Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel Level 3 GCE

Centre Number | Candidate Number

Paper reference **8MA0/21**

Mathematics

Advanced Subsidiary
PAPER 21: Statistics

You must have:
Mathematical Formulae and Statistical Tables, calculator

Total Marks

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

- 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.
- 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 30. There are 5 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.
- Good luck with your examination.

Turn over ▶

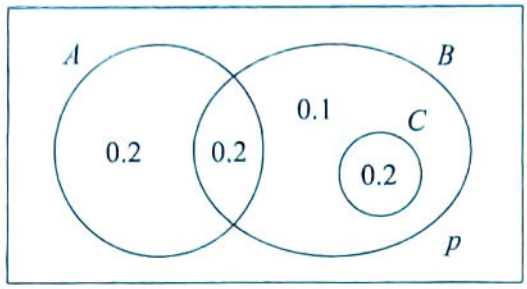
P66806A

1/1/1/1/1/

P 6 6 8 0 6 A 0 1 1 6

Pearson

1.



The Venn diagram, where p is a probability, shows the 3 events A, B and C with their associated probabilities.

(a) Find the value of p.

(1)

(b) Write down a pair of mutually exclusive events from A, B and C.

(1)

a) $0.2 + 0.2 + 0.1 + 0.2 + p = 1$

$\Rightarrow p = 1 - 0.7 \Rightarrow p = 0.3$.

b) A and C are mutually exclusive.

