Pearson Edexcel IAL (Further) Mathematics

Statistics 3

www.CasperYC.club/wst03

Past Paper Collection

Last updated: January 21, 2025

Comments and suggestions to DrYuFromShanghai@QQ.com

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel
International Advanced Level

Centre Number | | | | | |

Candidate Number | | | | |

Monday 12 October 2020

Afternoon (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 (Blue), 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 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 ▶

Leave blank

1. The random variable X has the discrete uniform distribution

$$P(X = x) = \frac{1}{\alpha} \quad \text{for } x = 1,\ 2,\ ...,\ \alpha$$

The mean of a random sample of size n, taken from this distribution, is denoted by $\bar{X}$

(a) Show that $2\bar{X}$ is a biased estimator of α

(2)

A random sample of 6 observations of X is taken and the results are given below.

8 7 3 7 2 9

(b) Use the sample mean to estimate the value of α

(2)

Leave blank

Question 1 continued

(Total 4 marks)

Q1

Leave blank

2. A university awards its graduates a degree in one of three categories, Distinction, Merit or Pass.

Table 1 shows information about a random sample of 200 graduates from three departments, Arts, Humanities and Sciences.

	Arts	Humanities	Sciences	Total
Distinction	22	32	38	92
Merit	15	30	13	58
Pass	18	15	17	50
Total	55	77	68	200

Table 1

Xiu wants to carry out a test of independence between the category of degree and the department.

Table 2 shows some of the values of $\frac{(O-E)^2}{E}$ for this test.

	Arts	Humanities	Sciences	Total
Distinction	0.43	0.33	1.44	2.20
Merit	0.06	2.63	2.29	4.98
Pass				

Table 2

(a) Complete Table 2

(4)

(b) Hence, complete Xiu's hypothesis test using a 5% level of significance. You should state the hypotheses, the degrees of freedom and the critical value used for this test.

(5)

Leave blank

Question 2 continued

Leave blank

Question 2 continued

Leave blank

Question 2 continued

Q2

(Total 9 marks)

Leave blank

3. Each of 7 athletes competed in a 200 metre race and a 400 metre race.

The table shows the time, in seconds, taken by each athlete to complete the 200 metre race.

Athlete	*A*	*B*	*C*	*D*	*E*	*F*	*G*
200 metre race (seconds)	23.4	23.1	22.9	23.7	27.6	24.4	24.1

The finishing order in the 400 metre race is shown below, with athlete *A* finishing in the fastest time.

A *B* *G* *C* *D* *F* *E*

(a) Calculate the Spearman's rank correlation coefficient between the finishing order in the 200 metre race and the finishing order in the 400 metre race.

(5)

(b) Stating your hypotheses clearly, test whether or not there is evidence of a positive correlation between the finishing order in the 200 metre race and the finishing order in the 400 metre race. Use a 5% level of significance.

(4)

The 7 athletes also competed in a long jump competition with the following results.

Athlete	*A*	*B*	*C*	*D*	*E*	*F*	*G*
Long jump (metres)	6.50	6.47	6.12	6.12	6.48	6.38	6.47

Yuliya wants to calculate the Spearman's rank correlation coefficient between the finishing order in the 200 metre race and the finishing order in the long jump for these athletes.

(c) Without carrying out any further calculations, explain how Yuliya should do this.

(2)

Leave blank

Question 3 continued

Leave blank

Question 3 continued

Leave blank

Question 3 continued

(Total 11 marks)

Q3

Leave blank

4. Luka wants to carry out a survey of students at his school.

He obtains a list of all 280 students.

(a) Explain how he can use this list to select a systematic sample of 40 students.

(2)

Luka is trying to make his own random number table. He generates 400 digits to put in his table. Figure 1 shows the frequency of each digit in his table.

Digit generated	0	1	2	3	4	5	6	7	8	9
Frequency	36	42	33	41	44	43	48	38	32	43

Figure 1

A test is carried out at the 10% level of significance to see if the digits Luka generates follow a uniform distribution.

For this test $\sum \frac{(O-E)^2}{E} = 5.9$

(b) Determine the conclusion of this test.

(3)

The digits generated by Luka are taken two at a time to form two-digit numbers.

Figure 2 shows the frequency of two-digit numbers in his table.

Two-digit numbers generated	00–19	20–39	40–59	60–79	80–99
Frequency	31	49	30	42	48

Figure 2

(c) Test, at the 10% level of significance, whether the two-digit numbers generated by Luka follow a uniform distribution. You should state the hypotheses, the degrees of freedom and the critical value used for this test.

(8)

There are 70 students in Year 12 at his school.

(d) State, giving a reason, the advice you would give to Luka regarding the use of his table of numbers for generating a simple random sample of 10 of the Year 12 students.

(2)

Leave blank

Question 4 continued

Leave blank

Question 4 continued

Leave blank

Question 4 continued

Q4

(Total 15 marks)

Leave blank

5. A greengrocer is investigating the weights of two types of orange, type *A* and type *B*. She believes that on average type *A* oranges weigh greater than 5 grams more than type *B* oranges. She collects a random sample of 40 type *A* oranges and 32 type *B* oranges and records the weight, *x* grams, of each orange.

The table shows a summary of her data.

	n	$\bar{x}$	$\sum x^2$
Type *A* oranges	40	140.4	790 258
Type *B* oranges	32	134.7	581 430

(a) Calculate unbiased estimates for the variance of the weights of the population of type *A* oranges and the variance of the weights of the population of type *B* oranges.

(3)

(b) Test, at the 5% level of significance, the greengrocer's belief. You should state the hypotheses and the critical value used for this test.

(7)

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

(2)

Leave blank

Question 5 continued

Leave blank

Question 5 continued

Leave blank

Question 5 continued

(Total 12 marks)

Q5

Leave blank

6. The number of toasters sold by a shop each week may be modelled by a Poisson distribution with mean 4

A random sample of 35 weeks is taken and the mean number of toasters sold per week is found.

(a) Write down the approximate distribution for the mean number of toasters sold per week from a random sample of 35 weeks.

(2)

The number of kettles sold by the shop each week may be modelled by a Poisson distribution with mean λ

A random sample of 40 weeks is taken and the mean number of kettles sold per week is found. The width of the 99% confidence interval for λ is 2.6

(b) Find an estimate for λ

(4)

A second, independent random sample of 40 weeks is taken and a second 99% confidence interval for λ is found.

(c) Find the probability that only one of these two confidence intervals contains λ

(2)

Leave blank

Question 6 continued

Q6

(Total 8 marks)

Leave blank

7. A company makes cricket balls and tennis balls.

The weights of cricket balls, C grams, follow a normal distribution

$$C \sim \text{N}(160, 1.25^2)$$

Three cricket balls are selected at random.

(a) Find the probability that their total weight is more than 475.8 grams.

