Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel Level 3 GCE

Time 2 hours

Paper reference **9ST0/02**

Statistics

Advanced
PAPER 2: Statistical Inference

You must have:
Statistical formulae and tables booklet
Calculator

Total Marks

Candidates may use any calculator allowed by Pearson regulations.
Calculators must not 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.
- Unless otherwise stated, inexact answers should be given to three significant figures.
- Unless otherwise stated, statistical tests should be carried out at the 5% significance level.

Information

- A booklet 'Statistical formulae and tables' is provided.
- There are 6 questions in this question paper. The total mark for this paper is 80.
- 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 ▶

P69433A

Q:1/1/1/

P 6 9 4 3 3 A 0 1 2 0

Pearson

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

1 The breaking strength of cables has usually been assessed by testing samples of cables until they break (testing to destruction).

A manufacturer of carbon-fibre-reinforced polymer (CFRP) cables wants to replace testing to destruction with a method for predicting breaking strength based on known characteristics of a cable.

The methods to be compared are a computer simulation model and a mathematical model.

The investigation involves selecting a sample of 18 CFRP cables. Nine of them have their breaking strengths predicted by a computer simulation model and nine by a mathematical model.

Each cable is then tested to destruction to determine its actual breaking strength.

The following are the percentage errors made in the predictions for each model.
A negative error means that the prediction is too low.

Previous research suggests that percentage errors such as these have skew distributions.

Computer model	Mathematical model
–7.5	4.4
–3.2	8.3
–3.3	8.2
–5	6.9
–1.5	10.1
–0.6	10.9
0.7	12.2
4.8	15.9
5.2	15.9

[Data source: https://doi.org/10.1590/1679-78254177]

Question 1 continued

(a) Perform a non-parametric test to decide whether there is evidence of a difference in the average prediction error of these models.

(9)

Question 1 continued

Khalid, a junior researcher, instead suggested evaluating the methods by comparing mean percentage error using a *t*-test.

(b) Make **two** comments on the suitability of this suggestion.

(2)

Lauren, the chief engineer, suggests that, to make the comparison, it would have been better to run both the computer and mathematical models using the same cables, and then testing each to destruction.

(c) State, with justification, whether or not you agree with Lauren's suggestion.

(2)

(d) Suggest a test that Lauren could use to compare the two methods.

(2)

(Total for Question 1 is 15 marks)

2 A large-scale study of the sleeping habits of students in India was carried out.

Students in the study were required to fill out questionnaires about their sleeping habits, which were then used to give a score on a standard Pittsburgh Sleep Quality Index (PSQI).

A student with a PSQI score more than a given value was classified as a 'poor sleeper'.

Amongst the study's reported findings were

- the PSQI scores for students had mean 6.45 and standard deviation 2.85
- 20% of students slept for less than 5 hours a night,
- 62.6% of students were classified as 'poor sleepers'.

[Source: ncbi.nlm.nih.gov/pmc/articles/PMC4659578]

Hamish, a university researcher, is conducting a study to compare the sleeping habits of students in India with those in the UK.

He suspects that the proportion of students who sleep for less than 5 hours a night is **smaller** in the UK than in India.

To investigate this, Hamish asks 40 students at his university how many hours they slept the previous night.

He finds that 6 of them reported sleeping for less than 5 hours.

(a) Making any necessary assumptions, use an exact test to decide whether this sample provides evidence to support Hamish's suspicion.

(5)

Question 2 continued

(b) Identify **two** limitations of Hamish's study.

(2)

Hamish also obtains the PSQI scores for a randomly selected sample of 105 UK students.

He finds that 80% of these students can be classified as 'poor sleepers'.

(c) Conduct a test to investigate whether there is evidence for a difference in the proportion of 'poor sleepers' between students in the UK and those in India.

(6)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 2 continued

The PSQI scores for these 105 students have mean 6.48 and standard deviation 1.71

(d) Construct a 95% confidence interval for the corresponding population mean PSQI score.

(3)

The 95% confidence interval for the mean PSQI score for students in India was found to be (6.07, 6.83).

(e) Using this information and the confidence interval calculated in (d), comment on the mean PSQI scores of the UK and India.

(2)

(Total for Question 2 is 18 marks)

3 A study was conducted into whether social media data could be used to understand tourists' preferences for nature-based experiences.

A large number of photographs taken by tourists at Kruger National Park over a given period, and then posted on Instagram and Flickr, was analysed.

The researchers divided the photographs into categories according to subject matter.

They then used two-sample z-tests to compare proportions of pictures posted on Instagram with those posted on Flickr, for each category.

The p-values from some of these tests are given in **Figure 1**.

All differences were calculated as (proportion on Instagram) – (proportion on Flickr).

Category	p-value
People active	$p < 0.0001$
People posing	$p < 0.01$
Birds	$p < 0.01$
Arthropods (insects, spiders, etc.)	$p < 0.01$
Reptiles	$p < 0.01$

Figure 1

[Data source: https://conbio.onlinelibrary.wiley.com/doi/pdf/10.1111/conl.12343]

(a) Interpret, in context, the p-value in **Figure 1** for 'People active'.

