MODEL ANSWERS

Write your name here

Surname	Other names

Pearson Edexcel Level 3 GCE

Centre Number | | | | | |

Candidate Number | | | | |

Further Mathematics

Advanced
Further Mathematics Option 2
Paper 4: Further Statistics 2

Sample Assessment Material for first teaching September 2017

Time: 1 hour 30 minutes

Paper Reference **9FM0/4E**

You must have:
Mathematical Formulae and Statistical Tables, calculator

Total Marks

Candidates may use any calculator permitted by Pearson regulations. Calculators must not have the facility for algebraic 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.
- Answers should be given to three significant figures unless otherwise stated.

Information

- A booklet 'Mathematical Formulae and Statistical Tables' is provided.
- There are 7 questions in this question paper. The total mark for this paper is 75.
- 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 ▶

S54445A

1/1/1/1/1/

S 5 4 4 4 5 A 0 1 2 5

Pearson

Answer ALL questions. Write your answers in the spaces provided.

1. The three independent random variables A, B and C each have a continuous uniform distribution over the interval [0, 5].

(a) Find the probability that A, B and C are all greater than 3

(3)

The random variable Y represents the maximum value of A, B and C.

The cumulative distribution function of Y is

$$F(y) = \begin{cases} 0 & y < 0 \\ \dfrac{y^3}{125} & 0 \leqslant y \leqslant 5 \\ 1 & y > 5 \end{cases}$$

(b) Using algebraic integration, show that Var $(Y) = 0.9375$

(4)

(c) Find the mode of Y, giving a reason for your answer.

(2)

(d) Describe the skewness of the distribution of Y. Give a reason for your answer.

(1)

(e) Find the value of k such that $P(k < Y < 2k) = 0.189$

(3)

a) continuous across [0,5] ⇒ p.d.f. height $\frac{1}{5}$

so $P(i>3) = \frac{2}{5}$, $i = A, B, C$

independent ⇒ multiply Ps together

$$\left(\frac{2}{5}\right)^3 = \frac{8}{125}$$

b) cumulative → p.d.f. by differentiation

$$f(y) = \frac{3y^2}{125}$$

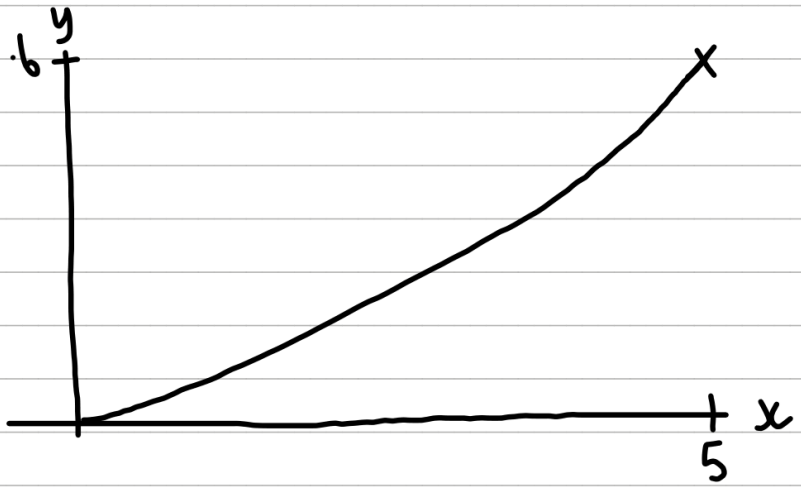
$$E(Y) = \int_0^5 \frac{3y^2 \cdot y}{125}\,dy = \left[\frac{3y^4}{500}\right]_0^5 = \frac{15}{4}$$

Question 1 continued

$$\text{Var}(Y) = \int_0^5 \left(\frac{3y^4}{125}\right) dy - \left(\frac{15}{4}\right)^2 = 0.9375$$

$$E(Y^2) - (E(Y))^2$$

c)



$\therefore$ mode = 5

d) mode > mean $\therefore$ negative skew

e) using F(y): $\frac{(2k)^3}{125} - \frac{k^3}{125} = 0.189$

$$\frac{7k^3}{125} = 0.189$$

$$\Rightarrow k = 1.5$$

Question 1 continued

Question 1 continued

(Total for Question 1 is 13 marks)

2. A researcher claims that, at a river bend, the water gradually gets deeper as the distance from the inner bank increases. He measures the distance from the inner bank, b cm, and the depth of a river, s cm, at 7 positions. The results are shown in the table below.

Position	A	B	C	D	E	F	G
Distance from inner bank b cm	100	200	300	400	500	600	700
Depth s cm	60	75	85	76	110	120	104

The Spearman's rank correlation coefficient between b and s is $\frac{6}{7}$

(a) Stating your hypotheses clearly, test whether or not the data provides support for the researcher's claim. Use a 1% level of significance.

