

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 4 June 2026

Afternoon (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 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 ►

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

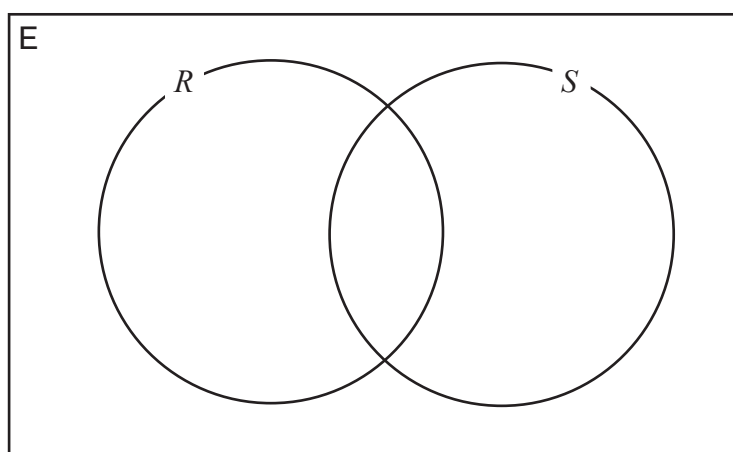
- 1 In a card game there is a deck of 108 cards. There are 8 black cards. The remaining cards are divided equally between 4 colours: red, blue, green and yellow.

Some of the cards have features that make them special. All the black cards and 6 of each of the other coloured cards are special.

[Source: <https://www.unorules.org/>]

- (a) Complete the Venn diagram below representing the numbers of cards in the deck which are either red (R) or special (S).

(3)



A single card is selected from the deck of cards.

- (b) Find the probability the card is:

(i) a special card,

(1)

(ii) neither a red nor a special card,

(1)

Question 1 continued

(iii) a red card given that it is not special.

(2)

(c) Show that R and S are **not** statistically independent.

You should give a numerical justification for your answer.

(2)

3 cards are selected at random from the deck of cards, without replacement.

(d) Find the probability that exactly 2 of them are red.

(2)

Question 1 continued

(e) Give a meaning, in context, of a card which is in a category defined as $_S R$

(2)

During a game, Sam has exactly one card left in his hand and it is now his turn. If he can play this card he will win the game.

Sam is only able to play his card if it is black or if it matches the colour of the last card played.

The last card played was blue.

20 cards have already been played, of which 6 were blue cards and 2 were black cards.

(f) Find the probability that Sam is able to play his card on this turn.

(2)

(Total for Question 1 is 15 marks)

DONOTWRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE

- 2 Systolic blood pressure is a measure of the pressure at which blood is pumped around a human body. It is measured in millimetres of mercury (mmHg). Typical values are between 80mmHg and 150mmHg.

A very large study of 70 000 randomly selected people was conducted. The results of the study were used to calculate the mean and standard deviation of the systolic blood pressure.

Here are the results:

- Mean=126.6mmHg
- Standard deviation=16.76mmHg

(a) Assuming that systolic blood pressure is normally distributed, find the probability that:

- (i) a randomly selected person from this study has a systolic blood pressure of more than 140 mmHg,

(1)

- (ii) the mean of a randomly selected sample of 100 people from this study is less than 128mmHg.

(3)

Question 2 continued

One person in the study had a systolic blood pressure of 179mmHg.

- (b) Using a suitable calculation, determine whether 179mmHg should be considered an outlier.