(4)

The weights of tennis balls, T grams, follow a normal distribution

$$T \sim \text{N}(60, 2^2)$$

Five tennis balls and two cricket balls are selected at random.

(b) Find the probability that the total weight of the five tennis balls and the two cricket balls is more than 625 grams.

(4)

A random sample of n tennis balls $T_1,\ T_2,\ T_3,\ ...,\ T_n$ is taken.

The random variable $Y = (n-1)T_1 - \sum_{r=2}^{n} T_r$

Given that $\text{P}(Y > 40) = 0.0838$ correct to 4 decimal places,

(c) find n.

(8)

Leave blank

Question 7 continued

Leave blank

Question 7 continued

Q7

(Total 16 marks)

TOTAL FOR PAPER: 75 MARKS

END

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel
International Advanced Level

Centre Number | Candidate Number

Thursday 21 January 2021

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 (Blue), 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 6 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 ▶

Leave blank

1. A journalist is going to interview a sample of 10 players from the 60 players in a local football club. The journalist uses the random numbers on page 27 of the formula booklet and starts at the top of the 10th column, where the first number is 96

The journalist worked down the 10th column to select 10 numbers. The first 3 numbers selected were: 33, 15 and 23

(a) Find the other 7 numbers to complete the sample of ten.

(2)

There are 24 girls and 36 boys who play football for the club.
The journalist labels the girls from 1 to 24 and the boys from 25 to 60

(b) Show how the journalist can use her 10 random numbers to select a stratified sample of 10 players from the club to interview.

(2)

The club provided the journalist with a list of the players in ascending order of ages, numbered 1 to 60.
The journalist uses the 10 random numbers to select a simple random sample of the players.

(c) State, giving a reason, a group of players who may not be represented in this sample.

(2)

Leave blank

Question 1 continued

Q1

(Total 6 marks)

Leave blank

2. A teacher believes that those of her students with strong mathematical ability may also have enhanced short-term memory. She shows a random sample of 11 students a tray of different objects for eight seconds and then asks them to write down as many of the objects as they can remember. The results, along with their percentage score in a recent mathematics test, are given in the table below.

Student	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*
No. of objects	8	11	9	15	17	6	10	14	12	13	5
% in maths test	30	62	57	80	75	43	65	51	48	55	32

(a) Calculate Spearman's rank correlation coefficient for these data. Show your working clearly.

(5)

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

(3)

The teacher shows these results to her class and argues that spending more time trying to improve their short-term memory would improve their mathematical ability.

(c) Explain whether or not you agree with the teacher's argument.

(1)

Leave blank

Question 2 continued

Q2

(Total 9 marks)

Leave blank

3. The students in a group of schools can choose a club to join. There are 4 clubs available: Music, Art, Sports and Computers. The director collected information about the number of students in each club, using a random sample of 88 students from across the schools. The results are given in Table 1 below.

	Music	**Art**	**Sports**	**Computers**
No. of students	14	28	27	19

Table 1

The director uses a chi-squared test to determine whether or not the students are uniformly distributed across the 4 clubs.

(a) (i) Find the expected frequencies he should use.

Given that the test statistic he calculated was 6.09 (to 3 significant figures)

(ii) use a 5% level of significance to complete the test. You should state the degrees of freedom and the critical value used.

(4)

The director wishes to examine the situation in more detail and takes a second random sample of 88 students. The director assumes that within each school, students select their clubs independently. The students come from 3 schools and the distribution of the students from each school amongst the clubs is given in Table 2 below.

Club / **School**	**Music**	**Art**	**Sports**	**Computers**
School *A*	3	10	9	8
School *B*	1	11	13	5
School *C*	11	6	7	4

Table 2

The director wishes to test for an association between a student's school and the club they choose.

(b) State hypotheses suitable for such a test.

(1)

(c) Calculate the expected frequency for School *C* and the Computers club.

(1)

The director calculates the test statistic to be 7.29 (to 3 significant figures) with 4 degrees of freedom.

(d) Explain clearly why his test has 4 degrees of freedom.

(2)

(e) Complete the test using a 5% level of significance and stating clearly your critical value.

(2)

Leave blank

Question 3 continued

Leave blank

Question 3 continued

Leave blank

Question 3 continued

(Total 10 marks)

Q3

Leave blank

4. The scores in a national test of seven-year-old children are normally distributed with a standard deviation of 18
A random sample of 25 seven-year-old children from town A had a mean score of 52.4

(a) Calculate a 98% confidence interval for the mean score of the seven-year-old children from town A.

(4)

An independent random sample of 30 seven-year-old children from town B had a mean score of 57.8
A local newspaper claimed that the mean score of seven-year-old children from town B was greater than the mean score of seven-year-old children from town A.

(b) Stating your hypotheses clearly, use a 5% significance level to test the newspaper's claim. You should show your working clearly.

(6)

The mean score for the national test of seven-year-old children is μ.

Considering the two samples of seven-year-old children separately, at the 5% level of significance, there is insufficient evidence that the mean score for town A is less than μ, and insufficient evidence that the mean score for town B is less than μ.

(c) Find the largest possible value for μ.

(4)

Leave blank

Question 4 continued

Leave blank

Question 4 continued

Leave blank

Question 4 continued

Q4

(Total 14 marks)

Leave blank

5. Chrystal is studying the lengths of pine cones that have fallen from a tree. She believes that the length, X cm, of the pine cones can be modelled by a normal distribution with mean 6 cm and standard deviation 0.75 cm.

She collects a random sample of 80 pine cones and their lengths are recorded in the table below.

Length, x cm	$x < 5$	$5 \leqslant x < 5.5$	$5.5 \leqslant x < 6$	$6 \leqslant x < 6.5$	$x \geqslant 6.5$
Frequency	6	14	24	26	10

(a) Stating your hypotheses clearly and using a 10% level of significance, test Chrystal's belief. Show your working clearly and state the expected frequencies, the test statistic and the critical value used.

(10)

Chrystal's friend David asked for more information about the lengths of the 80 pine cones. Chrystal told him that

$$\sum x = 464 \quad \text{and} \quad \sum x^2 = 2722.59$$

(b) Calculate unbiased estimates of the mean and variance of the lengths of the pine cones.

(3)

David used the calculations from part (b) to test whether or not the lengths of the pine cones are normally distributed using Chrystal's sample.
His test statistic was 3.50 (to 3 significant figures) and he did not pool any classes.

(c) Using a 10% level of significance, complete David's test stating the critical value and the degrees of freedom used.

(3)

(d) Estimate, to 2 significant figures, the proportion of pine cones from the tree that are longer than 7 cm.