(4)

(b) Without re-calculating the correlation coefficient, explain how the Spearman's rank correlation coefficient would change if

(i) the depth for G is 109 instead of 104

(ii) an extra value H with distance from the inner bank of 800 cm and depth 130 cm is included.

(3)

The researcher decided to collect extra data and found that there were now many tied ranks.

(c) Describe how you would find the correlation with many tied ranks.

(2)

a) $H_0: \rho = 0$, $H_1: \rho > 0$

using table, the critical value @ 1% level is 0.8929

$r_s = \frac{6}{7} < 0.8929$

$\therefore$ there is not significant evidence to reject H_0

$\rightarrow$ there is insufficient evidence to conclude that there is positive correlation between water depth & distance from the bank @ the 1% level

Question 2 continued

b)i. the ranks won't change, therefore the Spearman's rank remains the same

ii. Spearman's rank correlation coefficient will increase.

the ranks are the same for depth & distance → $d = 0$

n has increased

new point follows trend of large b ⇔ large s

$\therefore r_s$ will increase

c) give both values the mean of the values of the tied ranks

then use the PMCC

(Total for Question 2 is 9 marks)

3. A nutritionist studied the levels of cholesterol, X mg/cm^3, of male students at a large college. She assumed that X was distributed N(μ, σ^2) and examined a random sample of 25 male students. Using this sample she obtained unbiased estimates of μ and σ^2 as $\hat{\mu}$ and $\hat{\sigma}^2$

A 95% confidence interval for μ was found to be (1.128, 2.232)

(a) Show that $\hat{\sigma}^2 = 1.79$ (correct to 3 significant figures)

(4)

(b) Obtain a 95% confidence interval for σ^2

(3)

a) t-distribution used

95% confidence interval for μ uses t-value of 2.604

$$2.604 \times \frac{\hat{\sigma}}{\sqrt{25}} = \frac{1}{2}(2.232 - 1.128)$$

$$\hat{\sigma} = \frac{2.76}{2.064} = 1.3372\ldots$$

$$\hat{\sigma}^2 = 1.788\ldots$$
$$= 1.79 \text{ (3s.f.)}$$

b) $12.401 < \frac{24 \times 1.79}{\sigma^2} < 39.364$

$$1.09 < \sigma^2 < 3.46$$

Question 3 continued

(Total for Question 3 is 7 marks)

4. The times, x seconds, taken by the competitors in the 100 m freestyle events at a school swimming gala are recorded. The following statistics are obtained from the data.

	No. of competitors	Sample mean $\bar{x}$	$\sum x^2$
Girls	8	83.1	55 746
Boys	7	88.9	56 130

Following the gala, a mother claims that girls are faster swimmers than boys. Assuming that the times taken by the competitors are two independent random samples from normal distributions,

(a) test, at the 10% level of significance, whether or not the variances of the two distributions are the same. State your hypotheses clearly.

(7)

(b) Stating your hypotheses clearly, test the mother's claim. Use a 5% level of significance.

(6)

a) $H_0: \sigma_G^2 = \sigma_B^2$; $H_1: \sigma_G^2 \neq \sigma_B^2$

$S_B^2 = \frac{1}{6}(56130 - 7 \times 88.9^2) = \frac{807.53}{6} = 134.6$

($\frac{1}{N-1}$, Σx^2, $N\bar{x}^2$, $\bar{x}$)

$S_G^2 = \frac{1}{7}(55746 - 8 \times 83.1^2) = \frac{501.12}{7} = 71.58$

F-distribution: $\frac{S_B^2}{S_G^2} = 1.880$

critical value $F_{6,7} = 3.87$

calculated value isn't significant ⇒ variances can be treated as the same

b) $H_0: \mu_B = \mu_G$; $H_1: \mu_B > \mu_G$

pool variance estimates: $s^2 = \frac{6 \times 1346 + 7 \times 71.58}{13}$

$= 100.665\ldots$

Question 4 continued

test stat. $t = \dfrac{88.9 - 83.1}{S\sqrt{\frac{1}{7} + \frac{1}{8}}} = 1.12$

Critical value $t_{13}(0.05) = 1.771$

reject H_1, insufficient evidence to support mother's claim

Question 4 continued

Question 4 continued

(Total for Question 4 is 13 marks)

5. Scaffolding poles come in two sizes, long and short. The length L of a long pole has the normal distribution $N(19.6, 0.6^2)$. The length S of a short pole has the normal distribution $N(4.8, 0.3^2)$. The random variables L and S are independent.