(2)

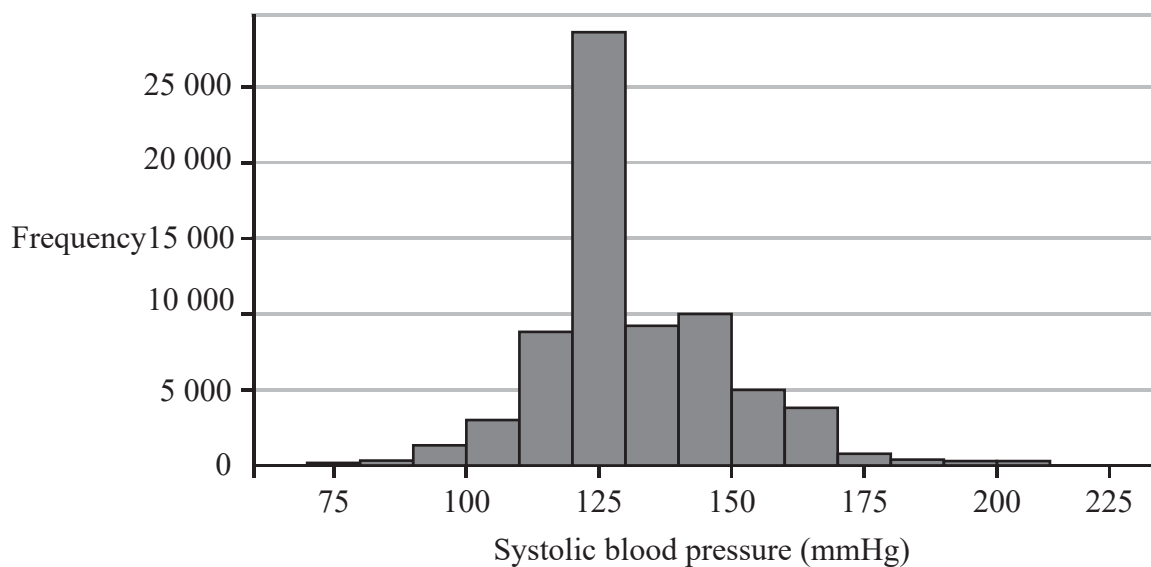
An investigation into the systolic blood pressure reading in (b) shows that the person had a heart condition which led to the high reading.

- (c) Explain whether or not the reading of 179mmHg should be removed from the data set. Give a reason for your answer.

(2)

Question 2 continued

Figure 1 shows the distribution of the systolic blood pressure for all 70 000 people in the study.



[Source: <https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset/data>]

Figure 1

- (d) Using **Figure 1**, give **one** criticism of the assumption that systolic blood pressure is normally distributed.

(1)

(Total for Question 2 is 9 marks)

DONOTWRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE

- 3 A large educational establishment wants to assess the usefulness of a particular mock examination paper, based on a comparison of results in that mock examination and results for the same students in the actual examination.

The establishment randomly samples a group of 12 students and matches their scores in the mock exam with their scores in the actual exam.

The results are shown in **Figure 2**

Student	Mock exam score	Actual exam score
A	48	46
B	97	87
C	70	69
D	79	65
E	43	16
F	53	40
G	87	71
H	66	61
I	72	55
J	92	67
K	94	86
L	98	97

[Source: <https://www.kaggle.com/datasets/nikhil7280/student-performance-multiple-linear-regression/data>]

Figure 2

- (a) Give **one** reason why it is appropriate to expect the correlation coefficient to be positive.

(1)

Question 3 continued

- (b) (i) Calculate the value of the product moment correlation coefficient for the data in **Figure 2**
- (ii) Interpret this value in context. (2)

