

Please check the examination details below before entering your candidate information

Candidate surname					Other names				
Centre Number					Candidate Number				

Pearson Edexcel Level 3 GCE

Tuesday 17 June 2025

Afternoon (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 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 ►

P76383A

©2025 Pearson Education Ltd.
Y:1/1/1/1/1/



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE



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

- 1 A study in 2021 investigated the number of hail days in a large European area covering Germany, France, Belgium and Luxembourg.

A hail day was defined as one in which a radar detected the conditions necessary for hail to occur anywhere within this area. Over the past 20 years, the mean number of hail days in a year in this area was 86

[Source: www.doi.org/10.5194/nhess-21-683-2021]

- (a) Find the probability that

- (i) in a randomly chosen year there would be at least 90 hail days,

(2)

- (ii) in a randomly chosen **two-year** period, there are no more than 160 hail days.

(2)

- (b) State **two** assumptions needed, in the context of this question, to calculate the probabilities in (a) and in **each** case give a reason why that assumption may **not** be appropriate.

(4)

(Total for Question 1 is 8 marks)

- 2 A small company based in the tourism industry is trying to decide where best to spend its advertising budget in order to attract more visitors.

It is comparing the visitor numbers to mainland UK from France and from Ireland in different periods of time.

Figure 1 shows the number of visitors from France and from Ireland for a randomly selected series of yearly quarters. Visitor numbers have been rounded to the nearest 1000

		Quarter						
		2019 Q2	2019 Q3	2021 Q1	2021 Q3	2022 Q1	2022 Q3	2022 Q4
Country	France	929 000	898 000	13 000	148 000	462 000	697 000	842 000
	Ireland	718 000	716 000	24 000	247 000	417 000	640 000	731 000

Figure 1

[Source: www.ons.gov.uk/peoplepopulationandcommunity/leisureandtourism]

In some quarters, the number of visitors from France exceeds that from Ireland, whilst in others the reverse is true. The data is also clearly impacted by changes in travel patterns following the pandemic.

You should assume that the difference between the numbers of visitors from France and Ireland is **not** symmetrically distributed.

- (a) Carry out a suitable hypothesis test to determine whether there is a difference between the average number of visitors to mainland UK from the two countries.

(6)



Question 2 continued

- (b) Explain the purpose of pairing the data in this way rather than simply examining the total number of visitors for a randomly selected period of time.

(2)

An employee of the company suggests increasing the sample size by using visitor data from earlier than 2019

- (c) Explain briefly why using earlier data may not help the company make its decision about where to spend its advertising budget.

(2)**(Total for Question 2 is 10 marks)**

- 3 Eric is investigating which popular marathons are the quickest for an average runner to race in. As part of his initial investigation, he comes across data about 150 random male finishers in the London Marathon 2022, shown in **Figure 2**

Time (minutes)	Number of runners
Less than 180	12
180 to 210	22
210 to 240	35
240 to 270	24
270 to 300	18
300 to 360	25
More than 360	14
Total	150

Figure 2

[Source : www.marathonguide.com/results/]

Using **Figure 2** it is not possible to calculate the mean and standard deviation of the finishing times of these athletes.

- (a) Give **two** reasons why this is not possible.

(2)



Question 3 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Question 3 continued

Eric also calculated the mean and standard deviation for a random sample of 150 men to complete the Berlin Marathon 2022. The mean was 272.82 minutes and the standard deviation was 73.24 minutes.

- (c) Use a normal distribution to carry out a hypothesis test to determine whether the mean completion time for the men's London Marathon 2022 is quicker than that for the men's Berlin Marathon 2022

(7)

Neither of the London nor Berlin marathons have a population distribution for men's completion times that is normally distributed.

- (d) Explain why it is still valid to use a normal distribution in the test in (c)

(2)

(Total for Question 3 is 21 marks)

- 4 In order to collect data, a polling firm has a large database of people who have volunteered to take part in its surveys. The firm records a broad range of socio-demographic data about these volunteers.

In any individual survey, the firm chooses volunteers so that the sample is representative of adults in the UK population who voted in previous elections in terms of age, gender, social class and education.

[Source: www.yougov.co.uk/about/panel-methodology]

- (a) Name the type of sampling the polling firm uses to select volunteers for its surveys.

(1)

The polling firm is investigating how voters are likely to vote in the next UK general election.

In one survey taken in August 2023, focussing only on those expressing an intention to vote, the poll finds that of a sample of 2103 respondents

- 547 are intending to vote for the Conservative Party
- 925 are intending to vote for the Labour Party
- the remainder are intending to vote for other parties.

[Source: www.yougov.co.uk/topics/politics/explore/issue/Voting_Intention]

The proportions in the sample exactly represent the proportions of those that intend to vote in the UK population.

- (b) Show that the proportion of voters intending to vote for other parties is 30%, to the nearest whole number.

