Please check the examination details below before entering your candidate information

Candidate surname StemUp Model Solution | Other names

Centre Number | Candidate Number

Pearson Edexcel International Advanced Level

Friday 13 June 2025

Afternoon (Time: 1 hour 30 minutes) | Paper reference **WST03/01**

Mathematics

International Advanced Subsidiary/Advanced Level
Statistics S3

You must have:
Mathematical Formulae and 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 8 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 ▶

P76201A

©2025 Pearson Education Ltd.
Y:1/1/1/

P 7 6 2 0 1 A 0 1 2 8

Pearson

stemup.co.uk

1: A golf club has 360 members. The captain of the golf club needs to select 20 players from the club to play in a charity tournament.

The 360 members of the golf club are listed by their randomly allocated membership numbers.

The captain decides to use systematic sampling to select the 20 players.

(a) Give a reason for using systematic sampling in this case.

(1)

The members are numbered from 001 to 360 and the 5th member on the list is randomly selected as the first player and then every xth member thereafter is selected.

(b) Write down the value of x

(1)

(c) Write down the probability that the sample includes both the 5th and the 6th member on the list.

(1)

(a) It reduces the potential for bias (favouritism) in selecting the players: the captain cannot choose players based on ability, as the members are listed in a random order of membership numbers.

(b) $x = \frac{360}{20} = 18$

x = 18

(c) The 6th member cannot be selected if the 5th is selected (5, 23, 41, ...)

Probability = 0

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

stemup.co.uk

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

Question 1 continued

(Total for Question 1 is 3 marks)

2: At a local talent show 2 judges were asked to rank 10 performances, in order of preference. The results are shown in the following table.

Performance	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Judge 1	1	3	5	7	9	10	2	6	8	4
Judge 2	2	1	3	7	9	4	5	8	10	6

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

(3)

(b) Stating your hypotheses clearly, test at the 1% level of significance, whether or not there is evidence of a positive correlation between the ranks of Judge 1 and the ranks of Judge 2

(4)

Performance	A	B	C	D	E	F	G	H	I	J
d	-1	2	2	0	0	6	-3	-2	-2	-2
d^2	1	4	4	0	0	36	9	4	4	4

(a) $\Sigma d^2 = 1 + 4 + 4 + 0 + 0 + 36 + 9 + 4 + 4 + 4 = 66$

$r_s = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)} = 1 - \frac{6 \times 66}{10(10^2 - 1)} = 1 - \frac{396}{990}$

$r_s = 0.6$

(b) H_0: $\rho_s = 0$ H_1: $\rho_s > 0$

$n = 10$, 1% one-tailed: critical value $r_s = 0.7455$ (CR: $r_s \geq 0.7455$)

$0.6 < 0.7455$ so not in the critical region: not significant, do not reject H_0

There is insufficient evidence of a positive correlation between the ranks of Judge 1 and the ranks of Judge 2.

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

stemup.co.uk

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 2 continued

(Total for Question 2 is 7 marks)

3: A garage sold 250 new cars which work using one of three types of engine: petrol, diesel or electric.

For each car, the garage recorded the month it first experienced an engine problem, if it occurred during the first year after being sold.

The results for the 250 cars are shown in the table.

	When the car first experienced an engine problem			
	Month 1–6	**Month 7–12**	**None**	**Total**
Petrol engine	12	25	95	132
Diesel engine	10	8	17	35
Electric engine	14	10	59	83
Total	36	43	171	250

The results are to be used to investigate whether or not there is an association between when the car first experienced an engine problem and the type of engine.

(a) Calculate the expected frequency for a **diesel** engine that first experienced an engine problem during **month 1–6**

(2)

Given that for the other 8 classes $\sum \frac{(O-E)^2}{E} = 7.444$ to 3 decimal places,

(b) test, at a 5% level of significance, whether or not there is evidence of an association between when the car first experienced an engine problem and the type of engine. You should state the hypotheses, the test statistic, the degrees of freedom and the critical value used for this test.

