

Please check the examination details below before entering your candidate information

Candidate surname					Other names				
Centre Number				Candidate Number					
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Pearson Edexcel Level 3 GCE

Time 1 hour 30 minutes

Paper reference **9FM0/4B**

Further Mathematics

Advanced

PAPER 4B: Further Statistics 2

June 2022, Shadow Set 1

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

Candidates may use any calculator permitted 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 the 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.
- There are 8 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 ►

1. Emmanuel is investigating a possible relationship between total August rainfall, p mm, and apple production y thousand tonnes, for orchards in a particular region. He uses 30 years of past data to produce the following summary statistics for a linear regression model, with apple production as the dependent variable.

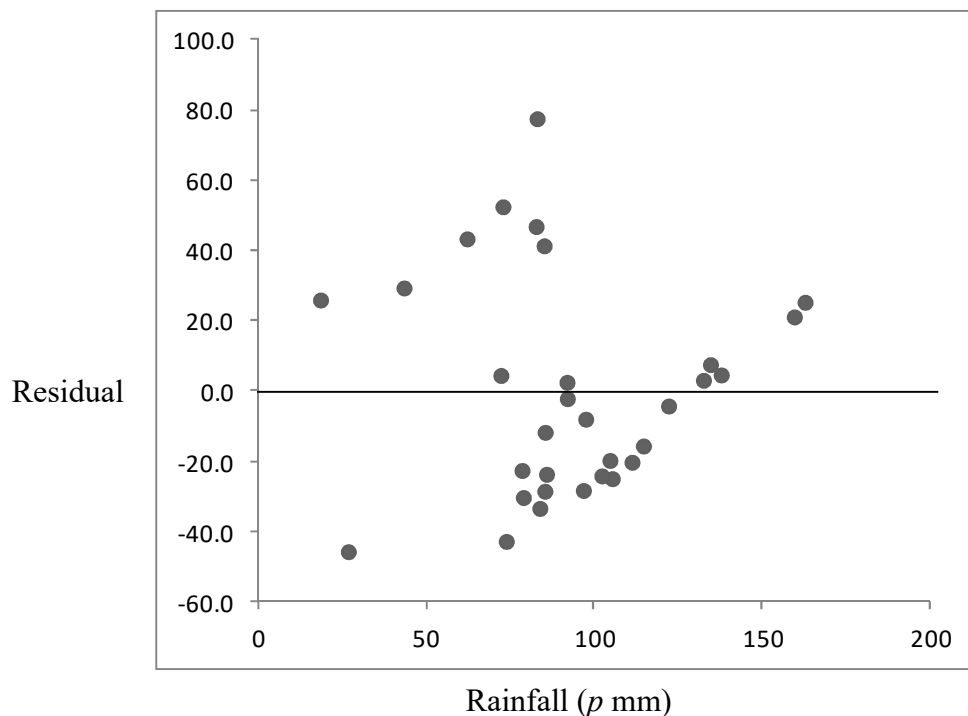
$$\text{Residual Sum of Squares (RSS)} = 28\,378 \qquad S_{pp} = 32\,257 \qquad S_{yy} = 30\,513$$

$$\text{least squares regression line:} \qquad \text{gradient} = 0.257 \qquad y\text{-intercept} = 108$$

- (a) Use the regression model to predict the apple production for a total August rainfall of 100 mm

(1)

He also produces the following residual plot for the data.



- (b) Explain what you understand by the term residual.

(1)

- (c) Calculate the product moment correlation coefficient between p and y

(2)

- (d) Explain why the linear model may not be a good fit for the data

- (i) with reference to your answer to part (c)
(ii) with reference to the residual plot.

(2)

Question 1 continued

Emmanuel also collects data on average August temperature, t °C, for each of these 30 years.

For a linear regression model of t on p the following summary statistic is found.

$$\text{Residual Sum of Squares (RSS)} = 86\,754$$

Emmanuel concludes that since this model has a larger RSS, there must be a weaker linear relationship between t and p than between y and p (where $\text{RSS} = 28\,378$)

- (e) State, giving a reason, whether or not you agree with the reasoning that led to Emmanuel's conclusion.

(1)

(Total for Question 1 is 7 marks)

-
2. A factory produces packets of cheese and onion flavour and packets of salt and vinegar flavour crisps. Independent samples, one of packets of cheese and onion flavour crisps and one of packets of salt and vinegar flavour crisps, are taken. The table shows information about the weights of the packets of cheese and onion flavour crisps, C grams, and the weights of the packets of salt and vinegar flavour crisps, S grams.

