Centre No.						Paper Reference	Surname	Initial(s)
Candidate No.						6 6 8 3 / 0 1	Signature	

Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Friday 18 May 2012 – Morning

Time: 1 hour 30 minutes

Examiner's use only

Team Leader's use only

Question Number	Leave Blank
1	
2	
3	
4	
5	
6	
7	
Total	

Materials required for examination	Items included with question papers
Mathematical Formulae (Pink)	Nil

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 or symbolic differentiation/integration, or have retrievable mathematical formulae stored in them.

Instructions to Candidates

In the boxes above, write your centre number, candidate number, your surname, initials and signature.
Check that you have the correct question paper.
Answer ALL the questions.
You must write your answer to each question in the space following the question.
Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
The marks for individual questions and the parts of questions are shown in round brackets: e.g. **(2)**.
There are 7 questions in this question paper. The total mark for this paper is 75.
There are 24 pages in this question paper. Any blank pages are indicated.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.
You should show sufficient working to make your methods clear to the Examiner.
Answers without working may not gain full credit.

Printer's Log. No.
P40105XA
W850/R6638/57570 5/5/2

P 4 0 1 0 5 X A 0 1 2 4

Turn over

PEARSON

Leave blank

1. A discrete random variable X has the probability function

$$P(X = x) = \begin{cases} k(1-x)^2 & x = -1, 0, 1 \text{ and } 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Show that $k = \frac{1}{6}$ **(3)**

(b) Find $E(X)$ **(2)**

(c) Show that $E(X^2) = \frac{4}{3}$ **(2)**

(d) Find $\text{Var}(1-3X)$ **(3)**

x	-1	0	1	2
$P(X=x)$	$4k$	k	0	k

$4k + k + k = 1$

$6k = 1$

$k = 1/6$

b/

x	-1	0	1	2
$P(X=x)$	$2/3$	$1/6$	0	$1/6$

$E(X) = -1(2/3) + 0(1/6) + 1(0) + 2(1/6) = -1/3$

c/

$E(X^2) = 1(2/3) + 4(1/6) = 4/3$

d/ $\text{Var}(X) = \frac{4}{3} - \left(-\frac{1}{3}\right)^2$

$= \frac{11}{9}$

$\text{Var}(1-3X) = \frac{11}{9} \times 9$

$= 11$

Leave blank

2. A bank reviews its customer records at the end of each month to find out how many customers have become unemployed, u, and how many have had their house repossessed, h, during that month. The bank codes the data using variables $x = \frac{u-100}{3}$ and $y = \frac{h-20}{7}$. The results for the 12 months of 2009 are summarised below.

$$\sum x = 477 \quad S_{xx} = 5606.25 \quad \sum y = 480 \quad S_{yy} = 4244 \quad \sum xy = 23\,070$$

(a) Calculate the value of the product moment correlation coefficient for x and y.
(3)

(b) Write down the product moment correlation coefficient for u and h.
(1)

The bank claims that an increase in unemployment among its customers is associated with an increase in house repossessions.

(c) State, with a reason, whether or not the bank's claim is supported by these data.
(2)

a) $r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$

$S_{xy} = \Sigma xy - \frac{\Sigma x \Sigma y}{n}$

$= 23070 - \frac{(477)(480)}{12}$

$= 3990$

$r = \frac{3990}{\sqrt{(5606.25)(4244)}}$

$= 0.818$ (3sf)

b) 0.818 (3sf)

c) There is a positive correlation so the banks claim is supported by the data.

Leave blank

3. A scientist is researching whether or not birds of prey exposed to pollutants lay eggs with thinner shells. He collects a random sample of egg shells from each of 6 different nests and tests for pollutant level, p, and measures the thinning of the shell, t. The results are shown in the table below.

p	3	8	30	25	15	12
t	1	3	9	10	5	6

[You may use $\sum p^2 = 1967$ and $\sum pt = 694$]

(a) Draw a scatter diagram on the axes on page 7 to represent these data.
(2)

(b) Explain why a linear regression model may be appropriate to describe the relationship between p and t.
(1)

(c) Calculate the value of S_{pt} and the value of S_{pp}.
(4)

(d) Find the equation of the regression line of t on p, giving your answer in the form $t = a + bp$.
(4)

(e) Plot the point $(\bar{p}, \bar{t})$ and draw the regression line on your scatter diagram.
(2)

The scientist reviews similar studies and finds that pollutant levels above 16 are likely to result in the death of a chick soon after hatching.

