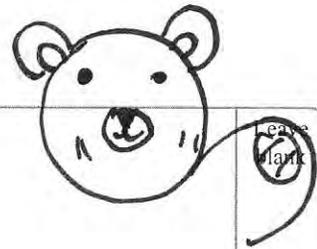1. The discrete random variable X can only take the values 1, 2, 3 and 4
   For these values the cumulative distribution function is defined by

$$F(x) = kx^2 \quad \text{for} \quad x = 1, 2, 3, 4$$

   where k is a constant.

   (a) Find the value of k.

   (2)

   (b) Find the probability distribution of X.

   (2)

a)

x	1	2	3	4
F(x)	k	$4k$	$9k$	$16k$

$16k = 1$

$\therefore k = \frac{1}{16}$

b)

x	1	2	3	4
F(x)	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{9}{16}$	$\frac{16}{16}$

x	1	2	3	4
P(x)	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{5}{16}$	$\frac{7}{16}$

2. Paul believes there is a relationship between the value and the floor size of a house. He takes a random sample of 20 houses and records the value, £v, and the floor size, s m^2

   The data were coded using $x = \dfrac{s - 50}{10}$ and $y = \dfrac{v}{100\,000}$ and the following statistics obtained.

   $\sum x = 441.5, \quad \sum y = 59.8, \quad \sum x^2 = 11\,261.25, \quad \sum y^2 = 196.66, \quad \sum xy = 1474.1$

   (a) Find the value of S_{xy} and the value of S_{xx}

   (3)

   (b) Find the equation of the least squares regression line of y on x in the form $y = a + bx$

   (3)

   The least squares regression line of v on s is $v = c + ds$

   (c) Show that $d = 1020$ to 3 significant figures and find the value of c

   (3)

   (d) Estimate the value of a house of floor size 130 m^2

   (2)

   (e) Interpret the value d

   (1)

   Paul wants to increase the value of his house. He decides to add an extension to increase the floor size by 31 m^2

   (f) Estimate the increase in the value of Paul's house after adding the extension.

   (1)

---

a) $S_{xy} = 1474.1 - \dfrac{441.5 \times 59.8}{20} = 154.015$

$S_{xx} = 11261.25 - 441.5^2 \div 20 = 1515.1375$

b) $b = \dfrac{S_{xy}}{S_{xx}} = \dfrac{154.015}{1515.1375} = 0.10165\ldots$

$a = \bar{y} - b\bar{x} = \dfrac{59.8}{20} - 0.10165\ldots \times \dfrac{441.5}{20} = 0.74606\ldots$

$y = 0.102x + 0.746$

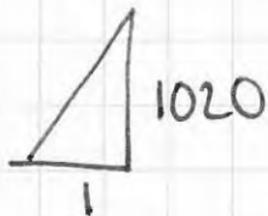c) $\dfrac{v}{100000} = 0.102\left(\dfrac{s-50}{10}\right) + 0.746$

$$V = 1016 \cdot 508(s-50) + 74607 \cdot 625$$
$$V = 1016 \cdot 508s + 23782 \cdot 20486$$
$$V = 1020s + 23800 \qquad (3sf)$$

d) $V = 1020 \times 130 + 23800 = £156400$

e)

an extra £1020 onto the value
for each extra

f) $1020 \times 31 = £31620$

3. A company employs 90 administrators. The length of time that they have been employed by the company and their gender are summarised in the table below.

Length of time employed, x years	Female	Male	
$x < 4$	9	16	25
$4 \leqslant x < 10$	14	20	34
$10 \leqslant x$	7	24	31
	30	60	90

One of the 90 administrators is selected at random.

(a) Find the probability that the administrator is female.

(1)

(b) Given that the administrator has been employed by the company for less than 4 years, find the probability that this administrator is male.

(2)

(c) Given that the administrator has been employed by the company for less than 10 years, find the probability that this administrator is male.

(2)

(d) State, with a reason, whether or not the event 'selecting a male' is independent of the event 'selecting an administrator who has been employed by the company for less than 4 years'.

(3)

a) $\dfrac{30}{90} = \dfrac{1}{3}$ b) $\dfrac{16}{25}$ c) $\dfrac{36}{59}$

d) $P(M \cap x < 4) = \dfrac{16}{90} = 0.1\dot{7}$

$P(M) = \dfrac{60}{90} = \dfrac{2}{3}$ $P(x < 4) = \dfrac{25}{90} = \dfrac{5}{18}$

$P(M) \times P(x < 4) = \dfrac{10}{54} = 0.185$

$\therefore P(M \cap x < 4) \neq P(M) \times P(x < 4)$

$\therefore$ NOT Independent.

4. A bag contains 19 red beads and 1 blue bead only.

Linda selects a bead at random from the bag. She notes its colour and replaces the bead in the bag. She then selects a second bead at random from the bag and notes its colour.

Find the probability that

(a) both beads selected are blue,

(1)

(b) exactly one bead selected is red.

(2)

In another bag there are 9 beads, 4 of which are green and the rest are yellow.

Linda selects 3 beads from this bag at random without replacement.

(c) Find the probability that 2 of these beads are yellow and 1 is green.

(3)

Linda replaces the 3 beads and then selects another 4 at random without replacement.

(d) Find the probability that at least 1 of the beads is green.

(3)

a) $\left(\frac{1}{20}\right)^2 = \frac{1}{400}$

b) RB or BR

$= 2 \times \frac{1}{20} \times \frac{19}{20} = \frac{19}{200}$

c) YYG or YGY or GYY

$= 3 \times \frac{5}{9} \times \frac{4}{8} \times \frac{4}{7} = \frac{10}{21}$

d) $\Rightarrow 1 - YYYY$

$= 1 - \frac{5}{9} \times \frac{4}{8} \times \frac{3}{7} \times \frac{2}{6} = \frac{121}{126}$

5. Police measure the speed of cars passing a particular point on a motorway.
The random variable X is the speed of a car.

X is modelled by a normal distribution with mean 55 mph (miles per hour).

(a) Draw a sketch to illustrate the distribution of X. Label the mean on your sketch.

(2)

The speed limit on the motorway is 70 mph. Car drivers can choose to travel faster than the speed limit but risk being caught by the police.

The distribution of X has a standard deviation of 20 mph.

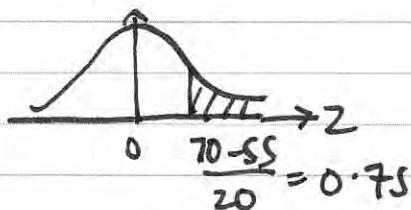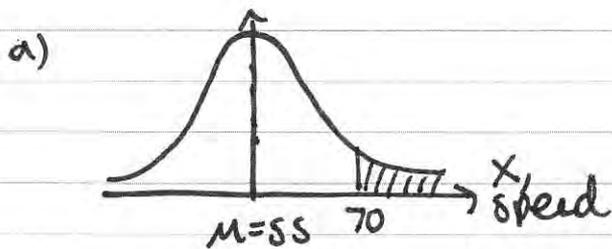(b) Find the percentage of cars that are travelling faster than the speed limit.

(3)

The fastest 1% of car drivers will be banned from driving.

(c) Show that the lowest speed, correct to 3 significant figures, for a car driver to be banned is 102 mph. Show your working clearly.

(3)

Car drivers will just be given a caution if they are travelling at a speed m such that

$$P(70 < X < m) = 0.1315$$

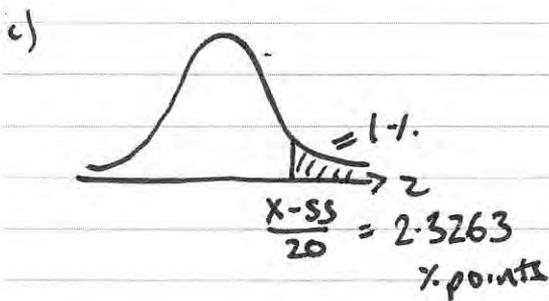(d) Find the value of m. Show your working clearly.

(4)

a)

b) $X \sim N(55, 20)$

$P(X > 70)$

$P\left(Z > \dfrac{70-55}{20}\right)$

$P(Z > 0.75)$

$= 1 - \Phi(0.75) = 0.2266$
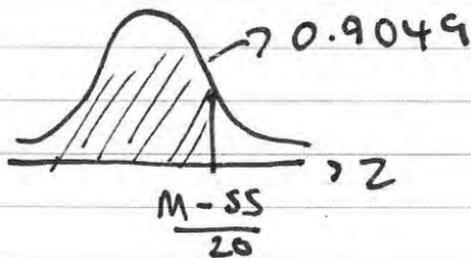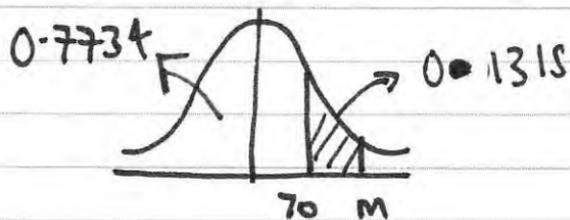
c)

