



Edexcel AS Further Mathematics

8FM0/24 - Further Statistics 2

Thursday 16 May 2019

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 2019 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

Bara is investigating whether the two judges of a skating competition are in agreement. The scores awarded to each of the eight skaters are:

Skater	A	B	C	D	E	F	G	H
Judge 1	71	70	72	62	63	61	57	53
Judge 2	73	71	67	64	62	56	52	53

- (a) Calculate Spearman's rank correlation coefficient between the ranks of the two judges. (4)
 (b) Test, at the 1% level of significance, whether the two judges are in agreement. (4)
 Judge 1 accidentally swapped the scores for skaters D and E: D should have scored 63 and E should have scored 62.
 (c) Without further calculation, explain how Spearman's rank correlation coefficient changes, giving a reason. (2)

Model solution

(a) Rank each judge's scores with rank 1 for the highest score.

Skater	A	B	C	D	E	F	G	H
Judge 1 rank	2	3	1	5	4	6	7	8
Judge 2 rank	1	2	3	4	5	6	8	7
d	1	1	-2	1	-1	0	-1	1
d ²	1	1	4	1	1	0	1	1

$$\sum d^2 = 10, \quad r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} = 1 - \frac{6(10)}{8(8^2 - 1)} = 0.880952 \dots$$

$$r_s = 0.881 \text{ (3 s.f.)}$$

(4)

(b) Let ρ be the population Spearman rank correlation coefficient.

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

For $n = 8$ at the 1% one-tailed level, the critical value is 0.8333. Since $0.880952 > 0.8333$, reject H_0 .

There is sufficient evidence at the 1% level that the two judges are in agreement.

(4)

(c) After the correction, D and E have the same Judge 1 ranks as their Judge 2 ranks. Thus their contributions to $\sum d^2$ fall from $1 + 1$ to $0 + 0$, so $\sum d^2$ decreases.

Spearman's rank correlation coefficient increases.

(2)

Question 2

Question 2

Lloyd's waiting time T , in minutes, has probability density function

$$f(t) = \begin{cases} \frac{t}{120}, & 4 \leq t \leq 16, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Show that its cumulative distribution function is

$$F(t) = \begin{cases} 0, & t < 4, \\ \frac{t^2}{240} - c, & 4 \leq t \leq 16, \\ 1, & t > 16, \end{cases}$$

where c is to be found.

(b) Find the exact probability that Lloyd waits between 5 and 10 minutes.

(c) Find the median of T .

(d) Find k such that $P(T < k) = \frac{2}{3}P(T > k)$, giving the answer to 3 significant figures.

Model solution

(a) For $4 \leq t \leq 16$,

$$F(t) = \int_4^t \frac{x}{120} dx = \left[\frac{x^2}{240} \right]_4^t = \frac{t^2}{240} - \frac{1}{15}.$$

$$F(t) = \begin{cases} 0, & t < 4, \\ \frac{t^2}{240} - \frac{1}{15}, & 4 \leq t \leq 16, \\ 1, & t > 16, \end{cases} \quad c = \frac{1}{15}.$$

(2)

(b)

$$P(5 \leq T \leq 10) = F(10) - F(5) = \left(\frac{100}{240} - \frac{1}{15} \right) - \left(\frac{25}{240} - \frac{1}{15} \right) = \frac{5}{16}.$$

$$\frac{5}{16}$$

(2)

(c) If m is the median, $F(m) = \frac{1}{2}$.

$$\frac{m^2}{240} - \frac{1}{15} = \frac{1}{2} \implies m^2 = 136 \implies m = 2\sqrt{34}.$$

$$2\sqrt{34} \text{ minutes} \approx 11.7 \text{ minutes}$$

(2)

(d) Put $p = P(T < k) = F(k)$. Continuity gives $P(T > k) = 1 - p$, so

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

$$\frac{k^2}{240} - \frac{1}{15} = \frac{2}{5} \implies k^2 = 112 \implies k = 4\sqrt{7}.$$

$$k = 4\sqrt{7} \text{ minutes} \approx 10.6 \text{ minutes (3 s.f.)}$$

(3)

www.mathswithmatt.co.uk

Question 3

Question 3

For 10 years, mean annual rainfall w cm and annual leek yield l tonnes per hectare were recorded. The summaries are

$$S_{wl} = 42.786, \quad S_{ww} = 9936.9, \quad \sum l^2 = 26.2326, \quad \sum l = 16.06.$$

(a) Find the product moment correlation coefficient between l and w . (2)

Dora codes $s = w - 6$ and $t = l - 20$.

