Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel Level 3 GCE

Centre Number | Candidate Number

Thursday 20 June 2019

Morning (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 permitted 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). Coloured pencils and highlighter pens must not be used.
- **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, statistical tests should be carried out at the 5% significance level.
- When a calculator is used, the answer should be given to three significant figures unless otherwise stated.

Information

- A booklet 'Statistical Formulae and Tables' is provided.
- There are 7 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 ▶

©2019 Pearson Education Ltd.

Pearson

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

1 As part of an investigation into marine pollution in a particular area, the mercury content in the muscles of a sample of 36 dolphins is measured.

These 36 dolphins have muscle mercury content with mean 4.4 μg/g and standard deviation 2.3 μg/g.

(a) Find a 99% confidence interval for the mean mercury content in the muscles of dolphins in this area.

(3)

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

(b) Why is it not necessary to assume that mercury content in dolphin muscle follows a normal distribution?

(2)

(c) What assumption is required for the interval calculated in part (a) to be valid?

(1)

(Total for Question 1 is 6 marks)

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

2 The following table shows the numbers of butterflies of two species, *Orange Tips* and *Peacocks*, found in comparable surveys in 7 successive years.

	Year						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Orange Tips	90	14	36	24	47	55	22
Peacocks	48	110	85	54	65	67	95

For these data, the requirements for using a procedure based on Pearson's product moment correlation coefficient are **not** met.

Calculate an alternative measure and carry out a hypothesis test to decide whether these data support the view that when *Orange Tips* are relatively abundant, *Peacocks* are relatively uncommon.

(7)

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 7 marks)

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

3 It is known that the speed of sound, at 20 °C at sea level, is 343 m/s through dry air and 1481 m/s through distilled water.

It is required to investigate whether a **new** procedure can be used to accurately measure the speed of sound as it passes through a range of different materials.

As an initial study, independent measurements of the speed of sound at 20 °C at sea level through dry air and through distilled water are taken using the new procedure. You may assume that these measurements follow normal distributions.

The 6 measurements (in m/s) taken through **dry air** are

342 340 337 335 339 343

(a) Test whether, on average, measurements using this new procedure differ significantly from the known speed of sound at 20 °C at sea level through **dry air**.

(7)

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

The 10 measurements of the speed of sound through **distilled water** are found to have a standard deviation of 1.25 m/s.

(b) Find the critical region, based on a random sample of 10 measurements, that could be used to test whether, on average, measurements using this new procedure differ significantly at the 5% level from the known speed of sound at 20 °C at sea level through **distilled water**.

Give your answers correct to an appropriate level of accuracy.

(4)

(Total for Question 3 is 11 marks)

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

4 In 2015, a nuclear energy advocacy group claimed that a majority of adults in the USA support nuclear energy.

In the same year, a Gallup telephone poll of a sample of 1025 adults in the USA found that 51% of respondents expressed support for nuclear energy.

[Source: Gallup.com]

Making any necessary assumptions, use a hypothesis test to show that the poll does not provide convincing evidence to support the nuclear energy advocacy group's claim.

You should **also** explain, in context, whether this conclusion means that the nuclear energy advocacy group's claim was definitely wrong.

(9)

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 9 marks)

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

5 Robert, a psychologist, is devising a diagnostic test to assess the ability of air traffic controllers to cope with unexpected stress. There are two versions of the test that are to be compared.

He suspects that scores using version B of the test will be, on average, lower than scores using version A.

To investigate this, Robert runs a trial of the two versions of the test. He takes a random sample of air traffic controllers and randomly assigns either version A or version B of the test to each controller. Given below are the scores Robert obtained.

Version A	Version B
53	62
71	28
62	31
42	35
80	42
36	63
70	79
88	

Preliminary trials of these two versions of the diagnostic test suggest that the scores are **unlikely to be normally distributed**.

(a) Carry out a hypothesis test to investigate whether these data support Robert's suspicion that scores using version B of the diagnostic test will be, on average, lower than scores using version A.

(9)

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

Question 5 continued

(b) Discuss briefly whether you think the scores for the two groups of air traffic controllers are independent of each other.

(3)

(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 Lara, a regional bank manager, is interested in studying the usage of the ATM located outside a particular branch during the evening.

She divides the hours between 6 pm and midnight into 10-minute intervals. She then counts the numbers of customers who use the ATM during each interval over a 5-day week.

The data are given in **Figure 1** below.

Number of customers (x)	**Number of intervals**	**Poisson probability $P(X = x)$**
0	73	0.3329
1	50	0.3662
2	30	0.2014
3	20	
4	7	
$\geqslant 5$	0	

Figure 1

Lara assumes that a Poisson distribution is an appropriate model for the number of customers arriving in a 10-minute interval.

(a) Using the data in **Figure 1**, find the mean number of customers per 10-minute interval. Use this mean to calculate the three missing probabilities in **Figure 1** giving your probabilities to 4 decimal places.

(4)

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

(b) Use a goodness-of-fit test to decide whether it was reasonable for Lara to use a Poisson distribution to model the number of customers using this ATM during the evening.

(9)

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

Question 6 continued

(c) What does this analysis tell you about the arrival pattern of customers at this ATM during the evening?

(3)

(Total for Question 6 is 16 marks)

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

BLANK PAGE

7 An experiment is conducted to compare the effects of 3 energy drinks, A, B and C, on endurance.

A sample of 6 volunteer subjects from a local gym are instructed to take each energy drink in a randomised order.

Thirty minutes after taking a drink, the subject's endurance is measured using the time to exhaustion while jogging at a medium intensity level on a treadmill. A suitable recovery time is allowed between each drink.

The times recorded, X minutes, with various totals, are given in the table below.

	Drink			
Subject	**A**	**B**	**C**	**Totals**
1	52	58	68	**178**
2	46	44	54	**144**
3	65	59	69	**193**
4	54	56	58	**168**
5	47	41	49	**137**
6	56	64	67	**187**
Totals	**320**	**322**	**365**	**1007**

Note that $\Sigma\Sigma x_{ij}^2 = 57\,555$

(a) Use the analysis of variance (ANOVA) to determine whether, on average, the 3 energy drinks differ in their effects on endurance.

(10)

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

Question 7 continued

(b) Describe the nature of any differences found in part (a).

(2)

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

(c) State the name of the experimental design used and describe the advantages of using such a design in this experiment.

(3)

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

(d) With reference to the ANOVA test carried out in part (a), discuss whether the blocking factor was effective. Use calculations to justify your answer.

(4)

(Total for Question 7 is 19 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

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