$\dfrac{X-55}{20} = 2.3263$

$X - 55 = 46 \cdot 526$

$X = 101 \cdot 526$

$\sim X = \underline{102}$

$\dfrac{X-55}{20} = 2.3263$

% points

d) $P(X < 70) = 0.7734$ from (b)

$\therefore P(70 < X < M) = 0.1315 \Rightarrow P(X < M) = 0.9049$

$P\left(Z < \frac{M-55}{20}\right) = 0.9049$

$\phi\left(\frac{M-55}{20}\right) = 0.9049 \Rightarrow \frac{M-55}{20} = 1.31$

$$M - 55 = 26.2$$

$$M = 81.2 \text{ mph}$$

6. The random variable X has a discrete uniform distribution and takes the values 1, 2, 3, 4

Find

(a) F(3), where F(x) is the cumulative distribution function of X,

(1)

(b) E(X).

(1)

(c) Show that $\text{Var}(X) = \dfrac{5}{4}$

(3)

The random variable Y has a discrete uniform distribution and takes the values

$$3, \ 3 + k, \ 3 + 2k, \ 3 + 3k$$

where k is a constant.

(d) Write down P(Y = y) for y = 3, 3 + k, 3 + 2k, 3 + 3k

(1)

The relationship between X and Y may be written in the form $Y = kX + c$ where c is a constant.

(e) Find Var(Y) in terms of k.

(2)

(f) Express c in terms of k.

(1)

a)

x	1	2	3	4
$P(x)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

x	1	2	3	4
$F(X)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{3}{4}$	$\frac{4}{4}$

$F(3) = \frac{3}{4}$

b) $E(X) = \frac{n+1}{2} = \frac{5}{2} = 2.5$

c) $V(X) = \frac{(n+1)(n-1)}{12} = \frac{5 \times 3}{12} = \frac{15}{12} = \frac{5}{4}$ ✓

d)

y	3	3+k	3+2k	3+3k
$P(y=y)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

e) $V(x) = \frac{5}{4}$

$V(Y) = V(uX + c) = u^2 V(x) = \frac{5}{4} u^2$

f)

X	1	2	3	4
uX+c	u+c	2u+c	3u+c	4u+c
Y	3	3+u	3+2u	3+3u

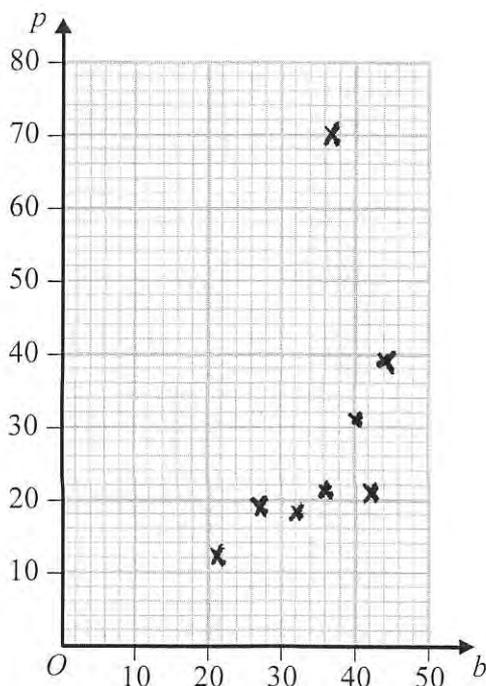$\therefore u + c = 3 \qquad \therefore c = 3 - u$

2

7. A doctor is investigating the correlation between blood protein, p, and body mass index, b.

He takes a random sample of 8 patients and the data are shown in the table below.

Patient	A	B	C	D	E	F	G	H
b	32	36	40	44	42	21	27	37
p	18	21	31	39	21	12	19	70

(a) Draw a scatter diagram of these data on the axes provided.

(2)

The doctor decides to leave out patient H from his calculations.

(b) Give a reason for the doctor's decision.

(1)

For the 7 patients A, B, C, D, E, F and G,

$$S_{bp} = 369, \quad S_{pp} = 490 \quad \text{and} \quad S_{bb} = 423\frac{5}{7}$$

(c) Find the product moment correlation coefficient, r, for these 7 patients.

