Please check the examination details below before entering your candidate information

Candidate surname StemUp Model Solution | Other names

Pearson Edexcel
International Advanced Level

Centre Number | Candidate Number

Thursday 23 May 2019

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.
- 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 ▶

P56167A

©2019 Pearson Education Ltd.

1/1/1/

P 5 6 1 6 7 A 0 1 2 4

Pearson

stemup.co.uk

Leave blank

Answer All questions. Write your answers in the spaces provided.

1. A head teacher wants to find out the students' opinions about the length of lessons. The head teacher wishes to survey the students using the three groups in the table below.

	Number of students
Group 1 (Y7 – Y9)	432
Group 2 (Y10 – Y11)	360
Group 3 (Y12 – Y13)	108

She decides to take a stratified sample of 50 students.

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

(3)

(b) Give one advantage of carrying out this survey using stratified sampling, rather than taking a simple random sample of the whole school.

(1)

(a) Total number of students = 432 + 360 + 108 = 900

Group 1: $\frac{432}{900} \times 50 = 24$ Group 2: $\frac{360}{900} \times 50 = 20$

Group 3: $\frac{108}{900} \times 50 = 6$

Number (label) the students in each group separately, e.g. Group 1 from 1 to 432, Group 2 from 1 to 360 and Group 3 from 1 to 108.
Use random numbers (e.g. a random number generator) to select a simple random sample of 24 students from Group 1, 20 students from Group 2 and 6 students from Group 3, ignoring any repeats.

(b) Each year group is fairly represented in the sample, so the sample reflects the population structure. This matters as opinions on lesson length may differ systematically between the age groups.

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

Leave blank

Question 1 continued

Q1

(Total 4 marks)

Leave blank

2. A holiday camp manager wishes to investigate the popularity of the three activities the camp offers to teenagers who attend. A random sample of 560 teenagers, who attended the camp, are asked which of the three activities is their favourite. The results are given in the table below.

		Activity		
		Water sports	**Bushcraft**	**Mountain activities**
Gender	**Boys**	142	96	90
	Girls	110	46	76

(a) Test, at the 5% level of significance, whether or not there is evidence of an association between the teenagers' favourite activity and their gender. Show your working clearly, stating your hypotheses and degrees of freedom. You should state your expected frequencies correct to 2 decimal places where necessary.

(9)

(b) State which of the three activities could be worth further investigation. Give a reason for your answer.

(1)

(a) H_0: There is no association between gender and favourite activity

H_1: There is an association between gender and favourite activity

Expected frequency $= \frac{\text{row total} \times \text{column total}}{\text{grand total}}$, e.g. $\frac{328 \times 252}{560} = 147.6$

Expected	Water sports	Bushcraft	Mountain	Total
Boys	147.60	83.17	97.23	328
Girls	104.40	58.83	68.77	232
Total	252	142	166	560

O	E	$(O - E)^2/E$
142	147.60	0.2124...
96	83.17	1.9787...
90	97.23	0.5374...
110	104.40	0.3003...
46	58.83	2.7974...
76	68.77	0.7598...

$X^2 = \Sigma\frac{(O - E)^2}{E} = 0.2124... + 1.9787... + ... + 0.7598... = 6.5862...$

$X^2 = 6.59$ (awrt 6.6)

Degrees of freedom $\nu = (2 - 1)(3 - 1) = 2$

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

Leave blank

Question 2 continued

Critical value $\chi^2_2(0.05) = 5.991$ (CR: $X^2 \geq 5.991$)

6.59 > 5.991 so in the critical region, significant, reject H_0

There is evidence of an association between the teenagers' gender and their favourite activity.

(b) Bushcraft, as its cells contribute the most to the X^2 value (1.98 + 2.80 = 4.78 out of 6.59): boys choose it more often and girls less often than expected.

Leave blank

Question 2 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

Leave blank

Question 2 continued

(Total 10 marks)

Q2

Leave blank

3. In a singing competition, 9 people are asked to sing individually. Jamil is blindfolded and asked to rank the 9 singers in order of increasing age by listening to their voices.

