Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel
International Advanced Level

Centre Number	Candidate Number

Sample Assessment Materials for first teaching September 2018

(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, 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

S59770A

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

S 5 9 7 7 0 A 0 1 2 8

Pearson

S3 specimen.

Leave blank

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

1. The names of the 720 members of a swimming club are listed alphabetically in the club's membership book. The chairman of the swimming club wishes to select a systematic sample of 40 names. The names are numbered from 001 to 720 and a number between 001 and w is selected at random. The corresponding name and every xth name thereafter are included in the sample.

(a) Find the value of w.

(1)

(b) Find the value of x.

(1)

(c) Write down the probability that the sample includes both the first name and the second name in the club's membership book.

(1)

(d) State one advantage and one disadvantage of systematic sampling in this case.

(2)

(a) $\frac{720}{40} = \underline{\underline{18}}$

(b) $= \underline{18}$

(C) 0 → first name is chosen randomly from first 18 members.

(d) <u>ADV</u>: – simple and easy to use

<u>Disadv</u> : Alphabetical list is not random so bias in introduced.

Leave blank

2. Nine dancers, Adilzhan (A), Bianca (B), Chantelle (C), Lee (L), Nikki (N), Ranjit (R), Sergei (S), Thuy (T) and Yana (Y), perform in a dancing competition.

Two judges rank each dancer according to how well they perform. The table below shows the rankings of each judge starting from the dancer with the strongest performance.

Rank	1	2	3	4	5	6	7	8	9
Judge 1	S	N	B	C	T	A	Y	R	L
Judge 2	S	T	N	B	C	Y	L	A	R

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

(5)

(b) Stating your hypotheses clearly, test at the 1% level of significance, whether or not the two judges are generally in agreement.

(4)

(a)

Dancer	Rank (J1)	Rank (J2)	d	d^2
A	6	8	2	4
B	3	4	1	1
C	4	5	1	1
L	9	7	2	4
N	2	3	1	1
R	8	9	1	1
S	1	1	0	0
T	5	2	3	9
Y	7	6	1	1
				22

$\Sigma d^2 = 22$.

$\therefore r_s = 1 - \frac{6(22)}{9(80)} = \underline{0.817}$

(b) $H_0: \rho = 0$ critical value: ± 0.7833

$H_1: \rho > 0$

$0.817 > 0.7833$

$\therefore$ Result is significant $\therefore$ Reject H_0. Evidence suggests judges are generally in agreement.

Leave blank

3. The number of accidents on a particular stretch of motorway was recorded each day for 200 consecutive days. The results are summarised in the following table.

Number of accidents	0	1	2	3	4	5
Frequency	47	57	46	35	9	6

(a) Show that the mean number of accidents per day for these data is 1.6
(1)

A motorway supervisor believes that the number of accidents per day on this stretch of motorway can be modelled by a Poisson distribution.

She uses the mean found in part (a) to calculate the expected frequencies for this model. Her results are given in the following table.

Number of accidents	0	1	2	3	4	5 or more
Frequency	40.38	64.61	r	27.57	11.03	s

(b) Find the value of r and the value of s, giving your answers to 2 decimal places.
(3)

(c) Stating your hypotheses clearly, use a 10% level of significance to test the motorway supervisor's belief. Show your working clearly.
(7)

(a) mean $= \dfrac{\text{total accidents}}{\text{total days}}$

$= \dfrac{0(47)+1(57)+2(46)+3(35)+4(9)+5(1)}{200}$

$= 1.6$

(b) $r = 200 \times P(X=2) = 200\left[\dfrac{e^{-1.6}(1.6)^2}{2!}\right]$

$= 51.69$

$s = 200 - (\Sigma E_i)$

$= 200 - 40.38 - 64.61 - 51.69 - 27.57 - 11.03$

$= 4.72$

(c) H_0: Poisson ($\lambda = 1.6$) is ~~not~~ a suitable model for these data.

H_1: Poisson ($\lambda = 1.6$) is not a suitable model for these data.

Remember, expected frequencies must be greater than 5 for the test statistic χ^2 to be approximated well by the chi-squared distribution(s).

So pool find two groups;

Question 3 continued

No accidents	0	1	2	3	≥ 4
E	40.38	64.61	51.69	27.57	15.75
O	47	57	46	35	15
$\frac{(O-E)^2}{E}$	1.0853	0.8963	0.6264	2.0024	0.0357

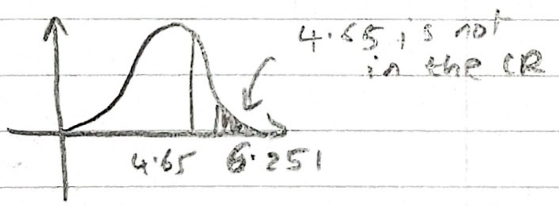
$$X^2 = \sum \frac{(O-E)^2}{E} = \underline{\underline{4.65}}$$

$\nu = 5-1-1 \underline{=3}$

subtract an addition 1 as the parameter λ was calculated.

