MODEL ANSWERS

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel Level 3 GCE

Centre Number	Candidate Number

Wednesday 20 May 2020

Afternoon	Paper Reference **8MA0/21**

Mathematics
Advanced Subsidiary
Paper 21: Statistics

You must have:
Mathematical Formulae and Statistical Tables (Green), 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.

Turn over ▶

P63935A

1/1/1/1/

P 6 3 9 3 5 A 0 1 1 6

Pearson

1.

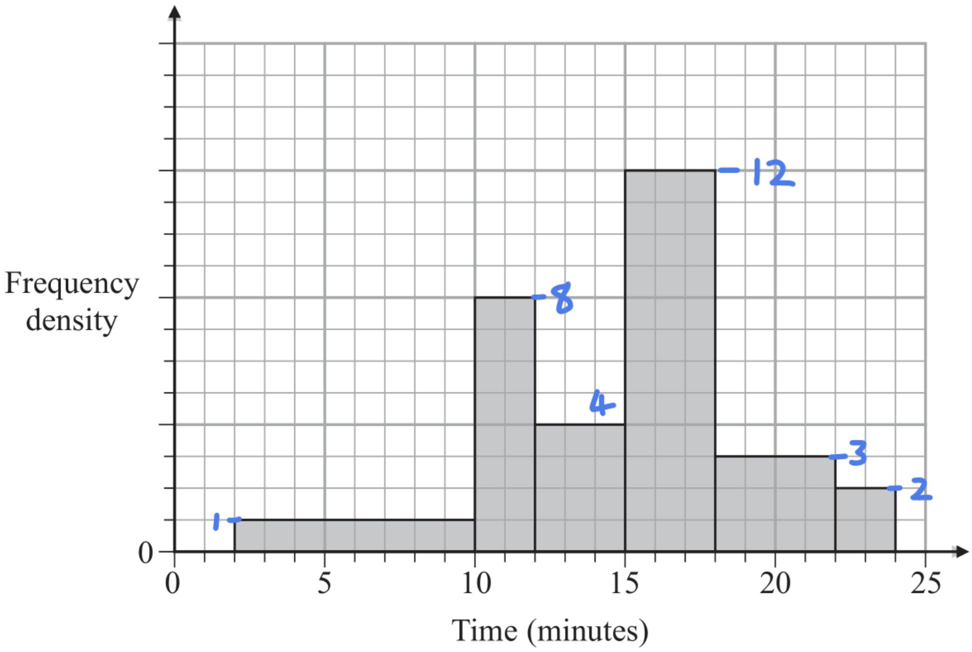

Figure 1

The histogram in Figure 1 shows the times taken to complete a crossword by a random sample of students.

The number of students who completed the crossword in more than 15 minutes is 78

Estimate the percentage of students who took less than 11 minutes to complete the crossword.

(4)

area above 15 represents 78 students:

$(12\times3+3\times4+2\times2)$ squares $\times k = 78$ students

area scale

$\therefore$ one square represents $k = \dfrac{78}{52}$

$= 1.5$

area below $11 = 8\times1 + 1\times8 = 16$

so $16\times1.5 = 24$ students below 11 minutes

BUT we are asked for a percentage so we need the total

Question 1 continued

number of students who did the crossword:

$$\underbrace{78+24}_{\text{save time - don't recalculate}} + \underbrace{1.5(1\times 8 + 3\times 4)}_{\text{scale}\times\text{no. squares}} = 132 \text{ students}$$

$$\Rightarrow \text{percentage} = \frac{24}{132}\times 100$$

$$= 18.181\ldots\%$$

$$\approx 18\%$$

(Total for Question 1 is 4 marks)

2. Jerry is studying visibility for Camborne using the large data set June 1987.

The table below contains two extracts from the large data set.

It shows the daily maximum relative humidity and the daily mean visibility.

Date	Daily Maximum Relative Humidity	Daily Mean Visibility
Units	%	
10/06/1987	90	5300
28/06/1987	100	0

(The units for Daily Mean Visibility are deliberately omitted.)

Given that daily mean visibility is given to the nearest 100,

(a) write down the range of distances in metres that corresponds to the recorded value 0 for the daily mean visibility.

(1)

Jerry drew the following scatter diagram, Figure 2, and calculated some statistics using the June 1987 data for Camborne from the large data set.