	Sample size	Mean weight of random sample (grams)	Known population standard deviation of weights (grams)
Cheese and onion	200	28.20	0.15
Salt and vinegar	150	28.15	0.09

- (a) Find a 95% confidence interval for the mean weight of packets of cheese and onion crisps. Give your answer to 4 significant figures.

(3)

Terrence claims that the mean weight of the population of packets of cheese and onion crisps is greater than the mean weight of the population of packets of salt and vinegar crisps. A test of Terrence's claim is carried out.

- (b) (i) Specify the approximate distribution of $\bar{C} - \bar{S}$ under the null hypothesis of the test.

(3)

- (ii) Explain the relevance of the large sample sizes to your answer to part (i).

(1)

- (c) Complete the hypothesis test using a 5% level of significance.

You should state your hypotheses and the value of your test statistic clearly.

(5)

(Total for Question 2 is 12 marks)

3. The random variable $X \sim N(2, 0.3^2)$ and the random variable $Y \sim N(10, 0.2^2)$

X and Y are independent random variables.

A random sample of a independent observations is taken from the distribution of X and one observation is taken from the distribution of Y

The random variable $W = X_1 + X_2 + X_3 + \dots + X_a + bY$ and has the distribution $N(78, 1.5^2)$

Find the value of a and the value of b

(6)

(Total for Question 3 is 6 marks)

4. A doctor believes that a four-week exercise programme can reduce the diastolic blood pressure of her patients. She takes a random sample of patients and records their diastolic blood pressure before the exercise programme and again after the exercise programme.

Patient	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>
Diastolic pressure before	95	93	67	88	72	69	74	87
Diastolic pressure after	80	70	74	67	70	81	65	75

- (a) Using a 5% level of significance, carry out an appropriate test of the doctor's belief. You should state your hypotheses, test statistic and critical value.

(7)

- (b) State the assumption made about the resting heart rates that was required to carry out the test.

(1)

(Total for Question 4 is 8 marks)

5. The concentration of phosphates in water samples is measured in parts per million (ppm). Samples of water were taken at two different sites and the concentration of phosphates was recorded.

For Site *A* the summary statistics are shown below.

	number of samples	s_A^2
Site <i>A</i>	11	5.93

For Site *B* there were 7 samples of air taken.

A test of the hypothesis $H_0 : \sigma_A^2 = \sigma_B^2$ against the hypothesis $H_1 : \sigma_A^2 \neq \sigma_B^2$ is carried out using a 10% level of significance.

- (a) State a necessary assumption required to carry out the test. (1)

Given that the assumption in part (a) holds,

- (b) find the set of values of s_B^2 that would lead to the null hypothesis being rejected, (4)

- (c) find a 95% confidence interval for the variance of the concentration of the air pollutant at Site *A*. (3)

(Total for Question 5 is 8 marks)

6. Finn and Gerald challenge each other to find an estimator for the mean, μ , of the continuous random variable X which has variance σ^2

$X_1, X_2, X_3, \dots, X_n$ are n independent observations taken from X

Finn's estimator is given by

$$F = \frac{1}{2^n - 1} \sum_{r=1}^n 2^{r-1} X_r$$

Gerald's estimator is given by

$$G = \frac{X_1 + X_2 + X_3}{4} + \frac{X_4 + X_5 + \dots + X_n}{4(n-3)}$$

- (a) Show that F and G are both unbiased estimators of μ (5)
- (b) (i) Find $\text{Var}(F)$
(ii) Find $\text{Var}(G)$ (7)

The winner of the challenge is the person who finds the better estimator.

- (c) Determine the winner of the challenge for large values of n .
Give reasons for your answer. (3)

(Total for Question 6 is 15 marks)

7. A square based cuboid is to have a volume of 400 cm^3

The height of the cuboid, H cm, follows a continuous uniform distribution over the interval $[4, 25]$

Find the expected value of the surface area of the cuboid.

Use algebraic integration, rather than your calculator, to evaluate any definite integrals. (7)

(Total for Question 7 is 7 marks)

8. The continuous random variable X has cumulative distribution function given by

$$F(x) = \begin{cases} 0 & x < 2 \\ 0.56x - 0.04x^2 - 0.96 & 2 \leq x \leq 7 \\ 1 & x > 7 \end{cases}$$

(a) Find the exact value of the median of X (2)

(b) Find $P(X < 5 \mid X > 3)$ (3)

The random variable $Y = \frac{1}{X}$

(c) Specify fully the cumulative distribution function of Y (4)

(d) Hence or otherwise find the mode of Y (3)

(Total for Question 8 is 12 marks)

TOTAL FOR PAPER IS 75 MARKS