SI JAN 10

1. Gary compared the total attendance, x, at home matches and the total number of goals, y, scored at home during a season for each of 12 football teams playing in a league. He correctly calculated:

$$S_{xx} = 1022\,500 \quad S_{yy} = 130.9 \quad S_{xy} = 8825$$

(a) Calculate the product moment correlation coefficient for these data. **(2)**

(b) Interpret the value of the correlation coefficient. **(1)**

Helen was given the same data to analyse. In view of the large numbers involved she decided to divide the attendance figures by 100. She then calculated the product moment correlation coefficient between $\frac{x}{100}$ and y.

(c) Write down the value Helen should have obtained. **(1)**

(a) PMCC $= \dfrac{S_{xy}}{\sqrt{S_{xx} \times S_{yy}}} = \dfrac{8825}{\sqrt{1022500 \times 130.9}} \Rightarrow r = 0.763$ (3sf)

(b) evidence to suggest that positive correlation exists $\Rightarrow$ the higher the total attendance the higher the total goals scored.

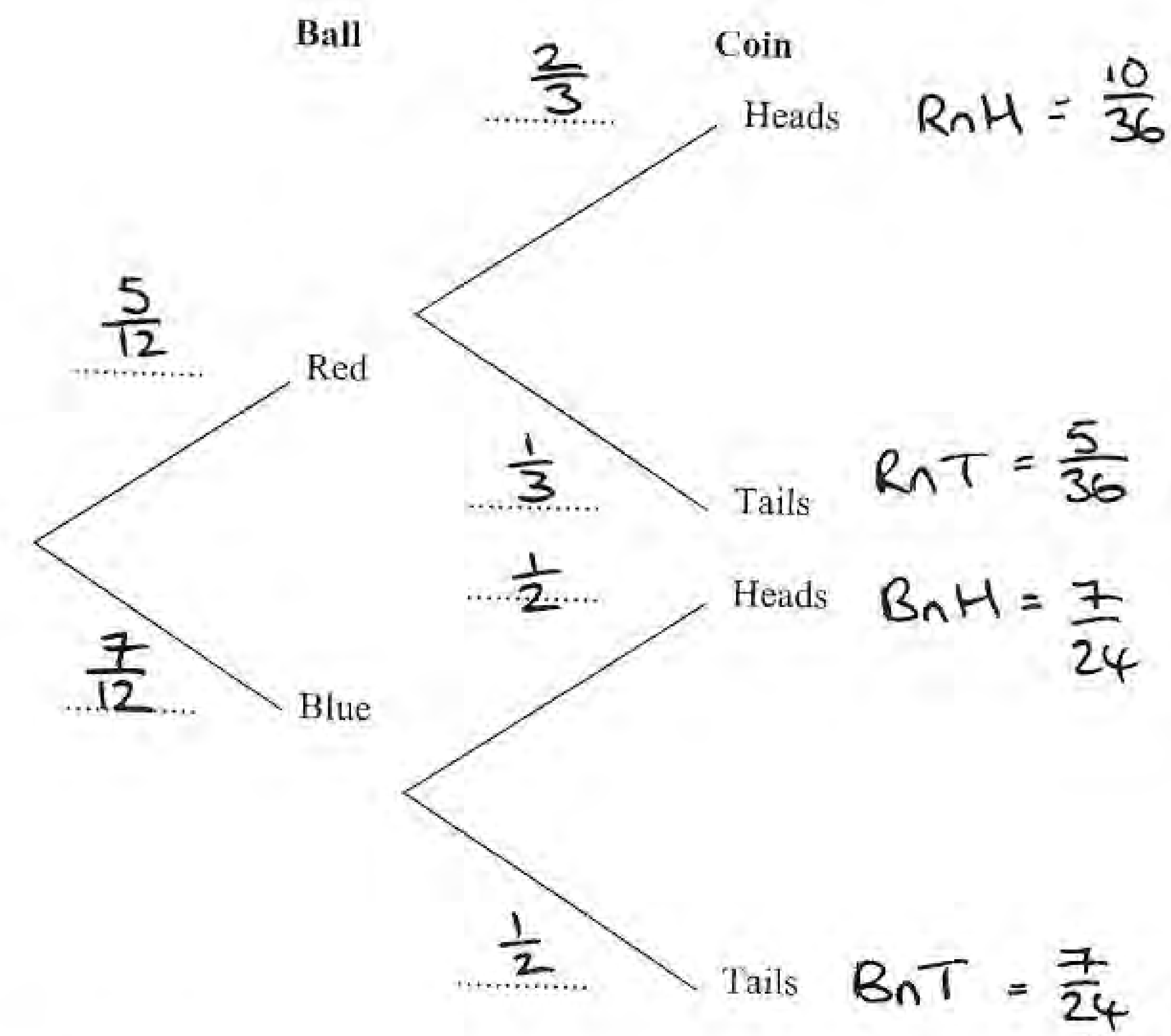
c) 0.763 coding does not affect PMCC

2. An experiment consists of selecting a ball from a bag and spinning a coin. The bag contains 5 red balls and 7 blue balls. A ball is selected at random from the bag, its colour is noted and then the ball is returned to the bag.