(2)

Leave blank

Question 5 continued

Leave blank

Question 5 continued

Leave blank

Question 5 continued

(Total 18 marks)

Q5

Leave blank

6. A potter makes decorative tiles in two colours, red and yellow. The length, R cm, of the red tiles has a normal distribution with mean 15 cm and standard deviation 1.5 cm. The length, Y cm, of the yellow tiles has the normal distribution $N(12, 0.8^2)$. The random variables R and Y are independent.

A red tile and a yellow tile are chosen at random.

(a) Find the probability that the yellow tile is longer than the red tile.

(4)

Taruni buys 3 red tiles and 1 yellow tile.

(b) Find the probability that the total length of the 3 red tiles is less than 4 times the length of the yellow tile.

(7)

Stefan defines the random variable $X = aR + bY$, where a and b are constants. He wants to use values of a and b such that X has a mean of 780 and minimum variance.

(c) Find the value of a and the value of b that Stefan should use.

(7)

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Q6

(Total 18 marks)

TOTAL FOR PAPER: 75 MARKS

END

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel
International Advanced Level

Centre Number | Candidate Number

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 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 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 6 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.
- Good luck with your examination.

Turn over ▶

1. A plant biologist claims that as the percentage moisture content of the soil in a field increases, so does the percentage plant coverage. He splits the field into equal areas labelled *A*, *B*, *C*, *D* and *E* and measures the percentage plant coverage and the percentage moisture content for each area. The results are shown in the table below.

	A	*B*	*C*	*D*	*E*
Coverage %	10	12	25	0	6
Moisture %	30	20	40	10	25

(a) Calculate Spearman's rank correlation coefficient for these data.

(5)

(b) Stating your hypotheses clearly, test at the 5% level of significance, whether or not these data provide support for the plant biologist's claim.

(4)

Question 1 continued

(Total for Question 1 is 9 marks)

2. A doctor believes that the diet of her patients and their health are not independent.

She takes a random sample of 200 patients and records whether they are in good health or poor health and whether they have a good diet or a poor diet. The results are summarised in the table below.

	Good health	**Poor health**
Good diet	86	8
Poor diet	91	15

Stating your hypotheses clearly, test the doctor's belief using a 5% level of significance. Show your working for your test statistic and state your critical value clearly.

(9)

Question 2 continued

(Total for Question 2 is 9 marks)

3. Components are manufactured such that their length in mm is normally distributed with mean μ and variance σ^2. Below is a 95% confidence interval for μ calculated from a random sample of components.

(11.52, 13.75)

Using the same random sample,

(a) find a 90% confidence interval for μ.

(6)

Four 90% confidence intervals are found from independent random samples.

(b) Calculate the probability that only 3 of these 4 intervals will contain μ.

(2)

Question 3 continued

(Total for Question 3 is 8 marks)

4. A college runs academic and vocational courses. The college has 1680 academic students and 2520 vocational students.

(a) Describe how a stratified sample of 70 students at the college could be taken.

(3)

All students at the college take a basic skills test. A random sample of 50 academic students has a mean score of 57 and a variance of 60. An independent random sample of 80 vocational students has a mean score of 62 with a variance of 70

(b) Stating your hypotheses clearly, test at the 5% level of significance, whether or not the mean basic skills score for vocational students is greater than the mean basic skills score for academic students.

(7)

(c) Explain the importance of the Central Limit Theorem to the test in part (b).

(2)

(d) State an assumption that is required to carry out the test in part (b).

(1)

All the academic students at the college take a basic skills course. Another random sample of 50 academic students and another independent random sample of 80 vocational students retake the basic skills test. The hypotheses used in part (b) are then tested again at the same level of significance.

The value of the test statistic z is now 1.54

(e) Comment on the mean basic skills scores of academic and vocational students after taking this course.

(2)

(f) Considering the outcomes of the tests in part (b) and part (e), comment on the effectiveness of the basic skills course.

(1)

Question 4 continued

Question 4 continued

Question 4 continued

(Total for Question 4 is 16 marks)

5. A researcher is looking into the effectiveness of a new medicine for the relief of symptoms. He collects random samples of 8 people who are taking the medicine from each of 50 different medical practices. The number of people who say that the medicine is a success, in each sample, is recorded. The results are summarised in the table below.

Number of successes	0	1	2	3	4	5	6	7	8
Number of practices	4	6	3	12	10	7	4	2	2

The researcher decides to model this data using a binomial distribution.

(a) State two necessary assumptions that the researcher made in order to use this model.

(2)

(b) Show that the mean number of successes per sample is 3.54

(2)

He decides to use this mean to calculate expected frequencies. The results are shown in the table below.

Number of successes	0	1	2	3	4	5	6	7	8
Expected frequency	0.47	2.96	8.23	13.07	f	8.23	3.27	0.74	g

(c) Calculate the value of f and the value of g. Give your answers to 2 decimal places.

(4)

(d) Stating your hypotheses clearly, test at the 10% level of significance, whether or not the binomial distribution is a suitable model for the number of successes in samples of 8 people.

(8)

Question 5 continued

Question 5 continued

Question 5 continued

(Total for Question 5 is 16 marks)

6. A baker produces bread buns and bread rolls. The weights of buns, B grams, and the weights of rolls, R grams, are such that $B \sim N(55, 1.3^2)$ and $R \sim N(51, 1.2^2)$

A bun and a roll are selected at random.

(a) Find the probability that the bun weighs less than 110% of the weight of the roll.

(5)

Two buns are chosen at random.

(b) Find the probability that their weights differ by more than 1 gram.

(6)

The baker sells bread in bags. Each bag contains either 10 buns or 11 rolls. The weight of an empty bag, S grams, is such that $S \sim N(3, 0.2^2)$

(c) Find the probability that a bag of buns weighs less than a bag of rolls.

(6)

Question 6 continued

Question 6 continued

Question 6 continued

Question 6 continued

(Total for Question 6 is 17 marks)

TOTAL FOR PAPER: 75 MARKS

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel International Advanced Level

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 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 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 ▶

Leave blank

1. A machine makes screws with a mean length of 30 mm and a standard deviation of 2.5 mm.

A manager claims that, following some repairs, the machine is now making screws with a mean length of less than 30 mm. The manager takes a random sample of 80 screws and finds that they have a mean length of 29.5 mm.

Use a suitable test, at the 5% level of significance, to determine whether there is evidence to support the manager's claim. State your hypotheses clearly.

(5)

Leave blank

Question 1 continued

Q1

(Total 5 marks)

Leave blank

2. Andy has some apple trees. Over many years she has graded each apple from her trees as *A*, *B*, *C*, *D* or *E* according to the quality of the apple, with *A* being the highest quality and *E* being the lowest quality.

She knows that the **proportion** of apples in each grade produced by her trees is as follows.

Grade	*A*	*B*	*C*	*D*	*E*
Proportion	4%	28%	52%	10%	6%

Raj advises Andy to add potassium to the soil around her apple trees.

Andy believes that adding potassium will not affect the distribution of grades for the quality of the apples.

