



Edexcel AS Further Mathematics

8FM0/24 - Further Statistics 2

Thursday 17 May 2018

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 2018 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

www.mathswithmatt.co.uk

Question 1

Question 1

The maths-test score m and physics-test score p of 16 students satisfy

$$\sum m = 392, \quad \sum p = 254, \quad \sum p^2 = 4748, \quad S_{mm} = 1846, \quad S_{mp} = 1115.$$

(a) Find the product moment correlation coefficient between m and p . (2)

(b) Find the equation of the linear regression line of p on m . (3)

Figure 1 is a residual plot.

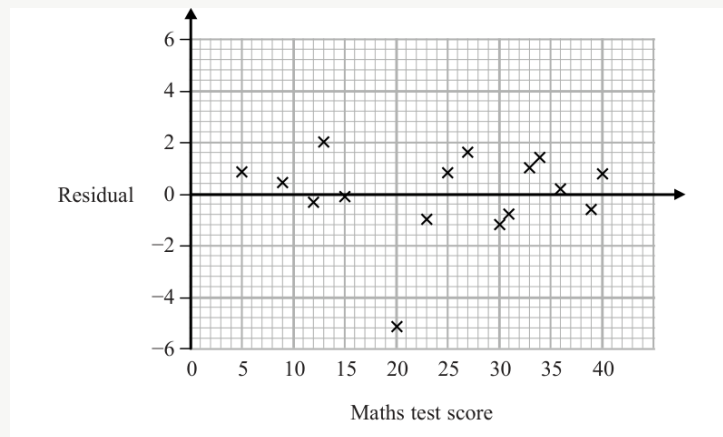


Figure 1

(c) Calculate the residual sum of squares (RSS). (2)

For the person with $m = 30$, (d) find the physics-test score. (2)

The data for the person with $m = 20$ are removed. (e) Suggest a reason why. (1)

(f) Without further calculation, state how the recalculated product moment correlation coefficient compares with that in (a), giving a reason. (1)

Model solution

(a)

$$S_{pp} = \sum p^2 - \frac{(\sum p)^2}{16} = 4748 - \frac{254^2}{16} = 715.75.$$

$$r = \frac{S_{mp}}{\sqrt{S_{mm}S_{pp}}} = \frac{1115}{\sqrt{1846(715.75)}} = 0.970014\dots$$

$$r = 0.970 \text{ (3 s.f.)}$$

(b) For the regression of p on m , the gradient is

$$b = \frac{S_{mp}}{S_{mm}} = \frac{1115}{1846} = 0.6040\dots$$

and, since $\bar{m} = 392/16 = 24.5$ and $\bar{p} = 254/16 = 15.875$,

$$a = \bar{p} - b\bar{m} = 15.875 - 0.6040\dots(24.5) = 1.076\dots$$

$$p = 1.08 + 0.604m \text{ (coefficients to 3 s.f.)}$$

(c)

$$RSS = S_{pp} - \frac{S_{mp}^2}{S_{mm}} = 715.75 - \frac{1115^2}{1846} = 42.2803 \dots$$

$$RSS = 42.3 \text{ (3 s.f.)}$$

(d) The residual at $m = 30$, read from Figure 1, is -1.2 . Thus

$$p = 1.08 + 0.604(30) - 1.2 = 18.0.$$

18 marks.

(e) The point has a residual far from 0 (about -5.2), so it may be an outlier/anomalous observation.

It is an outlier (has a large residual).

(f) Removing this point leaves points closer to the regression line, so the positive linear association is stronger.

The new value of r is closer to 1 (so is greater than 0.970).

The residual is observed score minus the score predicted by the regression line.

Question 2

Question 2

X has density $f_X(x) = \frac{1}{8}$ for $1 \leq x \leq 9$ and 0 otherwise. $Y = 5 - 2X$.

- (a) Name the distribution of X . (1)
- (b) Find $P(Y > 0)$. (2)
- (c) Find $E(Y)$. (2)
- (d) Find $P(Y < 0 \mid X < 7.5)$. (3)

Model solution

(a)

$X \sim U(1, 9)$, a continuous uniform distribution.