When a red ball is selected, a biased coin with probability $\frac{2}{3}$ of landing heads is spun.

When a blue ball is selected a fair coin is spun.

(a) Complete the tree diagram below to show the possible outcomes and associated probabilities.



(2)

Shivani selects a ball and spins the appropriate coin.

(b) Find the probability that she obtains a head. $P(R \cap H) + P(B \cap H) = \frac{10}{36} + \frac{7}{24} = \frac{41}{72}$ **(2)**

Given that Tom selected a ball at random and obtained a head when he spun the appropriate coin,

(c) find the probability that Tom selected a red ball. $P(R|H) = \dfrac{P(R \cap H)}{P(H)} = \dfrac{\frac{10}{36}}{\frac{41}{72}}$ **(3)**

$P(R|H) = \frac{20}{41}$

Shivani and Tom each repeat this experiment.

(d) Find the probability that the colour of the ball Shivani selects is the same as the colour of the ball Tom selects. RR or BB $= \frac{5}{12} \times \frac{5}{12} + \frac{7}{12} \times \frac{7}{12} = \frac{37}{72}$ **(3)**

3. The discrete random variable X has probability distribution given by

x	-1	0	1	2	3
$P(X = x)$	$\frac{1}{5}$	a	$\frac{1}{10}$	a	$\frac{1}{5}$

where a is a constant.

(a) Find the value of a. $\sum P(X=x)=1 \Rightarrow \frac{1}{5}+a+\frac{1}{10}+a+\frac{1}{5}=1$ **(2)**

$2a=\frac{1}{2} \Rightarrow a=\frac{1}{4}$

(b) Write down $E(X)$. **(1)**

(c) Find $\mathrm{Var}(X)$. **(3)**

b) $E(X)=-1\times\frac{1}{5}+1\times\frac{1}{10}+2\times\frac{1}{4}+3\times\frac{1}{5}$

$E(X)=1$

The random variable $Y = 6 - 2X$

(d) Find $\mathrm{Var}(Y)$. **(2)**

$E(X^2)=(-1)^2\times\frac{1}{5}+1^2\times\frac{1}{10}+2^2\times\frac{1}{4}+3^2\times\frac{1}{5}$

$E(X^2)=3.1$

(e) Calculate $P(X \geqslant Y)$. **(3)**

c) $V(X)=E(X^2)-E(X)^2=3.1-1^2=2.1$

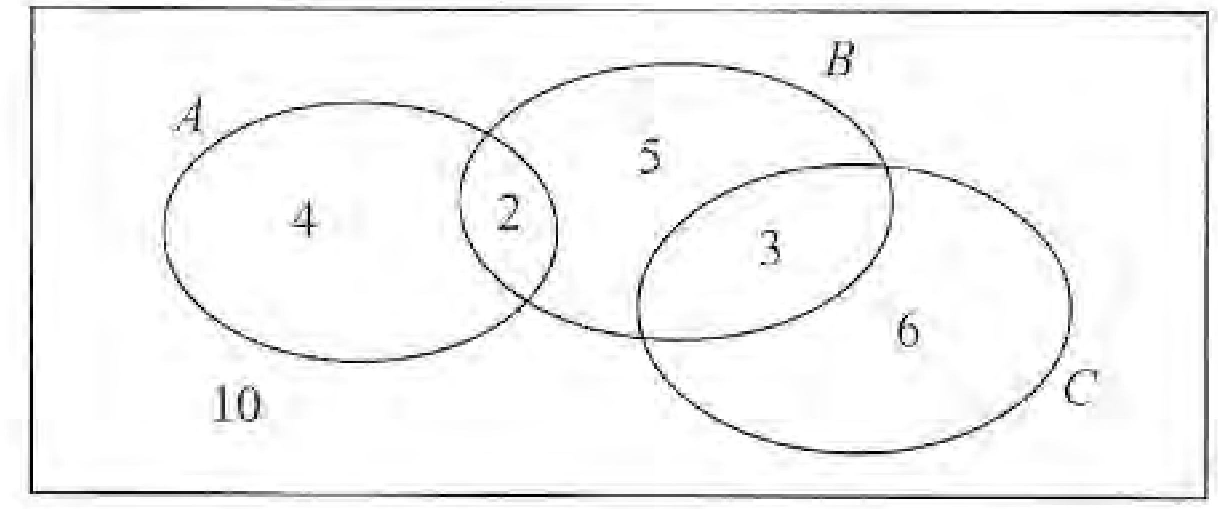
d) $V(Y)=V(6-2X)=(-2)^2\times V(X)=4\times 2.1=8.4$