A scatter diagram showing the 12 students' scores is shown in **Figure 3**

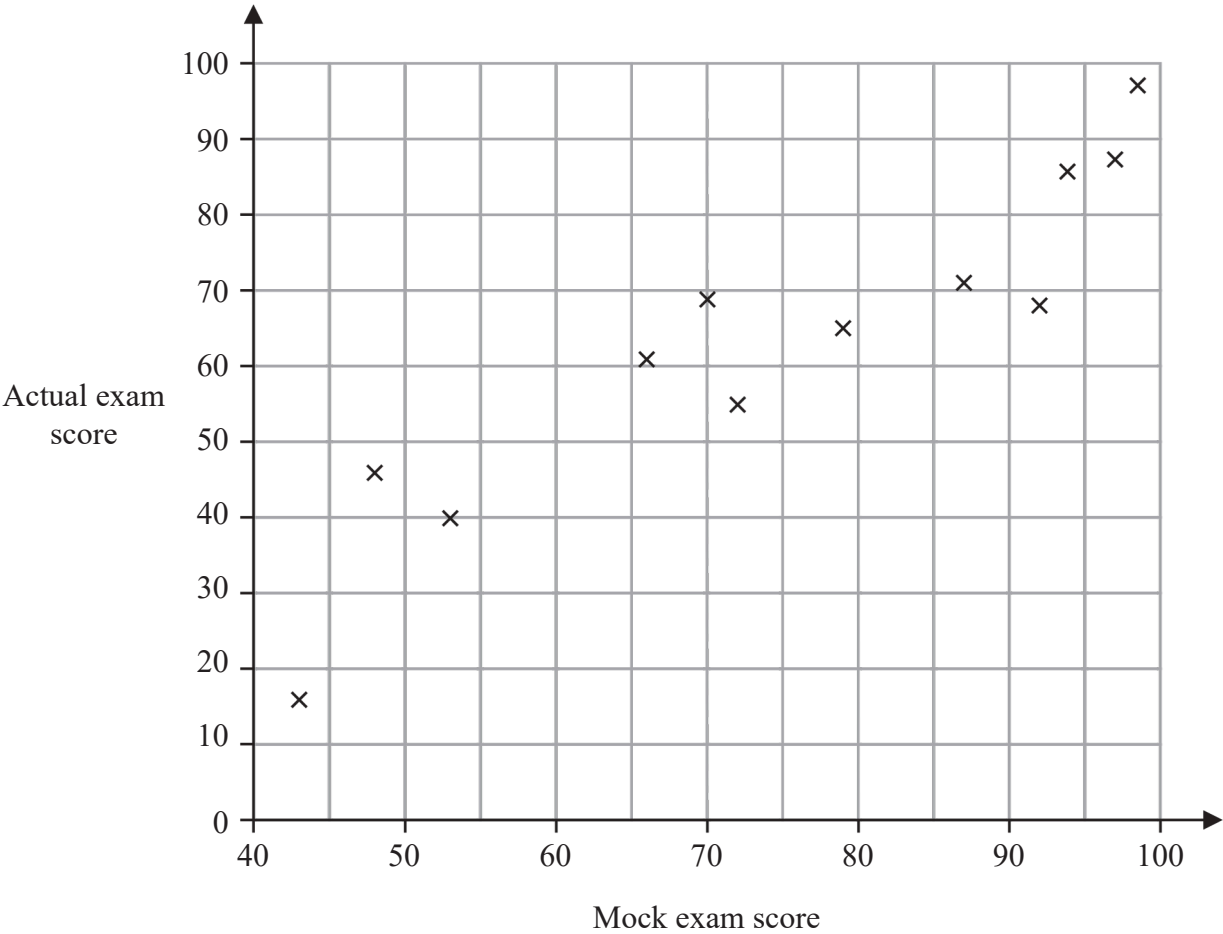


Figure 3

- (c) Give **one** reason why a scatter diagram of the data may be useful when selecting which correlation coefficient to use. (1)

Question 3 continued

The establishment wants to predict a student's actual exam score once it knows a student's mock score.

Using the data from the table, the establishment calculated the following equation of the least squares regression line, y on x

$$y = 15.63 + 1.05x$$

where y is the actual exam score and x is the mock exam score.

- (d) Use the regression line to estimate the actual exam score of a student who achieves a mock exam score of 60

(2)

- (e) Comment on the reliability of your estimate in (d)

(2)

- (f) Give an interpretation, in context, of the number 1.05 in the equation of the least squares regression line.

(2)

- (g) State why the regression equation cannot be used when a student scores less than 15 in the mock exam.

(1)

Question 3 continued

- (j) State an assumption in context that you need to make so that the probability in (i) is valid.

(1)

(Total for Question 3 is 18 marks)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DONOTWRITE IN THIS AREA

DONOTWRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE

- 4 In the 365 days of 2023 there were 148 earthquakes worldwide which were of magnitude 6 or more on the Richter scale.

