1. A meteorologist believes that there is a relationship between the height above sea level, h m, and the air temperature, t °C. Data is collected at the same time from 9 different places on the same mountain. The data is summarised in the table below.

h	1400	1100	260	840	900	550	1230	100	770
t	3	10	20	9	10	13	5	24	16

[You may assume that $\sum h = 7150$, $\sum t = 110$, $\sum h^2 = 7171500$, $\sum t^2 = 1716$, $\sum th = 64980$ and $S_{tt} = 371.56$]

(a) Calculate S_{th} and S_{hh}. Give your answers to 3 significant figures.

(3)

(b) Calculate the product moment correlation coefficient for this data.

(2)

(c) State whether or not your value supports the use of a regression equation to predict the air temperature at different heights on this mountain. Give a reason for your answer.

(1)

(d) Find the equation of the regression line of t on h giving your answer in the form $t = a + bh$.

(4)

(e) Interpret the value of b.

(1)

(f) Estimate the difference in air temperature between a height of 500 m and a height of 1000 m.

(2)

a) $S_{th} = 64980 - 7150 \times 110 \div 9 = -22408.\dot{8}$ $= -22400$ (3sf)

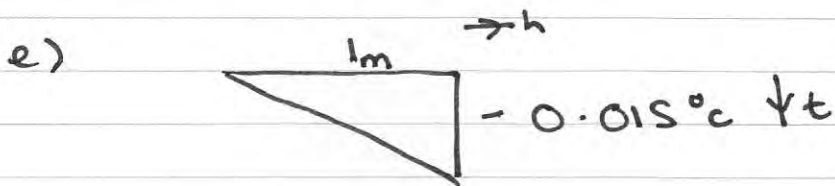$S_{hh} = 7171500 - 7150^2 \div 9 = 1491222.222 = 1490000$ (3sf)

b) PMCC $= \dfrac{-22408.\dot{8}}{\sqrt{1491222.222 \times 371.56}} = -0.952$

c) yes, r is close to -1, strong evidence to suggest negative correlation exists.

d) $b = \dfrac{S_{th}}{S_{hh}} = \dfrac{1491222.222}{-22408.\dot{8}}$ $\ddot{n} = -0.015027$

$$a = \bar{t} - b\bar{h} = \left(\frac{110}{9}\right) - (-0.015\ldots)\left(\frac{7150}{9}\right) = 24.1604\ldots$$

$$\therefore t = 24.2 - 0.015h$$

e)

$$\rightarrow h$$
$$1m$$
$$-0.015°c \;\uparrow t$$

for each Increase in height of 1m, the temp falls by 0.015 °c.

f) $-0.015 \times 500 = -7.5°c$

7.5°c difference.

2. The marks of a group of female students in a statistics test are summarised in Figure 1
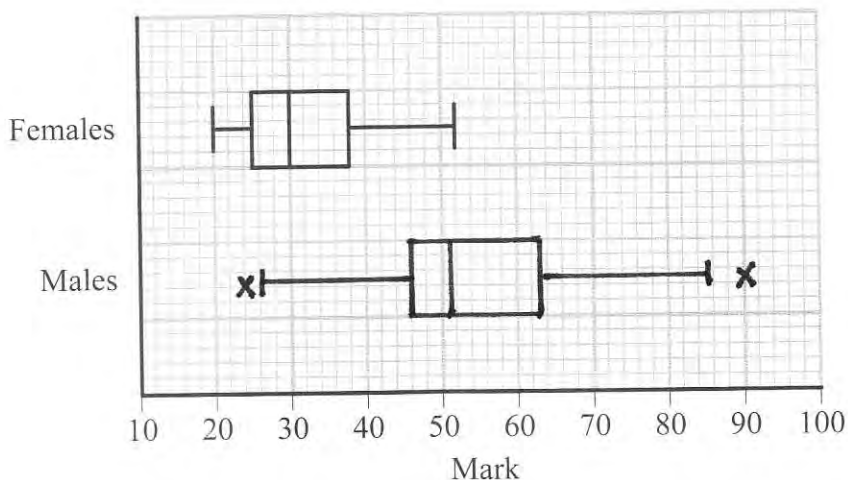

Figure 1

(a) Write down the mark which is exceeded by 75% of the female students.