A long pole and a short pole are selected at random.

(a) Find the probability that the length of the long pole is more than 4 times the length of the short pole. Show your working clearly.

(6)

Four short poles are selected at random and placed end to end in a row. The random variable T represents the length of the row.

(b) Find the distribution of T.

(3)

(c) Find $P(|L - T| < 0.2)$

(4)

a) define $X = L - 4S$ so $E(X) = 19.6 - 4 \times 4.8$

($E(L)$, $E(S)$)

$E(X) = 0.4$

$Var(X) = Var(L) + 4^2 Var(S)$

$= 0.6^2 + 16 \times 0.3^2$

$= 1.8$

$L > 4S \rightarrow L - 4S > 0 \rightarrow X > 0 \rightarrow$ want to find $P(X > 0)$

$P(X > 0) = \left[P\left(z > \frac{0 - 0.4}{\sqrt{1.8}} = -0.298\ldots\right)\right]$

shift by $E(X)$ & normalise by $\sigma(X)$ to get standard $\sim N(0,1)$ distribution

$= 0.617202$

b) $T = S_1 + S_2 + S_3 + S_4$

$T \sim N(19.2, 0.36)$ $\quad E(T) = 19.2$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 5 continued

$\text{Var}(T) = 0.6^2 = 0.36$

c) define $Y = L - T$

$E(Y) = E(L) - E(T) = 0.4$

$\text{Var}(Y) = \text{Var}(L) + \text{Var}(T) = 0.72$

modulus ⇒ +ve & −ve

we require $P(-0.2 < Y < 0.2) = 0.16708\ldots$

Question 5 continued

Question 5 continued

(Total for Question 5 is 13 marks)

6. A random sample of 10 female pigs was taken. The number of piglets, x, born to each female pig and their average weight at birth, m kg, was recorded. The results were as follows:

Number of piglets, x	4	5	6	7	8	9	10	11	12	13
Average weight at birth, m kg	1.50	1.20	1.40	1.40	1.23	1.30	1.20	1.15	1.25	1.15

(You may use $S_{xx} = 82.5$ and $S_{mm} = 0.12756$ and $S_{xm} = -2.29$)

(a) Find the equation of the regression line of m on x in the form $m = a + bx$ as a model for these results.

(2)

(b) Show that the residual sum of squares (RSS) is 0.064 to 3 decimal places.

(2)

(c) Calculate the residual values.

(2)

(d) Write down the outlier.

(1)

(e) (i) Comment on the validity of ignoring this outlier.

(ii) Ignoring the outlier, produce another model.

(iii) Use this model to estimate the average weight at birth if $x = 15$

(iv) Comment, giving a reason, on the reliability of your estimate.

(5)

a) $b = \frac{S_{xm}}{S_{xx}} = -0.0277576$

$a = \bar{m} - b\bar{x} = 1.278 + 0.0277576 \times 8.5 = 1.5139$

$m = 1.5139 - 0.02775x$

b) $RSS = 0.12756 - \frac{(-2.29)^2}{82.5}$ (S_{xm}, S_{xx})

$= 0.06399$

Question 6 continued

residuals

c)

x	m	$m = a + bx$	ε
4	1·50	1·4029	+ 0·0971
5	1·20	1·3752	− 0·1752
6	1·40	1·3474	+ 0·0526
7	1·40	1·3196	+ 0·0804
8	1·23	1·2919	− 0·0619
9	1·30	1·2641	+ 0·0359
10	1·20	1·2364	− 0·0364
11	1·15	1·2086	− 0·0586
12	1·25	1·1808	+ 0·0692
13	1·15	1·1531	− 0·0031

d) (5, 1·2) is an outlier (residual doesn't fit trend / order of magnitude)

e) i. it is valid data so its use is justified, however it doesn't fit the pattern of the residuals — it may contain an error that could confuse the results, so its removal is also justified

you can argue for either case as long as you support your argument

ii. $a = \bar{m} - b\bar{x}$
$= 1·28667 + 0·03765 \times 8·88889$
$= 1·6213$

$\rightarrow m = 1·6213 - 0·03765x$

iii. $m = 1·6213 - (0·03765 \times 15)$
$= 1·056$

Question 6 continued

iv. model is only reliable if we limit the values to the range given. 15 is outside the range, so the model is not reliable.