(b) Write down the PMCC between s and t , justifying your answer. (1)

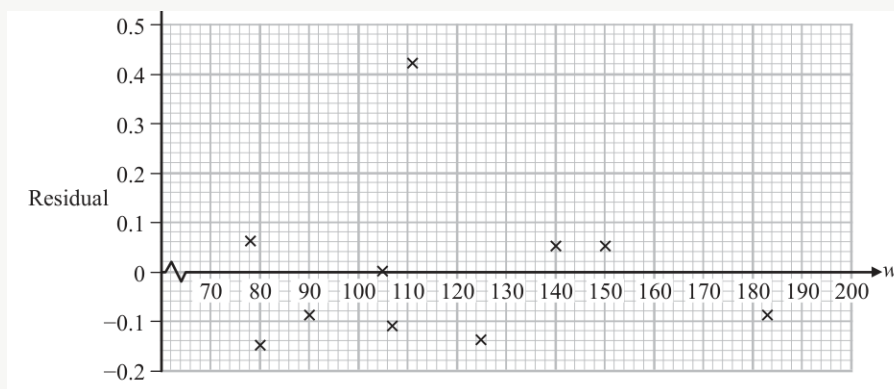
The regression line of t on s is $t = 0.00431s - 18.87$.

(c) Find the regression line of l on w in the form $l = a + bw$, giving a and b to 3 significant figures. (3)

(d) Use this line to estimate leek yield when $w = 100$ cm. (1)

(e) Calculate the residual sum of squares. (2)

The residual plot below shows the residual for each value of l .



(f) (i) State whether the plot suggests a linear regression model is suitable, giving a reason.

(ii) Other than collecting more data, suggest how the fit of the model in part (c) could be improved. (2)

Model solution

(a) First calculate S_{ll} :

$$S_{ll} = \sum l^2 - \frac{(\sum l)^2}{10} = 26.2326 - \frac{16.06^2}{10} = 0.44024.$$

$$r = \frac{S_{wl}}{\sqrt{S_{ww}S_{ll}}} = \frac{42.786}{\sqrt{9936.9(0.44024)}} = 0.64689 \dots$$

$$r = 0.647 \text{ (3 s.f.)}$$

(2)

(b) Subtracting constants from either variable does not change the PMCC.

$$0.647$$

(1)

(c) Substitute $t = l - 20$ and $s = w - 6$ into the given regression equation:

$$l - 20 = 0.00431(w - 6) - 18.87.$$

$$l = 0.00431w + 20 - 18.87 - 0.00431(6) = 0.00431w + 1.10414.$$

$$l = 1.10 + 0.00431w$$

(3)

(d)

$$l = 1.10414 + 0.00431(100) = 1.53514.$$

1.53 tonnes per hectare

(1)

(e) For a regression of l on w ,

$$RSS = S_{ll} - \frac{S_{wl}^2}{S_{ww}} = 0.44024 - \frac{42.786^2}{9936.9} = 0.2560 \dots$$

$$RSS = 0.256$$

(2)

(f)(i) The residuals are randomly scattered above and below zero, so there is no apparent systematic pattern to contradict a linear model.

A linear regression model is suitable.

(f)(ii) The plot suggests a possible outlier; remove that observation and recalculate the regression line.

(2)

Question 4

Question 4

$X \sim U[5, a]$, where a is a constant, and $\text{Var}(X) = \frac{27}{4}$.

(a) Show that $a = 14$.

(3)

The continuous random variable Y has density

$$f(y) = \begin{cases} \frac{1}{20}(2y - 3), & 2 \leq y \leq 6, \\ 0, & \text{otherwise.} \end{cases}$$

The random variable $T = 3(X^2 + X) + 2Y$.

(b) Show that $E(T) = \frac{9857}{30}$.

(7)

Model solution

(a) A continuous uniform distribution on $[5, a]$ has variance $(a - 5)^2/12$. Therefore

$$\frac{(a - 5)^2}{12} = \frac{27}{4} \implies (a - 5)^2 = 81 \implies a - 5 = \pm 9.$$

As $a > 5$, take $a - 5 = 9$.

$$a = 14$$

(3)

(b) Using the given density,

$$E(Y) = \int_2^6 y \left(\frac{2y - 3}{20} \right) dy = \frac{1}{20} \left[\frac{2y^3}{3} - \frac{3y^2}{2} \right]_2^6 = \frac{68}{15}.$$

Since $X \sim U[5, 14]$,

$$E(X) = \frac{5 + 14}{2} = \frac{19}{2}, \quad E(X^2) = \text{Var}(X) + [E(X)]^2 = \frac{27}{4} + \left(\frac{19}{2} \right)^2 = 97.$$

$$\begin{aligned} E(T) &= 3\{E(X^2) + E(X)\} + 2E(Y) \\ &= 3 \left(97 + \frac{19}{2} \right) + 2 \left(\frac{68}{15} \right) \\ &= \frac{9857}{30}. \end{aligned}$$

$$E(T) = \frac{9857}{30}$$

(7)