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 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 ▶

P66654A

©2021 Pearson Education Ltd.

1/1/

P 6 6 6 5 4 A 0 1 2 0

Pearson

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)

(a) 29, 34, 39, 06, 31, 13, 42

(b) Girls: $\frac{24}{60} \times 10 = 4$, Boys: $\frac{36}{60} \times 10 = 6$

Girls are numbered 01 to 24, so the numbers 15, 23, 06, 13 select the 4 girls numbered 15, 23, 6 and 13

Boys are numbered 25 to 60, so the numbers 33, 29, 34, 39, 31, 42 select the 6 boys numbered 33, 29, 34, 39, 31 and 42

This gives 4 girls and 6 boys, in proportion to 24 : 36

(c) The highest random number is 42, so no player numbered 43 to 60 can be chosen. The list is in ascending order of age, so the oldest players (numbers 43 to 60) may not be represented in the sample.

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 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)

(a) Rank 1 = highest in each row

Student	A	B	C	D	E	F	G	H	I	J	K
Objects rank	9	6	8	2	1	10	7	3	5	4	11
Maths rank	11	4	5	1	2	9	3	7	8	6	10
d	−2	2	3	1	−1	1	4	−4	−3	−2	1
d^2	4	4	9	1	1	1	16	16	9	4	1

$\Sigma d^2 = 4 + 4 + 9 + 1 + 1 + 1 + 16 + 16 + 9 + 4 + 1 = 66$

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 1 - \frac{6 \times 66}{11(11^2-1)} = 1 - \frac{396}{1320}$$

$r_s = 0.7$

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

Critical value ($n = 11$, 5% one-tail) = 0.5364

$0.7 > 0.5364$, so significant: reject H_0

There is evidence of a positive correlation between short-term memory and mathematical ability, which supports the teacher's belief.

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 2 continued

(c) I disagree. The data show positive correlation, but correlation does not imply causation: better short-term memory does not necessarily cause an increase in mathematical ability.

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)

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

(a)(i) Expected frequency = $\frac{88}{4}$ = 22 for each club

Music 22, Art 22, Sports 22, Computers 22

(ii) H_0: the students are uniformly distributed across the 4 clubs

H_1: the students are not uniformly distributed across the clubs

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

Critical value $\chi^2_3(5\%) = 7.815$

$6.09 < 7.815$, so not significant: do not reject H_0

Insufficient evidence to suggest the students are not uniformly distributed, so a uniform distribution across the clubs is suitable.

(b) H_0: the club chosen is independent of the student's school

H_1: the club chosen depends on (is associated with) the school

(c) School C total = 28, Computers total = 17, grand total = 88

$E = \frac{28 \times 17}{88} = 5.409... = 5.41$

(d) Music total = 15

Expected frequency for School C and Music = $\frac{28 \times 15}{88} = 4.77 < 5$

So the Music column must be combined with another column (e.g. Music with Computers), giving a 3×3 table.

$\nu = (3 - 1)(3 - 1) = 2 \times 2 = 4$ degrees of freedom

(e) Critical value $\chi^2_4(5\%) = 9.488$

$7.29 < 9.488$, so not significant: do not reject H_0

Insufficient evidence of an association between a student's school and the club they choose.

Leave blank

Question 3 continued

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

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

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)

(a) 98% CI: $\bar{x} \pm z \times \frac{\sigma}{\sqrt{n}}$, $z = 2.3263$

$52.4 \pm 2.3263 \times \frac{18}{\sqrt{25}} = 52.4 \pm 8.3747...$

$= (44.0253..., 60.7746...)$

98% CI = (44.0, 60.8)

(b) H_0: $\mu_A = \mu_B$ H_1: $\mu_B > \mu_A$

$z = \frac{57.8 - 52.4}{\sqrt{18^2/25 + 18^2/30}} = \frac{5.4}{\sqrt{23.76}} = \frac{5.4}{4.8744...} = 1.1078...$

$z = 1.11$

5% one-tail critical value: $z = 1.6449$

$1.11 < 1.6449$, so not significant: do not reject H_0

Insufficient evidence to support the newspaper's claim that the mean score in town B is greater than the mean score in town A.

(c) Not significant requires $\frac{\bar{x} - \mu}{18/\sqrt{n}} > -1.6449$

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 4 continued

Town A: $\frac{52.4 - \mu}{18/\sqrt{25}} > -1.6449 \Rightarrow \mu < 52.4 + 1.6449 \times \frac{18}{5} = 58.3216...$

Town B: $\frac{57.8 - \mu}{18/\sqrt{30}} > -1.6449 \Rightarrow \mu < 57.8 + 1.6449 \times \frac{18}{\sqrt{30}} = 63.2057...$

Both conditions hold when $\mu < 58.3216...$

Largest possible value of $\mu = 58.3$

Leave blank

Question 4 continued

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

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

Leave blank

Question 4 continued

(Total 14 marks)

Q4

Leave blank

5. Chrystal is studying the lengths of pine cones that have fallen from a tree. She believes that the length, Xcm, 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)