To test her belief Andy adds potassium to the soil around her apple trees. The following year she counts the number of apples in each grade. The number of apples in each grade is shown in the table below.

Grade	*A*	*B*	*C*	*D*	*E*
Frequency	9	71	136	21	3

Test Andy's belief using a 5% level of significance. Show your working clearly, stating your hypotheses, expected frequencies and degrees of freedom.

(8)

Leave blank

Question 2 continued

Q2

(Total 8 marks)

Leave blank

3. A cafe owner wishes to know whether the price of strawberry jam is related to the taste of the jam. He finds a website that lists the price per 100 grams and a mark for the taste, out of 100, awarded by a judge, for 9 different strawberry jams *A*, *B*, *C*, *D*, *E*, *F*, *G*, *H* and *I*. He then ranks the marks for taste and the prices.

The ranks are shown in the table below.

Rank	1	2	3	4	5	6	7	8	9
Price	*A*	*B*	*E*	*C*	*D*	*F*	*G*	*H*	*I*
Taste	*A*	*B*	*F*	*E*	*H*	*G*	*I*	*C*	*D*

(a) Calculate Spearman's rank correlation coefficient for these data.

(5)

(b) Test, at the 5% level of significance, whether or not there is a relationship between the price and the taste of these strawberry jams. State your hypotheses clearly.

(3)

A friend suggests that it would be better to use the price per 100 grams, *c*, and the mark for the taste, *m*, for each strawberry jam rather than rank them.

Given that

$$S_{cc} = 2.0455 \qquad S_{mm} = 243.5556 \qquad S_{cm} = 16.4943$$

(c) calculate the product moment correlation coefficient between the price and the mark for taste of these strawberry jams, giving your answer correct to 3 decimal places.

(2)

(d) Use your value of the product moment correlation coefficient to test, at the 5% level of significance, whether or not there is evidence of a positive correlation between the price and the mark for taste of these 9 strawberry jams.
State your hypotheses clearly.

(3)

(e) State which of the tests in parts (b) and (d) is more appropriate for the cafe owner to use. Give a reason for your answer.

(1)

Leave blank

Question 3 continued

Leave blank

Question 3 continued

Leave blank

Question 3 continued

(Total 14 marks)

Q3

Leave blank

4. A local village radio station, *LSB*, decides to survey adults in its broadcasting area about the programmes it produces.
LSB broadcasts to 4 villages A, B, C and D.
The number of households in each of the villages is given below.

Village	Number of households
A	41
B	164
C	123
D	82

LSB decides to take a stratified sample of 200 households.

(a) Explain how to select the households for this stratified sample.

(3)

One of the questions in the survey related to the age group of each member of the household and whether they listen to *LSB*. The data received are shown below.

	Age group		
	18–49	50–69	Older than 69
Listen to *LSB*	130	162	65
Do not listen to *LSB*	78	98	62

The data are to be used to determine whether or not there is an association between the age group and whether they listen to *LSB*.

(b) Calculate the expected frequencies for the age group 50–69 that

(i) listen to *LSB*

(ii) do not listen to *LSB*

(2)

Given that for the **other 4** classes $\sum \frac{(O-E)^2}{E} = 4.657$ to 3 decimal places,

(c) test at the 5% level of significance, whether or not there is evidence of an association between age and listening to *LSB*. Show your working clearly, stating the degrees of freedom and the critical value.

(6)

Leave blank

Question 4 continued

Leave blank

Question 4 continued

Leave blank

Question 4 continued

(Total 11 marks)

Q4

Leave blank

5. Assam produces bags of flour. The stated weight printed on the bags of flour is 3 kg. The weights of the bags of flour are normally distributed with standard deviation 0.015 kg.

Assam weighs a random sample of 9 bags of flour and finds their mean weight is 2.977 kg.

(a) Calculate the 99% confidence interval for the mean weight of a bag of flour. Give your limits to 3 decimal places.

(3)

Assam decides to increase the amount of flour put into the bags.

(b) Explain why the confidence interval has led Assam to take this action.

(1)

After the increase a random sample of n bags of flour is taken. The sample mean weight of these n bags is 2.995 kg. A 95% confidence interval for μ gave a lower limit of less than 2.991 kg.

(c) Find the maximum value of n.

(4)

Leave blank

Question 5 continued

Q5

(Total 8 marks)

Leave blank

6. Amala believes that the resting heart rate is lower in men who exercise regularly compared to men who do not exercise regularly. She measures the resting heart rate, h, of a random sample of 50 men who exercise regularly and a random sample of 40 men who do not exercise regularly. Her results are summarised in the table below.

	Sample size	$\sum h$	$\sum h^2$	**Unbiased estimate of the mean**	**Unbiased estimate of the variance**
Exercise regularly	50	3270	214 676	α	β
Do not exercise regularly	40	2832	201 660	70.8	29.6

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

(3)

(b) Test, at the 5% level of significance, whether there is evidence to support Amala's belief. State your hypotheses clearly.

(6)

(c) Explain the significance of the central limit theorem to the test in part (b).

(1)

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

(2)

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Leave blank

Question 6 continued

(Total 12 marks)

Q6

Leave blank

7. A company produces bricks.

The weight of a brick, B kg, is such that $B \sim \mathrm{N}\left(1.96, \sqrt{0.003}^2\right)$

Two bricks are chosen at random.

(a) Find the probability that the difference in weight of the 2 bricks is greater than 0.1 kg

(5)

A random sample of n bricks is to be taken.

(b) Find the minimum sample size such that the probability of the sample mean being greater than 2 is less than 1%

(5)

The bricks are randomly selected and stacked on pallets.

The weight of an empty pallet, E kg, is such that $E \sim \mathrm{N}\left(21.8, \sqrt{0.6}^2\right)$

The random variable M represents the total weight of a pallet stacked with 500 bricks.

The random variable T represents the total weight of a container of cement.

Given that T is independent of M and that $T \sim \mathrm{N}\left(774, \sqrt{1.8}^2\right)$

(c) calculate $\mathrm{P}(4T > 100 + 3M)$

(7)

Leave blank

Question 7 continued

Leave blank

Question 7 continued

Leave blank

Question 7 continued

Leave blank

Question 7 continued

Q7

(Total 17 marks)

TOTAL FOR PAPER: 75 MARKS

END

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel International Advanced Level

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 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 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 6 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. The weights, x kg, of each of 10 watermelons selected at random from Priya's shop were recorded. The results are summarised as follows

$$\sum x = 114.2 \qquad \sum x^2 = 1310.464$$

(a) Calculate unbiased estimates of the mean and the variance of the weights of the watermelons in Priya's shop.

(3)

Priya researches the weight of watermelons, for the variety she has in her shop, and discovers that the weights of these watermelons are normally distributed with a standard deviation of 0.8 kg

(b) Calculate a 95% confidence interval for the mean weight of watermelons in Priya's shop. Give the limits of your confidence interval to 2 decimal places.

