

**Paper Reference 9MA0/31
Pearson Edexcel
Level 3 GCE**

Mathematics

Advanced

PAPER 31: Statistics

Tuesday 20 June 2023 – Afternoon

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

YOU MUST HAVE

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

YOU WILL BE GIVEN

Data Booklet

Answer Booklet

Y72819A

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.

(continued on the next page)

Turn over

Instructions continued.

Answer the questions in the Answer Booklet – there may be more space than you need.

Do NOT write on the 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.

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. Refer to the diagram for Question 1 in the Data Booklet.**

It is a Venn Diagram.

The Venn diagram, where p and q are probabilities, shows the three events **A, **B** and **C** and their associated probabilities.**

- (a) Find $P(A)$**
(1 mark)

(continued on the next page)

1. continued.

The events **B** and **C** are independent.

(b) Find the value of **p** and the value of **q**

(3 marks)

(c) Find $P(A|B')$

(2 marks)

(Total for Question 1 is 6 marks)

Turn over

2. A machine fills packets with sweets and $\frac{1}{7}$ of the packets also contain a prize.

The packets of sweets are placed in boxes before being delivered to shops.

There are 40 packets of sweets in each box.

The random variable T represents the number of packets of sweets that contain a prize in each box.

(continued on the next page)

Turn over

2. continued.

(a) State a condition needed for T to be modelled by

$$\mathbf{B(40, \frac{1}{7})}$$

(1 mark)

(continued on the next page)

Turn over

2. continued.

A box is selected at random.

(b) Using $T \sim B(40, \frac{1}{7})$ find

**(i) the probability that the
box has exactly 6 packets
containing a prize,**

**(ii) the probability that the box
has fewer than 3 packets
containing a prize.**

(2 marks)

(continued on the next page)

Turn over

2. continued.

Kamil's sweet shop buys 5 boxes of these sweets.

(c) Find the probability that exactly 2 of these 5 boxes have fewer than 3 packets containing a prize.

(2 marks)

(continued on the next page)

Turn over

2. continued.

Kamil claims that the proportion of packets containing a prize is less than $\frac{1}{7}$

A random sample of 110 packets is taken and 9 packets contain a prize.

(continued on the next page)

Turn over

2. continued.

**(d) Use a suitable test to assess
Kamil's claim.**

You should

- **state your hypotheses clearly**
- **use a 5% level of significance**

(4 marks)

(Total for Question 2 is 9 marks)

Turn over

- 3. Refer to the table for Question 3 in the Data Booklet.**

Ben is studying the Daily Total Rainfall, x mm, in Leeming for 1987

He used all the data from the large data set and summarised the information in the table in the Data Booklet.

- (a) Explain how the data will need to be cleaned before Ben can start to calculate statistics such as the mean and standard deviation.**
- (2 marks)**

3. continued.

Using all 184 of these values, Ben estimates

$$\sum x = 390 \quad \text{and} \quad \sum x^2 = 4336$$

(b) Calculate estimates for

(i) the mean Daily Total Rainfall,

**(ii) the standard deviation of the
Daily Total Rainfall.**

(3 marks)

(continued on the next page)

Turn over

3. continued.

**Ben suggests using the statistic
calculated in part (b)(i) to estimate
the annual mean Daily Total Rainfall
in Leeming for 1987**

(continued on the next page)

Turn over

3. continued.

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

(i) give a reason why these data would not be suitable,

(ii) state, giving a reason, how you would expect the estimate in part (b)(i) to differ from the actual annual mean Daily Total Rainfall in Leeming for 1987

(2 marks)

(Total for Question 3 is 7 marks)

Turn over

4. A study was made of adult men from region **A** of a country.

It was found that their heights were normally distributed with a mean of **175.4 cm** and standard deviation **6.8 cm**

- (a) Find the proportion of these men that are taller than **180 cm**
(1 mark)

(continued on the next page)

4. continued.

