S1 May 2012.

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	$4k$	k	0	k

$\Sigma P = 1 \quad 4k + k + 0 + k = 1 \Rightarrow 6k = 1 \Rightarrow k = \frac{1}{6}$ #

b)

x	-1	0	2
p	$\frac{4}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$E(x) = -\frac{4}{6} + \frac{2}{6} = -\frac{2}{6} = -\frac{1}{3}$

c)

x^2	1	0	4
p	$\frac{4}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$E(x^2) = \frac{4}{6} + 0 + \frac{4}{6} = \frac{8}{6} = \frac{4}{3}$ #

d) $\text{Var}(1-3x) = (-3)^2 V(x) = 9(V(x))$

$V(x) = E(x^2) - E(x)^2 = \frac{4}{3} - \left(-\frac{1}{3}\right)^2 = \frac{11}{9}$

$\therefore V(1-3x) = 9 \times \frac{11}{9} = 11$

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) $S_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 3990$

$\text{PMCC} = \frac{S_{xy}}{\sqrt{S_{xx} \times S_{yy}}} = \frac{3990}{\sqrt{5606.25 \times 4244}}$

PMCC = 0.818 (3sf)

b) 0.818

c) as PMCC = 0.818 there is evidence to suggest positive correlation exists between the amount of customers unemployed and the amount of houses repossessed.

∴ The bank's claim is supported by the data.

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)

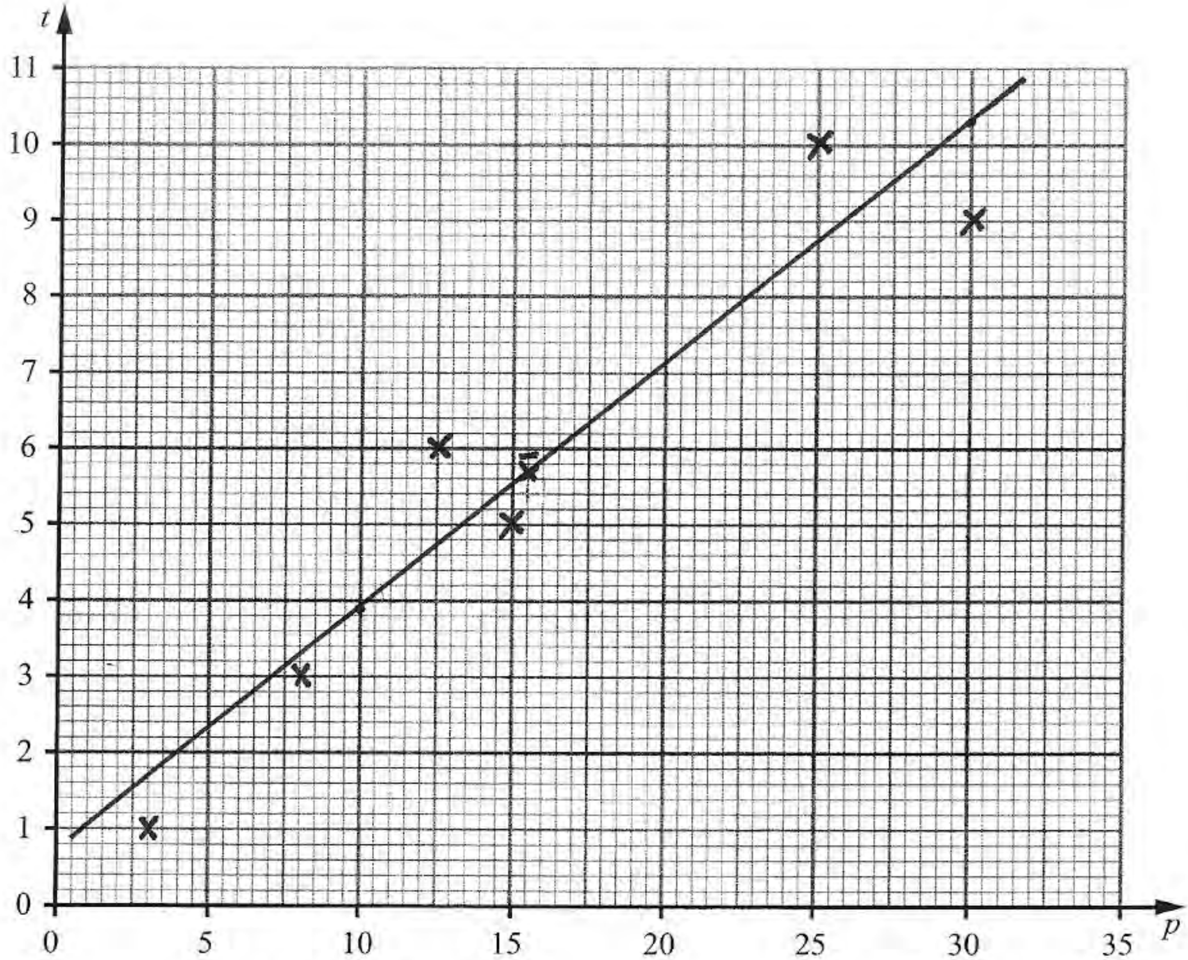
(e) Plot the point $(\bar{p}, \bar{t})$ and draw the regression line on your scatter diagram.