(4)

Priya claims that the confidence interval in part (b) suggests that nearly all of the watermelons in her shop weigh more than 10.5 kg

(c) Use your answer to part (b) to estimate the smallest proportion of watermelons in her shop that weigh less than 10.5 kg

(3)

Question 1 continued

Question 1 continued

Question 1 continued

(Total for Question 1 is 10 marks)

2. Secondary schools in a region conduct ability testing at the start of Year 7 and the start of Year 8. Each year a regional education officer randomly selects 240 Year 7 students and 240 Year 8 students from across the region. The results for last year are summarised in the table below.

	Mean score	**Variance of scores**
Year 7	101	38
Year 8	103	42

The regional education officer claims that there is no difference between the mean scores of these two year groups.

(a) Test the regional education officer's claim at the 1% significance level. You should state your hypotheses, test statistic and critical value clearly.

(7)

(b) Explain the significance of the Central Limit Theorem in part (a).

(1)

Question 2 continued

Question 2 continued

Question 2 continued

(Total for Question 2 is 8 marks)

3. A medical research team carried out an investigation into the metabolic rate, MR, of men aged between 30 years and 60 years.

A random sample of 10 men was taken from this age group.

The table below shows for each man his MR and his body mass index, BMI. The table also shows the rank for the level of daily physical activity, DPA, which was assessed by the medical research team.

Rank 1 was assigned to the man with the highest level of daily physical activity.

Man	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
MR (*x*)	6.24	5.94	6.83	6.53	6.31	7.44	7.32	8.70	7.88	7.78
BMI (*y*)	19.6	19.2	23.6	21.4	20.2	20.8	22.9	25.5	23.3	25.1
DPA rank	10	7	9	8	6	3	1	4	5	2

[You may use $S_{xy} = 15.1608$ $S_{xx} = 6.901\,81$ $S_{yy} = 45.304$]

(a) Calculate the value of the product moment correlation coefficient between MR and BMI for these 10 men.

(2)

(b) Use your value of the product moment correlation coefficient to test, at the 5% significance level, whether or not there is evidence of a positive correlation between MR and BMI.
State your hypotheses clearly.

(3)

(c) State an assumption that must be made to carry out the test in part (b).

(1)

(d) Calculate the value of Spearman's rank correlation coefficient between MR and DPA for these 10 men.

(4)

(e) Use a two-tailed test and a 5% level of significance to assess whether or not there is evidence of a correlation between MR and DPA.

(2)

Question 3 continued

Question 3 continued

Question 3 continued

(Total for Question 3 is 12 marks)

4. A survey was carried out with students that had studied Maths, Physics and Chemistry at a college between 2016 and 2020. The students were divided into two groups *A* and *B*.

(a) Explain how a sample could be obtained from this population using quota sampling.

(2)

The students were asked which of the three subjects they enjoyed the most.
The results of the survey are shown in the table.

	Subject enjoyed the most			
	Maths	Physics	Chemistry	Total
Group *A*	16	10	13	39
Group *B*	38	13	10	61
Total	54	23	23	100

(b) Test, at the 5% level of significance, whether the subject enjoyed the most is independent of group.
You should state your hypotheses, expected frequencies, test statistic and the critical value used for this test.

(8)

The Headteacher discovered later that the results were actually based on a random sample of 200 students but had been recorded in the table as percentages.

(c) For the test in part (b), state with reasons the effect, if any, that this information would have on

(i) the null and alternative hypotheses,

(ii) the critical value,

(iii) the value of the test statistic,

(iv) the conclusion of the test.

(4)

Question 4 continued

Question 4 continued

Question 4 continued

(Total for Question 4 is 14 marks)

5. Charlie is training for three events: a 1500 m swim, a 40 km bike ride and a 10 km run.

From past experience his times, in minutes, for each of the three events independently have the following distributions.

$S \sim N(41, 5.2^2)$ represents the time for the swim
$B \sim N(81, 4.2^2)$ represents the time for the bike ride
$R \sim N(57, 6.6^2)$ represents the time for the run

(a) Find the probability that Charlie's total time for a randomly selected swim, bike ride and run exceeds 3 hours.

(5)

(b) Find the probability that the time for a randomly selected swim will be at least 20 minutes quicker than the time for a randomly selected run.

(4)

Given that $P(S + B + R > t) = 0.95$

(c) find the value of t

(3)

A triathlon consists of a 1500 m swim, immediately followed by a 40 km bike ride, immediately followed by a 10 km run.

Charlie uses the answer to part (a) to find the probability that, in 6 successive independent triathlons, his time will exceed 3 hours on at least one occasion.

(d) Find the answer Charlie should obtain.

(3)

Jane says that Charlie should not have used the answer to part (a) for the calculation in part (d).

(e) Explain whether or not Jane is correct.

(2)

Question 5 continued

Question 5 continued

Question 5 continued

(Total for Question 5 is 17 marks)

6. A farmer sells strawberries in baskets. The contents of each of 100 randomly selected baskets were weighed and the results, given to the nearest gram, are shown below.

Weight of strawberries (grams)	Number of baskets
302 – 303	5
304 – 305	13
306 – 307	10
308 – 309	18
310 – 311	25
312 – 313	20
314 – 315	5
316 – 317	4

The farmer proposes that the weight of strawberries per basket, in grams, should be modelled by a normal distribution with a mean of 310 g and standard deviation 4 g.

Using his model, the farmer obtains the following expected frequencies.

Weight of strawberries (s, grams)	Expected frequency
$s \leqslant 303.5$	a
$303.5 < s \leqslant 305.5$	7.8
$305.5 < s \leqslant 307.5$	13.6
$307.5 < s \leqslant 309.5$	18.4
$309.5 < s \leqslant 311.5$	19.6
$311.5 < s \leqslant 313.5$	16.3
$313.5 < s \leqslant 315.5$	10.6
$s > 315.5$	b

(a) Find the value of a and the value of b. Give your answers correct to one decimal place.

(5)

Question 6 continues on page 23

Question 6 continued

Before $s \leqslant 303.5$ and $s > 315.5$ are included, for the remaining cells,

$$\sum \frac{(O - E)^2}{E} = 9.71$$

(b) Using a 5% significance level, test whether the data are consistent with the model. You should state your hypotheses, the test statistic and the critical value used.

(7)

An alternative model uses estimates for the population mean and standard deviation from the data given.

Using these estimated values no expected frequency is below 5

Another test is to be carried out, using a 5% significance level, to assess whether the data are consistent with this alternative model.

(c) State the effect, if any, on the critical value for this test. Give a reason for your answer.