[Source: <https://earthquake.usgs.gov/earthquakes/>]

It can be assumed that the number of earthquakes worldwide of magnitude 6 or more can be modelled by a Poisson distribution with mean 148 per year.

- (a) Calculate the standard deviation of the number of earthquakes worldwide of magnitude 6 or more using this model.

(1)

- (b) Find the probability that:

- (i) in a randomly chosen **one-year** period, there will be at least 150 earthquakes worldwide of magnitude 6 or more,

(1)

- (ii) in a randomly chosen **two-year** period, there will be no more than 260 earthquakes worldwide of magnitude 6 or more,

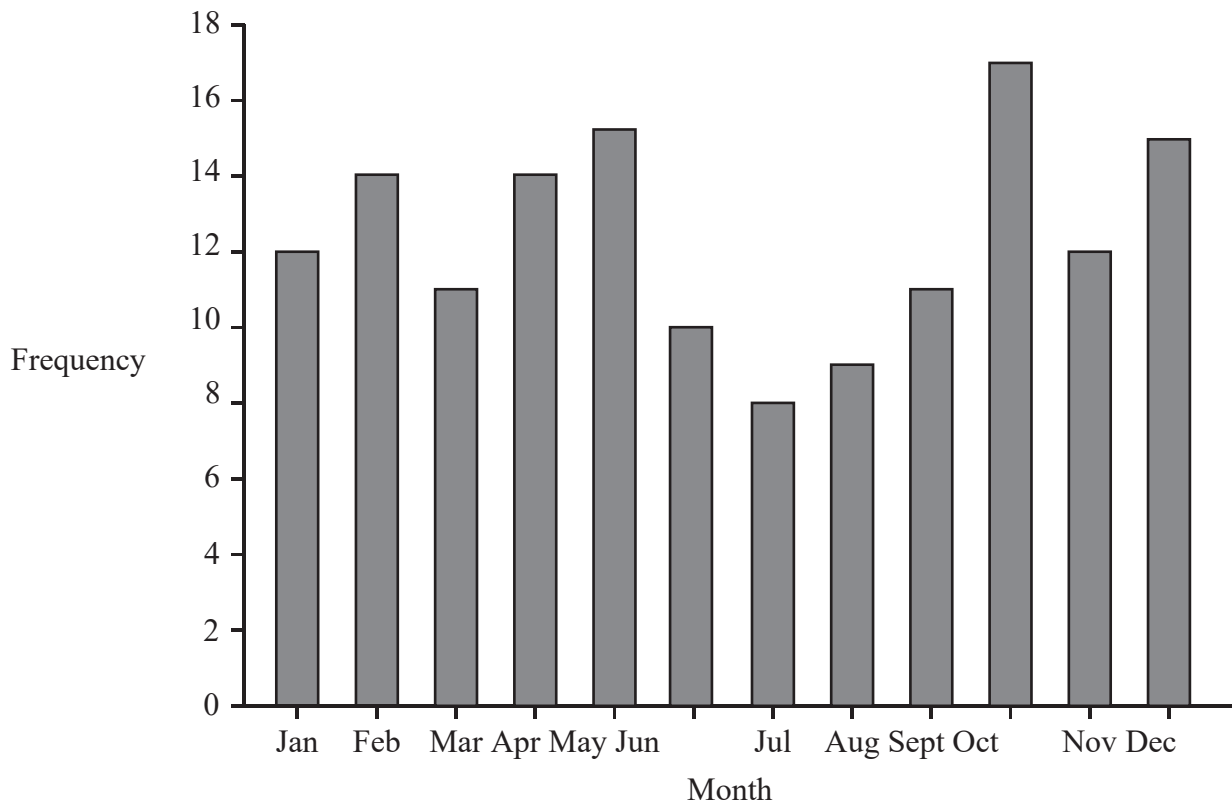
(2)

- (iii) there will be a gap of more than 6 days between the next pair of consecutive earthquakes worldwide of magnitude 6 or more.