(2)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 3 continued

The researchers also calculated Cohen's *d* values for the differences in proportions, which are interpreted in the same way as Cohen's *d* values for differences in means.

All differences were calculated as (proportion on Instagram) – (proportion on Flickr).

Note that a **negative** value of Cohen's *d* indicates that the proportion on Instagram is **smaller** than the proportion on Flickr.

These values, for the categories used in **Figure 1**, are summarised in **Figure 2**.

Category	Cohen's *d*
People active	0.82
People posing	1.49
Birds	–1.00
Arthropods (insects, spiders, etc.)	–1.50
Reptiles	–0.44

Figure 2

(b) Interpret, in context, the *d*-values in **Figure 2** for

(i) people posing,

(ii) reptiles.

(3)

Question 3 continued

(c) Provide an overall summary for a reader with limited statistical knowledge of the main findings from **Figure 1** and **Figure 2**.

(3)

(Total for Question 3 is 8 marks)

4 A farmer wanted to maximise the spring oat yield, tonnes per hectare, that she could get from a field.

The farmer was able to control two factors

- variety of oat seed planted: Aspen, Canyon, Delfin or Merlin
- concentration of fertiliser used: 10%, 20% or 30%

She wanted to investigate whether either of these factors might have an impact on the yield.

For her experiment, she split a field into four strips and randomly allocated one of the four different types of oat seeds for planting in each strip.

She then subdivided each strip into three smaller plots and randomly allocated one of the three different fertiliser concentrations to each plot during the growing season.

At harvest time, she measured the oat yield for each plot. **Figure 3** gives the plot yields in tonnes per hectare.

	Plot fertiliser concentration			
Oat seed variety	**10%**	**20%**	**30%**	**Total**
Aspen	5.5	5.8	5.7	**17**
Canyon	5.6	5.6	5.5	**16.7**
Delfin	5.5	5.8	5.6	**16.9**
Merlin	5.8	6.1	6.3	**18.2**
Total	**22.4**	**23.3**	**23.1**	**68.8**

Figure 3

Note that $\sum\sum x_{ij}^2 = 395.14$

Question 4 continued

(a) Making any necessary assumptions, perform a two-factor ANOVA to investigate for a difference between varieties **and** for a difference between fertiliser concentration.

(11)

Question 4 continued

(b) State **two** assumptions necessary to make the analysis in (a) valid.

(2)

(c) With numerical justification, what advice would you give to the farmer to maximise her yield with reference to fertiliser concentration and seed variety.

(2)

(Total for Question 4 is 15 marks)

5 Based on an earlier study, Sinead decided to investigate the extent to which people are influenced by other people's laughter.

Participants were asked to rate the 'funniness' of a set of 10 jokes.

The 10 jokes were videoed being told by a professional comedian. Random samples of participants were then shown the video and asked to rate the funniness of each joke on a scale of 1 (unfunny) to 7 (extremely funny).

A first sample of participants watched the video with no laughter track added.

A second sample watched it with a backing track of fake machine-generated laughter.

A third sample watched it with a backing track of real laughter generated by an audience.

The funniness ratings for each joke were then averaged for each of the three samples. The mean ratings obtained are given in **Figure 4**.

		No laughter	Fake laughter	Real laughter
Joke	**1**	2.3	1.9	1.0
	2	3.2	2.9	4.2
	3	2.1	4.2	3.8
	4	3.4	1.9	3.8
	5	2.5	3.5	3.8
	6	2.2	3.0	3.7
	7	1.3	1.9	2.5
	8	1.3	2.4	2.9
	9	1.4	3.5	4.3
	10	2.5	4.0	5.0

Figure 4

Sinead's belief is that jokes are perceived to be more funny if accompanied by the sound of laughter.

The test values calculated using the data in **Figure 4** are

$t = 1.93$ for 'Fake laughter' compared to 'No laughter' and

$t = 3.51$ for 'Real laughter' compared to 'No laughter'

Question 5 continued

(a) State the hypotheses being tested and decide whether these data support Sinead's belief.

(4)

(b) Conduct a *t*-test using data from **Figure 4** to investigate whether there is evidence of any difference between 'Fake laughter' and 'Real laughter' in terms of the perceived funniness of jokes.

(8)

Question 5 continued

(Total for Question 5 is 12 marks)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

6 A large gardening company conducted a survey on customer satisfaction by e-mailing a random sample of 400 of its customers, of whom 62 responded.

The company wanted to see if there was a relationship between whether it was the customer's first experience and the likelihood that the customer would use the company again.

The results of this survey are summarised in **Figure 5**.

		First experience		
		Yes	**No**	**Total**
Likelihood of using company again	**Definitely yes**	9	12	**21**
	Probably yes	18	2	**20**
	No	17	4	**21**
	Total	**44**	**18**	**62**

Figure 5

(a) Investigate whether there is evidence of an association between whether it was the customer's first experience and the likelihood that the customer would use the company again.

(8)

Question 6 continued

(b) Describe, with numerical justification, the nature of any association identified in your conclusion to the test in (a).

(2)

(c) Give **two** possible sources of bias in this investigation.

(2)

(Total for Question 6 is 12 marks)

TOTAL FOR PAPER IS 80 MARKS

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE

BLANK PAGE