(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)



$\Sigma p = 93 \quad \Sigma t = 34$ $\qquad p \to x \quad t \to y$

b) the points lie quite close to a straight line

c) $S_{pt} = \Sigma pt - \dfrac{(\Sigma p)(\Sigma t)}{n} = 694 - \dfrac{93 \times 34}{6} = 167$

$S_{pp} = \Sigma p^2 - \dfrac{(\Sigma p)^2}{n} = 1967 - \dfrac{93^2}{6} = 525.5$

d) $b = \dfrac{S_{pt}}{S_{pp}} = 0.317793$ $\qquad a = \left(\dfrac{34}{6}\right) - b\left(\dfrac{93}{6}\right) = 0.74088$

$\bar{y} = 5.7 \quad \bar{x} = 15.5$

$\bar{t} \qquad \bar{p}$

$t = 0.741 + 0.318p$

e) $p = 10, t = 3.9$ $\qquad p = 30, t = 10.3$

f) $p = 16 \quad t = 0.741 + 0.318 \times 16 = \underline{5.83}$

4.

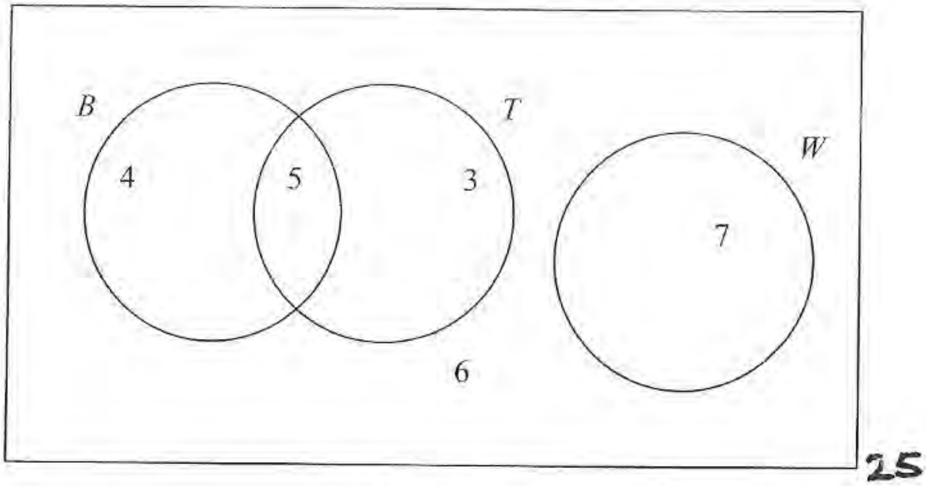


Figure 1

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.

B and W since $P(B \cap W) = 0$

(2)

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

$P(B \cap T) = \frac{5}{25} = \frac{1}{5}$ $\quad P(B) \times P(T) = \frac{9}{25} \times \frac{8}{25} = \frac{72}{625}$

(3)

$P(B \cap T) \neq P(B) \times P(T)$

NOT independent.

One person is chosen at random.

Find the probability that this person

(c) walks to work,

$P(W) = \frac{7}{25}$

(1)

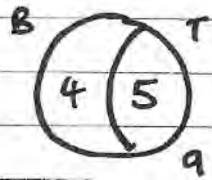
(d) travels to work by bicycle and train.