(1)

The marks of a group of male students in the same statistics test are summarised by the stem and leaf diagram below.

Mark	(2|6 means 26)	Totals
1	4	(1)
2	6	(1)
3	4 4 7	(3)
4	0 6 6 7 7 8	(6)
5	0 0 1 1 1 3 6 7 7	(9)
6	2 2 3 3 3 8	(6)
7	0 0 8	(3)
8	5	(1)
9	0	(1)

(b) Find the median and interquartile range of the marks of the male students.

(3)

An outlier is a mark that is

either more than 1.5 × interquartile range above the upper quartile

or more than 1.5 × interquartile range below the lower quartile.

(c) In the space provided on Figure 1 draw a box plot to represent the marks of the male students, indicating clearly any outliers.

(5)

(d) Compare and contrast the marks of the male and the female students.

(2)

a) 25

b) $n = 31$ Q_1 $\frac{1}{4}n = 7.75 \rightarrow 8^{th} = 46$

Q_2 $\frac{1}{2}n = 15.5 \rightarrow 16^{th} = 51$

Q_3 $\frac{3}{4}n = 23.25 \rightarrow 24^{th} = 63$

∴ median $= 51$, IQR $= 17$

c) lower limit $= 46 - 1.5 \times 17 = 20.5$, 14 is an outlier.

upper limit $= 63 + 1.5 \times 17 = 88.5$, 90 is an outlier.

d) Males achieved a higher mark on average

Male marks were more dispersed.

3. In a company the 200 employees are classified as full-time workers, part-time workers or contractors.
The table below shows the number of employees in each category and whether they walk to work or use some form of transport.

	Walk	Transport	total
Full-time worker	2	8	10
Part-time worker	35	75	110
Contractor	30	50	80
total	67	133	200

The events F, H and C are that an employee is a full-time worker, part-time worker or contractor respectively. Let W be the event that an employee walks to work.

An employee is selected at random.

Find

(a) $P(H)$

(2)

(b) $P([F \cap W]')$

(2)

(c) $P(W \mid C)$

(2)

Let B be the event that an employee uses the bus.
Given that 10% of full-time workers use the bus, 30% of part-time workers use the bus and 20% of contractors use the bus,
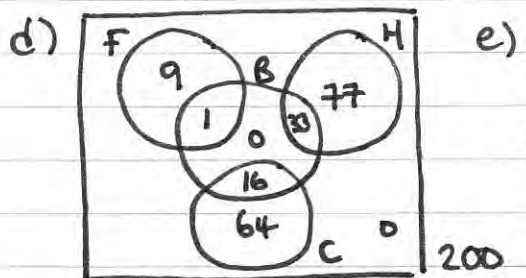
(d) draw a Venn diagram to represent the events F, H, C and B,

(4)

(e) find the probability that a randomly selected employee uses the bus to travel to work.

(2)

a) $P(H) = \dfrac{110}{200}$

b) $P(F \cap W) = \dfrac{2}{200}$ $\therefore$ $P([F \cap W]') = \dfrac{198}{200}$

c) $P(W \mid C) = \dfrac{30}{80}$

d)

e) $\dfrac{50}{200}$

4. The following table summarises the times, t minutes to the nearest minute, recorded for a group of students to complete an exam.

t	15.5	23	28	33	40.5	53
Time (minutes) t	11 – 20	21 – 25	26 – 30	31 – 35	36 – 45	46 – 60
Number of students f	62	88	16	13	11	10

(handwritten cumulative values: 0, 62, 150, 166, 179, 190, 200)

[You may use $\sum ft^2 = 134281.25$]

(a) Estimate the mean and standard deviation of these data.

(5)

(b) Use linear interpolation to estimate the value of the median.

(2)

(c) Show that the estimated value of the lower quartile is 18.6 to 3 significant figures.

(1)

(d) Estimate the interquartile range of this distribution.

(2)

(e) Give a reason why the mean and standard deviation are not the most appropriate summary statistics to use with these data.