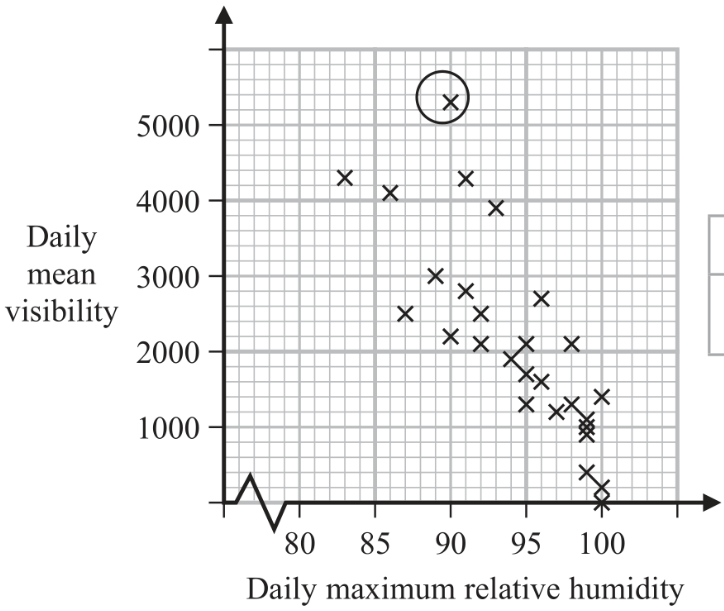

Figure 2

	Q_1	IQR
Daily mean visibility	1100	1600
Daily maximum relative humidity (%)	92	8

Jerry defines an outlier as a value that is more than 1.5 times the interquartile range above Q_3 or more than 1.5 times the interquartile range below Q_1.

(b) Show that the point circled on the scatter diagram is an outlier for visibility.

(2)

(c) Interpret the correlation between the daily mean visibility and the daily maximum relative humidity.

(1)

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

Jerry drew the following scatter diagram, Figure 3, using the June 1987 data for Camborne from the large data set, but forgot to label the x–axis.

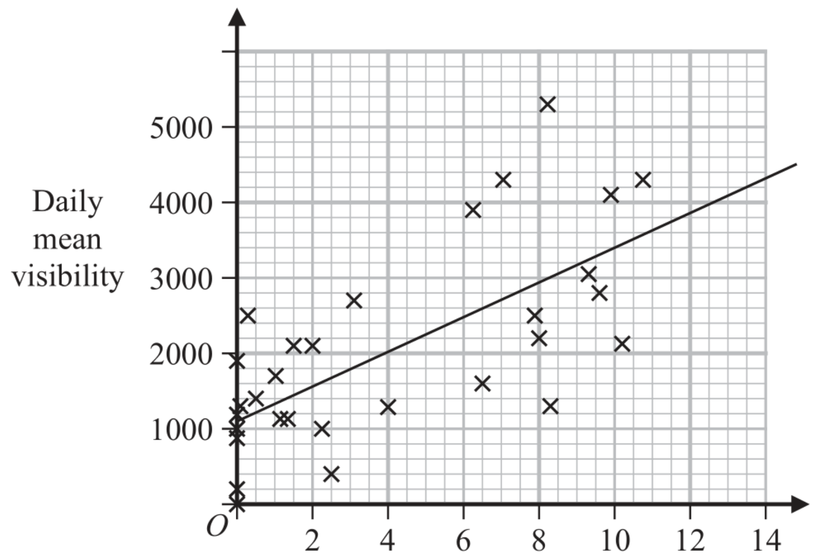

Figure 3

(d) Using your knowledge of the large data set, suggest which variable the x-axis on this scatter diagram represents.

familiarise yourself with how the quantities in the set are measured

(1)

a) rounded to nearest hundred decametres

any value 0-50dm rounds down to 0 ⇒ 0 to 500m

b) data point: (90, 5300)

value at which point becomes an outlier:

$Q_3 = Q_1 + IQR = 1100 + 1600 = 2700\text{m}$

$Q_3 + 1.5IQR = 2700 + 1.5 \times 1600 = 5100\text{m}$

$5300 > 5100 \Rightarrow$ outlier

c) as the humidity increases, the mean visibility decreases

d) we need something that increases visibility

x = hours of sunshine

Question 2 continued

Question 2 continued

(Total for Question 2 is 5 marks)

Turn over

3. In a game, a player can score 0, 1, 2, 3 or 4 points each time the game is played.

The random variable S, representing the player's score, has the following probability distribution where a, b and c are constants.

s	0	1	2	3	4
$\mathrm{P}(S = s)$	a	b	c	0.1	0.15

The probability of scoring less than 2 points is twice the probability of scoring at least 2 points.

Each game played is independent of previous games played.

John plays the game twice and adds the two scores together to get a total.

Calculate the probability that the total is 6 points.

(6)

$P(S<2) = 2 \times P(S \geq 2) \Rightarrow a+b = 2(c+0.1+0.15)$

$a+b = 2c+0.5$ ①

total probability $= 1 \Rightarrow a+b+c+0.1+0.15 = 1$

$a+b+c = 0.75$ ②

sub in ①: $2c+0.5+c = 0.75$

$3c = 0.25$

$c = \frac{1}{12}$

ways to score 6: $2+4$, $4+2$, $3+3$

↳ so we don't need a or b

$P(T.S. = 6) = \underbrace{\frac{1}{12} \times 0.15}_{P(2,4)} + \underbrace{0.15 \times \frac{1}{12}}_{P(4,2)} + \underbrace{0.1^2}_{P(3,3)}$

$= 0.035$

Question 3 continued

(Total for Question 3 is 6 marks)

4. A lake contains three different types of carp.

There are an estimated 450 mirror carp, 300 leather carp and 850 common carp.

