Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel Level 3 GCE

Centre Number | Candidate Number

Tuesday 25 June 2019

Morning (Time: 2 hours) | Paper Reference **9ST0/03**

Statistics

Advanced

Paper 3: Statistics in Practice

You must have:
Statistical Formulae and Tables booklet
Calculator

Total Marks

Candidates may use any calculator allowed by the 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, 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 ▶

R61176A

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

Pearson

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

1 A group of eggs in a bird's nest is called a clutch. The number of eggs in a clutch is called the clutch size.

Stéphanie is studying the size of tree swallow clutches in southern Québec, Canada. She surveys a large number of occupied nest boxes and records the sizes of the clutches. She calculates the probability that each clutch size occurs using her data.

Stéphanie's results are shown in **Figure 1**.

Clutch size x	P($X = x$)
0	0
1	0
2	0.006
3	0.02
4	0.096
5	0.377
6	0.417
7	0.073
8	0.011
≥ 9	0

Source: Adapted from https://doi.org/10.1111/jav.00725

Figure 1: Probabilities of tree swallow clutch sizes

(a) Find the probability that a tree swallow clutch, chosen at random, contains more than 6 eggs.

(1)

(b) Calculate the probability that if 3 tree swallow clutches are chosen at random, none of them contains more than 6 eggs.

(2)

(c) Find the mode, median and mean number of eggs in a tree swallow clutch.

(5)

Question 1 continued

(d) Calculate the standard deviation of the number of eggs in a tree swallow clutch.

(2)

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

Stéphanie's colleague Max believes that the clutch size of tree swallows can be modelled by a normal distribution.

(e) Use **Figure 1** and your answers to parts (c) and (d) to make three comments about Max's belief.

Make **at least one** comment in support of Max's belief and **at least one** comment against it.

(3)

(Total for Question 1 is 13 marks)

2 Antonius is a researcher into the medical condition IBS (Irritable Bowel Syndrome). He surveyed the quality of life of random samples of male and female patients with IBS.

His summarised results are shown in **Figure 2**.

Higher results indicate a higher quality of life.

	Sample size	Sample mean	Sample standard deviation
Male	18	58	12.6
Female	53	47.9	12.7

Source: PLoS ONE 12(7): e0181764

Figure 2: Results of quality of life survey of male and female patients with IBS

Antonius carried out a *t*-test comparing the quality of life of the male and female patients.

The *t*-test resulted in a pooled sample variance of $s_p^2 = 160.7$ and a *p*-value of 0.00471.

(a) Calculate the value of Cohen's *d* for the effect of male IBS patients' quality of life as compared to female IBS patients' quality of life.

(1)

(b) Use both the result of Antonius' *t*-test and the value of Cohen's *d* calculated in part (a) to draw conclusions about the difference in male and female IBS patients' quality of life. You may assume that quality of life survey results follow a normal distribution.