Jamil places them in the order *B C A G F I E H D*

The actual ages, in years, of the 9 singers are shown in the table below.

Singer	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
Age	15	23	32	50	45	22	19	63	54

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

(5)

(b) Stating your hypotheses clearly, test whether or not there is evidence to suggest that Jamil can rank singers in order of age by listening to them sing. Use a 5% level of significance.

(4)

(c) Explain, without doing any further calculations, how you would find the rank correlation if singer *F*'s age was recorded incorrectly and is 23 rather than 22

(2)

(a) Rank ages from youngest = 1:

Singer	A	B	C	D	E	F	G	H	I
Jamil rank	3	1	2	9	7	5	4	8	6
Age rank	1	4	5	7	6	3	2	9	8
d	2	−3	−3	2	1	2	2	−1	−2
d^2	4	9	9	4	1	4	4	1	4

$\Sigma d^2 = 4 + 9 + 9 + 4 + 1 + 4 + 4 + 1 + 4 = 40$

$r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 1 - \frac{6 \times 40}{9 \times 80} = 1 - \frac{240}{720} = 0.6666...$

$r_s = 0.667$ (3 s.f.)

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

$n = 9$, 5% one-tailed: critical value $r_s = 0.6000$ (CR: $r_s \geq 0.6$)

$0.667 > 0.6$ so r_s lies in the critical region, significant, reject H_0

There is evidence that Jamil can rank the singers in order of age by listening to their voices (positive rank correlation between his ranks and the singers' ages).

(c) Singers B and F would then both be 23, a tie for ranks 3 and 4, so each

P 5 6 1 6 7 A 0 8 2 4

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

Leave blank

Question 3 continued

is given the average rank 3.5. With tied ranks, the rank correlation is found by calculating the product moment correlation coefficient (pmcc) of the ranks, rather than using the formula with Σd^2.

Q3

(Total 11 marks)

Leave blank

4. The weights of packets of mints are normally distributed with standard deviation 8 grams. A random sample of 50 packets of mints has a total weight of 10.84 kg.

(a) Find a 98% confidence interval for the mean weight of packets of mints.

(4)

(b) Explain why the Central Limit theorem is not required for part (a).

(1)

Bindy believes that the mean weight of a packet of mints is 0.22 kg.

(c) Giving a reason for your answer, comment on whether Bindy's belief is supported by the data.

(1)

(a) $\bar{x} = \frac{10\,840}{50} = 216.8$ g 98% confidence: $z = 2.3263$

$\bar{x} \pm z\frac{\sigma}{\sqrt{n}} = 216.8 \pm 2.3263 \times \frac{8}{\sqrt{50}} = 216.8 \pm 2.6319...$

$= (214.168..., 219.431...)$

98% confidence interval = (214, 219) grams [= (0.214, 0.219) kg]

(b) We are told the weights of the packets are normally distributed, so the sample mean is exactly normally distributed for any sample size.

(c) 0.22 kg = 220 g lies outside the confidence interval (220 > 219.4), so Bindy's belief is not supported by the data.

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

Leave blank

Question 4 continued

Q4

(Total 6 marks)

stemup.co.uk

Leave blank

5. A fair six-sided die has the number 1 on three faces, the number 3 on two faces and the number 6 on one face. A group of 5 people each roll the die 60 times and the mean score for each person is recorded.

Calculate an estimate for the probability that exactly 2 of the 5 people have a mean score of more than 2.75

(8)

$P(X = 1) = \frac{1}{2}$, $P(X = 3) = \frac{2}{6} = \frac{1}{3}$, $P(X = 6) = \frac{1}{6}$

$E(X) = 1 \times \frac{1}{2} + 3 \times \frac{1}{3} + 6 \times \frac{1}{6} = 2.5$

$E(X^2) = 1 \times \frac{1}{2} + 9 \times \frac{1}{3} + 36 \times \frac{1}{6} = 9.5$

$Var(X) = 9.5 - 2.5^2 = 3.25$

By the Central Limit Theorem, $\bar{X} \approx N\left(2.5, \frac{3.25}{60}\right)$ i.e. $\bar{X} \approx N\left(2.5, \frac{13}{240}\right)$