(2)

Question 6 continued

(Total for Question 6 is 14 marks)

TOTAL FOR PAPER: 75 MARKS

END

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel International Advanced Level

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 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 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. The table below shows the number of televised tournaments won and the total number of tournaments won by the top 10 ranked darts players in 2020

Player's rank	Televised tournaments won	Total tournaments won
1	55	135
2	7	33
3	5	17
4	2	14
5	4	9
6	2	5
7	9	36
8	0	15
9	3	3
10	0	13

Michael did not want to calculate Spearman's rank correlation coefficient between player's rank and the rank in televised tournaments won because there would be tied ranks.

(a) Explain how Michael could have dealt with these tied ranks.

(1)

Given that the largest number of total tournaments won is ranked number 1

(b) calculate the value of Spearman's rank correlation coefficient between player's rank and the rank in total tournaments won.

(4)

(c) Stating your hypotheses and critical value clearly, test at the 5% level of significance, whether or not there is evidence of a positive correlation between player's rank and the rank in total tournaments won for these darts players.

(4)

Michael does not believe that there is a positive correlation between player's rank and the rank in total number of tournaments won.

(d) Find the largest level of significance, that is given in the tables provided, which could be used to support Michael's claim.
You must state your critical value.

(1)

Question 1 continued

Question 1 continued

Question 1 continued

(Total for Question 1 is 10 marks)

2. An experiment is conducted to compare the heat retention of two brands of flasks, brand A and brand B. Both brands of flask have a capacity of 750 ml.

In the experiment 750 ml of boiling water is poured into the flask, which is then sealed. Four hours later the temperature, in °C, of the water in the flask is recorded.

A random sample of 100 flasks from brand A gives the following summary statistics, where x is the temperature of the water in the flask after four hours.

$$\sum x = 7690 \qquad \sum (x - \bar{x})^2 = 669.24$$

(a) Find unbiased estimates for the mean and variance of the temperature of the water, after four hours, for brand A.

(3)

A random sample of 80 flasks from brand B gives the following results, where y is the temperature of the water in the flask after four hours.

$$\bar{y} = 75.9 \qquad s_y = 2.2$$

(b) Test, at the 1% significance level, whether there is a difference in the mean water temperature after four hours between brand A and brand B.
You should state your hypotheses, test statistic and critical value clearly.

(7)

(c) Explain why it is reasonable to assume that $\sigma^2 = s^2$ in this situation.

(1)

Question 2 continued

Question 2 continued

Question 2 continued

(Total for Question 2 is 11 marks)

3. The random variable X is normally distributed with unknown mean μ and known variance σ^2

A random sample of 25 observations of X produced a 95% confidence interval for μ of (26.624, 28.976)

(a) Find the mean of the sample.

(1)

(b) Show that the standard deviation is 3

(3)

The a % confidence interval using the 25 observations has a width of 2.1

(c) Calculate the value of a

(6)

(d) Find the smallest sample size, of observations from X, that would be required to obtain a 95% confidence interval of width at most 1.5

(4)

Question 3 continued

Question 3 continued

Question 3 continued

(Total for Question 3 is 14 marks)

4. Navtej travels to work by train. A train leaves the station every 7 minutes and Navtej's arrival at the station is independent of when the train is due to leave.

(a) Write down a suitable model for the distribution of the time, T minutes, that he has to wait for a train to leave.

(1)

(b) Find the mean and standard deviation of T

(3)

During a 10-week period, Navtej travels to work by train on 46 occasions.

(c) Estimate the probability that the mean length of time that he has to wait for a train to leave is between 3.4 and 3.6 minutes.

(5)

(d) State a necessary assumption for the calculation in part (c).

(1)

Question 4 continued

(Total for Question 4 is 10 marks)

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

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

(1)

(b) Explain what you understand by an unbiased estimator for μ

(1)

Two estimators for μ are U_1 and U_2 where

$$U_1 = 3X_1 - 2X_2 \qquad \text{and} \qquad U_2 = \frac{X_1 + 3X_2}{4}$$

(c) Show that both U_1 and U_2 are unbiased estimators for μ

(3)

The most efficient estimator among a group of unbiased estimators is the one with the smallest variance.

(d) By finding the variance of U_1 and the variance of U_2 state, giving a reason, the most efficient estimator for μ from these two estimators.

(4)

Question 5 continued

Question 5 continued

Question 5 continued

(Total for Question 5 is 9 marks)

6 A particular lift has a maximum load capacity of 700 kg.

The weights of men are normally distributed with mean 80 kg and standard deviation 10 kg.

The weights of women are normally distributed with mean 69 kg and standard deviation 5 kg.

You may assume that weights of people are independent.

(a) Find the probability that when 6 men and 3 women are in the lift, the load exceeds 700 kg.

(4)

A sign in the lift states: "Maximum number of people in the lift is c"

(b) Find the value of c such that the probability of the load exceeding 700 kg is less than 2.5% no matter the gender of the occupants.

(6)

Question 6 continued

Question 6 continued

Question 6 continued

(Total for Question 6 is 10 marks)

7 The following table shows observed frequencies, where x is an integer, from an experiment to test whether or not a six-sided die is biased.

Number on die	1	2	3	4	5	6
Observed frequency	$x + 6$	$x - 8$	$x + 8$	$x - 5$	$x + 4$	$x - 5$

A goodness of fit test is conducted to determine if there is evidence that the die is biased.

(a) Write down suitable null and alternative hypotheses for this test.

(1)

It is found that the null hypothesis is not rejected at the 5% significance level.

(b) Hence

(i) find the minimum value of x

(8)

(ii) determine the minimum number of times the die was rolled.

(2)

Question 7 continued

Question 7 continued

Question 7 continued

Question 7 continued

(Total for Question 7 is 11 marks)

TOTAL FOR PAPER: 75 MARKS

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel International Advanced Level

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 machine fills bottles with mineral water.

The machine is checked every day to ensure that it is working correctly. On a particular day a random sample of 100 bottles is taken. The volume of water, x millilitres, for each bottle is measured and each measurement is coded using

$$y = x - 1000$$

The results are summarised below

$$\sum y = 847 \qquad \sum y^2 = 13\ 510.09$$

(a) (i) Show that the value of the unbiased estimate of the mean of x is 1008.47

(ii) Calculate the unbiased estimate of the variance of x

(4)

The machine was initially set so that the volume of water in a bottle had a mean value of 1010 millilitres.

Later, a test at the 5% significance level is used to determine whether or not the mean volume of water in a bottle has changed. If it has changed then the machine is stopped and reset.

(b) Write down suitable null and alternative hypotheses for a 2-tailed test.

(1)

(c) Find the critical region for $\bar{X}$ in the above test.

(4)

(d) Using your answer to part (a) and your critical region found in part (c), comment on whether or not the machine needs to be stopped and reset.
Give a reason for your answer.

(2)

(e) Explain why the use of $\sigma^2 = s^2$ is reasonable in this situation.