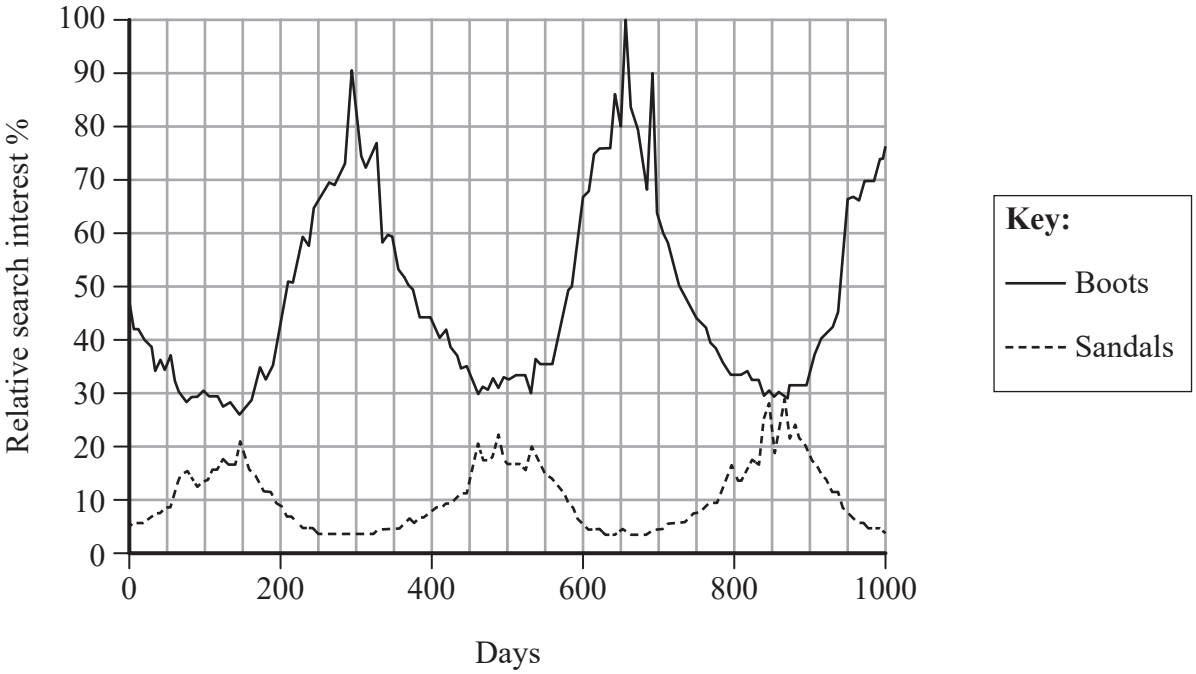
(2)

(Total for Question 2 is 3 marks)

3 The diagram, **Figure 3**, illustrates relative search interest % in boots and sandals using the Google search engine during a one-thousand-day period. The information is for the UK only.

The numbers on the vertical axis represent search interest relative to the highest point on the chart for the UK during this time period.

A value of 100% is the peak interest in boots or sandals, in the UK, during this time period.



Source: Google Trends

Figure 3

(a) Describe the pattern of search interest in boots and sandals that is illustrated in **Figure 3**.

Make **four different** comments.

(4)

Question 3 continued

(b) What useful information about search interest in boots and sandals, in the UK, during this time period, is not included in **Figure 3**?

Make **two different** comments.

(2)

(Total for Question 3 is 6 marks)

4 Following the Fukushima Daiichi nuclear power plant disaster in 2010, a group of scientists studied radiation shielding.

Before they started testing radiation shields, the scientists measured the background radiation within their laboratory. They used a Geiger counter which counted the number of radioactive particles detected in the laboratory in counts per minute.

The scientists modelled the background radiation with a Poisson distribution.

(a) Describe, in context, **three** assumptions that must be made about background radiation for the Poisson model to be suitable.

(3)

The Geiger counter recorded 84 cpm (counts per minute) in the laboratory.

Source: Adapted from DOI 10.1007/s12199-012-0314-6

(b) Calculate the mean and the standard deviation of the number of counts **per second** of radioactive particles in the laboratory when the Poisson distribution is used as a model.

(3)

(c) Find the probability that fewer than 4 counts of radioactive particles are recorded in any one second in the laboratory.

(2)

A Geiger counter clicks every time it detects a radioactive particle.

(d) State **fully** the distribution of the time (in seconds) between successive clicks of the Geiger counter in the laboratory.

(2)

(e) Using the model from part (d), find the probability that a wait of less than 2 seconds is recorded for the time between two randomly selected successive clicks of the Geiger counter in the laboratory.

(2)

(f) Calculate the expected time, in seconds, between successive clicks of the Geiger counter in the laboratory.

(1)

(Total for Question 4 is 13 marks)

5 Daisuke, a physicist, conducted a study into whether a 4 mm thick layer of cardboard is an effective shield against radiation from soil contaminated by a nuclear power plant disaster.