(7)

(a) $E = \frac{\text{row total} \times \text{column total}}{\text{grand total}} = \frac{35 \times 36}{250} = \frac{1260}{250}$

Expected frequency = 5.04

(b) H_0: there is no association between when the car first experienced an engine problem and the type of engine (they are independent)

H_1: there is an association between when the car first experienced an engine problem and the type of engine (they are not independent)

Diesel, month 1–6: $O = 10$, $E = 5.04$, $\frac{(O-E)^2}{E} = \frac{(10-5.04)^2}{5.04} = \frac{24.6016}{5.04} = 4.8813...$

$X^2 = \Sigma\frac{(O-E)^2}{E} = 7.444 + 4.8813... = 12.325...$

P 7 6 2 0 1 A 0 6 2 8
stemup.co.uk

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

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

Question 3 continued

Test statistic $X^2 = 12.3$ (3 s.f.)

Degrees of freedom $\nu = (3 - 1)(3 - 1) = 4$

Critical value $\chi^2_4(0.05) = 9.488$ (CR: $X^2 \geq 9.488$)

$12.3 > 9.488$ so in the critical region: significant, reject H_0

There is evidence of an association between when the car first experienced an engine problem and the type of engine.

Question 3 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

stemup.co.uk

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

Question 3 continued

(Total for Question 3 is 9 marks)

4: A biologist claims that the mean weight of grey squirrels is 200 grams more than the mean weight of red squirrels.

A research student believes that the mean weight of grey squirrels is greater than 200 grams more than the mean weight of red squirrels.

To test this, the research student collects and weighs a random sample of 80 grey squirrels and 80 red squirrels.

The results are summarised below, where

- x denotes the weights, in grams, of the grey squirrels
- y denotes the weights, in grams, of the red squirrels

$$\bar{x} = 500 \qquad s_x^2 = 1156 \qquad \sum y = 23\,200 \qquad \sum y^2 = 6\,741\,351$$

(a) Find an unbiased estimate of the mean and an unbiased estimate of the variance for the weights of the red squirrels.

(3)

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

(7)

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

(1)

(a) $\bar{y} = \frac{\Sigma y}{n} = \frac{23200}{80} = 290$

$s_y^2 = \frac{\Sigma y^2 - 80 \times 290^2}{80 - 1} = \frac{6741351 - 6728000}{79} = \frac{13351}{79} = 169$

Unbiased estimate of mean = 290 g, unbiased estimate of variance = 169 g^2

(b) H_0: $\mu_x - \mu_y = 200$ H_1: $\mu_x - \mu_y > 200$

$z = \frac{(500 - 290) - 200}{\sqrt{(1156/80 + 169/80)}} = \frac{10}{\sqrt{16.5625}} = \frac{10}{4.06971...}$

$z = 2.45718... = 2.46$ (3 s.f.)

5% one-tailed: critical value $z = 1.6449$ (CR: $Z \geq 1.6449$)

$2.46 > 1.6449$ so in the critical region: significant, reject H_0

There is sufficient evidence to support the research student's belief: the mean weight of grey squirrels is more than 200 g greater than the mean weight of red squirrels.

(c) I assumed that the sample variances are equal to the population variances, i.e. $s_x^2 = \sigma_x^2$ and $s_y^2 = \sigma_y^2$ (reasonable as both samples are large).

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

stemup.co.uk

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 4 continued

Question 4 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

stemup.co.uk

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

Question 4 continued

(Total for Question 4 is 11 marks)

5: The following table shows the number of male puppies born in 250 dog litters of size 5

Number of males	0	1	2	3	4	5
Number of litters	2	40	90	85	30	3

Jeff believes that a binomial distribution would be a suitable model for these data.

(a) Find the proportion of male puppies born in these litters of size 5

(2)

Jeff calculates expected frequencies, to 2 decimal places, as follows.

Number of males	0	1	2	3	4	5
Expected frequencies	r	41.92	79.91	76.16	36.30	6.92

(b) Find the value of r

(2)

The value of $\sum \frac{(O-E)^2}{E}$ for the given values in the table **excluding *r*** is 5.70 to 2 decimal places.

(c) Using a 5% significance level, test whether or not a binomial distribution is a suitable model for the number of male puppies born in these 250 litters.
You should state the hypotheses, the degrees of freedom and the critical value used.

(7)

(a) Total males = $0\times2 + 1\times40 + 2\times90 + 3\times85 + 4\times30 + 5\times3$

$= 0 + 40 + 180 + 255 + 120 + 15 = 610$

$\hat{p} = \frac{610}{250 \times 5} = \frac{610}{1250} = 0.488$

(b) $r = 250 - (41.92 + 79.91 + 76.16 + 36.30 + 6.92)$

$r = 250 - 241.21 = 8.79$

(c) H_0: a binomial distribution is a suitable model for the number of males

H_1: a binomial distribution is not a suitable model for the number of males

All expected frequencies are at least 5, so no cells need combining.