↑ extrapolation

Question 6 continued

(Total for Question 6 is 12 marks)

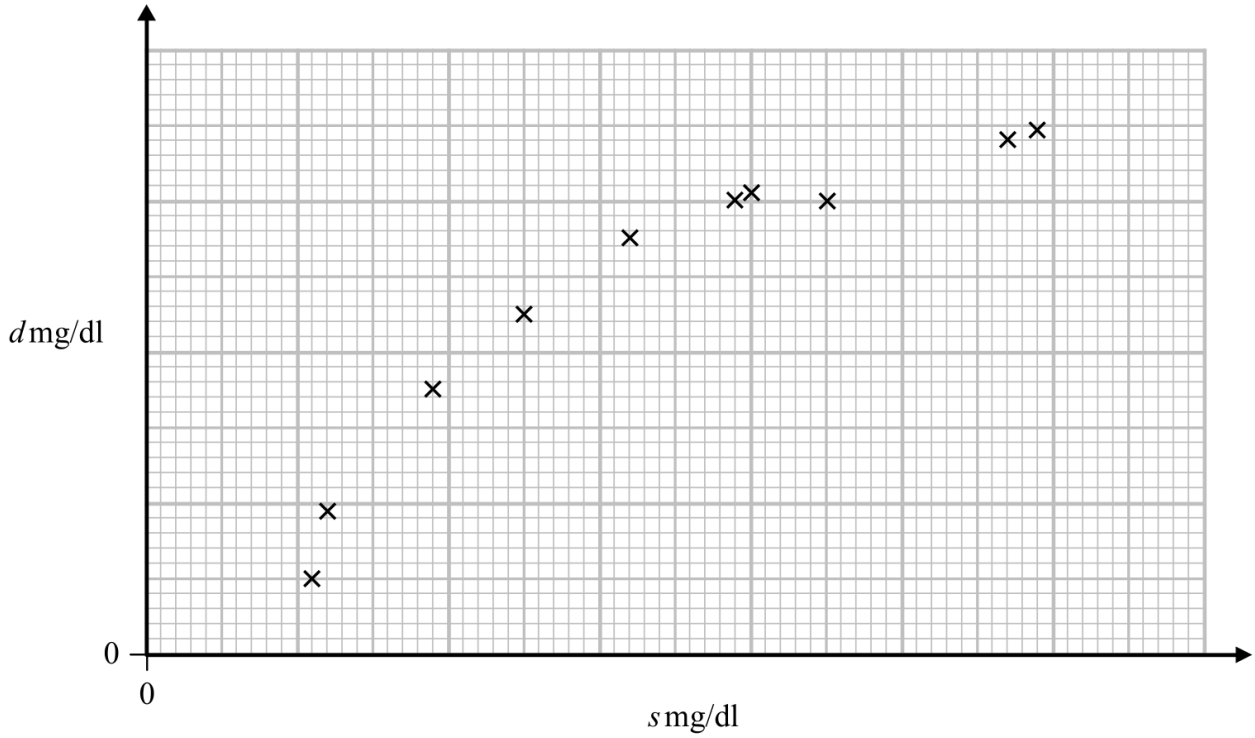
7. Over a period of time, researchers took 10 blood samples from one patient with a blood disease. For each sample, they measured the levels of serum magnesium, s mg/dl, in the blood and the corresponding level of the disease protein, d mg/dl. One of the researchers coded the data for each sample using $x = 10s$ and $y = 10(d - 9)$ but spilt ink over his work.

The following summary statistics and unfinished scatter diagram are the only remaining information.

$\sum d^2 = 1081.74$ $\quad S_{ds} = 59.524$

and

$\sum y = 64$ $\quad S_{xx} = 2658.9$



(a) Use the formula for S_{xx} to show that $S_{ss} = 26.589$

(3)

(b) Find the value of the product moment correlation coefficient between s and d.

(4)

(c) With reference to the unfinished scatter diagram, comment on your result in part (b).

(1)

a) $S_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{10}$

$x = 10s$

$S_{xx} = \Sigma(10s)^2 - \frac{(\Sigma 10s)^2}{10}$

Question 7 continued

$$2658.9 = 100\Sigma s^2 - \frac{100(\Sigma s)^2}{10}$$

$$2658.9 = 100 S_{xx} \Rightarrow S_{xx} = 26.589$$

b) $64 = \sum_1^{10} 10(d_i - 9)$ from $y = 10(d-9)$

$$= 10\sum_1^{10} d_i - 900$$

$$\sum_1^{10} d_i = 96.4$$

$$S_{dd} = 1081.74 - \frac{(96.4)^2}{10} = 152.444$$

$$r = 0.935$$

c) the linear correlation is significant (0.935 is close to 1), however the scatter diagram implies there is a non-linear trend between magnesium levels & the levels of the disease protein

Make sure you reference (b), the diagram, & the context ⇒ do r & the scatter plot match up?

Question 7 continued

Question 7 continued

(Total for Question 7 is 8 marks)

TOTAL FOR PAPER IS 75 MARKS