(1)

The person timing the exam made an error and each student actually took 5 minutes less than the times recorded above. The table below summarises the actual times.

Time (minutes) t	6 – 15	16 – 20	21 – 25	26 – 30	31 – 40	41 – 55
Number of students f	62	88	16	13	11	10

(f) Without further calculations, explain the effect this would have on each of the estimates found in parts (a), (b), (c) and (d).
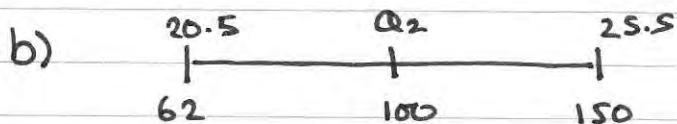
(3)

a) $\sum ft = 4837.5$ $\bar{t} = \dfrac{4837.5}{200} = 24.1875$

$= 24.2 \ (3sf)$

$S_{tt} = 134281.25 - 4837.5^2 \div 200$

$= 17274.21875$
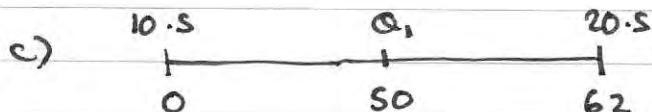
$\sigma_t = \sqrt{\dfrac{17274.2..}{200}} = 9.29 \ (3sf)$

b)

20.5 ——+—— 25.5
62   100   150   (Q_2 at 100)

$$\frac{Q_2 - 20.5}{5} = \frac{38}{88}$$

$$\therefore Q_2 = 22.7 \ (3sf)$$

c)

10.5 ——+—— 20.5
0   50   62   (Q_1 at 50)
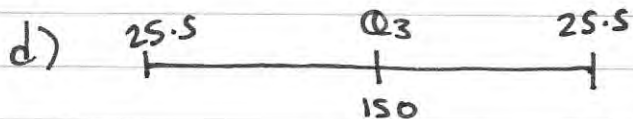
$$\frac{Q_1 - 10.5}{10} = \frac{50}{62}$$

$$\Rightarrow Q_1 = \frac{50}{62} \times 10 + 10.5$$

$$= 18.6 \ (3sf)$$
#

d)

25.5 ——+—— 25.5
150   (Q_3)

$$Q_3 = 25.5$$

$$\therefore IQR = 25.5 - 18.6 = \underline{6.94 \ (3sf)}$$

e)   data is heavily positively skewed

f)   mean, median, quartiles would all also reduce by 5 min but the standard deviation and IQR would not change.

5. A biased die with six faces is rolled. The discrete random variable X represents the score on the uppermost face. The probability distribution of X is shown in the table below.

x	1	2	3	4	5	6
$P(X=x)$	a	a	a	b	b	0.3

(a) Given that $E(X) = 4.2$ find the value of a and the value of b.

(5)

(b) Show that $E(X^2) = 20.4$

(1)

(c) Find $Var(5 - 3X)$

(3)

A biased die with five faces is rolled. The discrete random variable Y represents the score which is uppermost. The cumulative distribution function of Y is shown in the table below.

y	1	2	3	4	5
$F(y)$	$\dfrac{1}{10}$	$\dfrac{2}{10}$	$3k$	$4k$	$5k$

(d) Find the value of k.

(1)

(e) Find the probability distribution of Y.

(3)

Each die is rolled once. The scores on the two dice are independent.

(f) Find the probability that the sum of the two scores equals 2

(2)

a) $E(x) = a + 2a + 3a + 4b + 5b + 1.8 = 4.2$ $\Rightarrow$ $6a + 9b = 2.4$

$(\div 3)$ $2a + 3b = 0.8$

$\Sigma P = 1 \Rightarrow 3a + 2b + 0.3 = 1$ $3a + 2b = 0.7$

$4a + 6b = 1.6$
$9a + 6b = 2.1$
$-5a = -0.5$ $\therefore a = 0.1$ $b = 0.2$

b) $E(x^2) = a + 4a + 9a + 16b + 25b + 10.8$

$= 14a + 41b + 10.8$

$= 1.4 + 8.2 + 10.8 = 20.4$ #

c) $V(x) = E(x^2) - E(x)^2 = 20 \cdot 4 - 4 \cdot 2^2 = 2 \cdot 76$

$V(5-3x) = 9 \times V(x) = \underline{24 \cdot 84}$

d) $5h = 1 \quad \therefore h = 0.2$

e)