(1)

Question 1 continued

Question 1 continued

Question 1 continued

(Total for Question 1 is 12 marks)

2 The table shows the season's best times, x seconds, for the 8 athletes who took part in the 200 m final in the 2021 Tokyo Olympics. It also shows their finishing position in the race.

Athlete	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Season's best time	19.89	19.83	19.74	19.84	19.91	19.99	20.13	20.10
Finishing position	1	2	3	4	5	6	7	8

Given that the fastest season's best time is ranked number 1

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

(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 rank of the season's best time and the finishing position for these athletes.

(4)

Chris suggests that it would be better to use the actual finishing time, y seconds, of these athletes rather than their finishing position.

Given that

$$S_{xx} = 0.1286875 \qquad S_{yy} = 0.55275 \qquad S_{xy} = 0.225175$$

(c) calculate the product moment correlation coefficient between the season's best time and the finishing time for these athletes.
Give your answer correct to 3 decimal places.

(2)

(d) Use your value of the product moment correlation coefficient to test, at the 1% level of significance, whether or not there is evidence of a positive correlation between the season's best time and the finishing time for these athletes.

(2)

Question 2 continued

Question 2 continued

Question 2 continued

(Total for Question 2 is 12 marks)

3 A mobile phone company offers an insurance policy to its customers when they purchase a mobile phone. The company conducted a survey on the age of the customers and whether or not claims were made.

A random sample of 1200 customers from this company was investigated for 2020 and the results are shown in the table below.

		Claim made in 2020	No claim made in 2020	Total
Age	**17 – 20 years**	24	176	200
	21 – 50 years	48	652	700
	51 years and over	14	286	300
	Total	86	1114	1200

The data are to be used to determine whether or not making a claim is independent of age.

(a) Calculate the expected frequencies for the age group 51 years and over that

(i) made a claim in 2020

(ii) did not make a claim in 2020

(2)

The 4 classes of customers aged between 17 and 50 give a value of

$\sum \frac{(O-E)^2}{E} = 7.123$ correct to 3 decimal places.

(b) Test, at the 1% level of significance, whether or not making a claim is independent of age. Show your working clearly, stating your hypotheses, the degrees of freedom, the test statistic and the critical value used.

(7)

Question 3 continued

Question 3 continued

Question 3 continued

(Total for Question 3 is 9 marks)

4 A research student is investigating the number of children who are girls in families with 4 children.

The table below shows her results for 200 such families.

Number of girls	0	1	2	3	4
Frequency	15	68	69	38	10

The research student suggests that a binomial distribution with $p = \frac{1}{2}$ could be a suitable model for the number of children who are girls in a family of 4 children.

(a) Using her results and a 5% significance level, test the research student's claim.
You should state your hypotheses, expected frequencies, test statistic and the critical value used.

(8)

The research student decides to refine the model and retains the idea of using a binomial distribution but does not specify the probability that the child is a girl.

(b) Use the data in the table to show that the probability that a child is a girl is 0.45

(2)

The research student uses the probability from part (b) to calculate a new set of expected frequencies, none of which are less than 5

The statistic $\sum \frac{(O-E)^2}{E}$ is evaluated and found to be 2.47

(c) Test, at the 5% significance level, whether using a binomial distribution is suitable to model the number of children who are girls in a family of 4 children.
You should state your hypotheses and the critical value used.

(4)

Question 4 continued

Question 4 continued

Question 4 continued

(Total for Question 4 is 14 marks)

5 Claire grows strawberries on her farm. She wants to compare two brands of fertiliser, brand *A* and brand *B*.

She grows two sets of plants of the same variety of strawberries under the same conditions, fertilising one set with brand *A* and the other with brand *B*.

The yields per plant, in grams, from each set of plants are summarised below.

	Mean	**Standard deviation**	**Number of plants**
Fertiliser *A*	1377	17.8	50
Fertiliser *B*	1368	18.4	40

(a) Stating your hypotheses clearly, carry out a suitable test to assess whether the mean yield from plants using fertiliser *A* is greater than the mean yield from plants using fertiliser *B*.
Use a 1% level of significance and state your test statistic and critical value.

(7)

The total cost of fertiliser *A* for Claire's 50 plants was £75

The total cost of fertiliser *B* for Claire's 40 plants was £50

Claire sells all her strawberries at £3 per kilogram.

(b) Use this information, together with your answer in part (a), to advise Claire on which of the two brands of fertiliser she should use next year in order to maximise her expected profit per plant, giving a reason for your answer.

(3)

Question 5 continued

Question 5 continued

Question 5 continued

(Total for Question 5 is 10 marks)

6 A garden centre sells bags of stones and large bags of gravel.

The weight, X kilograms, of stones in a bag can be modelled by a normal distribution with unknown mean μ and known standard deviation 0.4

The stones in each of a random sample of 36 bags from a large batch is weighed. The total weight of stones in these 36 bags is found to be 806.4 kg

(a) Find a 98% confidence interval for the mean weight of stones in the batch.

(4)

(b) Explain why the use of the Central Limit theorem is not required to answer part (a)

(1)

The manufacturer of these bags of stones claims that bags in this batch have a mean weight of 22.5 kg

(c) Using your answer to part (a), comment on the claim made by the manufacturer.

(2)

The weight, Y kilograms, of gravel in a large bag can be modelled by a normal distribution with mean 850 kg and standard deviation 5 kg

A builder purchases 10 large bags of gravel.

(d) Find the probability that the mean weight of gravel in the 10 large bags is less than 848 kg

(3)

Question 6 continued

Question 6 continued

Question 6 continued

(Total for Question 6 is 10 marks)

7 At a particular supermarket, the times taken to serve each customer in a queue at a standard checkout may be modelled by a normal distribution with mean 240 seconds and standard deviation 20 seconds.

There is a queue of 3 customers at a standard checkout.

Making a reasonable assumption about the times taken to serve these customers,

(a) find the probability that the total time taken to serve the 3 customers will be less than 11 minutes.

(3)

(b) State the assumption you have made in part (a)

(1)

In the supermarket there is also an express checkout, which is reserved for customers buying 10 or fewer items. The time taken to serve a customer at this express checkout may be modelled by a normal distribution with mean 100 seconds and standard deviation 8 seconds.

On a particular day Jiang has 8 items to pay for and has to choose whether to join a queue of 3 customers waiting at a standard checkout or a queue of 7 customers waiting at the express checkout.

Using a similar assumption to that made in part (a),

(c) find the probability that the total time taken to serve the 3 customers at the standard checkout will exceed the total time taken to serve the 7 customers at the express checkout.

(4)

Question 7 continued

Question 7 continued

(Total for Question 7 is 8 marks)

TOTAL FOR PAPER IS 75 MARKS

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

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

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

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

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

(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 $\overline{X}$ denotes the sample mean.