$P(\bar{X} > 2.75) = P\left(Z > \frac{2.75 - 2.5}{\sqrt{(3.25 \div 60)}}\right) = P(Z > 1.0741...)$

$= 1 - 0.8586... = 0.14137...$ $(p = 0.1414)$

Let Y = number of the 5 people with mean score more than 2.75

$Y \sim B(5, 0.1414)$

$P(Y = 2) = \binom{5}{2}(0.14137...)^2(1 - 0.14137...)^3$

$= 10 \times (0.14137...)^2 \times (0.85862...)^3$

$= 0.12651... = 0.127$ (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

Leave blank

Question 5 continued

Q5

(Total 8 marks)

Leave blank

6. Baako grows two varieties of strawberry plant, A and B. The mean yield per strawberry plant of variety A is 280 g. Baako believes that variety B gives a higher yield per strawberry plant than variety A. The yield per strawberry plant of variety B has standard deviation 70 g.
Baako takes a random sample of 125 strawberry plants of variety B and finds the mean yield is 290 g.

(a) Test, at the 5% level of significance, Baako's claim. State your hypotheses clearly.

(5)

Ayodele claims that using fertiliser increases the yield per strawberry plant by more than 100 g. He takes a random sample of 500 strawberry plants of variety C and adds fertiliser to just 300 of them. The yields, x grams, are summarised in the table below.

	n	Mean yield ($\bar{x}$)	s
With fertiliser	300	410	90
Without fertiliser	200	290	60

(b) Test, at the 0.5% level of significance, Ayodele's claim. State your hypotheses and show your working clearly.

(7)

(c) (i) State an assumption that you have made about the allocation of plants to the "With fertiliser" group in carrying out the test in part (b).

(ii) State a further necessary assumption that you have made in carrying out the test in part (b).

(2)

(a) H_0: $\mu_B = 280$ H_1: $\mu_B > 280$

$z = \dfrac{290 - 280}{70 \div \sqrt{125}} = \dfrac{10}{6.2609...} = 1.5971... = 1.60$

One-tailed 5%: critical value $z = 1.6449$ (CR: $z \geqslant 1.6449$)

$1.60 < 1.6449$ so not in the critical region, not significant, accept H_0

There is insufficient evidence to support Baako's claim that variety B gives a higher mean yield per plant than variety A.

(b) H_0: $\mu_f - \mu_w = 100$ H_1: $\mu_f - \mu_w > 100$

(μ_f, μ_w = mean yield per plant with and without fertiliser)

$z = \dfrac{(410 - 290) - 100}{\sqrt{90^2/300 + 60^2/200}} = \dfrac{20}{\sqrt{27 + 18}} = \dfrac{20}{\sqrt{45}} = \dfrac{20}{6.7082...}$

$z = 2.9814... = 2.98$

One-tailed 0.5%: critical value $z = 2.5758$ (CR: $z \geqslant 2.5758$)

$2.98 > 2.5758$ so in the critical region, significant, reject H_0

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

stemup.co.uk

Leave blank

Question 6 continued

There is sufficient evidence to support Ayodele's claim that using fertiliser increases the mean yield per plant by more than 100 g.

(c)(i) The plants that receive fertiliser are allocated (chosen) at random from the 500 plants.

(ii) The samples are large enough for the CLT to apply, so the sample standard deviations are good (accurate) estimates of the population standard deviations, i.e. $s \approx \sigma$ for each group.

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Leave blank

Question 6 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

Leave blank

Question 6 continued

(Total 14 marks)

Q6

Leave blank

7. Jeff records the number of births announced in his local weekly newspaper each week for n consecutive weeks, where n is an integer. He decides the resulting data can be modelled as independent samples from a Poisson distribution with mean 2.8 and calculates the expected frequencies using this model.

The following table shows the observed frequencies and the expected frequencies, to 2 decimal places.

The observed and expected frequencies for 5 births are not given.

Number of births	Observed frequency (O)	Expected frequency (E)
0	8	5.84
1	27	16.35
2	25	22.88
3	16	21.36
4	12	14.95
5		
6	2	3.91
$\geqslant 7$	1	2.34

(a) Work out the observed and expected frequencies for 5 births.