y	1	2	3	4	5
$P(y)$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{4}{10}$	$\frac{2}{10}$	$\frac{2}{10}$

f) $P(sum = 2) = P(x=1, y=1) = 0 \cdot 1 \times \frac{1}{10} = \frac{1}{100}$

6. The weight, in grams, of beans in a tin is normally distributed with mean μ and standard deviation 7.8
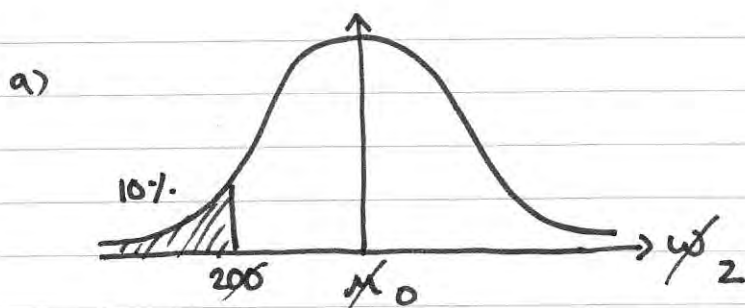
Given that 10% of tins contain less than 200 g, find

(a) the value of μ

(3)

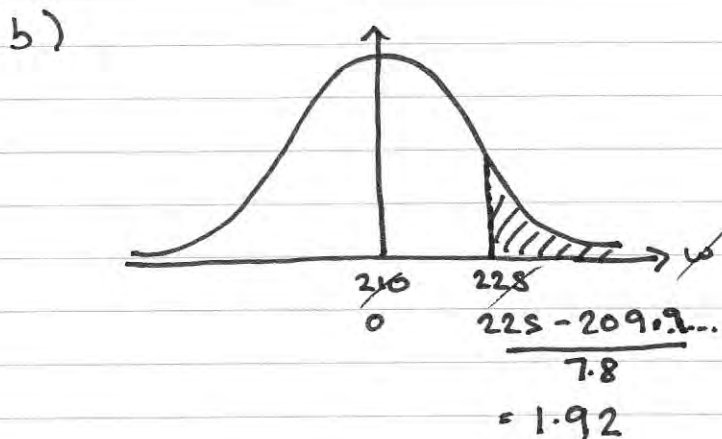(b) the percentage of tins that contain more than 225 g of beans.

(3)

The machine settings are adjusted so that the weight, in grams, of beans in a tin is normally distributed with mean 205 and standard deviation σ.

(c) Given that 98% of tins contain between 200 g and 210 g find the value of σ.
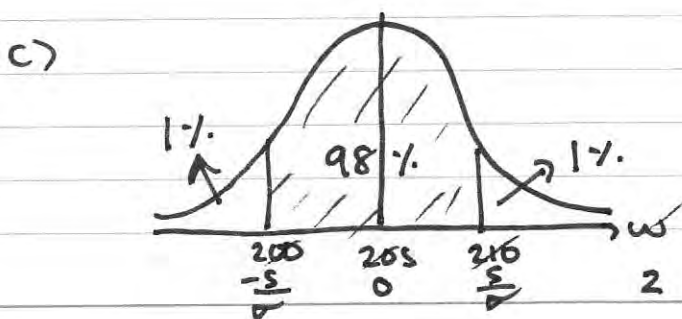
(4)

a)

$$\frac{200-\mu}{7.8} = -1.2816 \qquad \Rightarrow 200-\mu = -9.99648$$

$$\therefore \mu = 209.996 \quad (210)$$

b)

$$\frac{225-209.9...}{7.8}$$

$$= 1.92$$

$$\phi(1.92) = 0.9726$$

$$\therefore P(w > 225)$$

$$= 0.0274$$

c)

$$\therefore \frac{5}{\sigma} = 2.3263$$

$$\Rightarrow \sigma = 2.15$$