

Paper Reference 9MA0/31

Pearson Edexcel

Level 3 GCE

Mathematics

Advanced

PAPER 31: Statistics

Thursday 19 June 2025 – Afternoon

Time for 9MA0/31 and 9MA0/32: 2 hours

Question Paper

**THIS DOCUMENT MUST BE RETURNED TO
PEARSON AT THE END OF THE EXAMINATION.**

In the boxes below, write your name, centre number and candidate number.

Surname					
Other names					
Centre Number					
Candidate Number					

YOU MUST HAVE

**Mathematical Formulae and Statistical Tables (Green),
calculator, 9MA0/32**

YOU WILL BE GIVEN

Data Booklet

Answer Booklet

Candidates may use any calculator allowed by Pearson regulations. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulae stored in them.

INSTRUCTIONS

In the boxes on the Answer Booklet and on the Data Booklet, write 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 Answer Booklet or in the separate Data Booklet – there may be more space than you need.

Do NOT write on this Question Paper.

You should show sufficient working to make your methods clear.

Answers without working may not gain full credit.

**Values from statistical tables should be quoted in full.
If a calculator is used instead of tables the value should be given to an equivalent degree of accuracy.**

Inexact answers should be given to three significant figures unless otherwise stated.

Turn over

INFORMATION

A booklet ‘Mathematical Formulae and Statistical Tables’ is provided.

The total mark for this part of the examination is 50

There are 6 questions.

**The marks for EACH question are shown in brackets
– use this as a guide as to how much time to spend on each question.**

There may be spare copies of some diagrams.

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.

1. Bag **A** contains **5** red, **4** yellow and **3** green beads.

Bag **B** only contains red and yellow beads.

A bead is selected at random from bag **A** and a second bead is selected at random from bag **B**

Refer to the diagram for Question 1(a) in the Data Booklet.

It shows an incomplete tree diagram.

Given that the probability that both beads selected are yellow is $\frac{3}{16}$

- (a) (i) find the probability of selecting a yellow bead from bag **B**
(2 marks)

- (ii) Hence complete the tree diagram in the Data Booklet by finding the probability on each branch.
There are nine spaces to fill.
(2 marks)

(continued on the next page)

Turn over

1. continued.

(b) Find the exact probability that at least one of the two beads selected is yellow.

(3 marks)

The event **X** is that at least one of the beads selected is yellow.

The event **W** is that a green bead is selected.

(c) Find the exact value of $P(W|X)$

(2 marks)

(Total for Question 1 is 9 marks)

2. Runners in an athletics club can train with either coach **A** or coach **B** for the **400** metre race.

Coach **A** trains **120** runners for the **400** metre and records the best time, **x** seconds, for each runner.

The results are summarised by the following statistics

$$\sum x = 6612 \quad \sum x^2 = 364902$$

- (a) Calculate the mean of the best times for the runners trained by coach **A**
(1 mark)
- (b) Calculate the standard deviation of the best times for the runners trained by coach **A**
(2 marks)

(continued on the next page)

2. continued.

The mean and standard deviation for the best times of the **100** runners trained for the **400** metres by coach **B** are **55.1** seconds and **3.6** seconds respectively.

A **400** metre race consists of equal numbers of the fastest runners trained by coach **A** and by coach **B**

(c) State, giving a reason, which coach is more likely to have trained the winner.

(2 marks)

(Total for Question 2 is 5 marks)

- 3. A manufacturer makes components for the car industry.**

On average 1.5% of the components made are defective.

- (a) Using a suitable model, find the probability that in a random sample of 36 components there are**

(i) exactly 2 defective components,

(ii) more than 3 defective components.

(3 marks)

(continued on the next page)

3. continued.

Once every six months, the manufacturer tests whether the proportion of defective components has changed using hypotheses

$$H_0 : p = 0.015 \text{ and } H_1 : p \neq 0.015$$

A random sample of **260** components is taken and **8** are found to be defective.

(b) Using a **5%** level of significance, complete the test.

(3 marks)

(Total for Question 3 is 6 marks)

4. Kay is studying the variables Daily Total Sunshine (x) and Daily Total Rainfall (y) from the large data set for Leeming in 2015

Kay starts with 5th May and then selects every 10th day thereafter.

- (a) State the name of the sampling technique Kay uses.

(1 mark)

Kay wants to find the regression line of y on x for these data.

- (b) Using your knowledge of the large data set, explain how Kay might need to clean these data before finding the equation of the regression line.

(1 mark)

(continued on the next page)

4. continued.

The equation of the regression line Kay finds is

$$\mathbf{y = 0.741 + 0.199x}$$

(c) Using your knowledge of the large data set,

(i) state the units of the gradient of the regression line,

(ii) give an interpretation of the y -intercept of the regression line.

(2 marks)

Kay's teacher claimed that the greater the amount of sunshine in a day the lower the amount of rain there should be.

(d) State, giving a reason, whether the teacher's claim is true for Kay's data.

(1 mark)

(continued on the next page)

4. continued.

The teacher used all the data for these variables from the large data set for Leeming in 2015, as a sample.

The teacher calculated the product moment correlation coefficient for x and y to be -0.160

In a suitable test to determine whether there is evidence to support the teacher's claim, the p -value was 0.015

(e) Using a 5% level of significance, state the hypotheses and conclusion for this test.
(2 marks)

(f) For the test in part (e) describe

(i) the sample,

(ii) a possible population.

(2 marks)

(Total for Question 4 is 9 marks)

5. The heights of men in a tennis club are normally distributed with a mean of **183 cm** and a standard deviation of **4.9 cm**

A man from the tennis club is selected at random.

(a) Find the probability that the man's height is

(i) more than **186 cm**,

(ii) between **175 cm** and **185 cm**

(2 marks)

(continued on the next page)

5. continued.

The heights of women in the tennis club are normally distributed with a mean of μ cm and a standard deviation of σ cm

Given that 40% of the women are shorter than 170 cm and 15% are taller than 175 cm,

(b) find the value of μ and the value of σ

Show your working clearly.

(5 marks)

A man from the tennis club and a woman from the tennis club are selected at random.

(c) Find the probability that both players have heights between 170 cm and 175 cm

(3 marks)

(Total for Question 5 is 10 marks)

6. The discrete random variable **R** takes even integer values from **2** to **2n** inclusive.

The probability distribution of **R** is given by

$$P(R = r) = \frac{r}{k} \quad r = 2, 4, 6, \dots, 2n$$

where **k** is a constant.

- (a) Show that

$$k = n(n + 1)$$

(4 marks)

When

$$n = 20$$

- (b) find the exact value of

$$P(16 \leq R < 26)$$

(2 marks)

(continued on the next page)

6. continued.

When

$n = 20$, a random value g of R is taken and the quadratic equation in x

$$x^2 + gx + 3g = 5$$

is formed.

(c) Find the exact probability that the equation has no real roots.

(5 marks)

(Total for Question 6 is 11 marks)

TOTAL FOR STATISTICS IS 50 MARKS

END OF PAPER