(2)

The value of $\sum \frac{(O-E)^2}{E}$ for the given values for the number of births 0, 1, 2, 3 and 4 is 9.86

(b) Using a 5% significance level, test whether or not this Poisson model is suitable. State your hypotheses clearly.

(7)

(a) $n \times e^{-2.8} = 5.84 \Rightarrow n = \frac{5.84}{e^{-2.8}} = 96.03\ldots$ so $n = 96$

Observed (5) $= 96 - (8 + 27 + 25 + 16 + 12 + 2 + 1) = 96 - 91 = 5$

Expected (5) $= 96 \times \frac{e^{-2.8} \times 2.8^5}{5!} = 8.3725\ldots = 8.37$

(b) H_0: Po(2.8) is a suitable model for the number of births per week

H_1: Po(2.8) is not a suitable model for the number of births per week

E for ≥ 7 is $2.34 < 5$, so combine 6 and ≥ 7: $O = 2 + 1 = 3$,

$E = 3.91 + 2.34 = 6.25$

$X^2 = 9.86 + \frac{(5 - 8.37)^2}{8.37} + \frac{(3 - 6.25)^2}{6.25}$

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

stemup.co.uk

Leave blank

Question 7 continued

$= 9.86 + 1.3584... + 1.69 = 12.908... = 12.9$

7 classes (0, 1, 2, 3, 4, 5, ≥ 6), no parameters estimated:

Degrees of freedom $\nu = 7 - 1 = 6$

Critical value $\chi^2_6(0.05) = 12.592$ (CR: $X^2 \geq 12.592$)

$12.9 > 12.592$ so in the critical region, significant, reject H_0

The data are not consistent with a Poisson(2.8) model: the Poisson model with mean 2.8 is not suitable for the number of births per week.

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Leave blank

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

Leave blank

Question 7 continued

Q7

(Total 9 marks)

stemup.co.uk

Leave blank

8. The weight of coffee in large jars, L grams, is such that $L \sim N(300, 7^2)$
The weight of coffee in medium jars, M grams, is such that $M \sim N(200, 6^2)$

A large jar and a medium jar are selected at random.

(a) Calculate the probability that the weight of coffee in the large jar is greater than 90 g more than the weight of coffee in the medium jar.

(5)

The weight of coffee in small jars, S grams, is such that $S \sim N(70, \sigma^2)$

There is a probability of 0.975 that the weight of coffee in a randomly selected large jar is more than four times the weight of coffee in a randomly selected small jar.

(b) Calculate the value of σ giving your answer to 3 significant figures.

(8)

(a) Let $X = L - M$

$E(X) = 300 - 200 = 100$

$\text{Var}(X) = 7^2 + 6^2 = 49 + 36 = 85$, so $X \sim N(100, 85)$

$$P(X > 90) = P\left(Z > \frac{90 - 100}{\sqrt{85}}\right) = P(Z > -1.0846\ldots)$$

$= 0.86096\ldots = 0.861$

(b) Let $Y = L - 4S$

$E(Y) = 300 - 4 \times 70 = 20$

$\text{Var}(Y) = 7^2 + 4^2\sigma^2 = 49 + 16\sigma^2$, so $Y \sim N(20, 49 + 16\sigma^2)$

$P(L > 4S) = P(Y > 0) = 0.975$

$$P\left(Z > \frac{0 - 20}{\sqrt{49 + 16\sigma^2}}\right) = 0.975$$

$$\frac{0 - 20}{\sqrt{49 + 16\sigma^2}} = -1.96$$

$20^2 = 1.96^2(49 + 16\sigma^2)$

$$49 + 16\sigma^2 = \frac{400}{3.8416} = 104.123\ldots$$

$$\sigma^2 = \frac{104.123\ldots - 49}{16} = 3.4452\ldots$$

$\sigma = 1.8561\ldots = 1.86$ (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

Leave blank

Question 8 continued

Leave blank

Question 8 continued

Q8

(Total 13 marks)

TOTAL FOR PAPER: 75 MARKS

END

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

stemup.co.uk