1. Nina weighed a random sample of 50 carrots from her shop and recorded the weight, in grams to the nearest gram, for each carrot. The results are summarised below.

Weight of carrot	Frequency (f)	Weight midpoint (x grams)	fx
45–54	5	49.5	247.5
55–59	10	57	570
60–64	22	62	1364
65–74	13	69.5	903.5

(You may use $\sum fx^2 = 192\,102.5$) $\Sigma fx = 3085$

(a) Use linear interpolation to estimate the median weight of these carrots. **(2)**

(b) Find an estimate for the mean weight of these carrots. **(2)**

(c) Find an estimate for the standard deviation of the weights of these carrots. **(2)**

A carrot is selected at random from Nina's shop.

(d) Estimate the probability that the weight of this carrot is more than 70 grams. **(2)**

a) $m = 59.5 + \frac{10}{22} \times 5$

$= 61.77$

b) $\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{3085}{50} = 61.7$

d) Linear interpolation:

$\frac{f}{13} = \frac{74.5 - 70}{74.5 - 64.5}$

$f = 5.85$

$P(W > 70) = \frac{5.85}{50} = 0.117$

c) $\sigma^2 = \frac{\Sigma fx^2}{\Sigma f} - \left(\frac{\Sigma fx}{\Sigma f}\right)^2 = \frac{192102.5}{50} - 61.7^2$

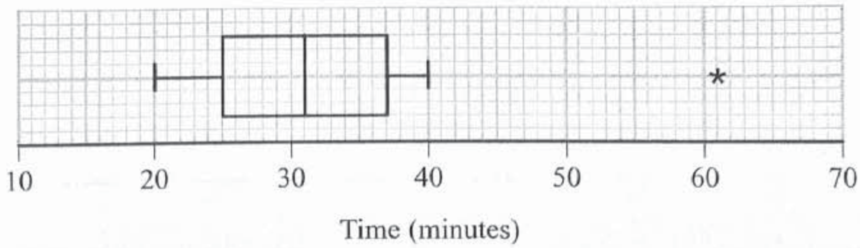
$\sigma = 5.93$

d) Normal approximation: $W \sim N(61.7, 5.93^2)$

$P(W > 70) = P\left(Z > \frac{70 - 61.7}{5.93}\right) = P(Z > 1.40) = 1 - \Phi(1.4)$

$= 0.0808$

2. The box plot shows the times, t minutes, it takes a group of office workers to travel to work.



(a) Find the range of the times.

$61 - 20 = 41$

(1)

(b) Find the interquartile range of the times.

$37 - 25 = 12$

(1)

(c) Using the quartiles, describe the skewness of these data. Give a reason for your answer.

Symmetrical because $Q_3 - Q_2 = Q_2 - Q_1$

(2)

Chetna believes that house prices will be higher if the time to travel to work is shorter. She asks a random sample of these office workers for their house prices £x, where x is measured in thousands, and obtains the following statistics

$$S_{xx} = 5514 \qquad S_{xt} = 10 \qquad S_{tt} = 1145.6$$

(d) Calculate the product moment correlation coefficient between x and t.

$$r = \frac{S_{xt}}{\sqrt{S_{xx} S_{tt}}} = \frac{10}{\sqrt{5314 \times 1143.6}} = 0.0040$$

(2)

(e) State, giving a reason, whether or not your correlation coefficient supports Chetna's belief.

r is close to zero. No correlation between time and price. Claim not supported

(2)

Adam and Betty are part of the group of office workers and they have both moved house.
Adam's time to travel to work changes from 32 minutes to 36 minutes.
Betty's time to travel to work changes from 38 minutes to 58 minutes.
Outliers are defined as values that are more than 1.5 times the interquartile range above the upper quartile.

(f) Showing all necessary calculations, determine how the box plot of times to travel to work will change and draw a new box plot on the grid on page 5.

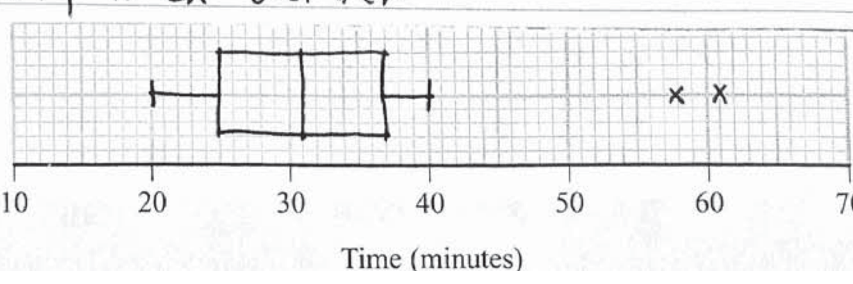
(3)

$Q_3 = 37$, $Q_2 = 31$, $Q_1 = 25$

Changes do not affect quartiles since they remain in the same intervals

Upper limit: $37 + 1.5(12) = 55$

$\Rightarrow$ Betty is an outlier



blank

3. At a school athletics day, the distances, in metres, achieved by students in the long jump are modelled by the normal distribution with mean 3.3 m and standard deviation 0.6 m

(a) Find an estimate for the proportion of students who jump less than 2.5 m

(3)

The long jump competition consists of 2 jumps. All the students can take part in the first jump and the 40% who jump the greatest distance in their first jump qualify for the second jump.