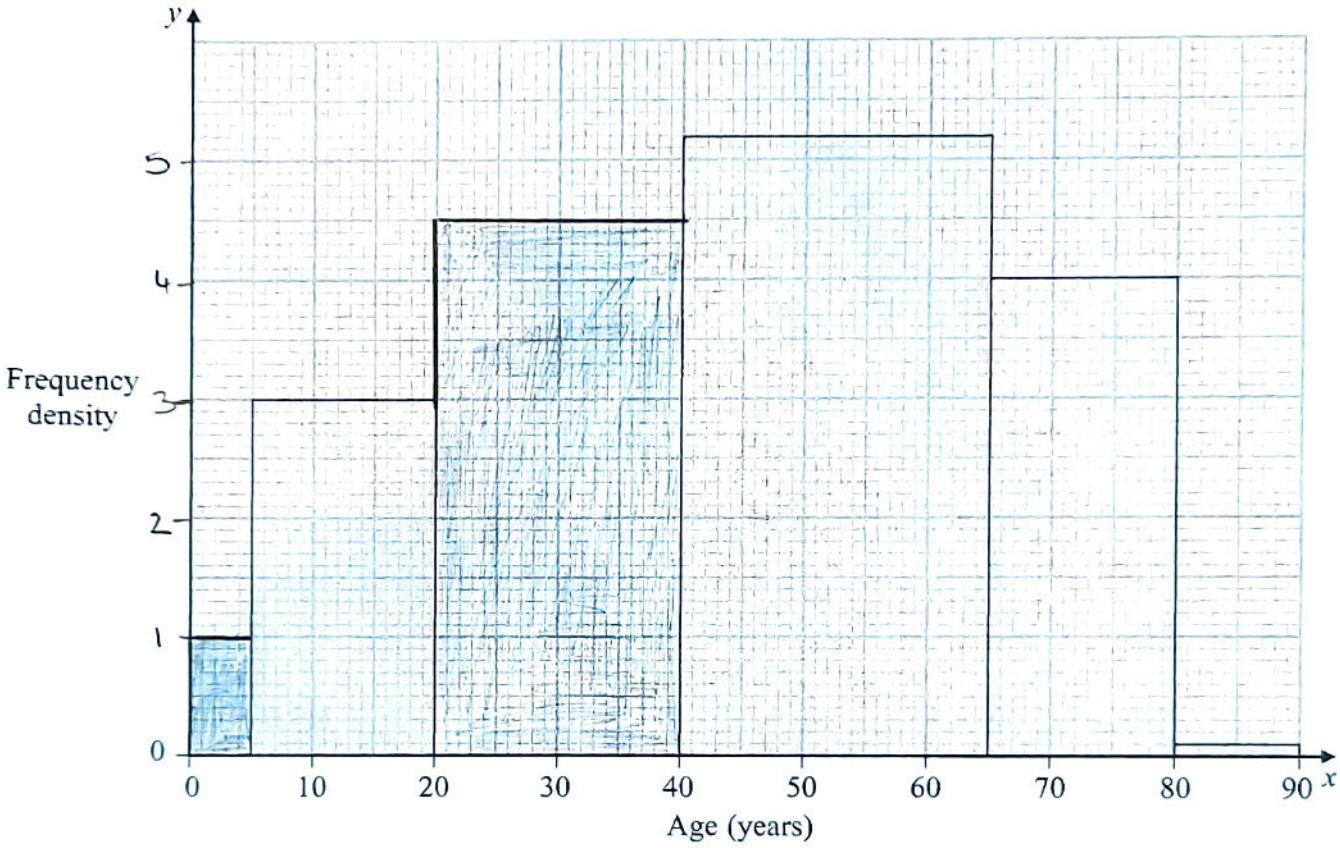
Question 1 continued

(Total for Question 1 is 2 marks)

2. The partially completed table and partially completed histogram give information about the ages of passengers on an airline.

There were no passengers aged 90 or over.

Age (x years)	$0 \leqslant x < 5$	$5 \leqslant x < 20$	$20 \leqslant x < 40$	$40 \leqslant x < 65$	$65 \leqslant x < 80$	$80 \leqslant x < 90$
Frequency	5	45	90	130	60	1



(a) Complete the histogram.

(3)

(b) Use linear interpolation to estimate the median age.

(4)

An outlier is defined as a value greater than $Q_3 + 1.5 \times$ interquartile range.

Given that $Q_1 = 27.3$ and $Q_3 = 58.9$

(c) determine, giving a reason, whether or not the oldest passenger could be considered as an outlier.

(2)

Question 2 continued

a) $5 \leq x < 20$: has a frequency density $= \frac{45}{15} = 3$.

So $0 \leq x < 5$ has $fd = \frac{5}{5} = 1$

$20 \leq x < 40$ has $fd = \frac{96}{20} = 4.5$.

b) $40 \leq x < 65$ has $fd = 5.2 = \frac{f}{25}$, $f = 130$.

$65 \leq x < 80$ has $fd = 4 = \frac{f}{15}$, $f = 60$.

Total passengers $= 331$.

$$\text{Median} = 60 + \frac{331 \times 0.5 - 140}{130} \times 25$$

$$= 44.9038\ldots$$

c) Interquartile range $= 58.9 - 27.3 = 31.6$.

Outlier $= 58.9 + 1.5 \times 31.6 = 106.3$

No passenger is over 90, so the oldest passenger is not an outlier.

Question 2 continued

Question 2 continued

(Total for Question 2 is 9 marks)

3. Helen is studying one of the qualitative variables from the large data set for Heathrow from 2015.

She started with the data from 3rd May and then took every 10th reading.

There were only 3 different outcomes with the following frequencies

Outcome	*A*	*B*	*C*
Frequency	16	2	1

(a) State the sampling technique Helen used.

(1)

(b) From your knowledge of the large data set

(i) suggest which variable was being studied,

(ii) state the name of outcome *A*.

(2)

George is also studying the same variable from the large data set for Heathrow from 2015. He started with the data from 5th May and then took every 10th reading and obtained the following

Outcome	*A*	*B*	*C*
Frequency	16	1	1

Helen and George decided they should examine all of the data for this variable for Heathrow from 2015 and obtained the following

Outcome	*A*	*B*	*C*
Frequency	155	26	3

(c) State what inference Helen and George could reliably make from their original samples about the outcomes of this variable at Heathrow, for the period covered by the large data set in 2015.

