Please check the examination details below before entering your candidate information

Candidate surname	Other names
StemUp Model Solution	

Pearson Edexcel Level 3 GCE

Centre Number	Candidate Number

Thursday 14 May 2020

Afternoon	Paper Reference **8FM0/24**

Further Mathematics

Advanced Subsidiary
Further Mathematics options
24: Further Statistics 2
(Part of option G only)

You must have:
Mathematical Formulae and Statistical Tables (Green), calculator

Total Marks

Candidates may use any calculator allowed by Pearson regulations. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulae stored in them.

Instructions

- Use **black** ink or ball-point pen.
- If pencil is used for diagrams/sketches/graphs it must be dark (HB or B).
- **Fill in the boxes** at the top of this page with your name, centre number and candidate number.
- Answer **all** questions and ensure that your answers to parts of questions are clearly labelled.
- Answer the questions in the spaces provided – *there may be more space than you need.*
- You should show sufficient working to make your methods clear. Answers without working may not gain full credit.
- Values from statistical tables should be quoted in full. If a calculator is used instead of tables the value should be given to an equivalent degree of accuracy.
- Inexact answers should be given to three significant figures unless otherwise stated.

Information

- A booklet 'Mathematical Formulae and Statistical Tables' is provided.
- The total mark for this part of the examination is 40. There are 4 questions.
- The marks for **each** question are shown in brackets – *use this as a guide as to how much time to spend on each question.*

Advice

- Read each question carefully before you start to answer it.
- Try to answer every question.
- Check your answers if you have time at the end.

Turn over ▶

P62659A

©2020 Pearson Education Ltd.

1/1/1/1/1/

P 6 2 6 5 9 A 0 1 1 6

Pearson

1. An estate agent in *Tornep* believes that houses further from the railway station are more expensive than those that are closer. She took a random sample of 22 three-bedroom houses in *Tornep* and calculated the product moment correlation coefficient between the house price and the distance from the station to be 0.3892

Stating your hypotheses clearly, use a 5% level of significance to test the estate agent's belief. State the critical region used in your test.

(3)

H_0: $\rho = 0$ H_1: $\rho > 0$

$n = 22$, 5% one-tailed critical value = 0.3598

Critical region: $r \geq 0.3598$

$0.3892 > 0.3598$, so r lies in the critical region: reject H_0

There is evidence to support the estate agent's belief that houses further from the railway station are more expensive.

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Question 1 continued

(Total for Question 1 is 3 marks)

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

2. Mary, Jahil and Dawn are judging the cakes in a village show. They have 5 features to consider and each feature is awarded up to 5 points. The total score the judges gave each cake are given in the table below.

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's scores and Jahil's scores.

(4)

(b) Calculate Spearman's rank correlation coefficient between Jahil's scores and Dawn's scores.

(3)

The judges discussed their interpretation of the points system and agreed that the first prize should go to cake *C*.

(c) Explain how different interpretations of the points system could give rise to the results in part (a) and part (b).

(2)

(a) Ranks (highest score = 1):

Cake	A	B	C	D	E	F	G	H	I
Mary	3	4	1	8	2	5	7	9	6
Jahil	2	5	3	9	1	4	6	8	7
d	1	−1	−2	−1	1	1	1	1	−1
d^2	1	1	4	1	1	1	1	1	1

$\Sigma d^2 = 1 + 1 + 4 + 1 + 1 + 1 + 1 + 1 + 1 = 12$

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)} = 1 - \frac{6 \times 12}{9 \times 80} = 1 - \frac{72}{720} = 0.9$$

(b) Ranks (highest score = 1, tied ranks averaged):

Cake	A	B	C	D	E	F	G	H	I
Jahil (j)	2	5	3	9	1	4	6	8	7
Dawn (k)	7.5	6	9	2	7.5	3	4.5	1	4.5

$\Sigma j = \Sigma k = 45, \quad \Sigma j^2 = 285, \quad \Sigma k^2 = 284, \quad \Sigma jk = 176$

$$S_{jj} = 285 - \frac{45^2}{9} = 60, \quad S_{kk} = 284 - \frac{45^2}{9} = 59$$

$$S_{jk} = 176 - \frac{45 \times 45}{9} = 176 - 225 = -49$$

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Question 2 continued

$r_s = \frac{S_{jk}}{\sqrt{(S_{jj}S_{kk})}} = \frac{-49}{\sqrt{(60 \times 59)}} = -0.823558...$

$r_s = -0.824$ (3 s.f.)

(c) Mary and Jahil awarded points for good features (a high score means a good cake), whereas Dawn awarded points for poor features (a low score means a good cake).

Both correlations are strong: Mary and Jahil positive (0.9) and Jahil and Dawn negative (−0.824), so in fact all three judges largely agree.

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 2 continued

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Question 2 continued

(Total for Question 2 is 9 marks)

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

3. The continuous random variable X has cumulative distribution function

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

where p and k are constants.

(a) Find the value of p and the value of k.

(4)

Given that $E(X) = \frac{119}{18}$

(b) show that $Var(X) = 2.05$ to 3 significant figures.