(b) Find an estimate for the minimum distance achieved in the first jump in order to qualify for the second jump.
Give your answer correct to 4 significant figures.

(3)

(c) Find an estimate for the median distance achieved in the first jump by those who qualify for the second jump.

(3)

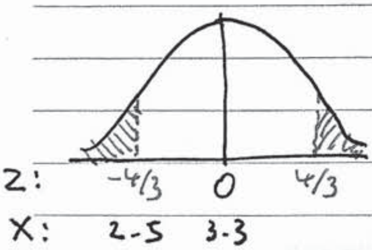
The distance of the second jump is independent of the distance of the first jump and is modelled with the same normal distribution. Students who jump a distance greater than 4.1 m in their second jump receive a certificate.

At the start of the long jump competition, a student is selected at random.

(d) Find the probability that this student will receive a certificate.

(3)

a) $X \sim N(3.3, 0.6^2)$ $\quad Z = \dfrac{X-\mu}{\sigma}$ $\quad P(X<2.5) = P(Z<-\frac{4}{3})$

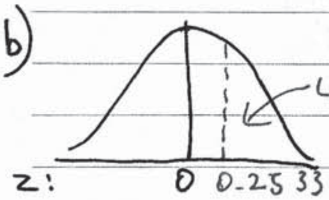


$= \dfrac{2.5-3.3}{0.6}$ $\quad = 1 - P(Z<\frac{4}{3})$

$= -\dfrac{4}{3}$ $\quad = 1 - 0.9082$

$= 0.0918$

b)

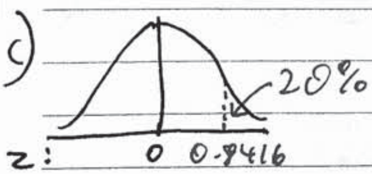


From percentage table,
$z = 0.2533$

$\Rightarrow 0.2533 = \dfrac{X-3.3}{0.6}$

$X = 3.452$ m

c)



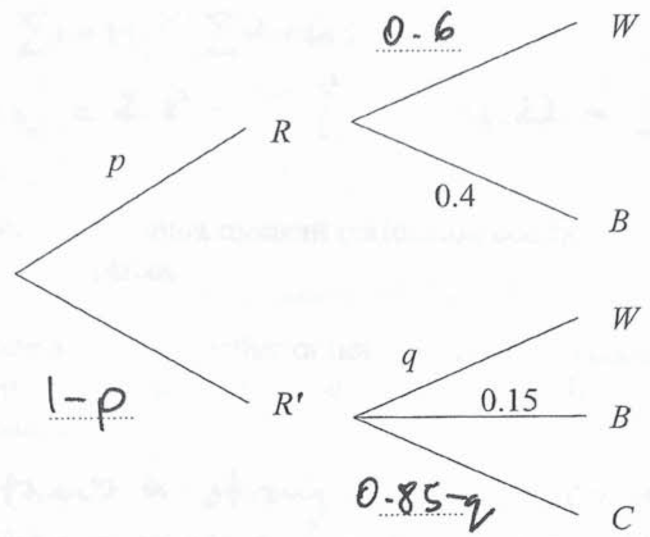
$0.8416 = \dfrac{X-3.3}{0.6}$

$X = 3.805$

d) $Z_2 = \dfrac{4.1-3.3}{0.6} = \dfrac{4}{3}$ $\quad P(X_2>4.1) = 0.0918$ $\quad P(\text{cert}) = 0.0918 \times 0.4$

$= 0.0367$

4. The partially completed tree diagram, where p and q are probabilities, gives information about Andrew's journey to work each day.



R represents the event that it is raining

W represents the event that Andrew walks to work

B represents the event that Andrew takes the bus to work

C represents the event that Andrew cycles to work

Given that $P(B) = 0.26$ $= (p \times 0.4) + (1-p)(0.15)$

(a) find the value of p

$0.11 = 0.25p$

$p = 0.44$

(3)

Given also that $P(R' \mid W) = 0.175$

(b) find the value of q

(4)

(c) Find the probability that Andrew cycles to work.

(2)

$P(C) = (1-0.44)(0.85-0.1) = 0.42$

Given that Andrew did not cycle to work on Friday,

(d) find the probability that it was raining on Friday.