Tim wishes to investigate the health of the fish in the lake.

He decides to take a sample of 160 fish.

(a) Give a reason why stratified random sampling cannot be used.

(1)

(b) Explain how a sample of size 160 could be taken to ensure that the estimated populations of each type of carp are fairly represented.

You should state the name of the sampling method used.

(2)

As part of the health check, Tim weighed the fish.

His results are given in the table below.

Weight (w kg)	Frequency (f)	Midpoint (m kg)
$2 \leqslant w < 3.5$	8	2.75
$3.5 \leqslant w < 4$	32	3.75
$4 \leqslant w < 4.5$	64	4.25
$4.5 \leqslant w < 5$	40	4.75
$5 \leqslant w < 6$	16	5.5

(You may use $\sum fm = 692$ and $\sum fm^2 = 3053$)

(c) Calculate an estimate for the standard deviation of the weight of the carp.

(2)

Tim realised that he had transposed the figures for 2 of the weights of the fish.

He had recorded in the table 2.3 instead of 3.2 and 4.6 instead of 6.4

(d) Without calculating a new estimate for the standard deviation, state what effect

(i) using the correct figure of 3.2 instead of 2.3

(ii) using the correct figure of 6.4 instead of 4.6

would have on your estimated standard deviation.

Give a reason for each of your answers.

(2)

a) it isn't possible to have a sampling frame

(we don't know the exact number of carp in the lake &

Question 4 continued

don't have a list of all of them)

b) use quota sampling.

estimated total: 1600 → 10x sample size

divide estimated numbers of each type by 10:

catch 85 common, 45 mirror, & 30 leather carp.

ignore any fish caught once the quota for that type is full.

c) $\sigma = \sqrt{\frac{\Sigma fm^2}{n} - \left(\frac{\Sigma fm}{n}\right)^2}$

$= \sqrt{\frac{3053}{160} - \left(\frac{692}{160}\right)^2}$

$= 0.6129\ldots$

we use the frequency in each class.

d) i. the data point stays in the same class, so this would not change the standard deviation.

ii. ~~iib this is~~ ~~116~~ this is outside the available classes, so does change the mean by a small amount. $6.4 - 4.6 = 1.8 \simeq 3\sigma$

so the estimate of σ will increase.

$\frac{692}{160} = 4.3\ldots$ so 4.6 is close to the mean, 6.4 is far from it

Question 4 continued

Question 4 continued

(Total for Question 4 is 7 marks)

5. Afrika works in a call centre.

She assumes that calls are independent and knows, from past experience, that on each sales call that she makes there is a probability of $\frac{1}{6}$ that it is successful.

Afrika makes 9 sales calls.

(a) Calculate the probability that at least 3 of these sales calls will be successful.

(2)

The probability of Afrika making a successful sales call is the same each day.

Afrika makes 9 sales calls on each of 5 different days.

(b) Calculate the probability that at least 3 of the sales calls will be successful on exactly 1 of these days.

(2)

Rowan works in the same call centre as Afrika and believes he is a more successful salesperson.

To check Rowan's belief, Afrika monitors the next 35 sales calls Rowan makes and finds that 11 of the sales calls are successful.

(c) Stating your hypotheses clearly test, at the 5% level of significance, whether or not there is evidence to support Rowan's belief.

(4)

a) independent, constant probability of success ⇒ prompts us to use the binomial distribution.

let C = number of successful calls.

$C \sim B(9, \frac{1}{6})$

no. attempts

probability

$P(C \geq 3) = 1 - P(C \leq 2) = 0.1782\ldots$ by calculator

b) we use the value of $P(C \geq 3)$ from (a) ← imagine we 'trial' the 9 calls on 5 days, with a new p

let X = the number of days when at least 3 calls succeed.

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

Question 5 continued

$X \sim (5, P(C \geq 31))$

$P(X=1) = 5 \times (0.1782) \times (1-0.1782)^4$

$= 0.4061\ldots$

c) 1. state your hypotheses

$H_0: p = \frac{1}{6}$ $H_1: p > \frac{1}{6}$

2. define your variables & calculate test statistic

let R = number of successful calls

$R \sim B(35, \frac{1}{6})$

$P(R \geq 11) = 1 - P(R \leq 10) = 0.02\ldots$

3. form conclusion

$0.02 < 0.05 \Rightarrow$ there is sufficient evidence to support that Rowan has more successful calls than Afrika. reject H_0.

always link to context of test

Question 5 continued

(Total for Question 5 is 8 marks)

TOTAL FOR STATISTICS IS 30 MARKS