(3)

Question 4 continued

Figure 4 shows the distribution of the earthquakes worldwide of magnitude 6 or more during 2023



[Source: <https://earthquake.usgs.gov/earthquakes/>]

Figure 4

A seismologist claims that in fact the Poisson distribution is not a good distribution to use to model the number of earthquakes worldwide of magnitude 6 or more.

(d) Comment on this claim using **Figure 4** and your knowledge of the Poisson distribution.

(4)

(Total for Question 4 is 14 marks)

- 5 The Human Resources team of a medium-sized company wants to evaluate the most recent round of training provided to all employees.

It is considering three different sampling methods to collect opinions from those that received training.

Method A

Hand a survey to all employees at the end of their last training session and ask them to complete and return it immediately.

Method B

Randomly select 50 of the employees using a simple random sampling technique and email them a survey immediately after the last training session.

Method C

Use a stratified sampling technique to randomly select 50 employees for a face-to-face interview in the week after the last training session. The sample will be selected so that the proportions of employees within each department of the company will be the same in the sample as in the company as a whole.

- (a) Explain the relative advantages and disadvantages of each of the three methods. Include within your answer a recommendation, with a reason, to the Human Resources team as to which to use.

(8)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DONOTWRITE IN THIS AREA

The Human Resources team decides to use Method C.

Figure 5 shows the number of employees that work in each department within the company.

Department	Number of employees
Sales	122
Operations	56
Customer Service	65
IT	42
Other	105
Total	390

Figure 5

Question 5 continued

(b) Calculate how many employees will be selected from the IT department using Method C.

(2)

One of the interview questions asks the employees to rate the statement below on a scale of 1 to 5 where 1 = “strongly disagree”, 2 = “disagree”, 3 = “neither agree nor disagree”, 4 = “agree” and 5 = “strongly agree”.

“The training provided me with the knowledge and skills necessary to perform my job more effectively.”

The mean of all the scores is 3.8

The company concludes that the majority of employees “agree” or “strongly agree” with the statement.

(c) State whether you agree with the company’s conclusion giving reasons for your answer.

(3)

(Total for Question 5 is 13 marks)

- 6 The UK Civil Aviation Authority reported that, between January 2023 and March 2023, there were 409 000 international flights scheduled to arrive into UK airports.

71 per cent of all 409 000 flights operated on time (within 15 minutes of scheduled arrival) and 6400 were cancelled.

[Source: <https://www.caa.co.uk/newsroom/news/cancelled-flights-and-flights-on-time-improve-at-uk-airports/>]

Helen is a frequent flyer and expects to book five flights into UK airports from international locations in the next year.

- (a) Using the information provided show that a suitable estimation for the probability of a flight being cancelled is 0.0156

(1)

- (b) Use suitable binomial distributions to find the probability that:

- (i) none of the flights that Helen books will be cancelled,

(1)

- (ii) at least one of the flights that Helen books will be cancelled,

(1)

- (iii) exactly two of the flights that Helen books will **not** operate on time.

(2)

Question 6 continued

- (c) Give **two** assumptions that need to be made in order for the binomial distributions used in (b) to be valid.

(2)

- (d) Explain why it would **not** be suitable for the binomial distributions in (b) to be approximated by the Normal distribution.

(1)

Question 6 continued

As part of the same report, the UK Civil Aviation Authority reported that Kenya Airways recorded the lowest on-time performance among airlines for the same months.

45 per cent of Kenya Airways scheduled flights arrived on time.

There were 180 flights from Kenya Airways that were scheduled to arrive into UK airports between January 2023 and March 2023

[Source: <https://www.caa.co.uk/newsroom/news/cancelled-flights-and-flights-on-time-improve-at-uk-airports/>]

One flight out of all the flights scheduled to arrive into UK airports in this period is chosen at random.

- (e) Calculate the probability that it is a Kenya Airways flight given that it was **not** on time.

(3)

(Total for Question 6 is 11 marks)

TOTAL FOR PAPER IS 80 MARKS