(3)

b) $P(R' \mid W) = \dfrac{P(R' \cap W)}{P(W)}$

$0.175 = \dfrac{(1-0.44)(q)}{(0.44)(0.6)+(1-0.44)(q)}$

$0.0462 + 0.098q = 0.56q$

$q = 0.1$

d) $P(R \mid C') = \dfrac{P(R \cap C')}{P(C')}$

$= \dfrac{0.44}{1-0.42}$

$= 0.7586$

5. Tomas is studying the relationship between temperature and hours of sunshine in *Seapron*. He records the midday temperature, t °C, and the hours of sunshine, s hours, for a random sample of 9 days in October. He calculated the following statistics

$$\sum s = 15 \qquad \sum s^2 = 44.22 \qquad \sum t = 127 \qquad S_{tt} = 10.89$$

(a) Calculate S_{ss} $= \Sigma s^2 - \frac{(\Sigma s)^2}{n} = 44.22 - \frac{15^2}{9} = 19.22$ **(2)**

Tomas calculated the product moment correlation coefficient between s and t to be 0.832 correct to 3 decimal places.

(b) State, giving a reason, whether or not this correlation coefficient supports the use of a linear regression model to describe the relationship between midday temperature and hours of sunshine. **(1)**

Yes, there's a strong positive correlation

(c) State, giving a reason, why the hours of sunshine would be the explanatory variable in a linear regression model between midday temperature and hours of sunshine. **(1)**

Temperature depends on sunshine

(d) Find S_{st} $\quad r = \frac{S_{st}}{\sqrt{S_{tt} S_{ss}}} \qquad S_{st} = 0.832\sqrt{10.89 \times 19.22} = 12.04$ **(3)**

(e) Calculate a suitable linear regression equation to model the relationship between midday temperature and hours of sunshine. **(4)**

(f) Calculate the standard deviation of s **(1)**

$\sigma = \sqrt{\frac{\Sigma s^2}{n} - \left(\frac{\Sigma s}{n}\right)^2} = \sqrt{\frac{44.22}{9} - \left(\frac{15}{9}\right)^2} = 1.4614$

Tomas uses this model to estimate the midday temperature in *Seapron* for a day in October with 5 hours of sunshine.

(g) State the value of Tomas' estimate. **(1)**

$t = 13.07 + 0.6264(5) = 16.20$

Given that the values of s are all within 2 standard deviations of the mean,

(h) comment, giving your reason, on the reliability of this estimate. **(2)**

e) $b = \frac{S_{st}}{S_{ss}} = \frac{12.04}{19.22} = 0.6264$

$a = \bar{t} - b\bar{s} = \frac{127}{9} - 0.6264\left(\frac{15}{9}\right) = 13.07$

$\therefore t = 13.07 + 0.6264s$

h) $\bar{s} + 2\sigma = \frac{15}{9} + 2(1.4614) = 4.59$ Not reliable because its extrapolating

6. A biased coin has probability 0.4 of showing a head. In an experiment, the coin is spun until a head appears. If a head has not appeared after 4 spins, the coin is not spun again. The random variable X represents the number of times the coin is spun.

For example, $X = 3$ if the first two spins do not show a head but the third spin does show a head. The coin would not then be spun a fourth time since the coin has already shown a head.

(a) Show that $P(X = 3) = 0.144$ **(1)**

$P(X=3) = P(TTH) = 0.6 \times 0.6 \times 0.4 = 0.144$

The table gives some values for the probability distribution of X

x	1	2	3	4
$P(X = x)$	0.4	0.24	0.144	0.216

(b) (i) Write down the value of $P(X = 1)$ 0.4

(ii) Find $P(X = 4)$ $1-(0.4+0.24+0.144) = 0.216$ **(3)**

(c) Find $E(X)$ $= 1(0.4)+2(0.24)+3(0.144)+4(0.216)$
$= 2.176$ **(2)**

(d) Find $\mathrm{Var}(X)$ $= 1^2(0.4)+2^2(0.24)+3^2(0.144)+4^2(0.216) - 2.176^2$
$= 1.3770$ **(3)**

The random variable H represents the number of heads obtained when the coin is spun in the experiment.

(e) Explain why H can only take the values 0 and 1 and find the probability distribution of H. Once a heads is obtained, tossing stops **(2)**

$H=0$ when $(TTTT)$: $0.6^4 = 0.1296$

h	0	1
$P(H=h)$	0.1296	0.8704

(f) Write down the value of

(i) $P(\{X = 3\} \cap \{H = 0\}) = 0$

(ii) $P(\{X = 4\} \cap \{H = 0\}) = 0.1296$ **(2)**

The random variable $S = X + H$

(g) Find the probability distribution of S **(4)**

H: $X=1, H=1, S=2$ (0.4)
TH: $X=2, H=1, S=3$ (0.24)
TTH: $X=3, H=1, S=4$ (0.144)
TTTH: $X=4, H=1, S=5$
TTTT: $X=4, H=0, S=4$ (0.1296)

$1-(0.4+0.24+0.144+0.1296) = 0.0864$
$0.144+0.1296 = 0.2736$

s	2	3	4	5
$P(S=s)$	0.4	0.24	0.2736	0.0864