



Edexcel AS Further Mathematics

8FM0/24 - Further Statistics 2

October 2020

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 2020 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	5
Question 4	7

www.mathswithmatt.co.uk

Question 1

Question 1

An estate agent believes that three-bedroom houses further from the railway station are more expensive. A random sample of 22 houses gives product moment correlation coefficient $r = 0.3892$. At the 5% level of significance, test this belief, stating the hypotheses and critical region.

Model solution

The claim is that price tends to increase as distance increases, so this is a test for a positive population correlation coefficient. It is therefore a one-tailed PMCC test.

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

For $n = 22$ at the 5% one-tailed level, the PMCC critical value is 0.3598. Thus the critical region is

$$r > 0.3598.$$

Since $0.3892 > 0.3598$, the observed value lies in the critical region, so H_0 is rejected.

There is significant evidence at the 5% level to support the estate agent's belief that

(3)

Question 2

Question 2

Cake	A	B	C	D	E	F	G	H	I
Mary	19	17	23	10	21	15	12	8	14
Jahil	22	18	21	10	24	20	16	12	15
Dawn	9	11	6	18	9	15	13	20	13

- (a) Calculate Spearman's rank correlation coefficient between Mary and Jahil.
 (b) Calculate Spearman's rank correlation coefficient between Jahil and Dawn.
 (c) Explain how different interpretations of the points system could give rise to the results in (a) and (b).

Model solution

(a) Spearman's coefficient compares ranks rather than the numerical scores. Rank 1 is assigned to the highest score, because a larger score represents a better cake for Mary and Jahil.

Cake	A	B	C	D	E	F	G	H	I
R_M	3	4	1	8	2	5	7	9	6
R_J	2	5	3	9	1	4	6	8	7
$d = R_M - R_J$	1	-1	-2	-1	1	1	1	1	-1
d^2	1	1	4	1	1	1	1	1	1

$$\sum d^2 = 12, \quad r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} = 1 - \frac{6(12)}{9(9^2 - 1)} = 0.9.$$

$$r_s = 0.900$$

(4)

(b) Dawn has tied scores, so tied observations must receive the mean of the ranks they would occupy. For example, the two scores of 13 occupy ranks 4 and 5, so each has rank 4.5.

Cake	A	B	C	D	E	F	G	H	I
R_J	2	5	3	9	1	4	6	8	7
R_D	7.5	6	9	2	7.5	3	4.5	1	4.5

Because of the ties, use the PMCC formula on these rank variables rather than the no-ties shortcut.

$$\bar{R}_J = \bar{R}_D = 5,$$

$$S_{JJ} = \sum (R_J - 5)^2 = 60, \quad S_{DD} = \sum (R_D - 5)^2 = 59,$$

$$S_{JD} = \sum (R_J - 5)(R_D - 5) = -49,$$

$$r_s = \frac{S_{JD}}{\sqrt{S_{JJ}S_{DD}}} = \frac{-49}{\sqrt{60 \times 59}} = -0.823558 \dots$$

$$r_s = -0.824(3\text{s.f.}).$$

(3)

(c) Mary and Jahil appear to have awarded more points for features they regarded as good: their rankings have a strong positive association. Dawn appears to have used high points for poor features (or low points for good features), reversing the order.

The strong positive Mary--Jahil correlation and strong negative Jahil--Dawn correlation can therefore still represent agreement about the cakes, but opposite meanings for a high score.

(2)

www.mathswithmatt.co.uk

Question 3

Question 3

The continuous random variable X has cumulative distribution function

$$F(x) = \begin{cases} 0, & x < 4, \\ px - k\sqrt{x}, & 4 \leq x \leq 9, \\ 1, & x > 9, \end{cases}$$

where p and k are constants.

(a) Find p and k .

Given that $E(X) = \frac{119}{18}$: (b) show that $\text{Var}(X) = 2.05$ to 3 significant figures; (c) write down the mode of X ; (d) find the exact a such that $P(X \leq a) = \frac{7}{27}$.

