Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel International Advanced Level

Wednesday 22 January 2025

Afternoon (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 ▶

P76395A

©2025 Pearson Education Ltd.
H:1/1/1/1/

P 7 6 3 9 5 A 0 1 2 8

Pearson

1. The table below gives information about the final positions of 10 football teams in the English Premier League and their average match day attendance for one season. The table also shows some of the average match day attendance ranks.

Team	Final position	Average match day attendance	Average match day attendance rank
Arsenal	5	59 776	2
Brighton and Hove Albion	9	30 988	
Chelsea	3	36 424	
Leicester City	8	32 440	
Liverpool	2	53 027	
Manchester City	1	53 739	
Manchester United	6	72 992	1
Tottenham Hotspur	4	56 523	
West Ham United	7	57 915	
Wolverhampton Wanderers	10	30 725	

(Source from: www.statista.com)

By completing the ranks for average match day attendance,

(a) calculate the value of the Spearman's rank correlation coefficient for these data. Show your working clearly.

(4)

(b) Stating your hypotheses clearly, test at the 1% level of significance, whether or not there is evidence of a positive correlation between the final position of a football team in the English Premier League and its average match day attendance.

(4)

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. The manager of a shop investigates the number of customers served by staff that work in the shop.
The manager records the number of customers being served in 100 consecutive five-minute time intervals.

The results are shown in the table below.

Number of customers	0	1	2	3	4	5	6 or more
Frequency	5	38	32	17	7	1	0

The manager plans to test whether or not a Poisson model is suitable for these data.

(a) Show that the sample mean number of customers served is 1.86

(1)

The manager uses the sample mean to calculate the following expected frequencies, correct to 3 decimal places.

Number of customers	0	1	2	3	4	5	6 or more
Expected frequency	15.567	28.955	26.928	16.696	7.763	2.888	r

(b) Explain why the value of r is 1.203

(1)

The manager combines cells 4, 5 and '6 or more', before continuing the test.

(c) Explain why it is necessary for the manager to do this.

(1)

After combining these cells, the manager calculates that $\sum \frac{(O-E)^2}{E} = 12.2$

correct to 3 significant figures.

(d) Carry out the manager's test at the 1% significance level.
You should clearly state the hypotheses, the degrees of freedom and the critical value used.

(4)

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

Question 2 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 2 continued

(Total for Question 2 is 7 marks)

3. A biologist is investigating the weights of rabbits in two different regions, region A and region B
The biologist collects random samples of 50 rabbits from region A and 40 rabbits from region B and records the weight, x kg, of each rabbit.

The table shows a summary of the biologist's data.

	Sample size	$\sum x$	$\sum x^2$	**Unbiased estimate of the mean**	**Unbiased estimate of the variance**
Region *A*	50	132.5	354.555	2.65	0.07
Region *B*	40	118	350.05	p	q

(a) Calculate the value of p and the value of q

(3)

The biologist believes that the mean weight of rabbits in region A is smaller than the mean weight of rabbits in region B

(b) Stating your hypotheses clearly, carry out a suitable test to assess the biologist's belief. Use a 5% level of significance and state your critical value.

(7)

(c) Explain how you have used the fact that the sample sizes are large in your answer to part (b)

(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 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 random variable X is normally distributed with unknown mean μ and known variance σ^2
A random sample of n observations of X produced a 99% confidence interval for μ of (0.900, 0.964)

(a) Show that the standard error used to find this confidence interval is 0.0124 to 3 significant figures.

(3)

Using the same random sample,

(b) find a 95% confidence interval for μ

Give your answers to 3 decimal places.

(4)

The r % confidence interval using the same random sample has a width of 0.04

(c) Calculate the value of r

(5)

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 12 marks)

5. A research company takes a sample of 120 students living in London and a sample of 80 students living in Birmingham. The company asks the students to state their favourite science subject from Biology, Chemistry and Physics.

The research company uses quota sampling rather than simple random sampling.

(a) (i) Give one advantage of using quota sampling.

(ii) Give one disadvantage of using quota sampling.

(2)

The results of the survey are shown in the table.

	Biology	**Chemistry**	**Physics**	**Total**
London	66	37	17	120
Birmingham	40	19	21	80

The data are to be used to determine whether or not a student's favourite science subject is independent of the place where they live.

(b) Calculate

(i) the expected frequency for the class Biology and London,

(ii) the expected frequency for the class Biology and Birmingham.

(3)

Before the two classes for Biology are included $\sum \frac{(O-E)^2}{E} = 4.549$

to 3 decimal places.

(c) Test at the 10% level of significance, whether or not a student's favourite science subject is independent of the place where they live.
Show your working clearly, stating your hypotheses, the degrees of freedom, the test statistic and the critical value used.

(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 5 continued

Question 5 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 5 continued

(Total for Question 5 is 12 marks)

6. The random variable X has a continuous uniform distribution over the interval $[2a + 3, 4a + 9]$ where a is a constant and $a > -3$
The mean of a random sample of size n, taken from this distribution, is denoted by $\bar{X}$

(a) Show that $\bar{X}$ is a biased estimator of a

(2)

(b) Hence find the bias, in terms of a, when $\bar{X}$ is used as an estimator of a

(1)

Given that $Y = c\bar{X} + d$ is an unbiased estimator of a where c and d are constants.

(c) find the value of c and the value of d

(3)

A random sample of 10 values of X is taken and the results are as follows

4.4 4.7 5.6 6.3 6.7 7.5 8.4 9.4 9.8 10.4

(d) Use the sample mean to estimate the maximum value that X can take.

(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 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 9 marks)

7. A company sells two types of board, short or long.

The weight of a short board, S kg, has distribution $N(7.7, 0.01^2)$

The weight of a long board, L kg, has distribution $N(20, 0.02^2)$

A random sample of 3 short boards and 4 long boards is taken.

(a) Find the probability that the total weight of the 3 short boards and the 4 long boards is more than 103.15 kg

(4)

A random sample of 2 long boards is taken.

(b) Find the probability that the weights of these 2 long boards differ by more than 0.01 kg

(5)

A random sample of $(n + 1)$ short boards is taken, with weights $S_1, S_2, S_3, ..., S_{n+1}$

The random variable T is defined as

$$T = nS_1 - \sum_{r=2}^{n+1} S_r$$

Given that, to 4 decimal places, $P(T > 2) = 0.0233$

(c) calculate the value of n

(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 15 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