(a) H_0: $N(6, 0.75^2)$ is a suitable model for the lengths of the pine cones

H_1: $N(6, 0.75^2)$ is not a suitable model for the lengths of the pine cones

$P(X < 5) = P\left(Z < \frac{5-6}{0.75}\right) = P(Z < -1.333) = 0.09121$

$E = 80 \times 0.09121 = 7.30$

$P(5 \leqslant X < 5.5) = P\left(-\frac{4}{3} \leqslant Z < -\frac{2}{3}\right) = 0.25249 - 0.09121 = 0.16128, \quad E = 12.90$

$P(5.5 \leqslant X < 6) = P\left(-\frac{2}{3} \leqslant Z < 0\right) = 0.24751, \quad E = 80 \times 0.24751 = 19.80$

By symmetry, $E(6 \leqslant X < 6.5) = 19.80$

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

Question 5 continued

$E(X \geqslant 6.5) = 80 - 7.30 - 12.90 - 2 \times 19.80 = 20.20$

Length	$x < 5$	$5 - 5.5$	$5.5 - 6$	$6 - 6.5$	$x \geqslant 6.5$
O_i	6	14	24	26	10
E_i	7.30	12.90	19.80	19.80	20.20
$(O - E)^2/E$	0.2305	0.0934	0.8906	1.9410	5.1500

$\chi^2 = \Sigma\dfrac{(O_i - E_i)^2}{E_i} = 0.2305 + 0.0934 + 0.8906 + 1.9410 + 5.1500 = 8.3055\ldots$

$\chi^2 = 8.31$

$\nu = 5 - 1 = 4 \Rightarrow$ critical value $\chi^2_4(10\%) = 7.779$

$8.31 > 7.779$, so significant: reject H_0

$N(6, 0.75^2)$ is not a suitable model for the lengths of the pine cones, so the data do not support Chrystal's belief.

(b) $\hat{\mu} = \bar{x} = \dfrac{464}{80} = 5.8$ cm

$s^2 = \dfrac{\Sigma x^2 - n\bar{x}^2}{n-1} = \dfrac{2722.59 - 80 \times 5.8^2}{79} = \dfrac{31.39}{79} = 0.39734\ldots = 0.397\ \text{cm}^2$

(c) $\nu = 5 - 1 - 2 = 2$ (μ and σ^2 both estimated)

Critical value $\chi^2_2(10\%) = 4.605$

$3.50 < 4.605$, so not significant: do not reject H_0

A normal distribution is a suitable model for the lengths of the pine cones.

(d) $X \sim N(5.8, 0.397)$: $P(X > 7) = P\left(Z > \dfrac{7 - 5.8}{\sqrt{0.39734\ldots}}\right) = P(Z > 1.9037\ldots)$

$= 1 - 0.97153 = 0.02847\ldots \approx 0.028$

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

Leave blank

Question 5 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

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

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)

(a) Let $D = Y - R$: $E(D) = 12 - 15 = -3$

$\text{Var}(D) = 0.8^2 + 1.5^2 = 0.64 + 2.25 = 2.89$, $D \sim N(-3, 1.7^2)$

$P(Y > R) = P(D > 0) = P\left(Z > \frac{0 - (-3)}{1.7}\right) = P(Z > 1.7647\ldots)$

$= 0.03880\ldots = 0.0388$

(b) $R_1 + R_2 + R_3 \sim N(45, 3 \times 1.5^2) = N(45, 6.75)$

$4Y \sim N(4 \times 12, 4^2 \times 0.8^2) = N(48, 10.24)$

Let $L = 4Y - (R_1 + R_2 + R_3)$: $E(L) = 48 - 45 = 3$

$\text{Var}(L) = 10.24 + 6.75 = 16.99$, $L \sim N(3, 16.99)$

$P(R_1 + R_2 + R_3 < 4Y) = P(L > 0) = P\left(Z > \frac{0 - 3}{\sqrt{16.99}}\right) = P(Z > -0.7278\ldots)$

$= 0.76663\ldots = 0.767$

(c) $E(X) = 15a + 12b = 780 \Rightarrow a = 52 - 0.8b$

$\text{Var}(X) = a^2 \times 1.5^2 + b^2 \times 0.8^2 = 2.25a^2 + 0.64b^2$

$= 2.25(52 - 0.8b)^2 + 0.64b^2$

$= 2.25(2704 - 83.2b + 0.64b^2) + 0.64b^2 = 2.08b^2 - 187.2b + 6084$

$\frac{d}{db}[\text{Var}(X)] = 4.16b - 187.2 = 0 \Rightarrow b = \frac{187.2}{4.16} = 45$

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 6 continued

$\frac{d^2}{db^2}[Var(X)] = 4.16 > 0$, so this gives the minimum variance

$a = 52 - 0.8 \times 45 = 52 - 36 = 16$

$a = 16, \quad b = 45$

Leave blank

Question 6 continued

Q6

(Total 18 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