Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel International Advanced Level

Friday 19 January 2024

Morning (Time: 1 hour 30 minutes)

Paper reference **WST03/01**

Mathematics

International Advanced Subsidiary/Advanced Level
Statistics S3

You must have:
Mathematical Formulae and Statistical 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 ▶

P75603A

©2024 Pearson Education Ltd.
S:1/1/1/

P 7 5 6 0 3 A 0 1 2 4

Pearson

1. Chen is treating vines to prevent fungus appearing. One month after the treatment, Chen monitors the vines to see if fungus is present.

The contingency table shows information about the type of treatment for a sample of 150 vines and whether or not fungus is present.

	Type of treatment		
	None	**Sulphur**	**Copper sulphate**
No fungus present	20	55	48
Fungus present	10	8	9

Test, at the 5% level of significance, whether or not there is any association between the type of treatment and the presence of fungus.
Show your working clearly, stating your hypotheses, expected frequencies, test statistic and critical value.

(8)

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 1 continued

Question 1 continued

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 1 continued

(Total for Question 1 is 8 marks)

2. A company has 800 employees.

The manager of the company is going to take a sample of 80 employees.

(a) Explain how this sample can be taken using systematic sampling.

(3)

The company has offices in London, Edinburgh and Cardiff. The table shows the number of employees in each city.

City	London	Edinburgh	Cardiff
Number of employees	430	250	120

The president of the company is going to take a sample of 100 employees to determine the average time employees spend in front of a computer each week.

(b) Explain how this sample can be taken using stratified sampling.

(3)

(c) Explain an advantage of using stratified sampling rather than simple random sampling.

(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 2 continued

(Total for Question 2 is 7 marks)

3. The table shows the annual tea consumption, t (kg/person), and population, p (millions), for a random sample of 7 European countries.

Country	A	B	C	D	E	F	G
Annual tea consumption, t (kg/person)	0.27	0.15	0.42	0.06	1.94	0.78	0.44
Population, p (millions)	5.4	5.8	9	10.2	67.9	17.1	8.7

(You may use $S_{tt} = 2.486$ $S_{pp} = 3026.234$ $S_{pt} = 83.634$)

Angela suggests using the product moment correlation coefficient to calculate the correlation between annual tea consumption and population.

(a) Use Angela's suggestion to test, at the 5% level of significance, whether or not there is evidence of any correlation between annual tea consumption and population.
State your hypotheses clearly and the critical value used.

(5)

Johan suggests using Spearman's rank correlation coefficient to calculate the correlation between the rank of annual tea consumption and the rank of population.

(b) Calculate Spearman's rank correlation coefficient between the rank of annual tea consumption and the rank of population.

(4)

(c) Use Johan's suggestion to test, at the 5% level of significance, whether or not there is evidence of a positive correlation between annual tea consumption and population.
State your hypotheses clearly and the critical value used.

(3)

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 3 continued

Question 3 continued

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 3 continued

(Total for Question 3 is 12 marks)

4. The number of jobs sent to a printer per hour in a small office is recorded for 120 hours. The results are summarised in the following table.

Number of jobs	0	1	2	3	4	5
Frequency	24	34	28	21	8	5

(a) Show that the mean number of jobs sent to the printer per hour for these data is 1.75

(1)

The office manager believes that the number of jobs sent to the printer per hour can be modelled using a Poisson distribution.

The office manager uses the mean given in part (a) to calculate the expected frequencies for this model. Some of the results are given in the following table.

Number of jobs	0	1	2	3	4	5 or more
Expected frequency	20.85	36.49	31.93	r	s	3.95

(b) Show that the value of s is 8.15 to 2 decimal places.

(1)

(c) Find the value of r to 2 decimal places.

(1)

The value of $\sum \frac{(O_i - E_i)^2}{E_i}$ for the first four frequencies in the table is 1.43

(d) Test, at the 5% level of significance, whether or not the number of jobs sent to the printer per hour can be modelled using a Poisson distribution.
Show your working clearly, stating your hypotheses, test statistic and critical value.

(7)

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

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

(Total for Question 4 is 10 marks)

5. A professor claims that undergraduates studying History have a typing speed of more than 15 words per minute faster than undergraduates studying Maths.

A sample is taken of 38 undergraduates studying History and 45 undergraduates studying Maths. The typing speed, x words per minute, of each undergraduate is recorded. The results are summarised in the table below.

	n	$\bar{x}$	s^2
Undergraduates studying History	38	56.3	27.2
Undergraduates studying Maths	45	39.8	18.5

(a) Use a suitable test, at the 5% level of significance, to investigate the professor's claim.
State clearly your hypotheses, test statistic and critical value.

(7)

(b) State two assumptions you have made in carrying out the test in part (a).

(2)

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 5 continued

(Total for Question 5 is 9 marks)

6. A random sample of 8 three-month-old golden retriever dogs is taken.

The heights of the golden retrievers are recorded.
Using this sample, a 95% confidence interval for the mean height, in cm, of three-month-old golden retrievers is found to be (45.72, 53.88)

(a) Find a 99% confidence interval for the mean height.
You may assume that the heights are normally distributed with known population standard deviation.

(5)

Some summary statistics for the weights, x kg, of this sample are given below.

$$\sum x = 91.2 \qquad \sum x^2 = 1145.16 \qquad n = 8$$

(b) Calculate unbiased estimates of the mean and the variance of the weights of three-month-old golden retrievers.

(3)

A further random sample of 24 three-month-old golden retrievers is taken.
The unbiased estimates of the mean and the variance of the weights, in kg, from this sample are found to be 10.8 and 17.64 respectively.

(c) Estimate the standard error of the mean weight for the combined sample of 32 three-month-old golden retrievers.

(7)

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 6 continued

Question 6 continued

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 6 continued

(Total for Question 6 is 15 marks)

7. Small containers and large containers are independently filled with fruit juice.

The amounts of fruit juice in small containers are normally distributed with mean 180 ml and standard deviation 4.5 ml

The amounts of fruit juice in large containers are normally distributed with mean 330 ml and standard deviation 6.7 ml

The random variable W represents the total amount of fruit juice in a random sample of 2 small containers minus the amount of fruit juice in 1 randomly selected large container.

$W \sim \mathrm{N}(a, b)$ where a and b are positive constants.

(a) Find the value of a and the value of b

(3)

(b) Find the probability that a randomly chosen large container of fruit juice contains more than 1.8 times the amount of fruit juice in a randomly chosen small container.

(5)

A random sample of 3 small containers of fruit juice is taken.

(c) Find the probability that the first container of fruit juice in this sample contains at least 5 ml more than the mean amount of fruit juice in all 3 small containers.

(6)

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 7 continued

Question 7 continued

(Total for Question 7 is 14 marks)

TOTAL FOR PAPER IS 75 MARKS

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