(f) Estimate the minimum thinning of the shell that is likely to result in the death of a chick.
(2)

b/ there appears to be a positive correlation

c/ $S_{pt} = \Sigma pt - \frac{\Sigma p \Sigma t}{n}$

$= 694 - \frac{(93)(34)}{6}$

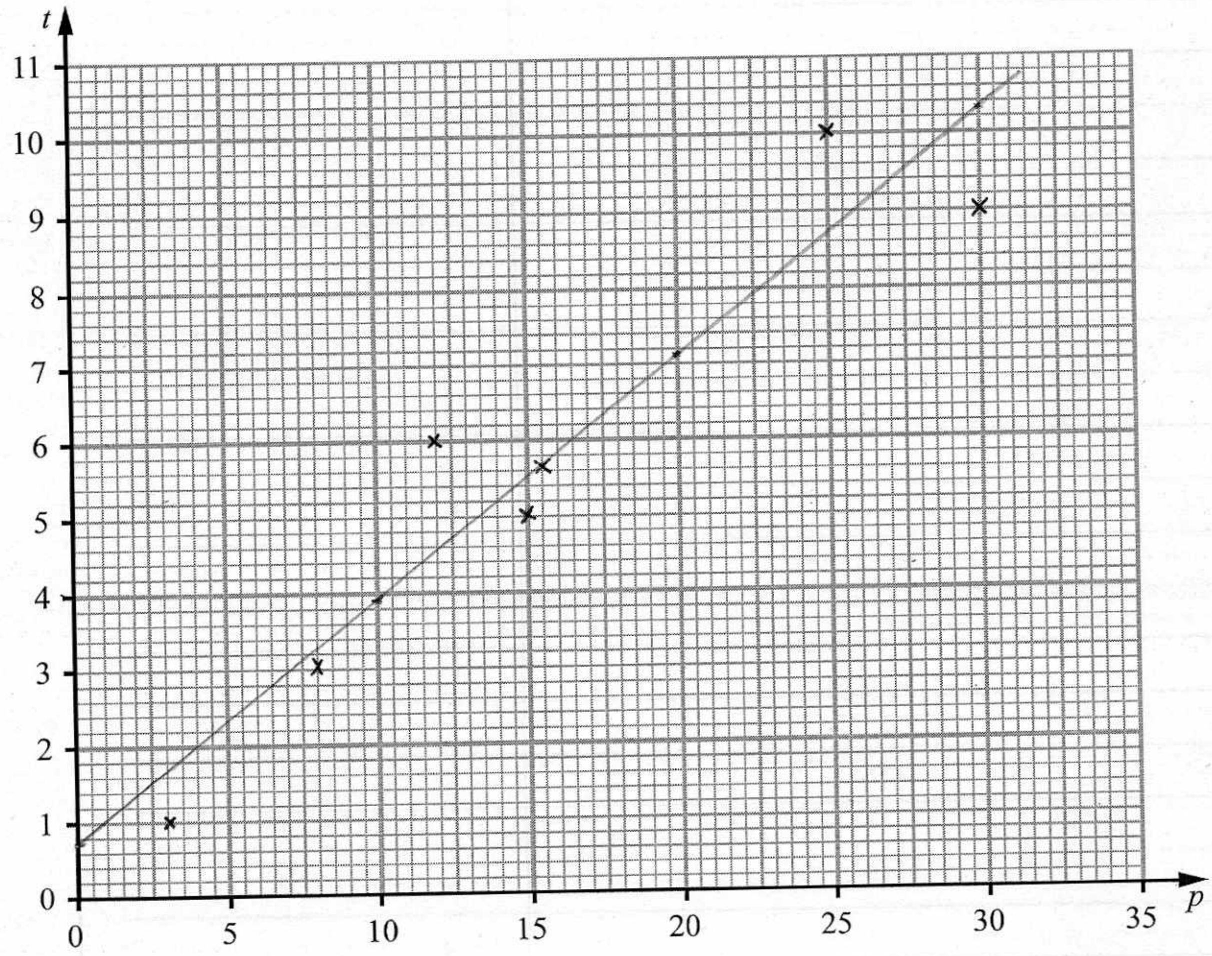
$= \underline{\underline{167}}$

$S_{pp} = \Sigma p^2 - \frac{(\Sigma p)^2}{n}$

$= 1967 - \frac{93^2}{6} = \underline{\underline{525.5}}$

Leave blank

Question 3 continued



d) $b = \frac{S_{pt}}{S_{pp}} = \frac{167}{525.5}$

$= \underline{0.318}$ 3sf

$a = \bar{t} - b\bar{p}$ $\qquad \bar{p} = 15.5 \qquad \bar{t} = \frac{17}{3}$

$= \underline{0.741}$ 3sf

$t = 0.741 + 0.318p$

f) $p = 16 \qquad t = 0.741 + 0.318(16)$

$= \underline{\underline{5.8}}$ 1dp

Leave blank

4.

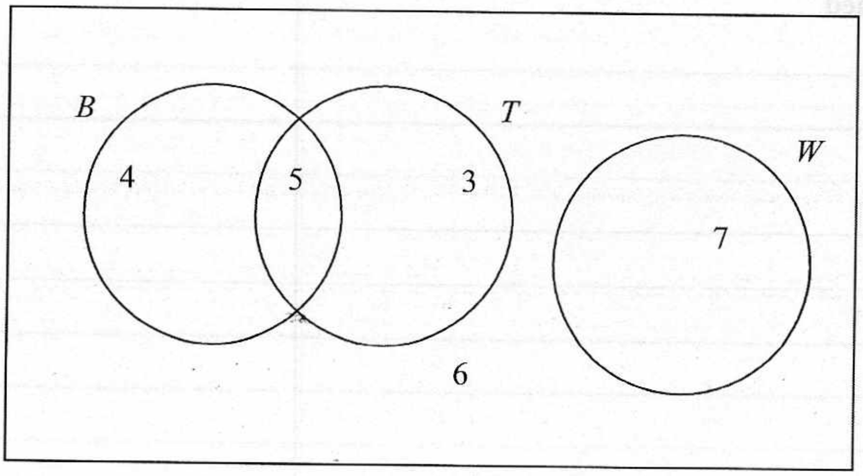


Figure 1

Figure 1 shows how 25 people travelled to work.

Their travel to work is represented by the events

B bicycle

T train

W walk

(a) Write down 2 of these events that are mutually exclusive. Give a reason for your answer. **(2)**

(b) Determine whether or not B and T are independent events. **(3)**

One person is chosen at random.

Find the probability that this person

(c) walks to work, **(1)**

(d) travels to work by bicycle and train. **(1)**

(e) Given that this person travels to work by bicycle, find the probability that they will also take the train. **(2)**

a/ B and W

b/ If independent $P(B) \times P(T) = P(B \cap T)$

Leave blank

Question 4 continued

$P(B) = 9/25$
$P(T) = 8/25$

$P(B) \times P(T) = \frac{72}{625}$

$P(B) \times P(T) \neq \frac{5}{25}$ $\therefore$ B and T are not independent.

c/ $7/25$

d/ $5/25$ $[1/5]$

e/ $5/9$ $\left[\frac{P(B \cap T)}{P(B)}\right]$

Leave blank

5.

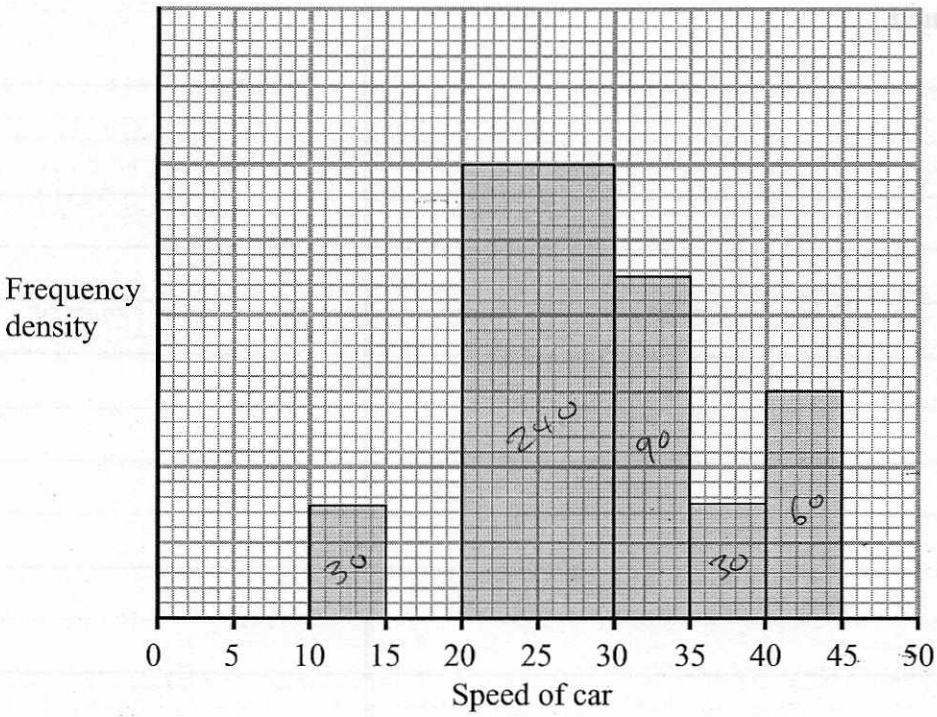


Figure 2

A policeman records the speed of the traffic on a busy road with a 30 mph speed limit. He records the speeds of a sample of 450 cars. The histogram in Figure 2 represents the results.

(a) Calculate the number of cars that were exceeding the speed limit by at least 5 mph in the sample. **(4)**

(b) Estimate the value of the mean speed of the cars in the sample. **(3)**

(c) Estimate, to 1 decimal place, the value of the median speed of the cars in the sample. **(2)**

(d) Comment on the shape of the distribution. Give a reason for your answer. **(2)**

(e) State, with a reason, whether the estimate of the mean or the median is a better representation of the average speed of the traffic on the road. **(2)**

a/ 22.5 boxes $\frac{450}{22.5}$ = 20 cars per box

~~90 + 30 + 60 = 180~~

30 + 60 = 90

Leave blank

Question 5 continued

b/

speed	frequency	fx
10 - 15	30	375
20 - 30	240	6000
30 - 35	90	2925
35 - 40	30	1125
40 - 45	60	2550
		12975

$\frac{12975}{450} = 28.8$ (3sf) mph

c/ 225^{th} car [Interpolation]

$20 + \frac{195}{240} \times 10 = 28.1$ (3sf) mph

d/ mean > median $\therefore$ positive skew

e/ there is a skew $\therefore$ the median is a better representation

[median is not affected by extreme values]

Turn over