e)

x	-1	0	1	2 ✓	3 ✓
y	8	6	4	2	0
				$x\geq y$	$x \geq y$

$P(X\geq Y)=P(X=2)+P(X=3)=\frac{1}{4}+\frac{1}{5}=\frac{9}{20}$

4. The Venn diagram in Figure 1 shows the number of students in a class who read any of 3 popular magazines A, B and C.



30

Figure 1

One of these students is selected at random.

(a) Show that the probability that the student reads more than one magazine is $\frac{1}{6}$. **(2)**

(b) Find the probability that the student reads A or B (or both). **(2)**

(c) Write down the probability that the student reads both A and C. **(1)**

Given that the student reads at least one of the magazines,

(d) find the probability that the student reads C. **(2)**

(e) Determine whether or not reading magazine B and reading magazine C are statistically independent. **(3)**

a) $P(A\cap B)+P(B\cap C)=\frac{2}{30}+\frac{3}{30}=\frac{5}{30}=\frac{1}{6}$ qed.

b) $P(A\cup B)=\frac{4+2+5+3}{30}=\frac{14}{30}=\frac{7}{15}$

c) $P(A\cap C)=0$ they are mutually exclusive

d) $P(C\mid A\cup B\cup C)=\frac{P(C\cap(A\cup B\cup C))}{P(A\cup B\cup C)}=\frac{\left(\frac{9}{30}\right)}{\left(\frac{20}{30}\right)}=\frac{9}{20}$

e) $P(B)=\frac{10}{30}=\frac{1}{3}$ $P(C)=\frac{9}{30}=\frac{3}{10}$ $P(B)\times P(C)=\frac{3}{30}=\frac{1}{10}$

$P(B\cap C)=\frac{3}{30}=\frac{1}{10}$ $P(B\cap C)=P(B)\times P(C)$ So Independent

5. A teacher selects a random sample of 56 students and records, to the nearest hour, the time spent watching television in a particular week.

median

Hours	1–10	11–20	21–25	26–30	31–40	41–59
Frequency	6	15	21 11 32	13	8	3
Mid-point	5.5	15.5	23	28	35.5	50

(a) Find the mid-points of the 21–25 hour and 31–40 hour groups. **(2)**

A histogram was drawn to represent these data. The 11–20 group was represented by a bar of width 4 cm and height 6 cm.

(b) Find the width and height of the 26–30 group. **(3)**

(c) Estimate the mean and standard deviation of the time spent watching television by these students. **(5)**

(d) Use linear interpolation to estimate the median length of time spent watching television by these students. **(2)**

The teacher estimated the lower quartile and the upper quartile of the time spent watching television to be 15.8 and 29.3 respectively.

(e) State, giving a reason, the skewness of these data. **(2)**

b)

F=15 A=24 cw=10, 6cm

10.5 — 20.5

$\div 2.5$

4cm

$A \propto F \Rightarrow A = k \times F \quad 24 = k \times 15 \Rightarrow k = 1.6$

Area = 1.6 × freq.

width = cw ÷ 2.5

F=13, cw=5, 30.5

width = 5 ÷ 2.5 = 2cm

Area = 1.6 × 13 = 20.8 cm^2

height = 20.8 ÷ 2 = 10.4cm

width = 2cm height = 10.4cm

c) $\Sigma fx = 6 \times 5.5 + 15 \times 15.5 + \ldots = 1316.5$

$\Sigma fx^2 = 6 \times 5.5^2 + 15 \times 15.5^2 + \ldots = 37378.25$

$\bar{x} = \frac{\Sigma fx}{n} = \frac{1316.5}{56} = 23.50892857 \simeq 23.5 \text{ (3sf)}$

$\sigma_x = \sqrt{\frac{\Sigma fx^2}{n} - \bar{x}^2} = \sqrt{\frac{37378.25}{56} - \left(\frac{1316.5}{56}\right)^2}$

