



Edexcel AS Further Mathematics

8FM0/24 - Further Statistics 2

Friday 16 May 2025

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 2025 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

In a singing competition, two judges ranked the vocal quality of eight singers, A to H.

Rank	1	2	3	4	5	6	7	8
Judge 1	C	B	A	G	E	F	H	D
Judge 2	B	A	C	G	D	E	F	H

(a) Calculate Spearman's rank correlation coefficient between the judges' ranks.

The judges also gave these marks out of 20 for vocal range:

Singer	A	B	C	D	E	F	G	H
Judge 1	18	13	16	12	11	7	13	6
Judge 2	15	17	14	14	9	8	12	8

(b) Without further calculations, explain how the Spearman's rank correlation coefficient between the judges' marks should be calculated.

Model solution

(a) Spearman's coefficient compares the two ranks for each same singer, so first put the ranks alongside each other. There are no tied ranks in this part.

Singer	A	B	C	D	E	F	G	H
Judge 1 rank	3	2	1	8	5	6	4	7
Judge 2 rank	2	1	3	5	6	7	4	8
$d = R_1 - R_2$	1	1	-2	3	-1	-1	0	-1
d^2	1	1	4	9	1	1	0	1

$$\sum d^2 = 18.$$

With $n = 8$ singers, use $r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$.

$$r_s = 1 - \frac{6(18)}{8(8^2 - 1)} = 1 - \frac{108}{504} = \frac{11}{14} = 0.785714 \dots$$

$$r_s = 0.786 \text{ (or } 11/14\text{)}.$$

(4)

(b) The marks contain ties (for example, Judge 1 gives 13 twice). Spearman's method therefore needs average ranks for each tied group; the usual shortcut involving $\sum d^2$ is not used without a tie correction. Rank each judge's marks, assigning average ranks to ties, then calculate the product moment correlation coefficient of the two sets of ranks.

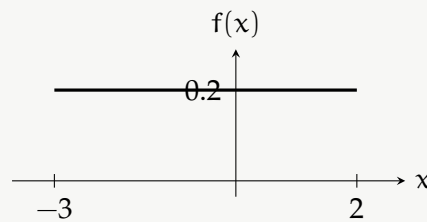
Rank both lists using average ranks for ties, then find the PMCC of the ranks.

(2)

Question 2

Question 2

The graph of the probability density function of the continuous uniform random variable X is shown.



- (a) Write down $P(X = 1)$. (b) Find $E(X)$. (c) Find $\text{Var}(X)$.
 (d) Sketch the cumulative distribution function for $-3 \leq x \leq 2$, labelling points where it touches or crosses the axes.
 (e) Find $P(X^2 > 1.96)$.

Model solution

(a) X is continuous, so the probability of one exact value is zero.

$$P(X = 1) = 0.$$

(1)

(b) A uniform distribution is centred at the midpoint of its interval.

$$E(X) = \frac{-3 + 2}{2} = -\frac{1}{2}.$$

$$E(X) = -\frac{1}{2}.$$

(1)

(c) For $X \sim U(a, b)$, $\text{Var}(X) = \frac{(b - a)^2}{12}$. Here the interval has width $2 - (-3) = 5$.

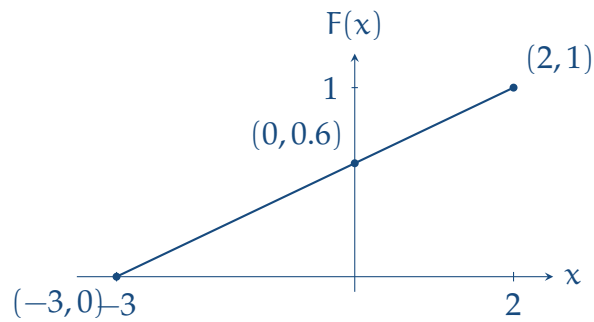
$$\text{Var}(X) = \frac{5^2}{12} = \frac{25}{12}.$$

$$\text{Var}(X) = \frac{25}{12}.$$

(1)

(d) A CDF accumulates area under the density. Since the density is constant at $1/5$ on $[-3, 2]$, the accumulated area increases linearly from 0 to 1.

$$F(x) = \frac{x - (-3)}{5} = \frac{x + 3}{5}, \quad -3 \leq x \leq 2.$$



A straight line from $(-3, 0)$ to $(2, 1)$, namely $F(x) = (x + 3)/5$ on this interval.

(3)

(e) First solve the inequality: $X^2 > 1.96$ means $|X| > \sqrt{1.96} = 1.4$. This gives two regions, both of which lie partly in the support $[-3, 2]$.

$$\begin{aligned} P(X^2 > 1.96) &= P(X < -1.4) + P(X > 1.4) \\ &= \frac{-1.4 - (-3)}{5} + \frac{2 - 1.4}{5} \\ &= \frac{1.6}{5} + \frac{0.6}{5} = \frac{11}{25}. \end{aligned}$$

$$P(X^2 > 1.96) = \frac{11}{25} = 0.44.$$

(3)

Question 3

Question 3

Dina collects data from 20 marathon runners, relating BMI x kg m^{-2} to average marathon speed v m s^{-1} . The summary statistics are

$$\bar{x} = 19.8, \quad \bar{v} = 5.2, \quad \sum xv = 2056.63, \quad S_{xx} = 15.78, \quad S_{vv} = 0.94.$$

(a) Show that $S_{xv} = -2.57$. (b) Find the PMCC. (c) At 5%, test whether there is negative correlation, stating hypotheses and critical value.

(d) Find the regression line of v on x , $v = a + bx$, to 2 d.p. (e) Find the RSS.

For runner A, $x = 20.9$ and the residual is 0.18; for runner B, $x = 19.7$ and the residual is -0.22 . (f) Determine who took the shorter time, showing working.

Model solution

(a) The corrected sum of products removes the effect of the sample means. Use $S_{xv} = \sum xv - \frac{(\sum x)(\sum v)}{n}$.

$$S_{xv} = 2056.63 - \frac{(19.8 \times 20)(5.2 \times 20)}{20} = 2056.63 - 2059.20 = -2.57.$$

$$S_{xv} = -2.57.$$

(1)

(b) The PMCC standardises the corrected sum of products by the spreads of both variables.

$$r = \frac{S_{xv}}{\sqrt{S_{xx}S_{vv}}} = \frac{-2.57}{\sqrt{(15.78)(0.94)}} = -0.667291 \dots$$

$$r = -0.667 \text{ (3 s.f.)}$$

(2)

(c) The question asks specifically about a negative population correlation, so use a one-tailed PMCC test.

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

For $n = 20$, the 5% one-tailed critical value from the PMCC table is -0.3783 (equivalently, the magnitude is 0.3783). Reject H_0 when $r < -0.3783$.

$$-0.667 < -0.3783.$$

Reject H_0 . At the 5% level there is evidence of negative correlation between BMI and average marathon speed.

(3)

(d) For the regression of v on x , the gradient is $b = S_{xv}/S_{xx}$. Then choose the intercept so that the line passes through $(\bar{x}, \bar{v})$.

$$b = \frac{-2.57}{15.78} = -0.162864\dots$$

$$a = 5.2 - (-0.162864\dots)(19.8) = 8.4247\dots$$

$$v = 8.42 - 0.16x.$$

(3)

(e) The RSS is the unexplained variation in v . For a linear regression it is $S_{vv}(1 - r^2)$.

$$\text{RSS} = 0.94(1 - (-0.667291\dots)^2) = 0.521\dots$$

$$\text{RSS} = 0.52 \text{ (approximately).}$$

(2)

(f) A residual is $v - \hat{v}$, so add it to the fitted speed to obtain each actual speed. Over the same marathon distance, the greater speed corresponds to the shorter time.

$$v_A = 8.42 - 0.16(20.9) + 0.18 = 5.256 \text{ m s}^{-1},$$

$$v_B = 8.42 - 0.16(19.7) - 0.22 = 5.048 \text{ m s}^{-1}.$$

Runner A has the greater average speed, so A took the shorter time.

(3)

Question 4

Question 4

The time T seconds for a computer process has CDF

$$F(t) = \begin{cases} 0, & t < 0, \\ at^4 - \frac{1}{8}t^3 + bt^2, & 0 \leq t \leq 4, \\ 1, & t > 4, \end{cases}$$

where a and b are constants.

(a) Find $P(T < 2)$ in terms of a and b . (b) Find the PDF for all t .

Given $P(T < 2) = 11/16$, (c) find the exact mode of T , showing all stages.

Model solution

(a) For a continuous variable, $P(T < 2) = F(2)$. Substitute $t = 2$ into the middle part because 2 lies in $[0, 4]$.

$$P(T < 2) = a(2)^4 - \frac{1}{8}(2)^3 + b(2)^2 = 16a - 1 + 4b.$$

$$P(T < 2) = 16a + 4b - 1.$$

(2)

(b) The density is the derivative of the CDF where it is differentiable; it is zero outside the possible time interval.

$$f(t) = \begin{cases} 4at^3 - \frac{3}{8}t^2 + 2bt, & 0 \leq t \leq 4, \\ 0, & \text{otherwise.} \end{cases}$$

$$f(t) = 4at^3 - \frac{3}{8}t^2 + 2bt \text{ for } 0 \leq t \leq 4, \text{ and } 0 \text{ otherwise.}$$

(2)

(c) First determine the constants. The given probability and the fact that a CDF must satisfy $F(4) = 1$ give two simultaneous equations.

$$16a + 4b - 1 = \frac{11}{16} \Rightarrow 16a + 4b = \frac{27}{16},$$

$$F(4) = 256a - \frac{1}{8}(64) + 16b = 1 \Rightarrow 256a + 16b = 9.$$

$$\begin{aligned} 64a + 16b &= \frac{27}{4}, \\ 256a + 16b &= 9, \end{aligned} \Rightarrow 192a = \frac{9}{4},$$

$$a = \frac{3}{256}, \quad 16b = 9 - 256\left(\frac{3}{256}\right) = 6, \text{ so } b = \frac{3}{8}.$$

The mode is where the PDF is greatest. Find stationary values of f , then compare them with the endpoints of its support.

$$f(t) = \frac{3}{64}t^3 - \frac{3}{8}t^2 + \frac{3}{4}t,$$

$$\frac{df}{dt} = \frac{9}{64}t^2 - \frac{3}{4}t + \frac{3}{4} = 0.$$

$$9t^2 - 48t + 48 = 0 \quad \Rightarrow \quad 3t^2 - 16t + 16 = 0 \quad \Rightarrow \quad (3t - 4)(t - 4) = 0.$$

$$t = \frac{4}{3} \quad \text{or} \quad t = 4.$$

At $t = 4$, the density is at the endpoint and is a minimum (indeed $f(4) = 0$); also $f(0) = 0$. Thus the interior stationary point $t = 4/3$ is the maximum density.

The exact mode of T is $\frac{4}{3}$ s.

(7)