(1)

a) Systematic sampling.

bi) Daily mean wind speed.

Question 3 continued

b ii) Light

c) Variable A occurs the most at around 84% of the time.

(Total for Question 3 is 4 marks)

4. A nursery has a sack containing a large number of coloured beads of which 14% are coloured red.

Aliya takes a random sample of 18 beads from the sack to make a bracelet.

(a) State a suitable binomial distribution to model the number of red beads in Aliya's bracelet.

(1)

(b) Use this binomial distribution to find the probability that

(i) Aliya has just 1 red bead in her bracelet,

(ii) there are at least 4 red beads in Aliya's bracelet.

(3)

(c) Comment on the suitability of a binomial distribution to model this situation.

(1)

After several children have used beads from the sack, the nursery teacher decides to test whether or not the proportion of red beads in the sack has changed.
She takes a random sample of 75 beads and finds 4 red beads.

(d) Stating your hypotheses clearly, use a 5% significance level to carry out a suitable test for the teacher.

(4)

(e) Find the p-value in this case.

(1)

a) Let R = number of red beads in Aliya's bracelet.

$R \sim B(18, 0.14)$.

bi) $P(R=1) = 0.19403\ldots = 0.194$ (3dp).

ii) $P(R \geqslant 4) = 1 - P(R \leqslant 3) = 1 - 0.76184\ldots$

$= 0.238$ (3dp)

c) It assumes that $p = 0.14$ does not change. This means the sack must have enough beads so that removing 18 at random does not appreciably affect the probability. If the sack does, then the model is suitable.

d) Null hypothesis: $H_0: p = 0.14$.

$H_1: p \neq 0.14$.

X = Number of beads in the sample.

$X \sim B(75, 0.14)$.

$P(X \leq 4) = 0.01506 < 0.025$

So significant enough to reject H_0.

There is evidence that the proportion of red beads has changed.

e) p-value $= 2 \times 0.01506 = 0.03012...$

Question 4 continued

Question 4 continued

(Total for Question 4 is 10 marks)

5. Two bags, **A** and **B**, each contain balls which are either red or yellow or green.

Bag **A** contains 4 red, 3 yellow and n green balls.
Bag **B** contains 5 red, 3 yellow and 1 green ball.

A ball is selected at random from bag **A** and placed into bag **B**.
A ball is then selected at random from bag **B** and placed into bag **A**.

The probability that bag **A** now contains an equal number of red, yellow and green balls is p.

Given that $p > 0$, find the possible values of n and p.

(5)

For bag A to end with an equal number of each colour it can either lose a red ball and gain a green ball, in which case $n=2$ or it can lose a green ball and gain a yellow ball, in which case $n=5$.

For $n=2$. A red is selected from A and a green selected from B.

$P(\text{Red}_A) = \frac{4}{4+3+2} = \frac{4}{9}$, $P(\text{Green}_B) = \frac{1}{6+3+1} = \frac{1}{10}$

(There would be 6 red balls in B after taken from A).

$P(\text{Red}_A \text{ and } \text{Green}_B) = \frac{4}{9} \times \frac{1}{10} = \frac{2}{45}$.

For $n=5$. A green taken from A and a yellow from B.

$P(\text{Green}_A) = \frac{5}{4+3+5} = \frac{5}{12}$, $P(\text{Yellow}_B) = \frac{3}{5+3+2} = \frac{3}{10}$

$P(\text{Green}_A \text{ and } \text{Yellow}_B) = \frac{5}{12} \times \frac{3}{10} = \frac{1}{8}$.

Possibilities: $n=2$ and $p=2/45$ or $n=5$ and $p=1/8$.

Question 5 continued

Question 5 continued

(Total for Question 5 is 5 marks)

TOTAL FOR STATISTICS IS 30 MARKS