(1)



Question 4 continued

X represents the number of people voting for other parties in a sample of 50 voters and can be modelled by a binomial distribution.

- (c) State **two** assumptions, in the context of the question, necessary to model X using this distribution. (2)

- (d) Find the probability that in a randomly selected sample of 50 voters, more than 20 of them will state an intention to vote for other parties. (2)

- (e) Using a suitable approximation, calculate the probability that 171 or more voters in a random sample of 500 are intending to vote for other parties. (4)

Question 4 continued

Another polling firm considers that the claim that 30% of voters intending to vote for other parties may **not** be correct.

It finds that in a randomly selected sample of 500 voters there are 171 who express an intention to vote for other parties.

- (f) Use your answer to (e) to comment on the claim that the proportion of voters intending to vote for other parties at the next UK general election is **not** equal to 30%

(2)

(Total for Question 4 is 12 marks)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE



- 5 Oseltamivir is the active ingredient in an antiviral medicine used to treat or prevent influenza (flu). The European Medicines Agency claims that a medicine containing oseltamivir reduced the duration of flu compared with a placebo (an inactive drug).

The European Medicines Agency report notes

- patients were blinded as to whether they were given the placebo or the medicine with oseltamivir
- effectiveness was measured by recording symptoms of flu
- in adults, the duration of flu was reduced from an average of 5.2 days for patients taking the placebo, to 4.2 days for patients taking the medicine

[Adapted from source: www.ema.europa.eu/]

- (a) Explain the purpose of patients being “blinded” in a medical trial.

(1)

An epidemiologist is trying to test the claim of the European Medicines Agency and carries out a trial with a random sample of 10 patients.

The epidemiologist gives each of the patients the medicine when they first report symptoms of flu and the patients each report how many days, x , their symptoms last. The results are summarised below.

$$\sum x = 47 \quad \sum x^2 = 233$$

- (b) Construct a 95% confidence interval for these data, assuming that the number of days the symptoms of flu last is normally distributed.

(6)



Question 5 continued

- (c) Explain why the selection of the epidemiologist's patients must be random.

(1)

- (d) By making reference to the confidence interval found in (b), comment on the European Medicines Agency's claim that the duration of the flu when taking the medicine was 4.2 days on average.

(2)

The epidemiologist also tests 8 patients in the same way using a placebo and finds the mean number of days that the flu lasts for this group to be 5.3 with a standard deviation of 1.05

A two-sample t -test was carried out to test whether there was a difference between the mean duration for those taking the medicine and for those taking the placebo. For this test a p -value of 0.137 was calculated.

- (e) Show that a suitable estimate for the pooled variance of the two groups of patients, correct to 3 significant figures, is 1.24

(1)

Question 5 continued

- (f) Calculate Cohen's d for the two groups and comment on the size of the effect that the medicine is having.

(2)

- (g) Write a short report about the use of this medicine to reduce the length of time that patients suffer from the symptoms of flu.

Use the value you calculated in (f) as well as the p -value from the epidemiologist's test and give consideration to the practical implications of giving a medicine to reduce the length of time a patient suffers from the symptoms of flu.

(4)

Question 5 continued

After publishing their report, the epidemiologist is given more funding to further investigate the effects of oseltamivir.

- (h) State **two** recommendations that you would make to the epidemiologist about the data collection in their next study.

(2)

(Total for Question 5 is 19 marks)

6 A scientist wishes to research the ratio of male badgers to female badgers living wild in the UK. To investigate this, the scientist intends to collect data in the following way.

- Choose 5 rural areas that badgers are known to inhabit
- Set up cameras to record the activity of the badgers in each of these areas over a period of 2 consecutive months
- Count the number of male and female badgers captured on the cameras using experts to ensure correct identification and that no badger is counted multiple times

(a) State **two** factors which may have an impact on the ratio of male to female badgers observed.

(2)

(b) Explain briefly why it is important for the scientist to declare the methodology before starting the data collection.

(2)



Question 6 continued

Of the 230 badgers observed, the scientist found that 108 of them were male.

The scientist then uses a one sample test of proportion to test the hypothesis that the ratio of male badgers to female badgers is 1 : 1

The hypothesis test is used to conclude that there is not enough evidence to say that there is anything other than a 1 : 1 ratio of male to female badgers in the UK.

Based on years of experience, the scientist suspects that the conclusion of the hypothesis test is incorrect.

(c) (i) State what type of statistical error the scientist believes has been made.

(1)

(ii) Suggest a way that the scientist could reduce the probability of this error.

(1)

(d) With reference to the way that the data was collected, make **two** criticisms of the conclusion of the research.

(2)

Question 6 continued

- (e) Suggest **two** improvements to the data collection method in order to better match the intention of the research.

(2)

(Total for Question 6 is 10 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