Model solution

(a) A CDF must join the constant pieces continuously. Use the two endpoints of the middle interval: $F(4) = 0$ and $F(9) = 1$.

$$4p - 2k = 0, \quad 9p - 3k = 1.$$

$$k = 2p \implies 9p - 6p = 1 \implies p = \frac{1}{3}, \quad k = \frac{2}{3}.$$

$$p = \frac{1}{3}, \quad k = \frac{2}{3}$$

(4)

(b) To find a variance, first obtain the density by differentiating the CDF on its support. Then use $\text{Var}(X) = E(X^2) - [E(X)]^2$.

$$f(x) = F'(x) = \frac{1}{3} - \frac{1}{3\sqrt{x}} = \frac{1}{3} \left(1 - \frac{1}{\sqrt{x}} \right), \quad 4 \leq x \leq 9.$$

$$\begin{aligned} E(X^2) &= \int_4^9 x^2 f(x) \, dx \\ &= \frac{1}{3} \int_4^9 (x^2 - x^{3/2}) \, dx \\ &= \frac{1}{3} \left[\frac{x^3}{3} - \frac{2x^{5/2}}{5} \right]_4^9 = \frac{2059}{45}. \end{aligned}$$

This is the second raw moment, so subtract the square of the given mean.

$$\text{Var}(X) = \frac{2059}{45} - \left(\frac{119}{18} \right)^2 = 2.048765 \dots = 2.05 \text{ (3 s.f.)}.$$

$$\text{Var}(X) = 2.05(3\text{s.f.}).$$

(6)

(c) The mode is where the density is greatest. Here $f(x) = \frac{1}{3}(1 - x^{-1/2})$ increases as x increases on $[4, 9]$.

The mode is 9.

(1)

(d) The required probability is a CDF value, and a must lie in $[4, 9]$, so substitute the middle expression for F .

$$\frac{a}{3} - \frac{2\sqrt{a}}{3} = \frac{7}{27}.$$

Let $y = \sqrt{a}$. This substitution turns the equation into a quadratic.

$$\frac{y^2}{3} - \frac{2y}{3} = \frac{7}{27} \implies 9y^2 - 18y - 7 = 0 \implies (y - \frac{7}{3})(y + \frac{1}{3}) = 0.$$

$$y = \frac{7}{3} \text{ since } y = \sqrt{a} \geq 0, \quad a = y^2 = \frac{49}{9}.$$

$$a = \frac{49}{9}.$$

(3)

www.maths

Question 4

Question 4

The diameter d (mm) and failure weight w (g) of wire are

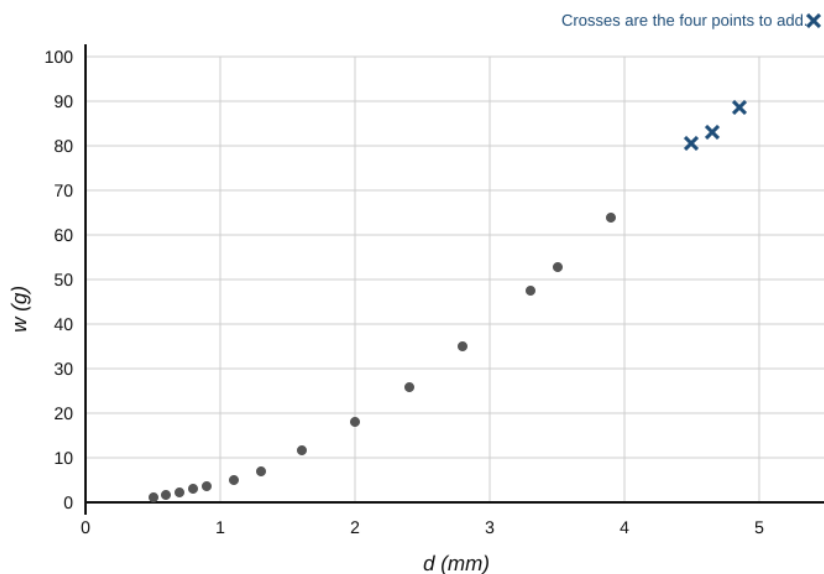
d	0.5	0.6	0.7	0.8	0.9	1.1	1.3	1.6	2	2.4	2.8	3.3	3.5	3.9	4.5	4.6	4.8	5.4
w	1.2	1.7	2.3	3.0	3.8	5.6	7.7	11.6	18	25.9	34.9	47.4	52.7	63.9	81	83.6	89.9	109.4

