Candidate surname | Other names

Pearson Edexcel
Level 3 GCE

Centre Number

Candidate Number

Tuesday 18 June 2019

Morning (Time: 1 hour 30 minutes) | Paper Reference **8ST0/01**

Statistics

Advanced Subsidiary
Paper 1

You must have:
Statistical Formulae and Tables booklet
Calculator

Total Marks

Candidates may use any calculator allowed by Pearson regulations. Calculators must not 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.
- Unless otherwise stated, statistical tests should be carried out at the 5% significance level.
- When a calculator is used, the answer should be given to three significant figures unless otherwise stated.

Information

- A booklet 'Statistical Formulae and Tables' is provided.
- There are 6 questions in this question paper. The total mark for this paper is 60.
- 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.
- If you change your mind about an answer, cross it out and put your new answer and any working underneath.

Turn over ▶

P 6 0 2 0 2 A 0 1 2 0

Pearson

1   Neelam has been told that a parcel she has ordered will be delivered tomorrow between 8am and 6pm and that she must be at home to sign for it.

Neelam believes that all the times between 8am and 6pm are equally likely to be the delivery time of her parcel.

Based on this belief, she draws a diagram of the probability distribution of the possible delivery times of the parcel. The diagram is shown in **Figure 1**.
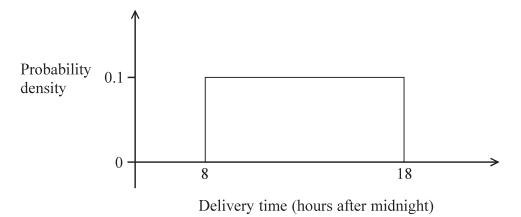

Figure 1: Probability density function of parcel delivery times

(a) Explain why

   (i) the height of Neelam's distribution is 0.1

   (2)

_____

_____

_____

_____

_____

_____

_____

   (ii) the mean and median of Neelam's distribution are the same.

   (1)

_____

_____

_____

_____

_____

_____

_____

Question 1 continued

(b) State the probability that Neelam's parcel will be delivered at exactly 10am.

(1)

(c) Find the probability that the parcel will arrive after 3pm.

(1)

Neelam requires the contents of the parcel in order to complete a job for a client. It will take 30 minutes to complete the job after the parcel arrives. She must complete the job before 6pm.

(d) Given that the parcel doesn't arrive before 3pm, calculate the probability that Neelam will be able to complete the job on time.

(2)

(Total for Question 1 is 7 marks)

2 As they get older, some people experience memory loss.

Eighteen participants, each over 65 years of age and who had been diagnosed with mild memory loss, were recruited to a trial of a new therapy. The therapy was designed to improve memory.

Of the participants, 9 were randomly assigned to the new therapy.

The other 9 participants only received standard GP advice on improving their memory.

Each participant was given a memory improvement score based on the difference between their score on a memory test before and after the trial.

A negative score indicates that the participant obtained a lower score on the memory test after the trial.

The participants' memory improvement scores are shown in **Figure 2**.

Memory improvement score	
Standard GP advice	**New therapy**
−0.75	−0.50
−0.25	0
0	0.50
0.25	0.75
0.50	1
1	1.25
1.25	1.5
1.25	2
1.75	2.25

Source: Adapted from experimental data

Figure 2: Memory improvement scores

(a) Use a non-parametric test to investigate whether the new therapy led to a greater improvement in average memory score.

You should state the name of the test that you use.

(8)

P60202A0520

Question 2 continued

(b) Considering both the results of your hypothesis test and the raw data in **Figure 2**,

- summarise the outcomes of the trial in a suitable form for presentation to a team of health and social care professionals

- suggest a next stage for the research.

(4)

(Total for Question 2 is 12 marks)

3 Gilbert runs a company that supplies plant bulbs. One of the bulb types he supplies is anemones. He stores the anemone bulbs in very large mixed batches. The batches contain bulbs that produce 68% purple flowers, 17% white flowers and 15% red flowers.

Priti, a park manager, wants to buy 3000 anemone bulbs for a display in a park. She requires that at least 2000 of the anemone bulbs produce purple flowers.

Use an appropriate **approximate** distribution to estimate the probability that a sample of 3000 anemone bulbs from Gilbert's company will **not** fulfil Priti's requirement.

(5)

(Total for Question 3 is 5 marks)

4 Katherine is investigating whether there is an association between the sex of a cat and its colour. She decides that she will investigate for this association using a χ^2 test.

Katherine summarises the information obtained for a sample of 136 cats in a table.

Her results are shown in **Figure 3**.

		Colour							
		Black and white	Black	Tabby	Ginger	Tortoise-shell	Grey	Mixed	Total
Sex	Female	26	21	17	4	4	2	3	**77**
	Male	25	9	11	11	0	2	1	**59**
	Total	**51**	**30**	**28**	**15**	**4**	**4**	**4**	**136**

Source: www.cats.org.uk

Figure 3: Sex and colour of cats

The table of expected frequencies for Katherine's χ^2 test is given in **Figure 4**.

		Colour						
		Black and white	Black	Tabby	Ginger	Tortoise-shell	Grey	Mixed
Sex	Female	28.88	16.99	15.85	8.49	2.26	2.26	2.26
	Male	22.13	13.01	12.15	6.51	1.74	1.74	1.74

Figure 4: Expected frequencies correct to 2 d.p.