(b)

$$P(Y > 0) = P(5 - 2X > 0) = P(X < 2.5) = \frac{2.5 - 1}{9 - 1} = \frac{3}{16}.$$

$$\frac{3}{16}.$$

(c) Since $E(X) = (1 + 9)/2 = 5$,

$$E(Y) = E(5 - 2X) = 5 - 2E(X) = 5 - 2(5) = -5.$$

$$-5.$$

(d) $Y < 0$ means $X > 2.5$. Therefore

$$P(Y < 0 \mid X < 7.5) = \frac{P(2.5 < X < 7.5)}{P(X < 7.5)} = \frac{(7.5 - 2.5)/8}{(7.5 - 1)/8} = \frac{10}{13}.$$

$$\frac{10}{13}.$$

The conditional probability restricts the uniform distribution to the interval $1 \leq X < 7.5$.

Question 3

Question 3

The high-jump heights (m) for competitors A–F are respectively

2.05, 1.93, 2.02, 1.96, 1.81, 2.02.

Their long-jump finishing order, furthest first, is C, A, F, D, B, E.

(a) Calculate Spearman's rank correlation coefficient. (4)

(b) State hypotheses and test at the 5% level for positive correlation between the high-jump and long-jump results. (4)

The PMCC between the high-jump height and long-jump length is 0.678.

(c) Test at the 5% level for evidence of positive correlation. (2)

(d) State the condition required for the test in (c) to be valid. (1)

(e) Explain what the conclusions in (b) and (c) suggest about the relationship. (1)

Model solution

(a) Using rank 1 for the best result, with tied high-jump heights given their mean rank,

Competitor	A	B	C	D	E	F
high-jump rank h	1	5	2.5	4	6	2.5
long-jump rank l	2	5	1	4	6	3

Hence $\sum h^2 = 90.5$, $\sum l^2 = 91$ and $\sum hl = 89$. Applying the PMCC formula to these ranks gives

$$r_s = \frac{6(89) - 21(21)}{\sqrt{[6(90.5) - 21^2][6(91) - 21^2]}} = 0.8985 \dots$$

$$r_s = 0.899 \text{ (3 s.f.)}$$

(b)

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

For $n = 6$ and a one-tailed 5% test, the critical value is 0.8286. Since $0.899 > 0.8286$, the statistic is in the critical region, so reject H_0 .

There is significant evidence at the 5% level of positive rank correlation between high-jump and long-jump results.

(c) For the PMCC test, $H_0 : \rho = 0$ and $H_1 : \rho > 0$. The one-tailed 5% critical value for $n = 6$ is 0.7293. Since $0.678 < 0.7293$, do not reject H_0 .

There is insufficient evidence at the 5% level of positive linear correlation.

(d)

The paired data must come from a bivariate normal distribution.

(e)

There is positive association between the ranks, but the relationship does not appear to follow a linear pattern.

Spearman's coefficient detects monotonic association through ranks; the PMCC test here is testing linear correlation.

www.mathswithmatt.co.uk

Question 4

Question 4

The continuous random variable X has cumulative distribution function

$$F(x) = \begin{cases} 0, & x < 3, \\ c - 4.5x^n, & 3 \leq x \leq 9, \\ 1, & x > 9, \end{cases}$$

where c is a positive constant and n is an integer.

(a) Showing all stages, find c and n . (7)

(b) Find the lower quartile of X . (2)

Model solution

(a) A CDF has $F(3) = 0$ and $F(9) = 1$. Hence

$$c - 4.5(3)^n = 0, \quad c - 4.5(9)^n = 1.$$

Eliminating c gives

$$4.5(3)^n = 1 + 4.5(9)^n.$$

Let $u = 3^n$, so that $9^n = u^2$. Then

$$4.5u = 1 + 4.5u^2, \quad 4.5u^2 - 4.5u + 1 = 0.$$

Solving,

$$u = \frac{4.5 \pm \sqrt{4.5^2 - 4(4.5)(1)}}{2(4.5)} = \frac{1}{3} \text{ or } \frac{2}{3}.$$

Thus $3^n = \frac{1}{3}$ gives $n = -1$, whereas $3^n = \frac{2}{3}$ does not give an integer n . Therefore $n = -1$, and

$$c = 4.5(3)^{-1} = 1.5.$$

$$n = -1, \quad c = 1.5.$$

(b) Let q be the lower quartile. Then $F(q) = 0.25$, so using $c = 1.5$ and $n = -1$,

$$1.5 - 4.5q^{-1} = 0.25.$$

Thus $4.5/q = 1.25$, giving $q = 3.6$.

$$\text{Lower quartile} = 3.6.$$

The endpoint values of the CDF determine its constants, and a lower quartile has cumulative probability 0.25.