$\sigma_x = 10.7144308\ldots \simeq 10.7 \text{ (3sf)}$

d) $\frac{1}{2}n = 28 \Rightarrow 21–25$

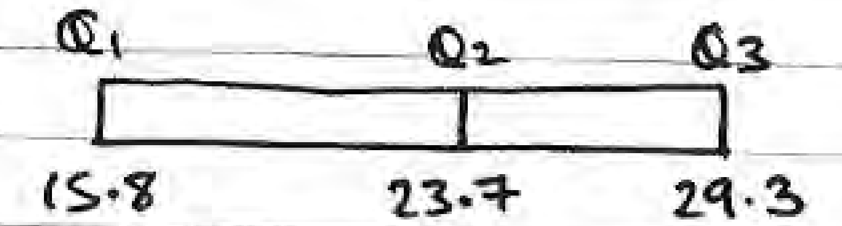
20.5 — Q_2 — 25.5

21 — 28 — 32

$\frac{Q_2 - 20.5}{5} = \frac{7}{11}$

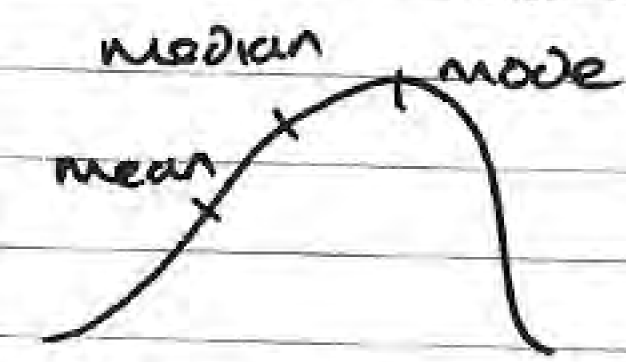
$Q_2 = 23.6\dot{8}\dot{1} \simeq 23.7 \text{ (3sf)}$

e)



$Q_2 - Q_1 > Q_3 - Q_2$
(7.9)   (5.6)

Slightly negatively skewed.



negative skew

mean < median
(23.5)   (23.7)

Slight negative skew.

6. A travel agent sells flights to different destinations from *Beerow* airport. The distance d, measured in 100 km, of the destination from the airport and the fare £f are recorded for a random sample of 6 destinations.

Destination	A	B	C	D	E	F
d	2.2	4.0	6.0	2.5	8.0	5.0
f	18	20	25	23	32	28

$\Sigma d = 27.7$
$\Sigma f = 146$

[You may use $\sum d^2 = 152.09$ $\sum f^2 = 3686$ $\sum fd = 723.1$]

(a) Using the axes below, complete a scatter diagram to illustrate this information. **(2)**

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

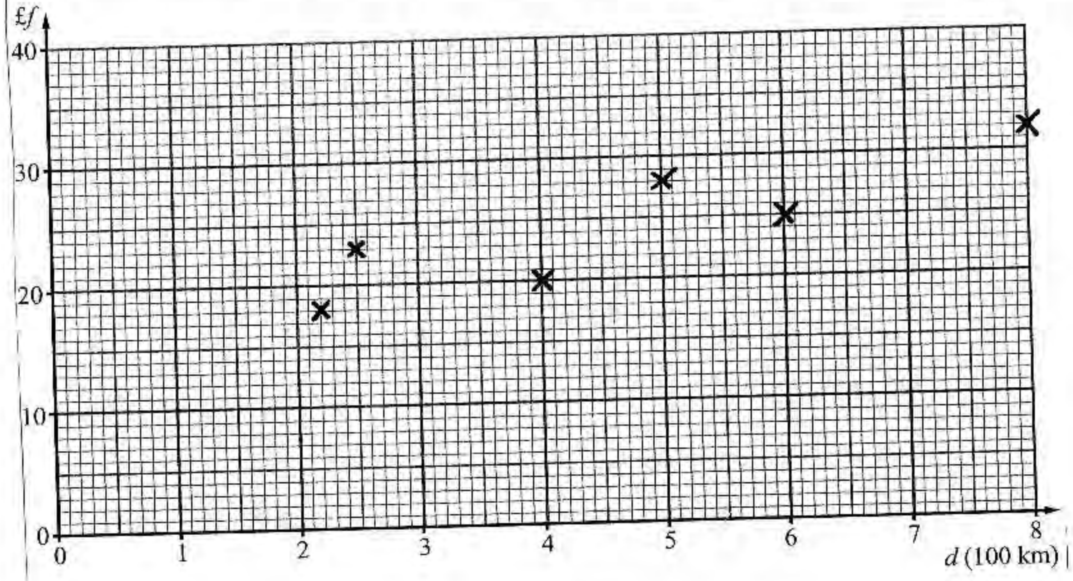
(c) Calculate S_{dd} and S_{fd} **(4)**

(d) Calculate the equation of the regression line of f on d giving your answer in the form $f = a + bd$. **(4)**

(e) Give an interpretation of the value of b. **(1)**

Jane is planning her holiday and wishes to fly from *Beerow* airport to a destination t km away. A rival travel agent charges 5p per km.

