

Please check the examination details below before entering your candidate information

Candidate surname		Other names	
Centre Number		Candidate Number	

Pearson Edexcel Level 3 GCE

Thursday 5 June 2025

Morning (Time: 2 hours)	Paper reference	9ST0/01
-------------------------	-----------------	----------------

Statistics

Advanced

PAPER 1: Data and Probability

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 9 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 ►

P76381A

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



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

- 1 The stem and leaf diagram below shows the number of points scored by the Los Angeles Lakers basketball team in the first 19 games of the 2019–2020 season.

Points	9 5 represents 95 points	Totals
9	5 5 9	(3)
10	2 3 4 9	(4)
11	2 4 4 8 9	(5)
12	0 0 0 2 3 5	(6)
13	0	(1)

[Source: www.basketball-reference.com]

- (a) Find the median and the interquartile range for these data.

(3)

Figure 1 shows a box and whisker diagram containing information about the number of points scored by the Los Angeles Lakers basketball team in the first 19 games of the 2020–2021 season.

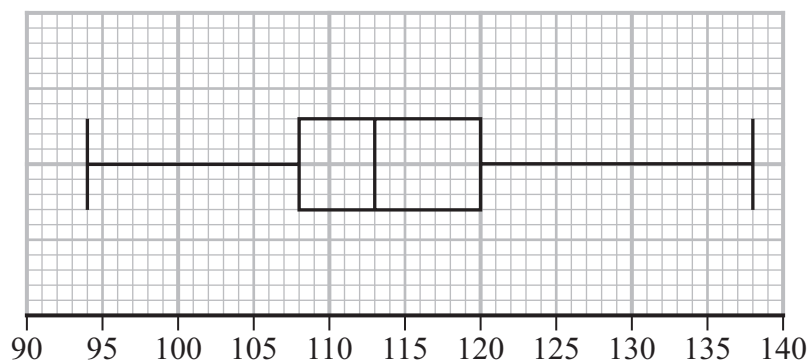


Figure 1

[Source: www.basketball-reference.com]



DO NOT WRITE IN THIS AREA

- 2 The figures below show the amount, in £ millions, of UK bilateral aid given to Pakistan for each year of the years 2014 to 2019

305 331 402 463 374 266

[Source: dfid.gov.uk]

The table gives a summary of the amount, in £ millions, of UK bilateral aid given to Ethiopia for each of the years 2014 to 2019

Mean	Standard deviation
295.5	36.1

[Source: dfid.gov.uk]

By comparing the means and standard deviations of the bilateral aid given to each country, Brian reaches a conclusion that the UK has given more consistent bilateral aid to Pakistan than to Ethiopia, between 2014 and 2019

Assess the validity of this conclusion.

(5)



DO NOT WRITE IN THIS AREA

(Total for Question 2 is 5 marks)



- 3 Ali, Bob and Colin travel independently each day from their homes to the same workplace.

All three adults can choose to walk or cycle or drive to work.

The table shows the probabilities of their independent daily choices.

	Walk	Cycle	Drive
Ali	0.65	0.1	0.25
Bob	0.45	0.4	0.15
Colin	0.2	0.5	0.3

- (a) Find the probability that, on any given day, at least 2 of the 3 walk to work.

(4)



Question 3 continued

Dina, a friend of Ali, never drives to work.

The probability that Dina walks to work when Ali walks to work is 0.7

The probability that Dina cycles to work when Ali cycles to work is 0.9

(b) Find the probability that, on any given day, Dina and Ali each travel to work by

(i) the same method,

(3)

(ii) a different method.

(1)

(Total for Question 3 is 8 marks)

4 Vicenti is a statistician and has been asked to classify the following variables.

A = The number of garden pea seeds that germinate from a pack of 30 seeds, given that the probability that a garden pea seed germinates is 0.8

B = The birth weight of newborn babies in the UK, given that the mean birth weight is 3.4 kg

C = The number of cars that arrive at a toll booth in a given minute between 8 am and 6 pm, given that the average number per minute is 5

Write a report to advise Vicenti whether each of the variables can be successfully modelled by

- the Poisson distribution
- the binomial distribution
- neither of these distributions

Your report should state, where possible, the value of any parameters and any necessary assumptions, giving a reason for your choice of Poisson, binomial or neither distribution for each variable.

(8)



DO NOT WRITE IN THIS AREA

(Total for Question 4 is 8 marks)

- 5 Chris works for an estate agent and was asked to investigate whether the average annual salary of regions in England influences the average price of a house in the same regions.

As part of this investigation, Chris searched the internet for data and found that in 2022 the Office for National Statistics published the information shown in Figure 2

[Source: www.ons.gov.uk]

Region	Median house price (£) (<i>P</i>)	Median gross annual workplace-based earnings (£) (<i>E</i>)
North East	145 000	29 521
North West	190 000	30 716
Yorkshire and The Humber	185 000	30 000
East Midlands	226 000	30 326
West Midlands	225 000	31 601
East	328 000	32 539
London	525 000	41 866
South East	370 000	34 431
South West	295 000	31 339

Figure 2

- (a) Suggest a suitable diagram that Chris could draw to see if there is an association between the median house price and the median gross annual workplace-based earnings. Give a reason for your answer.