(a) Complete the scatter diagram. (b) Find the regression line of w on d . (c) Comment on using this linear model for $0.5 \leq d \leq 5.4$. Given $r = 0.987$: (d) calculate its RSS. With $u = d^2$, $S_{wu} = 5721.625$, $S_{uu} = 1482.619$, and $\sum u = 157.57$: (e) give a physical reason for using u ; (f) find the regression line of w on u ; (g) calculate its RSS; (h) choose the better model; (i) estimate the failure weight when $d = 3$ mm.

Model solution

(a) The first fourteen points are already plotted. The remaining observations are the four points below; plotting these completes the scatter diagram.

$(4.5, 81)$, $(4.6, 83.6)$, $(4.8, 89.9)$, $(5.4, 109.4)$.



Plot these four points on the given axes.

(1)

(b) The requested regression is w on d , so take w as the response variable and fit a line of the form $w = a + bd$. Calculator linear regression gives

$$w = 21.5d - 17.7.$$

$$w = 21.5d - 17.7$$

(2)

(c) A linear model would be appropriate only if the plotted points lay approximately along a straight line. The scatter is curved: the line would overestimate in the middle and underestimate towards the ends.

The linear model is not appropriate for prediction across this interval.

(1)

(d) RSS measures the unexplained variation in w . For a simple linear regression, $RSS = S_{ww}(1 - r^2)$. First calculate the corrected sum of squares.

$$S_{ww} = \sum w^2 - \frac{(\sum w)^2}{18} = 45178.68 - \frac{643.6^2}{18} = 22166.404 \dots$$

$$RSS = 22166.404 \dots (1 - 0.987^2) = 570.33 \dots$$

$$RSS \approx 570 \text{ g}^2.$$

(2)

(e) A wire's strength is related to its cross-sectional area. For a circular wire, area is proportional to d^2 , so $u = d^2$ has a physical basis.

A thicker wire has greater cross-sectional area, proportional to d^2 , and should support more weight.

(1)

(f) For regression of w on u , the gradient is S_{wu}/S_{uu} . Then choose the intercept so that the line passes through $(\bar{u}, \bar{w})$.

$$c = \frac{S_{wu}}{S_{uu}} = \frac{5721.625}{1482.619} = 3.85913 \dots$$

$$f = \bar{w} - c\bar{u} = \frac{643.6}{18} - 3.85913 \dots \left(\frac{157.57}{18} \right) = 1.973 \dots$$

$$w = 1.97 + 3.86u.$$

$$w = 1.97 + 3.86u, \text{ where } u = d^2.$$

(3)

(g) For this regression, the unexplained sum of squares can be found by subtracting the explained sum of squares from S_{ww} .

$$RSS = S_{ww} - \frac{S_{wu}^2}{S_{uu}} = 22166.404 \dots - \frac{5721.625^2}{1482.619} = 85.8824 \dots$$

$$RSS \approx 85.9 \text{ g}^2.$$

(2)

(h) The model with the smaller RSS leaves less variation unexplained by the fitted line.

Robert's w – on – u model is better, since 85.9 is much smaller than 570.

(1)

(i) Use the better, quadratic-in-diameter model. For $d = 3$, $u = d^2 = 9$.

$$w = 1.973 \dots + 3.85913 \dots (9) = 36.705 \dots$$

The estimated failure weight is 36.7 g (approximately).

(1)

www.mathswithmatt.co.uk