$\therefore$ Critical value $= \chi^2_3 \,(10\%) = \underline{6.251}$

$4.65 < 6.215$



$\therefore$ Result is insignificant $\therefore$ Accept H_0. Evidence suggests that poisson is a suitable model - supervisors belief is correct.

Leave blank

4. A farm produces potatoes. The potatoes are packed into sacks.
The weight of a sack of potatoes is modelled by a normal distribution with mean 25.6 kg and standard deviation 0.24 kg

(a) Find the probability that two randomly chosen sacks of potatoes differ in weight by more than 0.5 kg

(6)

Sacks of potatoes are randomly selected and packed onto pallets.

The weight of an empty pallet is modelled by a normal distribution with mean 20.0 kg and standard deviation 0.32 kg

Each full pallet of potatoes holds 30 sacks of potatoes.

(b) Find the probability that the total weight of a randomly chosen full pallet of potatoes is greater than 785 kg

(5)

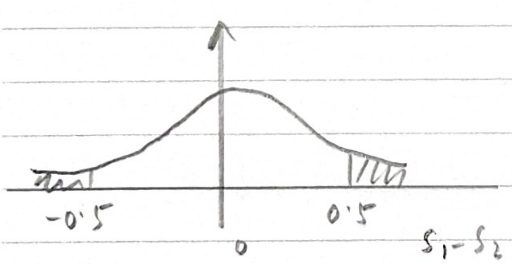
(a) let S = weight of a sack of potatoes

$S \sim N(25.6, 0.24^2)$

P(required) $= P(|S_1 - S_2| > 0.5)$

$(S_1 - S_2) \sim N(0, 2(0.24^2))$

$P(|S_1 - S_2| > 0.5) = 2P(S_1 - S_2 > 0.5)$ due to symetry.



$= 2\left[P\left(Z > \frac{0.5 - 0}{\sqrt{2(0.24^2)}}\right)\right]$

$= 2P(Z > 1.47)$

$= 2[1 - P(Z < 1.47)]$

$= 2(1 - 0.9292)$

$= \underline{0.1416}$

(b) let P = pallet (empty), $P \sim N(20, 0.32^2)$

let Full pallet $= (S_1 + \dots S_{30} + P) = F$.

$E(F) = 30E(S) + E(P) = 30(25.6) + 20$

$= \underline{788}$

$Var(F) = 30 Var(S) + Var(P) = 30(0.24^2) + 0.32^2$

$= \underline{1.8304}$

So $F \sim N(788, 1.8304)$

p(required) $= P(F > 785)$

$P\left(Z > \frac{785 - 788}{\sqrt{1.8304}}\right)$

Leave blank

Question 4 continued

$= P(Z > -2.22)$

$= P(Z < 2.22)$

$= \underline{0.9868}$

Leave blank

5. A Head of Department at a large university believes that gender is independent of the grade obtained by students on a Business Foundation course. A random sample was taken of 200 male students and 160 female students who had studied the course.

The results are summarised below.

		Male	Female
Grade	Distinction	18.5%	27.5%
	Merit	63.5%	60.0%
	Unsatisfactory	18.0%	12.5%

Stating your hypotheses clearly, test the Head of Department's belief using a 5% level of significance. Show your working clearly.

(12)

H_0: There is no association between grade and gender.

H_1: There is an association between grade and gender

Working out actual and observed values:

Male/Distinction = 200 × 0.185 = 37
Male/Merit = 200 × 0.635 = 127
Male/Unsatisfactory = 200 × 0.18 = 36.

Female/Distinction = 160 × 0.275 = 44
Female/merit = 160 × 0.600 = 96
Female/Unsatisfactory = 160 × 0.125 = 20

Observed:

	Male	Female
Distinction	37	44
Merit	127	96
Unsatisfactory	36	20

Expected no. = $\frac{\text{Row total} \times \text{column total}}{\text{Grand total}}$

Expected:

	Male	Female	
Distinction	45	36	81
Merit	123.89	99.11	223
Unsatisfactory	31.11	24.89	56
	200	160	$\boxed{360}$

O_i	E_i	$\frac{(O-E)^2}{E}$
37	45	1.42
44	36	1.78
127	123.89	0.08
96	99.11	0.1
36	31.11	0.77
20	24.89	0.96
		5.11

$\chi^2 = \Sigma \frac{(O-E)^2}{E} = 5.1$

$V = (\text{rows}-1)(\text{columns}-1) = (3-1)(2-1) = 2$

$\therefore$ Critical value = $\chi^2(5\%) = 5.991$

$5.1 < 5.991$

$\therefore$ Result is insignificant $\therefore$ accept H_0. Evidence suggests that Head's Belief is correct.