He measured the radiation dose in microSieverts per hour (μSv/h), for containers of contaminated soil, both without and with a dry 4 mm thick cardboard cover.

Daisuke's summarised results are shown in **Figure 4**.

	Sample size	Sample mean	Sample standard deviation
Without dry 4 mm thick cardboard cover	5	13.0	0.10
With dry 4 mm thick cardboard cover	5	9.99	0.08

Source: DOI 10.1007/s12199-012-0314-6

Figure 4

(a) Test at the **1% significance level**, whether the dry 4 mm thick cardboard cover has decreased the radiation dose from the contaminated soil by **more than 2 μSv/h** on average.

You may assume that doses of radiation follow a normal distribution.

(9)

Question 5 continued

Exposure to a radiation dose of 11.4 μSv/h or more, for the duration of one year, is considered dangerous to human health.

(b) Write a short report about the use of cardboard, in general, as a radiation shield.

Use your conclusion in part (a) and give consideration to the practical implications.

(4)

(Total for Question 5 is 13 marks)

6 Andrew works in the human resources department of a very large company. He has been given the task of assessing the impact of a new procedure for submitting work expenses claims.

Andrew has been given the following information about claims for **one department** of the company.

In the 6 months **before** the introduction of the new procedure, the number of claims received was 522, of which 43 were rejected because they were submitted incorrectly.

In the 3 months **after** the introduction of the new procedure, the number of claims received was 315, of which 19 were rejected because they were submitted incorrectly.

(a) Using Andrew's data, conduct a test to investigate whether the proportion of claims rejected because they were submitted incorrectly has changed, at the company, since the introduction of the new procedure.

(8)

Question 6 continued

(b) Give **three** reasons, in context, why the result of the test in part (a) might be invalid.

Consider the method of data collection and assumptions of the test used.

(3)

(c) Andrew has been told to conduct a questionnaire survey on the opinions of the company's employees about the new expenses claim procedure.

Advise Andrew on how he could successfully collect an appropriate sample of completed questionnaires. Take into account the following information.

- Andrew has been given only 10 working days to complete his survey and analyse the results.
- A spreadsheet containing records of employee roles and contact details is available to him.
- The more senior roles within the company have fewer people in them.
- The more senior roles within the company make more expenses claims.

(4)

Question 6 continued

(Total for Question 6 is 15 marks)

Turn over ▶

7 A group of researchers found that skin healing cells called fibroblasts are more active during the day than during the night.

The researchers then collected data from the records of a specialist hospital burns unit of the time x, in days, that 26 similar burn injuries, in adults, took to heal (to 95% wound reduction).

They categorised the data by the time of day at which the injury occurred.

Their results are shown in **Figure 5**.

	Time of day at which the burn injury occurred					
	00:00 – 03:59	**04:00 – 07:59**	**08:00 – 11:59**	**12:00 – 15:59**	**16:00 – 19:59**	**20:00 – 23:59**
	29	27	12	15	16	20
	38	30	15	17	17	21
	42	32	16	18	19	23
			18	20	19	25
				21	20	
					21	
					23	
Total	109	89	61	91	135	89

(Data source: DOI: 10.1126/scitranslmed.aal2774)

Figure 5: Time taken in days for burn injuries to heal

The data produced the following summary statistic

$$\sum\sum x^2 = 13\,962$$

You may assume that the times taken for burns to heal follow a normal distribution.

(a) Carry out a hypothesis test to investigate whether there is any difference between the mean times taken for a burn to heal based on the time of day at which the burn occurred.

Interpret your conclusion fully. Include comments on the specific differences, if any, between the mean times taken for a burn to heal for the different times of day at which the burn occurred.

(13)

Question 7 continued

Question 7 continued

(b) Identify **two** disadvantages of the research **method** used to collect the data for the analysis in part (a).

(2)

(c) Make **two** recommendations about data collection for the next stage of the research.

(2)

(Total for Question 7 is 17 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