(6)

(c) Write down the mode of X.

(1)

(d) Find the exact value of the constant a such that $P(X \leqslant a) = \frac{7}{27}$

(3)

(a) $F(4) = 0$: $4p - k\sqrt{4} = 0 \Rightarrow 4p - 2k = 0 \Rightarrow k = 2p$

$F(9) = 1$: $9p - k\sqrt{9} = 1 \Rightarrow 9p - 3k = 1$

$9p - 3(2p) = 1 \Rightarrow 3p = 1$

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

(b) $F(x) = \frac{1}{3}x - \frac{2}{3}x^{1/2} \Rightarrow f(x) = F'(x) = \frac{1}{3} - \frac{1}{3}x^{-1/2} = \frac{1}{3}(1 - x^{-1/2})$

$E(X^2) = \int_4^9 \frac{1}{3}x^2(1 - x^{-1/2})\,dx = \frac{1}{3}\int_4^9 (x^2 - x^{3/2})\,dx$

$= \frac{1}{3}\left[\frac{x^3}{3} - \frac{2x^{5/2}}{5}\right]_4^9 = \frac{1}{3}\left[\left(\frac{729}{3} - \frac{2 \times 243}{5}\right) - \left(\frac{64}{3} - \frac{2 \times 32}{5}\right)\right]$

$= \frac{1}{3}\left[145.8 - \frac{128}{15}\right] = \frac{1}{3} \times \frac{2059}{15} = \frac{2059}{45} \quad (= 45.7555\ldots)$

$Var(X) = E(X^2) - [E(X)]^2 = \frac{2059}{45} - \left(\frac{119}{18}\right)^2$

$= 45.7555\ldots - 43.7083\ldots = 2.048765\ldots$

$Var(X) = 2.05$ (3 s.f.) as required

(c) $f(x) = \frac{1}{3}\left(1 - \frac{1}{\sqrt{x}}\right)$ is increasing on $4 \leq x \leq 9$, so the maximum is at $x = 9$

Mode = 9

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

Question 3 continued

(d) $F(a) = \frac{7}{27}$: $\frac{1}{3}a - \frac{2}{3}\sqrt{a} = \frac{7}{27} \Rightarrow a - 2\sqrt{a} - \frac{7}{9} = 0$

Let $y = \sqrt{a}$: $y^2 - 2y - \frac{7}{9} = 0 \Rightarrow (y-1)^2 = \frac{16}{9}$

$y - 1 = \pm\frac{4}{3} \Rightarrow y = \frac{7}{3}$ or $y = -\frac{1}{3}$ (not valid since $\sqrt{a} \geqslant 0$)

$\sqrt{a} = \frac{7}{3} \Rightarrow a = \frac{49}{9}$ (which lies in $4 \leqslant a \leqslant 9$)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 3 continued

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Question 3 continued

(Total for Question 3 is 14 marks)

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

4. Some students are investigating the strength of wire by suspending a weight at the end of the wire. They measure the diameter of the wire, d mm, and the weight, w grams, when the wire fails. Their results are given in the following table.

	These 14 points are plotted on page 13														Not yet plotted			
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**

The first 14 points are plotted on the axes on page 13.

(a) On the axes on page 13, complete the scatter diagram for these data.

(1)

(b) Use your calculator to write down the equation of the regression line of w on d.

(2)

(c) With reference to the scatter diagram, comment on the appropriateness of using this linear regression model to make predictions for w for different values of d between 0.5 and 5.4

(1)

The product moment correlation coefficient for these data is $r = 0.987$ (to 3 significant figures).

(d) Calculate the residual sum of squares (RSS) for this model.

(2)

Robert, one of the students, suggests that the model could be improved and intends to find the equation of the line of regression of w on u, where $u = d^2$
He finds the following statistics

$$S_{wu} = 5721.625 \qquad S_{uu} = 1482.619 \qquad \sum u = 157.57$$

(e) By considering the physical nature of the problem, give a reason to support Robert's suggestion.

(1)

(f) Find the equation of the regression line of w on u.

(3)

(g) Find the residual sum of squares (RSS) for Robert's model.

(2)

(h) State, giving a reason based on these calculations, which of these models better describes these data.

(1)

