt{3}$. Also $f''(x) = -3x/8 < 0$ there, so it is a maximum; the endpoint values are $f(1) = 1/2$ and $f(3) = 0$.

The mode is $\sqrt{3}$.

(3)

Question 3

Question 3

For 9 randomly selected professional male tennis players, height h is measured in metres and maximum serve speed v in km/h. The summary statistics are

$$\sum h = 17.63, \quad \sum v = 2174.9, \quad \sum v^2 = 526407.8, \quad S_{hh} = 0.0487, \quad S_{hv} = 5.1376.$$

(a) Calculate the product moment correlation coefficient between h and v . (2)

(b) Explain whether the answer to (a) is consistent with a linear model. (1)

(c) Find the regression line of v on h , in the form $v = a + bh$, giving a and b to 1 decimal place. (3)

Pat's sum of residuals for the 9 players is 1.04.

(d) Without calculation, explain why this is a mistake. (1)

For the player of height 1.96 m, Pat misread the residual as 2.27.

(e) Find that player's maximum serve speed, in km/h. (3)

Model solution

(a) The PMCC is $r = S_{hv} / \sqrt{S_{hh} S_{vv}}$. The supplied $\sum v$ and $\sum v^2$ first allow S_{vv} to be found.

$$S_{vv} = \sum v^2 - \frac{(\sum v)^2}{n} = 526407.8 - \frac{2174.9^2}{9} = 831.132 \dots$$

$$r = \frac{5.1376}{\sqrt{0.0487 \times 831.132 \dots}} = 0.80753 \dots$$

$$r = 0.808(\text{to 3 d.p.}).$$

(2)

(b) A correlation close to +1 indicates a strong positive linear association. The value 0.808 is reasonably close to 1.

Yes; the strong positive correlation is consistent with a linear model.

(1)

(c) For the regression of v on h , the gradient is $b = S_{hv} / S_{hh}$. The line passes through $(\bar{h}, \bar{v})$, which then gives the intercept.

$$b = \frac{5.1376}{0.0487} = 105.49 \dots$$

$$a = \bar{v} - b\bar{h} = \frac{2174.9}{9} - 105.49 \dots \left(\frac{17.63}{9} \right) = 35.0 \dots$$

$$v = 35.0 + 105.5h.$$

(3)

(d) A least-squares regression line with an intercept always has residuals that sum to zero. Therefore a non-zero total cannot be correct.

The residual should sum to zero, not 1.04.

(1)

(e) The stated total 1.04 includes the misread residual. Replacing 2.27 by the correct residual must make the total zero, so the correct residual is $2.27 - 1.04 = 1.23$. Since residual means observed minus predicted, add it to the regression estimate.

$$\hat{v} = 35.0 + 105.5(1.96) = 241.78, \quad v = \hat{v} + 1.23 = 243.01.$$

The maximum service speed is 243 km/h (to 3 s.f.).

(3)

www.mathswithmatt.co.uk

Question 4

Question 4

The random variable X has a continuous uniform distribution on $[-3, k]$. Given that

$$P(-4 < X < 2) = \frac{1}{3},$$

(a) find k . (3)

A computer generates a random number Y , where Y has a continuous uniform distribution on $[a, b]$, $E(Y) = 6$, and $\text{Var}(Y) = 192$. The computer generates 5 random numbers.

(b) Calculate the probability that at least 2 of the 5 numbers are greater than 7.5. (6)

Model solution

(a) The lower endpoint of the distribution is -3 , so $-4 < X < 2$ includes exactly the same possible values as $-3 < X < 2$. For a uniform distribution, probability is interval length divided by total length.

$$\frac{2 - (-3)}{k - (-3)} = \frac{1}{3} \implies \frac{5}{k + 3} = \frac{1}{3}.$$

$$k + 3 = 15 \implies k = 12.$$

$$k = 12.$$

(3)

(b) For $Y \sim U(a, b)$, use the given mean and variance to determine the endpoints before finding a single-number probability.

$$\frac{a + b}{2} = 6 \implies a + b = 12,$$

$$\frac{(b - a)^2}{12} = 192 \implies b - a = 48.$$

Solving these simultaneous equations gives

$$a = -18, \quad b = 30.$$

The probability that one generated number exceeds 7.5 is its length to the right of 7.5, divided by the full interval length.

$$p = P(Y > 7.5) = \frac{30 - 7.5}{30 - (-18)} = \frac{15}{32}.$$

Let R be the number, out of five independent generated numbers, that exceed 7.5. This is a binomial count, so $R \sim B(5, \frac{15}{32})$. "At least 2" is conveniently found by subtracting the cases $R = 0$ and $R = 1$ from 1.

$$\begin{aligned} P(R \geq 2) &= 1 - P(R = 0) - P(R = 1) \\ &= 1 - \left(\frac{17}{32}\right)^5 - 5 \left(\frac{15}{32}\right) \left(\frac{17}{32}\right)^4 \\ &= 0.7710 \dots \end{aligned}$$

$$P(\textit{at least 2 exceed 7.5}) = 0.771(\text{to 3d.p.}).$$

(6)

www.mathswithmatt.co.uk