Number of males 0: $\frac{(O-E)^2}{E} = \frac{(2-8.79)^2}{8.79} = \frac{46.1041}{8.79} = 5.2450...$

$X^2 = \Sigma\frac{(O-E)^2}{E} = 5.70 + 5.2450... = 10.945...$

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

stemup.co.uk

Question 5 continued

Test statistic $X^2 = 10.95$ (2 d.p.)

Cells = 6, p estimated: $\nu = 6 - 1 - 1 = 4$

Critical value $\chi^2_4(0.05) = 9.488$ (CR: $X^2 \geq 9.488$)

$10.95 > 9.488$ so in the critical region: significant, reject H_0

There is significant evidence to suggest that a binomial distribution is not a suitable model for the number of male puppies born in these litters.

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 5 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

stemup.co.uk

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 5 continued

(Total for Question 5 is 11 marks)

stemup.co.uk

6: A population has an unknown mean μ and an unknown variance σ^2

A random sample represented by $X_1, X_2, \ldots, X_{20}$ is taken from the population.

(a) Explain why $\frac{1}{2}X_1 + \frac{3}{4}X_{20}$ is a statistic.

(1)

(b) Show that $R = \frac{1}{2}X_1 + \frac{3}{4}X_{20}$ is a biased estimator of μ

(2)

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

(1)

Given that $S = \frac{aX_1 - bX_2}{2}$ and $T = \frac{\sum_{i=1}^{14} aX_i + \sum_{i=15}^{20} bX_i}{20}$ are each used as an unbiased estimator of μ, where a and b are constants,

(d) find the value of a and the value of b

(4)

(a) It is calculated solely from the sample observations X_1 and X_{20} and contains no unknown population parameters (μ or σ).

(b) $E(R) = \frac{1}{2}E(X_1) + \frac{3}{4}E(X_{20}) = \frac{1}{2}\mu + \frac{3}{4}\mu = \frac{5}{4}\mu$

$E(R) = \frac{5}{4}\mu \neq \mu$, so R is a biased estimator of μ

(c) Bias $= E(R) - \mu = \frac{5}{4}\mu - \mu = \frac{1}{4}\mu$

(d) $E(S) = \frac{aE(X_1) - bE(X_2)}{2} = \frac{(a-b)\mu}{2} = \mu \Rightarrow a - b = 2$

$E(T) = \frac{14a\mu + 6b\mu}{20} = \mu \Rightarrow 14a + 6b = 20$

Substitute $a = b + 2$: $14(b + 2) + 6b = 20 \Rightarrow 20b + 28 = 20$

$20b = -8 \Rightarrow b = -0.4$, $a = -0.4 + 2 = 1.6$

$a = 1.6$, $b = -0.4$

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

stemup.co.uk

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 6 continued

(Total for Question 6 is 8 marks)

stemup.co.uk

7: The random variable X is such that $X \sim \mathrm{N}(\mu, \sigma^2)$

The mean of a random sample of n observations of X is denoted by $\bar{X}_n$

Given that

$$\mathrm{P}(\bar{X}_n < 45.2) = 0.3446 \text{ and } \mathrm{P}(\bar{X}_n > 75.2) = 0.0179$$

and using the statistical tables provided,

(a) show that $\sigma = 12\sqrt{n}$

(6)

(b) Hence, find the value of μ

(2)

(c) Find $\mathrm{P}(\bar{X}_n > 59)$

(3)

(a) $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$. $P(\bar{X}_n < 45.2) = 0.3446 \Rightarrow \frac{45.2 - \mu}{\sigma/\sqrt{n}} = -0.4$

So $\mu = 45.2 + \frac{0.4\sigma}{\sqrt{n}}$ ... (1)

$P(\bar{X}_n > 75.2) = 0.0179 \Rightarrow P(\bar{X}_n < 75.2) = 0.9821 \Rightarrow \frac{75.2 - \mu}{\sigma/\sqrt{n}} = 2.1$

So $\mu = 75.2 - \frac{2.1\sigma}{\sqrt{n}}$ ... (2)

(1) = (2): $45.2 + \frac{0.4\sigma}{\sqrt{n}} = 75.2 - \frac{2.1\sigma}{\sqrt{n}} \Rightarrow \frac{2.5\sigma}{\sqrt{n}} = 30$

$\frac{\sigma}{\sqrt{n}} = \frac{30}{2.5} = 12 \Rightarrow \sigma = 12\sqrt{n}$ as required

(b) $\mu = 45.2 + 0.4 \times \frac{12\sqrt{n}}{\sqrt{n}} = 45.2 + 0.4 \times 12 = 45.2 + 4.8$

$\mu = 50$

(c) Standard error $= \frac{\sigma}{\sqrt{n}} = 12$, so $\bar{X}_n \sim N(50, 12^2)$

$P(\bar{X}_n > 59) = P\left(Z > \frac{59 - 50}{12}\right) = P(Z > 0.75)$

$= 1 - 0.7734 = 0.2266 = 0.227$ (3 s.f.)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

stemup.co.uk

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 7 continued

Question 7 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

stemup.co.uk

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 7 continued

(Total for Question 7 is 11 marks)

8: The weights, Skg, of bags of sand are such that

$$S \sim \mathrm{N}(18, 0.03^2)$$

Three bags of sand are selected at random.

(a) Using standardisation, calculate the probability that their total weight is more than 54.1 kg.

(4)

The weights, Ckg, of bags of cement are such that

$$C \sim \mathrm{N}(25, 0.03^2)$$

Two bags of cement are selected at random.

(b) Using standardisation, calculate the probability that their weights differ by more than 0.02 kg.

(5)

The weights, Pkg, of pallets are such that

$$P \sim \mathrm{N}(15, 0.2^2)$$

The random variable T represents the total weight, in kg, of 28 bags of sand with 5 bags of cement and a single pallet of weight P_1

(c) Using standardisation, calculate the probability that $T < 30P_1 + 190$

(6)

(a) $X = S_1 + S_2 + S_3$

$E(X) = 3 \times 18 = 54$, $\mathrm{Var}(X) = 3 \times 0.03^2 = 0.0027$, $X \sim N(54, 0.0027)$

$P(X > 54.1) = P\left(Z > \frac{54.1 - 54}{\sqrt{0.0027}}\right) = P(Z > 1.92450\ldots)$

$= 1 - 0.9726 = 0.0274$ (calculator $0.027146\ldots$) ≈ 0.027

(b) $Y = C_1 - C_2$

$E(Y) = 25 - 25 = 0$, $\mathrm{Var}(Y) = 0.03^2 + 0.03^2 = 0.0018$, $Y \sim N(0, 0.0018)$

$P(Y > 0.02) = P\left(Z > \frac{0.02 - 0}{\sqrt{0.0018}}\right) = P(Z > 0.47140\ldots)$

$= 1 - 0.6808 = 0.3192$ (calculator $0.318675\ldots$)

$P(|C_1 - C_2| > 0.02) = 2 \times P(Y > 0.02) = 2 \times 0.3192$

$= 0.6384$ (calculator $0.63735\ldots$) ≈ 0.638

(c) $T = P_1 + C_1 + \ldots + C_5 + S_1 + \ldots + S_{28}$

$T < 30P_1 + 190 \Leftrightarrow T - 30P_1 < 190$

Let $R = T - 30P_1 = C_1 + \ldots + C_5 + S_1 + \ldots + S_{28} - 29P_1$

$E(R) = 5 \times 25 + 28 \times 18 - 29 \times 15 = 125 + 504 - 435 = 194$

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

stemup.co.uk

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

Question 8 continued

$\text{Var}(R) = 5 \times 0.03^2 + 28 \times 0.03^2 + 29^2 \times 0.2^2$

$= 0.0045 + 0.0252 + 33.64 = 33.6697, \quad R \sim N(194, 33.6697)$

$$P(R < 190) = P\left(Z < \frac{190 - 194}{\sqrt{33.6697}}\right) = P(Z < -0.68935\ldots)$$

$= 1 - 0.7549 = 0.2451$ (calculator $0.24530\ldots$) ≈ 0.245

Question 8 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

stemup.co.uk

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 8 continued

Question 8 continued

(Total for Question 8 is 15 marks)

TOTAL FOR PAPER IS 75 MARKS

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

stemup.co.uk