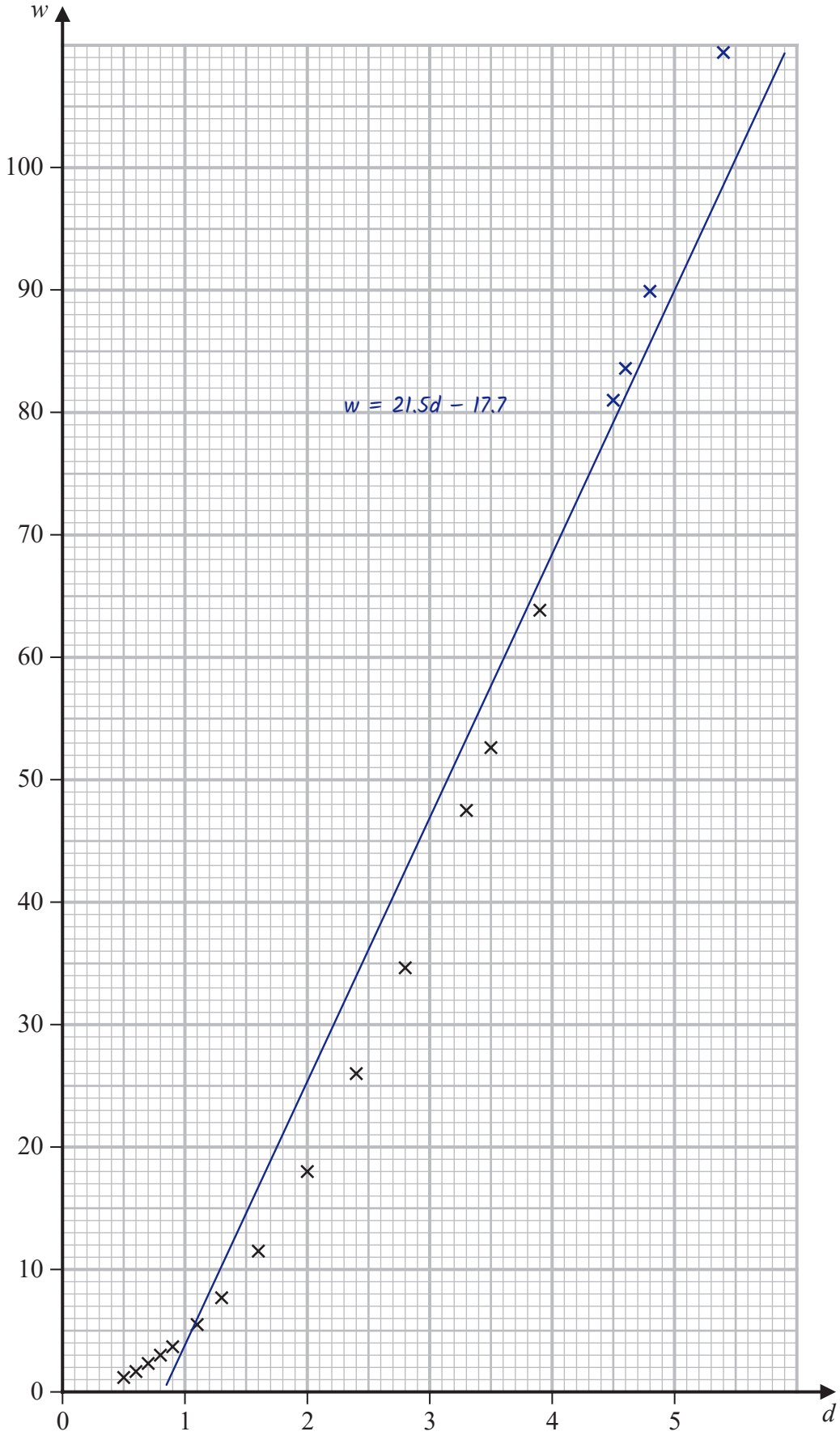
(i) Hence estimate the weight at which a piece of wire with diameter 3 mm will fail.

(1)

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Question 4 continued



Question 4 continued

(a) Points plotted on the scatter diagram on page 13.

(b) $w = 21.5d - 17.7$

(c) Not appropriate: the points lie on a curve, not a straight line. The line (drawn on the diagram) overestimates w for middle values of d and underestimates w at the ends, so predictions would be unreliable.

(d) $\Sigma w = 643.6, \quad \Sigma w^2 = 45178.68$

$$S_{ww} = \Sigma w^2 - \frac{(\Sigma w)^2}{n} = 45178.68 - \frac{643.6^2}{18} = 22166.404...$$

$S_{dd} = 46.565, \quad S_{dw} = 1002.806...$

$$RSS = S_{ww} - \frac{(S_{dw})^2}{S_{dd}} = 22166.404... - \frac{1002.806...^2}{46.565} = 570.329...$$

$RSS = 570$ (3 s.f.)

(e) A thicker wire should be stronger: the strength of the wire depends on its cross-sectional area, which is proportional to d^2.

(f) gradient $c = \frac{S_{wu}}{S_{uu}} = \frac{5721.625}{1482.619} = 3.859133...$

intercept $f = \bar{w} - c\bar{u} = \frac{643.6}{18} - 3.859133... \times \frac{157.57}{18} = 1.973127...$

$w = 1.97 + 3.86u$

(g) $RSS = S_{ww} - \frac{(S_{wu})^2}{S_{uu}} = 22166.404... - \frac{5721.625^2}{1482.619} = 85.888...$

$RSS = 85.9$ (3 s.f.)

(h) Robert's model (w on u) better describes the data, since its RSS (85.9) is much smaller than the RSS for the linear model in d (570).

(i) $u = 3^2 = 9$: $w = 1.973127... + 3.859133... \times 9 = 36.705...$

The wire will fail at a weight of approximately 36.7 grams.

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Question 4 continued

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Question 4 continued

(Total for Question 4 is 14 marks)

TOTAL FOR FURTHER STATISTICS 2 IS 40 MARKS

END

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA