Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel International Advanced Level

Tuesday 13 June 2023

Morning (Time: 1 hour 30 minutes)

Paper reference **WST03/01**

Mathematics

International Advanced Subsidiary/Advanced Level
Statistics S3

You must have:
Mathematical Formulae and Statistics Tables (Yellow), calculator

Total Marks

Candidates may use any calculator permitted by Pearson regulations. Calculators must not have the facility for symbolic 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 the 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.
- If you change your mind about an answer, cross it out and put your new answer and any working underneath.

Turn over

1. (a) State two conditions under which it might be more appropriate to use Spearman's rank correlation coefficient rather than the product moment correlation coefficient.

(2)

A random sample of 10 melons was taken from a market stall. The length, in centimetres, and maximum diameter, in centimetres, of each melon were recorded.

The Spearman's rank correlation coefficient between the results was –0.673

(b) Test, at the 5% level of significance, whether or not there is evidence of a correlation. State clearly your hypotheses and the critical value used.

(4)

The product moment correlation coefficient between the results was –0.525

(c) Test, at the 5% level of significance, whether or not there is evidence of a **negative** correlation.
State clearly your hypotheses and the critical value used.

(3)

Question 1 continued

Question 1 continued

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 9 marks)

2. A business accepts cash, bank cards or mobile apps as payment methods.

The manager wishes to test whether or not there is an association between the payment amount and the payment method used.

The manager takes a random sample of 240 payments and records the payment amount and the payment method used.

The manager's results are shown in the table.

		Payment amount		
		Under £50	**£50 to £150**	**Over £150**
Payment method	**Cash**	23	19	18
	Bank card	21	32	31
	Mobile app	16	39	41

Using these results,

(a) calculate the expected frequencies for the payment amount under £50 that

(i) use cash

(ii) use a bank card

(iii) use a mobile app

(3)

Given that for the other 6 classes $\sum \frac{(O-E)^2}{E} = 2.4048$ to 4 decimal places,

(b) test, at the 5% level of significance, whether or not there is evidence for an association between the payment amount and the payment method used.
You should state the hypotheses, the test statistic, the degrees of freedom and the critical value used for this test.

(7)

Question 2 continued

Question 2 continued

Question 2 continued

(Total for Question 2 is 10 marks)

3. A random sample of 2 observations, X_1 and X_2, is taken from a population with unknown mean μ and unknown variance σ^2

(a) Explain why $\dfrac{X_1 - X_2}{\sigma}$ is not a statistic.

(1)

$$S = \frac{3}{5}X_1 + \frac{5}{7}X_2$$

(b) Show that S is a biased estimator of μ

(2)

(c) Hence find the bias, in terms of μ, when S is used as an estimator of μ

(1)

Given that $Y = aX_1 + bX_2$ is an unbiased estimator of μ, where a and b are constants,

(d) find an equation, in terms of a and b, that must be satisfied.

(2)

(e) Using your answer to part (d), show that $\text{Var}(Y) = (2a^2 - 2a + 1)\sigma^2$

(3)

Question 3 continued

Question 3 continued

Question 3 continued

(Total for Question 3 is 9 marks)

4. It is suggested that the delay, in hours, of certain flights from a particular country may be modelled by the continuous random variable, T, with probability density function

$$\mathrm{f}(t) = \begin{cases} \dfrac{2}{25}t & 0 \leqslant t < 5 \\ 0 & \text{otherwise} \end{cases}$$

(a) Show that for $0 \leqslant a \leqslant 4$

$$\mathrm{P}(a \leqslant T < a + 1) = \frac{1}{25}(2a + 1)$$

(3)

A random sample of 150 of these flights is taken. The delays are summarised in the table below.

Delay (t hours)	**Frequency**
$0 \leqslant t < 1$	10
$1 \leqslant t < 2$	13
$2 \leqslant t < 3$	24
$3 \leqslant t < 4$	35
$4 \leqslant t < 5$	68

(b) Test, at the 5% significance level, whether the given probability density function is a suitable model for these delays.
You should state your hypotheses, expected frequencies, test statistic and the critical value used.

(8)

Question 4 continued

Question 4 continued

Question 4 continued

(Total for Question 4 is 11 marks)

5. The continuous random variable X is normally distributed with

$$X \sim \mathrm{N}(\mu, 5^2)$$

A random sample of 10 observations of X is taken and $\bar{X}$ denotes the sample mean.

(a) Show that a 90% confidence interval for μ, in terms of $\bar{x}$, is given by

$$(\bar{x} - 2.60, \bar{x} + 2.60)$$

(3)

The continuous random variable Y is normally distributed with

$$Y \sim \mathrm{N}(\mu, 3^2)$$

A random sample of 20 observations of Y are taken and $\bar{Y}$ denotes the sample mean.

(b) Find a 95% confidence interval for μ, in terms of $\bar{y}$

(3)

(c) Given that X and Y are independent,

(i) find the distribution of $\bar{X} - \bar{Y}$

(ii) calculate the probability that the two confidence intervals from part (a) and part (b) do not overlap.

(7)

Question 5 continued

Question 5 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 5 continued

(Total for Question 5 is 13 marks)

6. Roxane, a scientist, carries out an investigation into the fat content of different brands of crisps.

Roxane took random samples of different brands of crisps and recorded, in grams, the fat content (x) of a 30 gram serving.

The table below shows some results for just two of these brands.

Brand	$\sum x$	$\sum x^2$	$\bar{x}$	s	**Sample size**
A	350	1753.9744	5.0	0.24	70
B	331.5	1694.65	α	β	65

(a) Calculate the value of α and the value of β

(3)

Roxane claims that these results show that the crisps from brand A have a lower fat content than the crisps from brand B, as the mean fat content for brand A is, statistically, significantly less than the mean fat content for brand B.

(b) Stating your hypotheses clearly, carry out a suitable test, at the 5% level of significance, to assess Roxane's claim.
You should state your test statistic and critical value.

(7)

(c) For the test in part (b), state whether or not it is necessary to assume that the fat content of crisps is normally distributed. Give a reason for your answer.

(2)

(d) State an assumption you have made in carrying out the test in part (b).

(1)

Question 6 continued

Question 6 continued

Question 6 continued

(Total for Question 6 is 13 marks)

7. The random variable X is defined as

$$X = 4A - 3B$$

where A and B are independent and

$$A \sim \mathrm{N}(15, 5^2) \qquad B \sim \mathrm{N}(10, 4^2)$$

(a) Find $\mathrm{P}(X < 40)$

(4)

The random variable C is such that $C \sim \mathrm{N}(20, \sigma^2)$

The random variables C_1, C_2 and C_3 are independent and each has the same distribution as C

The random variable D is defined as

$$D = \sum_{i=1}^{3} C_i$$

Given that $\mathrm{P}(A + B + D < 76) = 0.2420$ and that A, B and D are independent,

(b) showing your working clearly, find the standard deviation of C

(6)

Question 7 continued

Question 7 continued

(Total for Question 7 is 10 marks)

TOTAL FOR PAPER IS 75 MARKS