$P(B \cap T) = \frac{5}{25}$

(1)

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

(2)



$P(T \mid B) = \frac{5}{9}$

5.

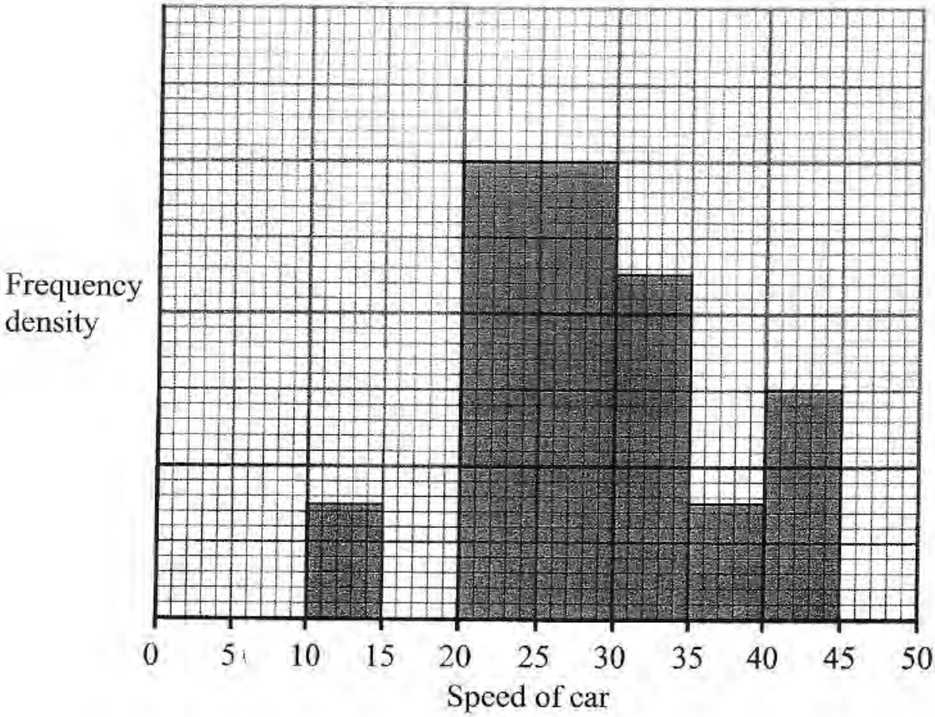


Figure 2

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)

5) Area $\propto$ freq $\Rightarrow$ Area $= k \times$ freq.

Total area $= 5\times7.5 + 10\times30 + 5\times22.5 + 5\times7.5 + 5\times15$

$= 562.5$

$\therefore\ 562.5 = k \times 450 \Rightarrow k = 1.25$

Area > 35 $= 5\times7.5 + 5\times15 = 112.5$

freq $= \dfrac{112.5}{1.25} = \underline{90}$

b)

speed	freq		x
10 - 15	30	30	12.5
20 - 30	240	270	25
30 - 35	90		32.5
35 - 40	30		37.5
40 - 45	60		42.5

mean $= \dfrac{\Sigma fx}{n}$

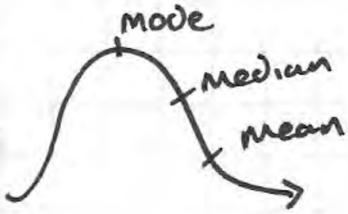
$= \dfrac{12975}{450} = 28.8$

c) $\frac{1}{2}n = 225$

20	Q_2	30
30	225	270

$\dfrac{Q_2 - 20}{10} = \dfrac{195}{240} \Rightarrow Q_2 = \dfrac{195}{240} \times 10 + 20 = 28.1$

d)



mean (28.8) > median (28.1) positive skew

e) median as the data is skewed

6. The heights of an adult female population are normally distributed with mean 162 cm and standard deviation 7.5 cm.

(a) Find the probability that a randomly chosen adult female is taller than 150 cm.
(3)

Sarah is a young girl. She visits her doctor and is told that she is at the 60th percentile for height.

(b) Assuming that Sarah remains at the 60th percentile, estimate her height as an adult.
(3)

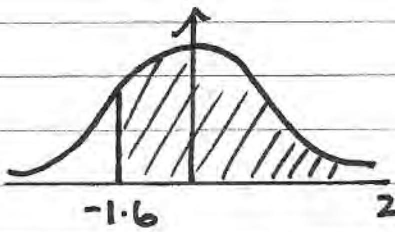
The heights of an adult male population are normally distributed with standard deviation 9.0 cm.