(a) Explain, with a reason, what adjustment Katherine must make to the data in **Figure 3** and **Figure 4** before she can carry out her χ^2 test.

(2)

....

....

....

....

....

....

....

....

....

....

....

....

....

Question 4 continued

(b) Making the necessary adjustment that you stated in part (a), complete Katherine's χ^2 test.

Make the assumption that Katherine's sample is random.

(7)

(c) By considering contributions to the χ^2 test statistic in part (b), identify, in context, the most notable fact concerning the association between the sex of a cat and its colour.

(2)

It is later discovered that Katherine obtained her data from a cat rescue charity's online database.

This database requires a geographic region to be selected before displaying any results. Katherine obtained the data for the 136 cats in her sample by selecting her own home city as the region.

(d) Explain two different sources of bias that might be present in Katherine's data.

(2)

(e) Comment on whether the assumption made in part (b) is reasonable.

(1)

(Total for Question 4 is 14 marks)

5 In order to monitor the gopher tortoise population in a particular area of Florida, a survey was taken of tortoise habitats to find the location and habitation status of tortoise burrows.

Three different tortoise habitat types were identified:

- **flatwoods** populated mainly with pines and saw palmetto,

- areas of **hardwood** trees such as oak, mahogany and maple,

- low **shrubby** areas with few trees.

Burrows were classified as either potentially occupied or abandoned.
Those that were potentially occupied were identified as either active or inactive.

The results of the survey are summarised in **Figure 5**.

		Burrow habitation status			
		Potentially occupied		Abandoned	
		Active B_1	Inactive B_2	C	Total
Habitat type	Flatwood F	17	9	24	**50**
	Hardwood H	8	12	6	**26**
	Shrubby S	43	8	18	**69**
	Total	**68**	**29**	**48**	**145**

Source: www.fnai.org

Figure 5: Results of tortoise survey

B_1 denotes the event 'the burrow was potentially occupied and active'.

B_2 denotes the event 'the burrow was potentially occupied and inactive'.

C denotes the event 'the burrow was abandoned'.

F denotes the event 'the burrow was in a flatwood habitat'.

H denotes the event 'the burrow was in a hardwood habitat'.

S denotes the event 'the burrow was in a shrubby habitat'.

(a) For a randomly selected tortoise burrow, find

    (i) $P(F')$

                                                        (1)

    (ii) $P(B_1 \cap H)$

                                                        (1)

(b) State, with justification, whether the events B_1 and H are statistically independent.

                                                        (3)

(c) For a randomly selected tortoise burrow, find $P(B_1 \cup B_2 \mid F)$

(2)

(d) In the context of the question, explain what $P(B_1 \cup B_2 \mid F)$ represents.

(1)

Question 5 continued

(e) State in which of the three habitat types it is most likely that a tortoise burrow was potentially occupied.

You should quote calculations to justify your answer.

(3)

(Total for Question 5 is 11 marks)

6 Femi works for a campaigning charity based in London. He is seeking evidence to support more resources being put into education and training for 16–18 year olds.

He finds some government data that gives two measures for each of the 32 London boroughs. These are:

X = percentage of 16–18 year olds not in education, employment or training (NEETs)

Y = number of first time youth offenders ($<$ 18 years old) per 10,000 of the population.

Femi starts his analysis by using statistical software to draw a scatter diagram of this data.

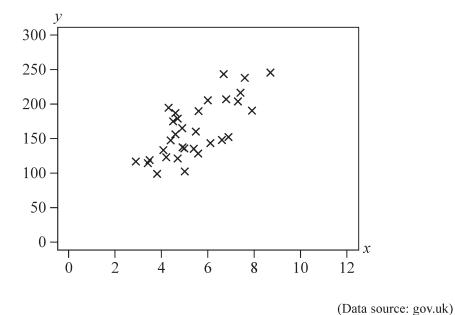The scatter diagram is given in **Figure 6**.

(Data source: gov.uk)

Figure 6

(a) Give a reason why it is sensible for Femi to start his analysis by drawing a scatter diagram.

(1)

Question 6 continued

Femi decides to calculate the least squares regression line in the form $y = a + bx$.

He obtains the equation $y = 48 + 21.2x$

(b) With reference to the aims of his research, explain why Femi might want to show a dependency of y on x.

(1)

(c) Explain what is wrong with Femi assuming a dependency of y on x.

(1)

(d) Interpret each of the values of a and b in context.

(2)

Femi states on his charity's website that his data and analysis show that:

- *If all 16-18 year olds were in education, employment or training, youth crime could be reduced by about 80%.*

(e) Show how Femi might have used his regression equation and scatter diagram to obtain the figure of 80%.

(2)

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Femi also states on his charity's website that:

- *Young people in London who are not given opportunities to take part in education or training are turning to crime.*

- *We need to increase funding for sixth forms and colleges to make sure that these young people have a future.*

(f) Make **four** criticisms of Femi's **three** statements.

(4)

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Question 6 continued

(Total for Question 6 is 11 marks)

TOTAL FOR PAPER IS 60 MARKS

BLANK PAGE

P60202A02020