(f) Find the range of values of t for which the first travel agent is cheaper than the rival. **(2)**



b) crosses are following a linear pattern, crosses would be fairly close to a line of best fit.

Evidence to suggest positive correlation, the further the destination the higher the fare.

c) $S_{dd} = \Sigma d^2 - \frac{(\Sigma d)^2}{n} = 152.09 - \frac{27.7^2}{6} = 24.208\dot{3}$

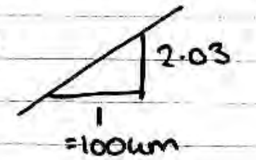
$S_{fd} = \Sigma fd - \frac{(\Sigma f)(\Sigma d)}{n} = 723.1 - \frac{(27.7)(146)}{6} = 49.0\dot{6}$

$S_{dd} \simeq \underline{24.2}$ (3sf) $S_{fd} \simeq \underline{49.1}$ (3sf)

d) $b = \frac{S_{fd}}{S_{dd}} = 2.026850285 \simeq 2.03$ (3sf)

$a = \bar{f} - b\bar{d} = \frac{146}{6} - 2.0268... \times \frac{27.7}{6} = 14.97604131$

$a \simeq 15.0$ (3sf) $\underline{f = 14.98 + 2.03d}$

e) 

an extra cost of £2.03 for each additional 100km.

f) $f_2 = 5d$ 5p per km $\Rightarrow$ £5.00 per 100km

$f_1 = f_2 \Rightarrow 5d = 14.98 + 2.03d \Rightarrow 2.97d = 14.98$

$d = 5.04377 \simeq 5.04$ (3sf)

the first travel agent is cheaper when $\underline{t > 504\text{km}}$

7. The distances travelled to work, D km, by the employees at a large company are normally distributed with $D \sim N(30, 8^2)$.

$\mu = 30 \quad \sigma = 8$

(a) Find the probability that a randomly selected employee has a journey to work of more than 20 km.

(3)

(b) Find the upper quartile, Q_3, of D.

(3)

(c) Write down the lower quartile, Q_1, of D.

(1)

An outlier is defined as any value of D such that $D < h$ or $D > k$ where

$$h = Q_1 - 1.5 \times (Q_3 - Q_1) \quad \text{and} \quad k = Q_3 + 1.5 \times (Q_3 - Q_1)$$

(d) Find the value of h and the value of k.

(2)

An employee is selected at random.

(e) Find the probability that the distance travelled to work by this employee is an outlier.

(3)

a) $P(D > 20) \Rightarrow P\left(Z > \frac{20-30}{8}\right) = P(Z > -1.25)$

$= P(Z < 1.25) = \Phi(1.25) = 0.8944$

b) $P(D < Q_3) = 0.7500 \Rightarrow P\left(Z < \frac{Q_3 - 30}{8}\right) = 0.7500$

$\Phi\left(\frac{Q_3 - 30}{8}\right) = 0.7500$

$\Phi(0.67) = 0.7486$
$\Phi(0.68) = 0.7517$ $\Rightarrow \Phi(0.675) \simeq 0.7500$

using (0.67) is fine. $\frac{Q_3 - 30}{8} = 0.675$

$\underline{Q_3 = 35.4}$

c) Normal distribution $\Rightarrow P(D > Q_3) = P(D < Q_1)$

$Q_1 = 30 - 5.4 = \underline{24.6}$

d) $Q_3 - Q_1 = IQR = 35.4 - 24.6 = 10.8$

$h = 24.6 - 1.5 \times 10.8 = 8.4$
$k = 35.4 + 1.5 \times 10.8 = 51.6$

e) outlier if $D > k$ or $D < h$

Normal distribution $\Rightarrow P(D > 51.6) = P(D < 8.4)$

$2 \times P(D > 51.6) = 2 \times P\left(Z > \frac{51.6 - 30}{8}\right) = 2 \times P(Z > 2.7)$

$= 2 \times (1 - 0.9965) = 2 \times 0.0035$

$= \underline{0.007}$