(2)

(d) Without any further calculations, state how r would differ from your answer in part (c) if it was calculated for all 8 patients.
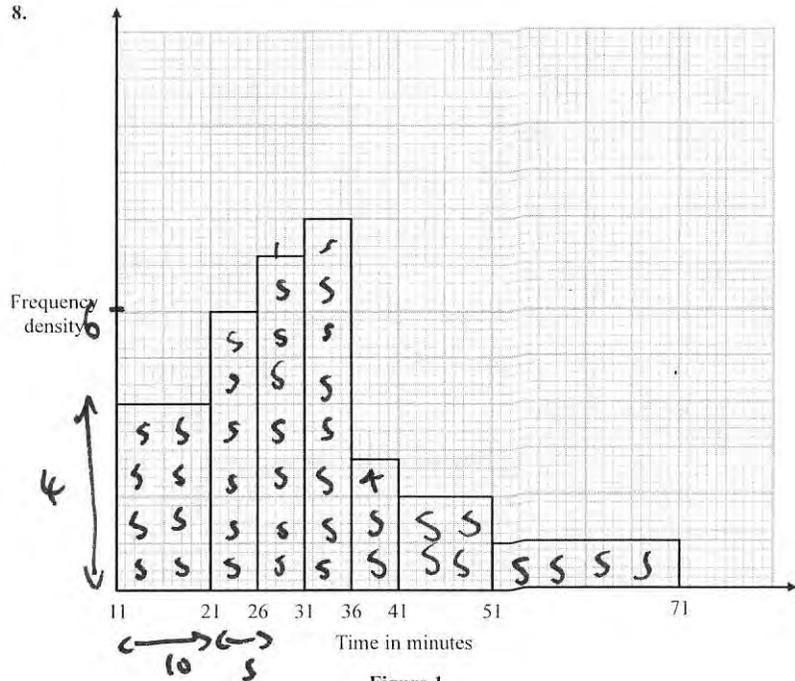
(1)

b) H is clearly an outlier, which doesn't fit into the pattern of the other results. If included it will skew the data.

c) $r = \dfrac{S_{bp}}{\sqrt{S_{bb} \times S_{pp}}} = \dfrac{0.81}{2}$

d) PMCC would reduce

8.

Figure 1

Frequency density

Time in minutes

The histogram in Figure 1 summarises the times, in minutes, that 200 people spent shopping in a supermarket.

(a) Give a reason to justify the use of a histogram to represent these data.

(1)

Given that 40 people spent between 11 and 21 minutes shopping in the supermarket, estimate

(b) the number of people that spent between 18 and 25 minutes shopping in the supermarket,

(3)

(c) the median time spent shopping in the supermarket by these 200 people.

(2)

The mid-point of each bar is represented by x and the corresponding frequency by f.

(d) Show that $\sum fx = 6390$

(2)

d) $\sum fx = 16 \times 40 + 23.5 \times 30 + 28.5 \times 36 + 33.5 \times 40 + 38.5 \times 14 + 46 \times 20 + 61 \times 20$

$= 6390$

Question 8 continued

Given that $\sum fx^2 = 238\,430$

(e) for the data shown in the histogram, calculate estimates of

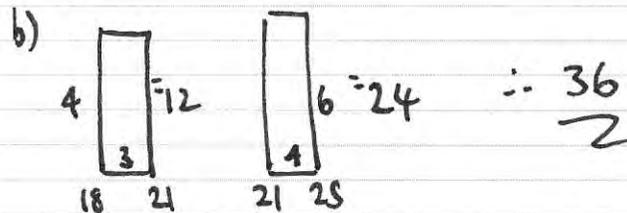(i) the mean,

(ii) the standard deviation.

(3)

A coefficient of skewness is given by $\dfrac{3(\text{mean} - \text{median})}{\text{standard deviation}}$

(f) Calculate this coefficient of skewness for these data.

(1)

The manager of the supermarket decides to model these data with a normal distribution.

(g) Comment on the manager's decision. Give a justification for your answer.

(2)

a) Continuous data

b)

4 $=12$ 6 $=24$ ∴ $\dfrac{36}{2}$

3   4

18 21   21 25

c)

	f
11-21	40
21-26	30
26-31	36
31-36	40
36-41	14
41-51	20
51-71	20

40
70 → Median
106

$\sum f = 200$

26   Q_2 31
├────┼────┤
70   100   106

$\dfrac{Q_2 - 26}{5} = \dfrac{30}{36}$

∴ $Q_2 = 30.2$