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

$$\left(\overline{x} - 2.60, \overline{x} + 2.60\right)$$

(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 $\overline{Y}$ denotes the sample mean.

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

(3)

(c) Given that X and Y are independent,

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

(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

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

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 ▶

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)

Question 1 continued

Question 1 continued

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)

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)

Question 3 continued

Question 3 continued

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)

Question 4 continued

Question 4 continued

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)

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)

Question 6 continued

Question 6 continued

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.

$W \sim \text{N}(a, b)$ 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)

Question 7 continued

Question 7 continued

(Total for Question 7 is 14 marks)

TOTAL FOR PAPER IS 75 MARKS

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel International Advanced Level

Thursday 13 June 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.

1. The names of the 400 employees of a company are listed alphabetically in a book.

The chairperson of the company wishes to select a sample of 8 employees.

The chairperson numbers the employees from 001 to 400

(a) Describe how the list of numbers can be used to select a systematic sample of 8 employees.

(2)

(b) State one disadvantage of systematic sampling in this case.

(1)

(c) Write down the probability that the sample includes both the first name (employee 001) and the last name (employee 400) in the list.

(1)

Question 1 continued

(Total for Question 1 is 4 marks)

2. Aarush is asked to estimate the price of 7 kettles and rank them in order of decreasing price.

Aarush's order of decreasing price is *D A F C B G E*

The actual prices of the 7 kettles are shown in the table below.

Kettle	*A*	*B*	*C*	*D*	*E*	*F*	*G*
Price (£)	99.99	14.99	34.97	49.99	19.97	29.99	8.99

(a) Calculate Spearman's rank correlation coefficient between Aarush's order and the actual order.

Use a rank of 1 for the highest priced kettle.
Show your working clearly.

(4)

(b) Using a 5% level of significance, test whether or not there is evidence to suggest that Aarush is able to rank kettles in order of decreasing price.

You should state your hypotheses and critical value.

(4)

(c) Explain why Aarush did not use the product moment correlation coefficient in this situation.

(1)

Aarush discovered that kettle *A*'s price was recorded incorrectly and should have been £49.99 rather than £99.99

(d) Explain what effect this has on the rankings for the price.

(1)

Question 2 continued

Question 2 continued

Question 2 continued

(Total for Question 2 is 10 marks)

3. The volume of water in a bottle has a normal distribution with **unknown** mean, μ millilitres, and **known** standard deviation, σ millilitres.

A random sample of 150 of the bottles of water gave a 95% confidence interval for μ of

$$(327.84, 329.76)$$

(a) Using the confidence interval given, test whether or not $\mu = 328$

State your hypotheses clearly and write down the significance level you have used.

(3)

A second random sample, of 200 of these bottles of water, had a mean volume of 328 millilitres.

(b) Calculate a 98% confidence interval for μ based on this second sample.

You must show all steps in your working.
(*Solutions relying entirely on calculator technology are not acceptable.*)

(6)

Using five different random samples of 200 of these bottles of water, five 98% confidence intervals for μ are to be found.

(c) Calculate the probability that more than 3 of these intervals will contain μ

(3)

Question 3 continued

Question 3 continued

Question 3 continued

(Total for Question 3 is 12 marks)

4. The manager of a company making ice cream believes that the proportions of people in the population who prefer vanilla, chocolate, strawberry and other are in the ratio 10 : 5 : 2 : 3

The manager takes a random sample of 400 customers and records their age and favourite ice cream flavour. The results are shown in the table below.

		Ice cream flavour				
		Vanilla	Chocolate	Strawberry	Other	Total
Age	Child	95	25	13	25	158
	Teenager	57	20	17	36	130
	Adult	36	50	10	16	112
	Total	188	95	40	77	400

(a) Use the data in the table to test, at the 5% level of significance, the manager's belief. You should state your hypotheses, test statistic, critical value and conclusion clearly.

(8)

A researcher wants to investigate whether or not there is a relationship between the age of a customer and their favourite ice cream flavour. In order to test whether favourite ice cream flavour and age are related, the researcher plans to carry out a χ^2 test.

(b) Use the table to calculate expected frequencies for the group

(i) teenagers whose favourite ice cream flavour is vanilla,

(ii) adults whose favourite ice cream flavour is chocolate.

(2)

(c) Write down the number of degrees of freedom for this χ^2 test.

(1)

Question 4 continued

Question 4 continued

Question 4 continued

(Total for Question 4 is 11 marks)

5. A manager of a large company is investigating the time it takes the company's employees to complete a task.

The manager believes that the mean time for full-time employees to complete the task is more than a minute quicker than the mean time for part-time employees to complete the task.

The manager collects a random sample of 605 full-time employees and 45 part-time employees and records the times, t minutes, it takes each employee to complete the task.

The results are summarised in the table below.

	n	$\bar{t}$	s^2
Full-time employees	605	5.6	9
Part-time employees	45	7.0	4

(a) Test, at the 5% level of significance, the manager's claim.

You should state your hypotheses, test statistic, critical value and conclusion clearly.

(8)

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

(2)

The company increases the size of the sample of part-time employees to 46
The time taken to complete the task by the extra employee is 8 minutes.

(c) Find an unbiased estimate of the variance for the sample of 46 part-time employees.

(4)

Question 5 continued

Question 5 continued

Question 5 continued

(Total for Question 5 is 14 marks)

6. The weights of bags of carrots, C kg, are such that $C \sim N(1.2, 0.03^2)$

Three bags of carrots are selected at random.

(a) Calculate the probability that their total weight is more than 3.5 kg.

(4)

The weights of bags of potatoes, R kg, are such that $R \sim N(2.3, 0.03^2)$

Two bags of potatoes are selected at random.

(b) Calculate the probability that the difference in their weights is more than 0.05 kg.

(5)

The weights of trays, T kg, are such that $T \sim N(2.5, \sqrt{0.1}^2)$

The random variable G represents the total weight, in kg, of a single tray packed with 10 bags of potatoes where G and T are independent.

(c) Calculate $P(G < 2T + 20)$

(7)

Question 6 continued

Question 6 continued

Question 6 continued

(Total for Question 6 is 16 marks)

7. The continuous random variable D is uniformly distributed over the interval $[x - 1, x + 5]$ where x is a constant.

A random sample of n observations of D is taken, where n is large.

(a) Use the Central Limit Theorem to find an approximate distribution for $\overline{D}$
Give your answer in terms of n and x where appropriate.

(3)

The n observations of D have a sample mean of 24.6

Given that the lower bound of the 99% confidence interval for x is 22.101 to 3 decimal places,

(b) find the value of n
Show your working clearly.

(5)

Question 7 continued

Question 7 continued

Question 7 continued

Question 7 continued

(Total for Question 7 is 8 marks)

TOTAL FOR PAPER IS 75 MARKS