Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel Level 3 GCE

Centre Number | Candidate Number

Time 1 hour 30 minutes

Paper reference **9FM0/4B**

Further Mathematics

Advanced

PAPER 4B: Further Statistics 2

October 2021 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 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.
- 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 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.
- Good luck with your examination.

Turn over ▶

© Pearson Education Ltd.

1 Maz works at a go-kart race track and is investigating whether people drive faster around the track at different times of the day. He collects information from 9 participants. He wants to calculate the Spearman's rank correlation coefficient for the 9 participants so he ranks the time of day (a lower rank representing an earlier time in the day) and the lap times (a lower rank representing a lower time).

Participant	**Time of day (hh:mm)**	**Lap time (s)**	**Time of day rank**	**Lap time rank**
A	13:00	36.3	1	5
B	13:30	55.7	2	9
C	13:45	36.5	3	*t*
D	14:30	34.7	4	3
E	14:45	36.5	5	*t*
F	15:15	34.9	*s*	4
G	15:15	36.5	*s*	*t*
H	15:30	34.5	8	2
I	16:00	33.1	9	1

(a) Write down the value of s and the value of t

(2)

The full product moment correlation coefficient (pmcc) formula is used with the ranks to calculate the Spearman's rank correlation coefficient instead of $r_s = 1 - \frac{6\sum d^2}{n\left(n^2-1\right)}$ and the value obtained is −0.6383 to 4 significant figures.

(b) Explain why the full pmcc formula is used to carry out the calculation.

(1)

(c) Stating your hypotheses clearly, test whether or not there is evidence to suggest that the later in the day a participant races, the faster they drive around the track.
Use a 5% level of significance.

(4)

© Pearson Education Ltd.

Question 1 continued

(Total for Question 1 is 7 marks)

© Pearson Education Ltd.

2 A company produces two types of smartphone, the *Mercury Plus* and the *Jupiter Power*. The standard deviation of the battery life of the *Mercury Plus* phones is 64.2 minutes and the standard deviation of the battery life of the *Jupiter Power* is 83.3 minutes.

Nish claims that the mean battery life of the *Mercury Plus* phones is more than 60 minutes less than the mean battery life of the *Jupiter Power* phones.

A random sample of 120 *Mercury Plus* phones is found to have a mean battery life of 3525.2 minutes. A random sample of 100 *Jupiter Power* phones is found to have a mean battery life of 3601.4 minutes.

(a) Stating your hypotheses clearly, use a suitable test to assess Nish's belief. Use a 1% level of significance.

(6)

(b) Explain how the hypothesis test in part (a) would be carried out differently if the variances of the battery lives of smartphones were unknown.

(1)

The battery lives for the smartphones are later discovered to follow a skewed distribution.

(c) Describe the effect this would have on the calculations in the hypothesis test in part (a). Give a reason for your answer.

(2)

© Pearson Education Ltd.

Question 2 continued

© Pearson Education Ltd.

Question 2 continued

© Pearson Education Ltd.

Question 2 continued

(Total for Question 2 is 9 marks)

© Pearson Education Ltd.

3 The continuous random variable X has cumulative distribution function given by

$$\mathrm{F}(x) = \begin{cases} 0 & x < 5 \\ 1.125 - \dfrac{225}{8x^2} & 5 \leq x \leq 15 \\ 1 & x > 15 \end{cases}$$

(a) Find $\mathrm{P}(\{X < 9\} \cup \{X > 12\})$

(2)

(b) Find the median of X.

(2)

(c) Find $\mathrm{E}(X^4)$

(3)

(d) (i) Sketch the probability density function of X.

(ii) Describe the skewness of the distribution of X.

(3)

© Pearson Education Ltd.

Question 3 continued

© Pearson Education Ltd.

Question 3 continued

© Pearson Education Ltd.

Question 3 continued

(Total for Question 3 is 10 marks)

© Pearson Education Ltd.

4 A environmental scientist is investigating the relationship between the distance downstream of a factory, x metres, and the dissolved oxygen content of river water, y parts per million.

From a random sample of 30 specimens of river water in the UK, the following results were obtained

$$S_{xy} = -157.4 \qquad S_{yy} = 19.98 \qquad \sum x = 479.8 \qquad \sum y = 140.2$$

The product moment correlation coefficient for these data is found to be −0.7399

(a) Interpret the value of this correlation coefficient.

(1)

(b) Show that the equation of the regression line of y on x can be written as

$$y = 5.78 - 0.0695x$$

(4)

The random variable W represents the dissolved oxygen content of river water in parts per **thousand**.

