Pearson Edexcel International Advanced Level

Friday 12 June 2026

Afternoon (Time: 1 hour 30 minutes)

Paper reference **WST03/01A**

Mathematics

International Advanced Subsidiary/Advanced Level
Statistics S3
Question paper

You must have:
Answer book (sent separately)
Do not return this question paper with the answer book

Turn over

P81331A

P:1/1/1/

Pearson

1. In a singing competition, 8 singers are judged separately by a jury and by an audience. The jury gives each singer a score out of 30 and the audience ranks the singers from most favourite (1) to least favourite (8). The scores given by the jury and the rankings given by the audience are shown in the table.

Singer	A	B	C	D	E	F	G	H
Jury score	28	25	29	24	16	19	27	23
Audience rank	2	1	3	4	7	8	5	6

(a) Calculate Spearman's rank correlation coefficient between the jury's ranking and the audience's ranking.
Use a rank of 1 for the singer with the highest score.
Show your working clearly.

(4)

(b) Test, at the 5% level of significance, whether or not there is evidence of a positive correlation between the rankings of the jury and the rankings of the audience.
State your hypotheses and critical value clearly.

(4)

It was later discovered that the score for singer F was recorded incorrectly and is corrected to 16

(c) State what effect this has on the rankings for the scores given by the jury.

(1)

(Total for Question 1 is 9 marks)

2. A marine biologist investigates the length of salmon in two regions, A and B.

A random sample of 60 salmon from region A gives the following summary statistics, where x cm is the length of each salmon.

$$\sum x = 4530 \qquad \sum x^2 = 350761$$

(a) Find unbiased estimates for the mean **and** variance of the length of salmon from region A.

(3)

A random sample of 80 salmon from region B gives the following results, where y cm is the length of each salmon.

$$\bar{y} = 82.3 \qquad s_y = 17.4$$

The biologist claims that the mean length of salmon in region B is more than 2 cm greater than the mean length of salmon in region A.

(b) Use a suitable test, at the 5% level of significance, to assess the biologist's claim. You should state your hypotheses, test statistic and critical value clearly.

(7)

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

(1)

(Total for Question 2 is 11 marks)

3. A random sample of 3 observations, X_1, X_2 and X_3 are taken from a population with unknown mean $\mu (\mu \neq 0)$ and unknown variance σ^2

(a) Explain why $\dfrac{X_1 + X_2 - X_3}{\sigma}$ is not a statistic.

(1)

Given that

$$S = \frac{2}{5}X_1 + X_2 - \frac{1}{3}X_3$$

(b) show that S is a biased estimator of μ

(2)

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

(1)

Given that $Y = aX_1 + bX_2 - 3X_3$ is an unbiased estimator of μ, where a and b are constants,

(d) find, in terms of a and b, an equation that must be satisfied.

(2)

Using the answer to part (d),

(e) show that

$$\text{Var}(Y) = (pa^2 + qa + r)\,\sigma^2$$

where p, q and r are constants to be found.

(3)

(f) Hence, or otherwise, find the smallest possible value of $\text{Var}(Y)$, giving your answer in terms of σ^2

(2)

(Total for Question 3 is 11 marks)

4. A pet shop sells bags of birdseed.

The weight, B kilograms, of birdseed in a bag can be modelled by a Normal distribution with unknown mean μ and known standard deviation 0.9

The birdseed in each of 28 bags in a random sample from a large batch is weighed. The total weight of birdseed in these 28 bags is found to be 335.6 kg

For this batch,

(a) find a 98% confidence interval for the mean weight of birdseed in a bag.

(4)

(b) Explain why the use of the Central Limit theorem is not required to answer part (a)

(1)

The manufacturer claims that the birdseed in each bag in this batch has a mean weight of 12 kg

(c) Using the answer to part (a), comment on the claim made by the manufacturer.

(1)

The pet shop also sells bags of nuts.
The weight, T kilograms, of nuts in a bag can be modelled by a Normal distribution with mean 5 kg and standard deviation 0.3 kg

A customer purchases n bags of nuts.

For these n bags $\mathrm{P}\left(\overline{T} < 5.07\right) = 0.8829$ to 4 decimal places.

(d) Find the value of n

(4)

(Total for Question 4 is 10 marks)

5. A large supermarket chain conducts a survey. An interviewer stands outside one of the supermarkets and interviews customers until a set number of customers in three different age categories have been interviewed.

(a) (i) Name the type of sampling method used.

(ii) In this case, state one possible disadvantage of using this sampling method rather than simple random sampling.

(2)

In the interview, the customer is asked about their age group and the type of product they usually buy. The results are shown below.

		Type of product			
		Value	**Standard**	**Luxury**	**Total**
Age (years)	**Under 30**	29	23	8	60
	30–50	16	29	30	75
	51 and over	20	26	19	65
	Total	65	78	57	200

A chi-squared test is carried out using these results to investigate whether or not there is an association between the type of product bought and the age of the customer.

(b) Calculate the expected frequencies for the age group **Under 30** who bought

(i) value products,

(ii) luxury products.

(2)

The value of $\sum \frac{(O-E)^2}{E}$ for the **other seven classes** is 6.456 to 3 decimal places.

(c) Carry out a suitable test, at the 1% level of significance, to determine whether or not there is an association between the type of product bought and the age of the customer.
Show your working clearly, stating your hypotheses, the degrees of freedom, the test statistic and the critical value used.

(7)

(Total for Question 5 is 11 marks)

6. Local runners take part in a 5 km race around a park. The finishing times for adult males, M minutes, and the finishing times for adult females, F minutes, are such that

$$M \sim \mathrm{N}(24, 9) \text{ and } F \sim \mathrm{N}(29, 14)$$

An adult male runner and an adult female runner are chosen at random.

Using standardisation,

(a) find the probability that the finishing time for the adult male runner is more than 2 minutes faster than the finishing time for the adult female runner.

(5)

Two adult male runners are selected at random.

Using standardisation,

(b) find the probability that their finishing times differ by more than one minute.

(6)

Two adult female runners and one adult male runner are selected at random.

Using standardisation,

(c) find the probability that the total finishing time of the two adult female runners is more than double the finishing time of the adult male runner.

(5)

(Total for Question 6 is 16 marks)

7. The continuous random variable X is uniformly distributed over the interval

$$[k - 6, k + 10]$$

where k is a constant.

A random sample of 80 observations of X is taken.

Given that $\bar{X} = \frac{1}{80}\sum_{i=1}^{80} X_i$

(a) use the Central Limit Theorem to find an approximate distribution for $\bar{X}$
Give your answer in terms of k where appropriate.

(3)

Given that the 80 observations of X have a sample mean of 17.5

(b) find a 99% confidence interval for the lower quartile of the distribution of X

(4)

(Total for Question 7 is 7 marks)

TOTAL FOR PAPER IS 75 MARKS