Mathematics
Advanced
PAPER 31: Statistics

Time: 2 hours

YOU MUST HAVE
Mathematical Formulae and Statistical Tables (Green), calculator

YOU WILL BE GIVEN
Data Booklet
Answer Booklet

Pearson

Candidates may use any calculator permitted 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 on the separate data sheets ‒ 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.

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 data sheets in case you need them.

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. George throws a ball at a target **15** times.

   Each time George throws the ball, the probability of the ball hitting the target is **0·48**

   The random variable **X** represents the number of times George hits the target in **15** throws.

   (a) Find

         (i) $P(X = 3)$

         (ii) $P(X \geq 5)$

   (3 marks)

   George now throws the ball at the target **250** times.

   (b) Use a normal approximation to calculate the probability that he will hit the target more than **110** times.

   (3 marks)

(Total for Question 1 is 6 marks)

Turn over

2. A manufacturer uses a machine to make metal rods.

The length of a metal rod, L cm, is normally distributed with

* a mean of **8 cm**
* a standard deviation of x **cm**

Given that the proportion of metal rods less than **7·902 cm** in length is **2·5%**

(a) show that $x = 0·05$ to **2** decimal places.

(2 marks)

(b) Calculate the proportion of metal rods that are between **7·94 cm** and **8·09 cm** in length.

(1 mark)

2. continued.

The COST of producing a single metal rod is

20 pence

A metal rod

- where **L < 7·94** is SOLD for scrap for **5** pence
- where **7·94 ≤ L ≤ 8·09** is SOLD for **50** pence
- where **L > 8·09** is shortened for an extra COST of **10** pence and then SOLD for **50** pence

(c) Calculate the expected profit per **500** of the metal rods.

Give your answer to the nearest pound.

(5 marks)

(continued on the next page)

2.   continued.

The same manufacturer makes metal hinges in large batches.

The hinges each have a probability of $0 \cdot 015$ of having a fault.

A random sample of 200 hinges is taken from each batch and the batch is accepted if fewer than 6 hinges are faulty.

The manufacturer's aim is for 95% of batches to be accepted.

(d)   Explain whether the manufacturer is likely to achieve its aim.

(4 marks)

(Total for Question 2 is 12 marks)

3. **Dian uses the large data set to investigate the Daily Total Rainfall, r mm, for Camborne.**

    (a) **Write down how a value of $0 < r \leq 0 \cdot 05$ is recorded in the large data set.**

    (1 mark)

    Dian uses the data for the 31 days of August 2015 for Camborne and calculates the following statistics

    $n = 31$

    $\sum r = 174 \cdot 9$

    $\sum r^2 = 3523 \cdot 283$

    (b) **Use these statistics to calculate**

        (i) **the mean of the Daily Total Rainfall in Camborne for August 2015,**

        (ii) **the standard deviation of the Daily Total Rainfall in Camborne for August 2015**

    (3 marks)

(continued on the next page)

3. **continued.**

Dian believes that the mean Daily Total Rainfall in August is less in the South of the UK than in the North of the UK.

The mean Daily Total Rainfall in Leuchars for August **2015** is **1·72 mm** to **2** decimal places.

(c) State, giving a reason, whether this provides evidence to support Dian's belief.

(2 marks)

Dian uses the large data set to estimate the proportion of days with no rain in Camborne for **1987** to be **0·27** to **2** decimal places.

(d) Explain why the distribution $B(14, 0·27)$ might NOT be a reasonable model for the number of days without rain for a **14**−day summer event.

(1 mark)

(Total for Question 3 is 7 marks)

4. A dentist knows from past records that **10%** of customers arrive late for their appointment.

A new manager believes that there has been a change in the proportion of customers who arrive late for their appointment.

A random sample of **50** of the dentist's customers is taken.

(a) Write down

- a null hypothesis corresponding to no change in the proportion of customers who arrive late
- an alternative hypothesis corresponding to the manager's belief

(1 mark)

(continued on the next page)

4. continued.

(b) Using a **5%** level of significance, find the critical region for a two−tailed test of the null hypothesis in (a)

You should state the probability of rejection in each tail, which should be less than **0·025**

(3 marks)

(c) Find the actual level of significance of the test based on your critical region from part (b)

(1 mark)

The manager observes that **15** of the **50** customers arrived late for their appointment.

(d) With reference to part (b), comment on the manager's belief.

(1 mark)

(Total for Question 4 is 6 marks)

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

   A company has 1825 employees.

   The employees are classified as professional, skilled or elementary.

   The table in the Data Booklet shows

   • **the number of employees in each classification**
   • **the two areas, A or B, where the employees live**

   An employee is chosen at random.

   Find the probability that this employee

   (a) **is skilled,**

   (1 mark)

   (b) **lives in area B and is not a professional.**

   (1 mark)

(continued on the next page)

5.   continued.

Refer to the diagram for Question 5(c) in the
Data Booklet.

It shows an incomplete Venn diagram.

Some classifications of employees are more likely
to work from home.

- **65%** of professional employees in both area **A**
  and area **B** work from home
- **40%** of skilled employees in both area **A** and
  area **B** work from home
- **5%** of elementary employees in both area **A** and
  area **B** work from home
- Event **F** is that the employee is a professional
- Event **H** is that the employee works from home
- Event **R** is that the employee is from area **A**

(c)   Using this information, complete the Venn
      diagram in the Data Booklet.

      (4 marks)

5.   **continued.**

   (d)   Find $P(R' \cap F)$

   (1 mark)

   (e)   Find $P([H \cup R]')$

   (1 mark)

   (f)   Find $P(F \mid H)$

   (2 marks)

(Total for Question 5 is 10 marks)

Turn over

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

It shows a scatter diagram.

Anna is investigating the relationship between exercise and resting heart rate.

She takes a random sample of 19 people in her year at school and records for each person

- their resting heart rate, h beats per minute
- the number of minutes, m, spent exercising each week

Her results are shown on the scatter diagram in the Data Booklet.

(a) Interpret the nature of the relationship between h and m

(1 mark)

(continued on the next page)

6. continued.

Anna codes the data using the formulae

$$x = \log_{10} m$$
$$y = \log_{10} h$$

The product moment correlation coefficient between

x and y is $-0\cdot897$

(b) Test whether or not there is significant evidence

of a negative correlation between x and y

You should

- state your hypotheses clearly
- use a **5%** level of significance
- state the critical value used

(3 marks)

Turn over

6. **continued.**

The equation of the line of best fit of y on x is

$$y = -0 \cdot 05x + 1 \cdot 92$$

(c) Use the equation of the line of best fit of y on x to find a model for h on m in the form

$$h = am^k$$

where a and k are constants to be found.

(5 marks)

(Total for Question 6 is 9 marks)

TOTAL FOR STATISTICS IS 50 MARKS

END OF PAPER