(c) Write down the equation of the regression line of w on x for these data in the form $w = a + bx$

(1)

(d) Show that the residual sum of squares (RSS) for the model for y and x is 9.04 correct to two decimal places.

(1)

One of the river water specimens was collected at 3.33 m downstream of a factory and had 8 ppm of dissolved oxygen content.

(e) (i) Calculate this river water specimen's contribution to the residual sum of squares. Give your answer as a percentage

(2)

(ii) Comment on the data for this river water specimen in light of your answer to part (e)(i).

(1)

© Pearson Education Ltd.

Question 4 continued

© Pearson Education Ltd.

Question 4 continued

© Pearson Education Ltd.

Question 4 continued

(Total for Question 4 is 10 marks)

© Pearson Education Ltd.

5 The continuous random variable X is uniformly distributed over the interval $[\delta, 2\delta]$, where δ is an unknown constant.

Three independent observations, X_1, X_2 and X_3, are taken of X and the following estimators for δ are proposed.

$$A = \frac{X_1 + X_2 + X_3}{3}$$

$$B = \frac{5X_1 + 6X_2 + 7X_3}{30}$$

$$C = \frac{13X_1 - 4X_2 + 13X_3}{30}$$

(a) Calculate the bias of A, the bias of B and the bias of C.

(5)

(b) By calculating the variances, explain which of B or C is the better estimator for δ

(4)

(c) Find an unbiased estimate for δ

(1)

© Pearson Education Ltd.

Question 5 continued

© Pearson Education Ltd.

Question 5 continued

© Pearson Education Ltd.

Question 5 continued

(Total for Question 5 is 10 marks)

© Pearson Education Ltd.

6 Agnetha is collecting information on the effect of 'sit-stand' desks on sitting time at work. She takes a random sample of 12 workers who use a conventional office desk and a random sample of 15 workers who use a 'sit-stand' desk.

The sitting times, s minutes, in a normal working day is recorded and summarised in the table below.

You may assume the sitting times for each type of desk may be modelled by independent normal distributions.

	Sample size	$\sum s$	$\sum s^2$
Conventional desk	13	4 085	1 328 577
'Sit-stand'" desk	17	3 612	794 603

(a) Test, at the 2% level of significance, whether or not there is evidence that the variance of the sitting times for those who use a conventional office desk is different from the variance of the sitting times for those who use a 'sit-stand' desk.
You should state your hypotheses clearly.

(7)

The k% confidence interval for the variance of sitting times for those that use the conventional office desk is (1 925.88 , 10 205.34)

(b) Find the value of k

(3)

(c) Calculate a 95% confidence interval for the difference between the mean sitting times of those who use a conventional office desk and the mean sitting times of those who use a 'sit-stand' desk.

(5)

© Pearson Education Ltd.

Question 6 continued

© Pearson Education Ltd.

Question 6 continued

© Pearson Education Ltd.

Question 6 continued

(Total for Question 6 is 15 marks)

© Pearson Education Ltd.

7 Guy is one of several barbers in a barber shop. The time taken for Guy to complete a haircut for a client, H minutes, and the time taken for Guy to complete a shave for a client, S minutes, each follow independent normal distributions.

$$H \sim \mathrm{N}(28.9\ ,\ 4.32^2) \qquad\qquad S \sim \mathrm{N}(14.2,\ 3.24^2)$$

(a) Find the distribution of

(i) $H + S$

(ii) the total time taken to complete 2 haircuts.

(3)

On a particular day, the barber shop has 20 scheduled clients. 19 of them request haircuts and 1 requests a shave. The barbers are randomly assigned to the clients and Guy is assigned two of the clients.

(b) Find the probability that the total time taken to fulfil both client's requests is at most 50 minutes.

(3)

Over a month, there are a large number of clients visiting the barber shop. Guy is randomly assigned m haircuts and 4 shaves. The random variable W is given by

$$W = \left(\sum_{i=1}^{m} H_i\right) - n\left(\sum_{j=1}^{4} S_j\right)$$

where n is a positive integer.

Given that the middle 95% of the distribution of W lies between 158.0264 and 378.1736 minutes.

(c) find the value of m and the value of n

(8)

Question 7 continued

© Pearson Education Ltd.

© Pearson Education Ltd.

Question 7 continued

© Pearson Education Ltd.

Question 7 continued

© Pearson Education Ltd.

Question 7 continued

(Total for Question 7 is 14 marks)

TOTAL FOR PAPER IS 75 MARKS

© Pearson Education Ltd.