A student claimed that the mean height of adult men from region B of this country was different from the mean height of adult men from region A

A random sample of 52 adult men from region B had a mean height of 177.2 cm

The student assumed that the standard deviation of heights of adult men was 6.8 cm both for region A and region B

(continued on the next page)

Turn over

4. continued.

(b) Use a suitable test to assess the student's claim.

You should

- **state your hypotheses clearly**
- **use a 5% level of significance**

(4 marks)

(c) Find the p-value for the test in part (b)

(1 mark)

(Total for Question 4 is 6 marks)

Turn over

5. Refer to the table for Question 5 in the Data Booklet.

Tisam is playing a game.

She uses a ball, a cup and a spinner.

The random variable X represents the number the spinner lands on when it is spun.

The probability distribution of X is given in the table in the Data Booklet where a , b , c and d are probabilities.

(continued on the next page)

Turn over

5. continued.

To play the game

- **the spinner is spun to obtain a value of X**
- **Tisam then stands X cm from the cup and tries to throw the ball into the cup**

(continued on the next page)

Turn over

5. continued.

The event **S** represents the event that Tisam successfully throws the ball into the cup.

To model this game Tisam assumes that

- $P(S | \{X = x\}) = \frac{k}{x}$ where **k** is a constant
- $P(S \cap \{X = x\})$ should be the same whatever value of **X** is obtained from the spinner

(continued on the next page)

Turn over

5. continued.

Using Tisam's model,

(a) show that

$$c = \frac{8}{5}b$$

(2 marks)

**(b) find the probability distribution
of X**

(5 marks)

(continued on the next page)

Turn over

5. continued.

**Nav tries, a large number of times,
to throw the ball into the cup from a
distance of 100 cm**

**He successfully gets the ball in the
cup 30% of the time.**

- (c) State, giving a reason, why
Tisam's model of this game is not
suitable to describe Nav playing
the game for all values of X
(1 mark)**

(Total for Question 5 is 8 marks)

Turn over

- 6. Refer to the diagram for Question 6 in the Data Booklet.**

It is a histogram.

A medical researcher is studying the number of hours, T , a patient stays in hospital following a particular operation.

The histogram in the Data Booklet summarises the results for a random sample of 90 patients.

- (a) Use the histogram to estimate**
 $P(10 < T < 30)$
(2 marks)

6. continued.

For these 90 patients the time spent in hospital following the operation had

- a mean of 14.9 hours**
- a standard deviation of 9.3 hours**

Tomas suggests that T can be modelled by
 $N(14.9, 9.3^2)$

(b) With reference to the histogram, state, giving a reason, whether or not Tomas' model could be suitable.

(1 mark)

(continued on the next page)

Turn over

6. continued.

Xiang suggests that the frequency polygon based on this histogram could be modelled by a curve with equation

$$y = kxe^{-x} \quad 0 \leq x \leq 4$$

where

- **X is measured in tens of hours**
- **k is a constant**

(continued on the next page)

Turn over

6. continued.

(c) Use algebraic integration to show that

$$\int_0^n x e^{-x} dx = 1 - (n + 1)e^{-n}$$

(4 marks)

**(d) Show that, for Xiang's model,
 $k = 99$ to the nearest integer.**

(3 marks)

(continued on the next page)

Turn over

6. continued.

(e) Estimate

$P(10 < T < 30)$ using

(i) Tomas' model of

$$T \sim N(14.9, 9.3^2)$$

(1 mark)

(ii) Xiang's curve with equation

**$y = 99xe^{-x}$ and the answer
to part (c)**

(2 marks)

(continued on the next page)

Turn over

6. continued.

**The researcher decides to use
Xiang's curve to model
 $P(a < T < b)$**

- (f) State one limitation of Xiang's
model.
(1 mark)**

(Total for Question 6 is 14 marks)

TOTAL FOR STATISTICS IS 50 MARKS

END OF PAPER