Leave blank

6. As part of an investigation, a random sample was taken of 50 footballers who had completed an obstacle course in the early morning. The time taken by each of these footballers to complete the obstacle course, x minutes, was recorded and the results are summarised by

$$\sum x = 1570 \quad \text{and} \quad \sum x^2 = 49\,467.58$$

(a) Find unbiased estimates for the mean and variance of the time taken by footballers to complete the obstacle course in the early morning.
(4)

An independent random sample was taken of 50 footballers who had completed the same obstacle course in the late afternoon. The time taken by each of these footballers to complete the obstacle course, y minutes, was recorded and the results are summarised as

$$\bar{y} = 30.9 \quad \text{and} \quad s_y^2 = 3.03$$

(b) Test, at the 5% level of significance, whether or not the mean time taken by footballers to complete the obstacle course in the early morning, is greater than the mean time taken by footballers to complete the obstacle course in the late afternoon.
State your hypotheses clearly.
(7)

(c) Explain the relevance of the Central Limit Theorem to the test in part (b).
(1)

(d) State an assumption you have made in carrying out the test in part (b).
(1)

(a) mean $= \frac{\Sigma x}{n} = \frac{1570}{50} = \underline{31.4}$

$(\hat{\sigma}^2) = s^2 = \frac{1}{n-1}\left(\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right)$

$= \frac{1}{49}\left(49467.58 - \frac{(1570)^2}{50}\right)$

$= \underline{3.46}$

(b) $H_0: \mu_x = \mu_y$. where x refers to morning samples and y refers to late afternoon.

$H_1: \mu_x > \mu_y$

critical value: ± 1.6449 (5% - 1 tail)

Test statistic $= \dfrac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{n_y}}}$

$= \dfrac{31.4 - 30.9}{\sqrt{\frac{3.46}{50} + \frac{3.03}{50}}}$ s^2 from (a)

$= \underline{1.39}$

$1.39 < 1.6449$

$\therefore$ Result is insignificant. Accept H_0. $\therefore$ no difference in mean times to complete obstacle course in morning / late afternoon

Leave blank

Question 6 continued

(c) Allows us to assume $\bar{x}$ and $\bar{Y}$ (sample means) are normally distributed as n is large.

(d) Sample variance = pop. variance

$s^2 = \sigma^2$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Leave blank

7. A fair six-sided die is labelled with the numbers 1, 2, 3, 4, 5 and 6
The die is rolled 40 times and the score, S, for each roll is recorded.

(a) Find the mean and the variance of S.

(2)

(b) Find an approximation for the probability that the mean of the 40 scores is less than 3

(3)

(a) Discrete uniform Distribution.

s	1	2	3	4	5	6.
$P(S=s)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$.

$E(S) = \frac{a+b}{2} = \frac{1+6}{2} = \underline{3.5}$

$Var(S) = \frac{n^2-1}{12} = \frac{6^2-1}{12} = \frac{35}{12}$

(b) $P(\bar{S} < 3) = p(\text{required})$

$\bar{S} \sim N(3.5, 35/12(40))$ By C.L.T.

$\Rightarrow P(\bar{S} < 3) = P\left(Z < \frac{3-3.5}{\sqrt{\frac{35}{12(40)}}}\right)$

$= p(Z < -1.85) = 1 - p(Z < 1.85)$

$= \underline{\underline{0.0322}}$

Leave blank

8. A factory produces steel sheets whose weights Xkg, are such that $X \sim N(\mu, \sigma^2)$

A random sample of these sheets is taken and a 95% confidence interval for μ is found to be (29.74, 31.86)

(a) Find, to 2 decimal places, the standard error of the mean.
(3)

(b) Hence, or otherwise, find a 90% confidence interval for μ based on the same sample of sheets.
(3)

Using four different random samples, four 90% confidence intervals for μ are to be found.

(c) Calculate the probability that at least 3 of these intervals will contain μ.
(3)

(a) $\bar{x} = \frac{29.74 + 31.86}{2} = 30.8$

Standard error $= \frac{\sigma}{\sqrt{n}}$

$\bar{x} + (1.96) \cdot \frac{\sigma}{\sqrt{n}} = 31.86.$

$1.96 \cdot \frac{\sigma}{\sqrt{n}} = 31.86 - \bar{x}$

$\therefore \frac{\sigma}{\sqrt{n}} = \frac{31.86 - 30.8}{1.96} = \underline{0.54}$

(b) 90%

$\underline{CI}: \left[\bar{x} \pm 1.6449 \frac{\sigma}{\sqrt{n}}\right]$

$= \left[30.8 \pm 1.6449(0.54)\right]$

$= [29.91, 31.69]$

(c) $X \sim B[4, 0.9]$ where x is the no. of confidence intervals containing μ

$p(X \geq 3) = p(X = 3) + p(X = 4)$

$= \binom{4}{3}(0.9)^3(0.1) + 0.9^4 = \underline{\underline{0.948}}$