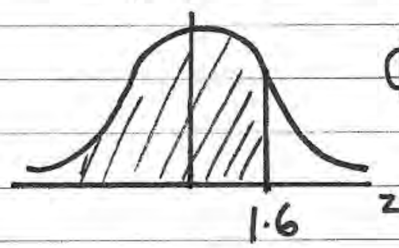
Given that 90% of adult males are taller than the mean height of adult females,

(c) find the mean height of an adult male.
(4)

$h \sim N(162, 7.5^2)$

$P(h > 150) \Rightarrow P\left(Z > \frac{150-162}{7.5}\right) = P(Z > -1.6)$

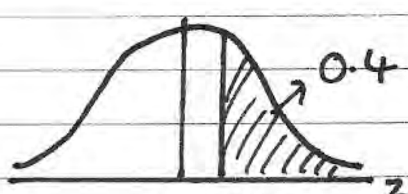




$\phi(1.6)$

$= \underline{0.9452}$

b) $P(h < x) = 0.6 \Rightarrow P\left(Z < \frac{x-162}{7.5}\right) = 0.6$



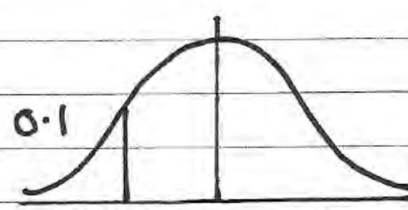
$\frac{x-162}{7.5} = 0.2533$

$x = 0.2533 \times 7.5 + 162$

$x = \underline{163.9 \text{ cm}}$

c) $P(M > 162) = 0.9$

$P\left(Z > \frac{162-\mu}{9}\right) = 0.9$



$\frac{162-\mu}{9} = -1.2816$

$\therefore \mu = 173.5$

7. A manufacturer carried out a survey of the defects in their soft toys. It is found that the probability of a toy having poor stitching is 0.03 and that a toy with poor stitching has a probability of 0.7 of splitting open. A toy without poor stitching has a probability of 0.02 of splitting open.

(a) Draw a tree diagram to represent this information.

(3)

(b) Find the probability that a randomly chosen soft toy has exactly one of the two defects, poor stitching or splitting open.

(3)

The manufacturer also finds that soft toys can become faded with probability 0.05 and that this defect is independent of poor stitching or splitting open. A soft toy is chosen at random.

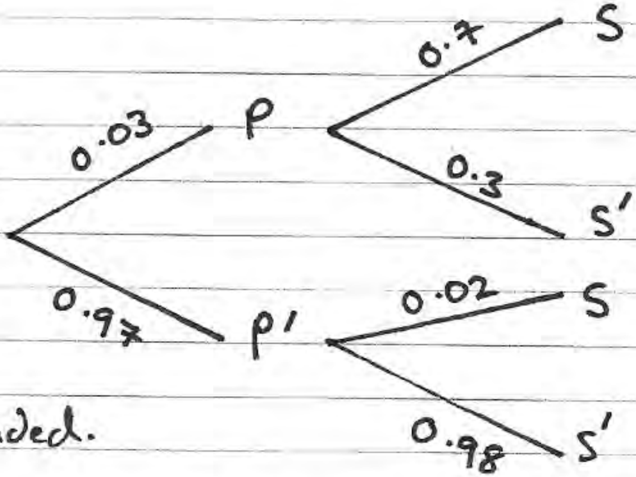
(c) Find the probability that the soft toy has none of these 3 defects.

(2)

(d) Find the probability that the soft toy has exactly one of these 3 defects.

(4)

P - Poor Stitching S - Splits open



b)

$PS' = 0.009$

$P'S = 0.0194$ +

0.0284

F - Faded.

c) $0.97 \times 0.98 \times 0.95 = 0.90307$

d) either 1 defect from P and S and no defects in fading

or no defects from P and S and a fading defect.

$0.0284 \times 0.95 = 0.9784$

$+ 0.97 \times 0.98 \times 0.05$ $+$ 0.04753

$0.02698 + 0.04753 = \underline{0.07451}$