

Write your name here	
Surname	Other names
Pearson Edexcel	Centre Number
Level 3 GCE	Candidate Number
Mathematics	
Advanced	
Paper 3: Statistics	
Friday 15 June 2018 – Afternoon	Paper Reference
Time: 2 hours	9MA0/03
You must have: Mathematical Formulae and Statistical Tables, calculator	Total Marks

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. 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.
- Answers should be given to three significant figures unless otherwise stated.

Information

- A booklet 'Mathematical Formulae and Statistical Tables' is provided.
- There are 10 questions in this question paper. The total mark for this paper is 100.
- 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 ►

P58350A

©2018 Pearson Education Ltd.

1/1/1/




Pearson

SECTION A: STATISTICS

Answer ALL questions. Write your answers in the spaces provided.

1. Helen believes that the random variable C , representing cloud cover from the large data set, can be modelled by a discrete uniform distribution.

(a) Write down the probability distribution for C .

(2)

(b) Using this model, find the probability that cloud cover is less than 50%

(1)

Helen used all the data from the large data set for Hurn in 2015 and found that the proportion of days with cloud cover of less than 50% was 0.315

(c) Comment on the suitability of Helen's model in the light of this information.

(1)

(d) Suggest an appropriate refinement to Helen's model.

(1)

$$a/ \quad \begin{array}{c|c|c|c|c|c|c|c|c|c} C & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ \hline P(C=c) & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \end{array}$$

$$b/ \quad P(C < 4) = \frac{4}{9}$$

c/ The proportion of days with cover of less than 50% is lower than expected. Suggests the model may not be suitable.

d/ Helen should use a non uniform model changing depending on month / location.



2. Tessa owns a small clothes shop in a seaside town. She records the weekly sales figures, £ w , and the average weekly temperature, $t^{\circ}\text{C}$, for 8 weeks during the summer. The product moment correlation coefficient for these data is -0.915

(a) Stating your hypotheses clearly and using a 5% level of significance, test whether or not the correlation between sales figures and average weekly temperature is negative. (3)

(b) Suggest a possible reason for this correlation. (1)

Tessa suggests that a linear regression model could be used to model these data.

(c) State, giving a reason, whether or not the correlation coefficient is consistent with Tessa's suggestion. (1)

(d) State, giving a reason, which variable would be the explanatory variable. (1)

Tessa calculated the linear regression equation as $w = 10\,755 - 171t$

(e) Give an interpretation of the gradient of this regression equation. (1)

$$a/ \quad H_0: \rho = 0 \quad H_1: \rho < 0$$

$$\text{critical value} = -0.6215$$

$$r < -0.6215 \quad \therefore \text{we reject } H_0 \text{ and accept } H_1$$

There is evidence to suggest a negative correlation

b/ As the temperature increases people are more likely to do outdoors activities and not shop.

c/ It is consistent with Tessa's suggestion because r is close to -1 .

d/ The temperature (t) sales depend on the temperature. Change in t causes change in w .

e/ For every additional degree change in temp. sales go down by £171.



3. In an experiment a group of children each repeatedly throw a dart at a target. For each child, the random variable H represents the number of times the dart hits the target in the first 10 throws.

Peta models H as $B(10, 0.1)$

- (a) State two assumptions Peta needs to make to use her model.

(2)

- (b) Using Peta's model, find $P(H \geq 4)$

(1)

For each child the random variable F represents the number of the throw on which the dart first hits the target.

Using Peta's assumptions about this experiment,

- (c) find $P(F = 5)$

(2)

Thomas assumes that in this experiment no child will need more than 10 throws for the dart to hit the target for the first time. He models $P(F = n)$ as

$$P(F = n) = 0.01 + (n - 1) \times \alpha$$

where α is a constant.

- (d) Find the value of α

(4)

- (e) Using Thomas' model, find $P(F = 5)$

(1)

- (f) Explain how Peta's and Thomas' models differ in describing the probability that a dart hits the target in this experiment.

(1)

a) The probability of hitting the target always remains 0.1
All throws are independent

$$\begin{aligned} b) \quad P(H \geq 4) &= 1 - P(H \leq 3) \\ &= 1 - 0.9872048016 \\ &= 0.0127951984 \end{aligned}$$

Binomial CD
 $x=3$
 $n=10$
 $p=0.1$

$$\begin{aligned} c) \quad P(F = 5) &= 0.9 \times 0.9 \times 0.9 \times 0.9 \times 0.1 \\ &= 0.06561 \end{aligned}$$



Question 3 continued

n	1	2	3	4	5	6	7	8	9	10
$P(F=n)$	0.01	$0.01 + a$	$0.01 + 2a$	$0.01 + 3a$	$0.01 + 4a$	$0.01 + 5a$	$0.01 + 6a$	$0.01 + 7a$	$0.01 + 8a$	$0.01 + 9a$

$$0.1 + 45a = 1$$

$$45a = 0.9$$

$$a = \underline{\underline{0.02}}$$

$$\begin{aligned} e/ \quad P(F=5) &= 0.01 + 4(0.02) \\ &= \underline{\underline{0.09}} \end{aligned}$$

f/ In Peta's model the probability of hitting the target remains constant.
In Thomas's model probability increases with each throw.



0 5 8 3 5 0 1 0 7 2 6

4. Charlie is studying the time it takes members of his company to travel to the office. He stands by the door to the office from 08 40 to 08 50 one morning and asks workers, as they arrive, how long their journey was.

(a) State the sampling method Charlie used.

(1)

(b) State and briefly describe an alternative method of non-random sampling Charlie could have used to obtain a sample of 40 workers.

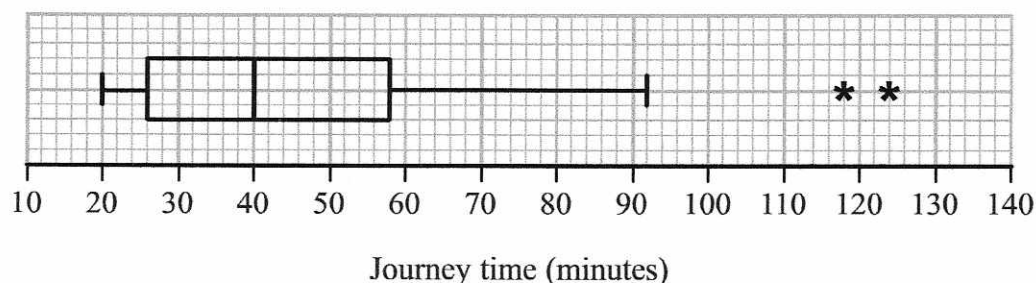
(2)

Taruni decided to ask every member of the company the time, x minutes, it takes them to travel to the office.

(c) State the data selection process Taruni used.

(1)

Taruni's results are summarised by the box plot and summary statistics below.



$$n = 95 \quad \sum x = 4133 \quad \sum x^2 = 202\,294$$

(d) Write down the interquartile range for these data.

(1)

(e) Calculate the mean and the standard deviation for these data.

(3)

(f) State, giving a reason, whether you would recommend using the mean and standard deviation or the median and interquartile range to describe these data.

(2)

Rana and David both work for the company and have both moved house since Taruni collected her data.

Rana's journey to work has changed from 75 minutes to 35 minutes and David's journey to work has changed from 60 minutes to 33 minutes.

Taruni drew her box plot again and only had to change two values.

(g) Explain which two values Taruni must have changed and whether each of these values has increased or decreased.

(3)



Question 4 continued

a/ opportunity sampling (or convenience)

b/ Quota sampling - Ask the first 20 men and 20 women how long their journey was

c/ A census

d/ $58 - 26 = \underline{32 \text{ minutes}}$

e/
$$\begin{aligned} \text{mean} &= \frac{\sum x}{n} \\ &= \frac{4133}{95} \\ &= 43.50526316 \text{ mins} \end{aligned}$$

$$\begin{aligned} \sigma &= \sqrt{\frac{202294}{95} - (43.5)^2} \\ &= 15.38514231 \text{ mins} \end{aligned}$$

f/ The distribution is not symmetrical (skew) so the IQR and median should be used.

g/ The median and upper quartile must have decreased

75 and 60 would have been between upper quartile and highest value. 35 and 33 are between lower quartile and median.

$\therefore$ median and UQ shift down by 2 values/people



5. The lifetime, L hours, of a battery has a normal distribution with mean 18 hours and standard deviation 4 hours.

Alice's calculator requires 4 batteries and will stop working when any one battery reaches the end of its lifetime.

- (a) Find the probability that a randomly selected battery will last for longer than 16 hours. (1)

At the start of her exams Alice put 4 new batteries in her calculator. She has used her calculator for 16 hours, but has another 4 hours of exams to sit.

- (b) Find the probability that her calculator will not stop working for Alice's remaining exams. (5)

Alice only has 2 new batteries so, after the first 16 hours of her exams, although her calculator is still working, she randomly selects 2 of the batteries from her calculator and replaces these with the 2 new batteries.

- (c) Show that the probability that her calculator will not stop working for the remainder of her exams is 0.199 to 3 significant figures. (3)

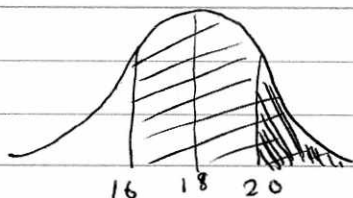
After her exams, Alice believed that the lifetime of the batteries was more than 18 hours. She took a random sample of 20 of these batteries and found that their mean lifetime was 19.2 hours.

- (d) Stating your hypotheses clearly and using a 5% level of significance, test Alice's belief. (5)

a/ Normal CD

$$P(L > 16) = 0.6914624617$$

b/ $P(L > 20)$ given $L > 16$



$$P(L > 20) = 0.3085375383$$

$$P(L > 20 | L > 16) = \frac{0.3085375383}{0.6914624617} = 0.446210...$$

each battery. ↗



Question 5 continued

$$("0.4462")^4 = 0.0396421...$$

$$= \underline{\underline{0.0396}} \text{ (3sf)}$$

c/ $P(L > 4) = 0.999767...$

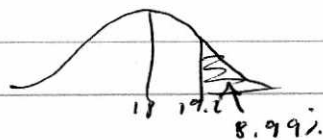
$$("0.999")^2 \times ("0.446")^2$$

$$= \underline{\underline{0.199}} \text{ (3sf)}$$

d/ $H_0: \mu = 18 \quad H_1: \mu > 18$

$$\bar{L} \sim N\left(18, \left(\frac{4}{\sqrt{20}}\right)^2\right)$$

$$P(\bar{L} > 19.2) = 0.08985624744 \quad (\text{Binomial CD})$$



$$8.99 > 5$$

Accept H_0 . There is not enough evidence to suggest $\mu > 18$.