(2)



Question 5 continued

- (b) For these data, find the value of Pearson's product-moment correlation coefficient between P (median house price) and E (median gross annual workplace-based earnings).

Interpret this value in context.

(2)

- (c) For these data, find the equation of the least squares regression line in the form

$$P = a + bE$$

Interpret your value of b in context.

(3)

Question 5 continued

- (d) Explain why Chris should not use the equation in part (c) to estimate the median house price in Wales, which had a median gross annual workplace-based earnings of £30 596 in 2022

(1)

Chris suggests that there is clear evidence that the average annual salary of a region in England influences the average price of a house in the same region.

- (e) Explain, in context, why Chris may be wrong.

(2)

(Total for Question 5 is 10 marks)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE



6 *Boiled in the Bag Ltd* is a company manufacturing boiled sweets.

These are sold in packets that contain 50 sweets, on average, with each packet containing a variety of colours of boiled sweets.

The number of green boiled sweets in a randomly selected packet can be modelled by a binomial distribution with parameter p equal to 0.3

(a) State **two** assumptions necessary for this distribution to be a suitable model.

(2)

(b) Find the probability that a randomly chosen packet of boiled sweets contains

(i) less than or equal to 16 green sweets,

(1)

(ii) exactly 6 green sweets,

(1)

(iii) more than 9 but fewer than 20 green sweets.

(2)



Question 6 continued

Boiled in the Bag Ltd establishes a production process in which the correct proportions of each colour of sweet are made and brought together for mixing prior to being packed in packets of 50 sweets.

Jaymini is the quality control manager at *Boiled in the Bag Ltd*. She draws a sample of 8 packets every hour throughout the workday in the factory and notes the number of green sweets per bag, finding a mean of 15.4 and a variance of 20.1

She concludes that there is a problem with the production process.

(c) Comment on the validity of this conclusion, giving reasons for your answer.

(4)

(Total for Question 6 is 10 marks)

- 7 A trial was carried out to investigate the effect of topical olive oil application on the symptoms of infant colic. A randomised, double-blind, placebo-controlled clinical trial was conducted.

Details of the design of the trial were stated as follows:

The study was carried out on 80 infants, one to three months old, randomly allocated into two groups, receiving abdominal massage with olive oil three times a day in the intervention group ($n = 40$) and the same procedure with liquid paraffin in the placebo group ($n = 40$), for 14 successive days.

During this period, the parents recorded the episodes of colic, duration of colic [and] crying intensity. These variables were assessed and compared at the beginning and on the 7th and 14th days of the trial.

[Source: British Medical Journal, 2023]

- (a) Explain, in the context of this trial, the purpose of

- (i) randomly allocating into two groups,

(2)

- (ii) a double-blind trial,

(2)



Question 7 continued

(iii) a placebo-controlled trial.

(2)

(b) Explain why the variables were assessed and compared

(i) at the beginning of the trial,

(1)

(ii) on the 7th day of the trial,

(1)

(iii) on the 14th day of the trial.

(1)

Question 7 continued

The results of the trial are summarised below.

The mean crying duration changed from 4.05 to 1.41 hours per day (65% decrease) in the olive oil group and from 3.85 to 1.60 hours per day (58% decrease) in the paraffin oil group.

The mean episodes of crying were reduced from 5.79 to 2.51 episodes per day (57% decrease) in the olive group and from 6.01 to 3.01 episodes per day (50% decrease) in the paraffin oil group.

Furthermore, the intensity of crying was decreased from 9.04 to 4.48 in the olive oil group and from 9.0 to 4.77 in the paraffin oil group.

[Data extracted from: British Medical Journal, 2023]

- (c) Using these results, assess the effect of topical olive oil application on the symptoms of infant colic.

(3)

(Total for Question 7 is 12 marks)



8 A certain disease affects 0.1% of the population.

The National Institute for Health and Care Excellence (NICE) is investigating a possible diagnostic test for the disease that gives a positive response, in the presence of the disease, with probability 0.95

It also gives a positive response with probability 0.02 if the disease is not present.

- (a) Find the probability that a person has the disease, given a positive response to the test.

(5)

Researchers at NICE write an internal report stating that the test is clearly unsatisfactory.

- (b) Using your answer to part (a), explain, with a reason, whether you agree with this finding.

(2)

(Total for Question 8 is 7 marks)

9 A random variable X has a normal distribution.

(a) Describe **two** features of the distribution of X

(2)

The time, T minutes, taken by a tyre fitter to change a tyre is normally distributed with mean 22.5 minutes and standard deviation 2.5 minutes.

(b) Find the probability that it takes a tyre fitter less than 28 minutes to change a tyre.

(1)

Charlie takes a car to the tyre fitter that needs all four tyres changed.

(c) Find the probability that

(i) it takes at least 5 minutes longer to change the second tyre than it does to change the third tyre,

(4)



Question 9 continued

- (ii) it takes over 2 minutes more to change the first tyre than the mean time to change all four tyres.

(6)

(Total for Question 9 is 13 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

BLANK PAGE



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE

