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

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 ▶

P63153A

©2021 Pearson Education Ltd.

1/1/1/1/1/1/1/1/1/

P 6 3 1 5 3 A 0 1 2 0

Pearson

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)

(a) Rank 1 = highest value in each row

Area	A	B	C	D	E
Coverage rank	3	2	1	5	4
Moisture rank	2	4	1	5	3
d	1	−2	0	0	1
d^2	1	4	0	0	1

$\Sigma d^2 = 1 + 4 + 0 + 0 + 1 = 6$

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 1 - \frac{6 \times 6}{5(5^2-1)} = 1 - \frac{36}{120}$$

$r_s = 0.7$

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

Critical value ($n = 5$, 5% one-tail): $r_s = 0.9$, critical region $r_s \geq 0.9$

$0.7 < 0.9$, so $r_s = 0.7$ does not lie in the critical region:

not significant, do not reject H_0

There is insufficient evidence of a positive correlation between moisture content and plant coverage, so the data do not support the biologist's claim.

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 9 marks)

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

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)

H_0: diet and health are independent (not associated)

H_1: diet and health are not independent (associated)

Expected frequency $E = \frac{\text{row total} \times \text{column total}}{\text{grand total}}$, e.g. $\frac{94 \times 177}{200} = 83.19$

Expected	Good health	Poor health	Total
Good diet	83.19	10.81	94
Poor diet	93.81	12.19	106
Total	177	23	200

O	86	8	91	15
E	83.19	10.81	93.81	12.19
$(O-E)^2/E$	0.0949	0.7304	0.0842	0.6478

$\chi^2 = \Sigma\frac{(O-E)^2}{E} = 0.0949 + 0.7304 + 0.0842 + 0.6478 = 1.557...$

$\chi^2 = 1.56$

$\nu = (2-1)(2-1) = 1$

Critical value $\chi^2_1(5\%) = 3.841$

$1.557 < 3.841$, so χ^2 does not lie in the critical region:

insufficient evidence to reject H_0

Diet and health are independent: there is no association between diet and health, so the doctor's belief is not supported by these data.

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 9 marks)

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

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)

(a) $\bar{x} = \frac{11.52 + 13.75}{2} = 12.635$

95%: $z = 1.96$, $\frac{\sigma}{\sqrt{n}} = \frac{13.75 - 12.635}{1.96} = \frac{1.115}{1.96} = 0.56887...$

90%: $z = 1.6449$

90% CI $= 12.635 \pm 1.6449 \times 0.56887... = 12.635 \pm 0.93573...$

$= (11.699..., 13.5707...)$

90% CI $= (11.7, 13.6)$

(b) P(an interval contains μ) $= 0.9$

X = number of the 4 intervals containing μ, $X \sim B(4, 0.9)$

$P(X = 3) = 4 \times 0.9^3 \times 0.1 = 0.2916$

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 8 marks)

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

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)

(a) Number (label) the academic students 1 to 1680 and the vocational students 1 to 2520

Academic: $\frac{1680}{4200} \times 70 = 28$, Vocational: $\frac{2520}{4200} \times 70 = 42$

Use random numbers (a random number generator) to select 28 academic students and 42 vocational students (simple random sample from each group).

(b) $H_0: \mu_v - \mu_a = 0$ $H_1: \mu_v - \mu_a > 0$

(μ_v = mean score of vocational students, μ_a = mean score of academic students)

$se = \sqrt{\left(\frac{70}{80} + \frac{60}{50}\right)} = \sqrt{2.075} = 1.44048...$

$z = \frac{62 - 57}{\sqrt{70/80 + 60/50}} = \frac{5}{1.44048...} = 3.4710...$

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

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

Question 4 continued

$z = 3.47$

Critical value (5% one-tail): $z = 1.6449$

$3.47 > 1.6449$, so significant: reject H_0

There is evidence that the mean basic skills score for vocational students is greater than the mean basic skills score for academic students.

(c) The sample sizes are large, so by the Central Limit Theorem the sample mean scores $\bar{X}_a$ for academic students and $\bar{X}_v$ for vocational students are each approximately normally distributed, so a z test can be used even though the distribution of scores is unknown.

(d) Both samples are large enough that the sample variances are good estimates of the population variances, i.e. $s^2 \approx \sigma^2$ for both groups.

(e) $1.54 < 1.6449$, so the test is no longer significant: insufficient evidence to reject H_0

There is now insufficient evidence that the mean basic skills score for vocational students is greater than the mean score for academic students: the academic students have improved their mean basic skills score.

(f) The course was effective: the significant difference in (b) is no longer significant in (e), as the academic students' mean score has improved.

Question 4 continued

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 16 marks)

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

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)

(a) Relief of symptoms is either a success or a failure for each person.
The probability that the medicine is a success is constant for each person, and the samples (people) from the different practices are independent.

(b) $\text{Mean} = \dfrac{0\times4 + 1\times6 + 2\times3 + 3\times12 + 4\times10 + 5\times7 + 6\times4 + 7\times2 + 8\times2}{50}$

$= \dfrac{0 + 6 + 6 + 36 + 40 + 35 + 24 + 14 + 16}{50} = \dfrac{177}{50} = 3.54$

(c) $np = 3.54 \Rightarrow p = \dfrac{3.54}{8} = 0.4425, \quad X \sim B(8, 0.4425)$

$f = 50 \times \binom{8}{4} \times 0.4425^4 \times 0.5575^4 = 50 \times 70 \times 0.0038339... \times 0.096597...$

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

$f = 12.9628... = 12.96$

$g = 50 \times 0.4425^8 = 0.07349... = 0.07$

(d) H_0: a binomial distribution is a suitable model

H_1: a binomial distribution is not a suitable model

Combine classes so that every $E \geqslant 5$: {0, 1, 2} and {5, 6, 7, 8}

Successes	0, 1, 2	3	4	5, 6, 7, 8
O	13	12	10	15
E	11.66	13.07	12.96	12.31
$(O - E)^2/E$	0.1540	0.0876	0.6760	0.5878

$\chi^2 = \Sigma\frac{(O - E)^2}{E} = 0.1540 + 0.0876 + 0.6760 + 0.5878 = 1.50546...$

$\chi^2 = 1.51$

$\nu = 4 - 2 = 2$ (4 classes, p estimated)

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

$1.51 < 4.605$, so not significant: insufficient evidence to reject H_0

The data are consistent with a binomial distribution: the binomial distribution is a suitable model for the number of successes.

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

Question 5 continued

(Total for Question 5 is 16 marks)

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

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)

(a) Let $W = B - 1.1R$

$E(W) = 55 - 1.1 \times 51 = 55 - 56.1 = -1.1$

$Var(W) = 1.3^2 + 1.1^2 \times 1.2^2 = 1.69 + 1.7424 = 3.4324$

$W \sim N(-1.1, 3.4324)$

$$P(B < 1.1R) = P(W < 0) = P\left(Z < \frac{0 - (-1.1)}{\sqrt{3.4324}}\right) = P(Z < 0.5937\ldots)$$

$= 0.7224$

(b) Let $X = B_1 - B_2$

$E(X) = 55 - 55 = 0, \quad Var(X) = 2 \times 1.3^2 = 3.38, \quad X \sim N(0, 3.38)$

$$P(|X| > 1) = 2 \times P(X > 1) = 2 \times P\left(Z > \frac{1 - 0}{\sqrt{3.38}}\right) = 2 \times P(Z > 0.5439\ldots)$$

$= 2 \times (1 - 0.7054) = 2 \times 0.2946$

$= 0.5892$

(c) $V = B_1 + B_2 + \ldots + B_{10} + S$ (bag of buns)

$Y = R_1 + R_2 + \ldots + R_{11} + S$ (bag of rolls)

$E(V) = 10 \times 55 + 3 = 553, \quad Var(V) = 10 \times 1.3^2 + 0.2^2 = 16.94$

$E(Y) = 11 \times 51 + 3 = 564, \quad Var(Y) = 11 \times 1.2^2 + 0.2^2 = 15.88$

$V \sim N(553, 16.94)$ and $Y \sim N(564, 15.88)$

Let $D = Y - V$: $E(D) = 564 - 553 = 11$

$Var(D) = 16.94 + 15.88 = 32.82, \quad D \sim N(11, 32.82)$

$$P(V < Y) = P(D > 0) = P\left(Z > \frac{0 - 11}{\sqrt{32.82}}\right) = P(Z > -1.920\ldots)$$

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

Question 6 continued

$= P(Z < 1.92) = 0.9726$

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

Question 6 continued

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

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

Question 6 continued

Question 6 continued

(Total for Question 6 is 17 marks)

TOTAL FOR PAPER: 75 MARKS

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA