

SI JAN 11

Leave blank

1. A random sample of 50 salmon was caught by a scientist. He recorded the length l cm and weight w kg of each salmon.

The following summary statistics were calculated from these data.

$$\sum l = 4027 \quad \sum l^2 = 327754.5 \quad \sum w = 357.1 \quad \sum lw = 29330.5 \quad S_{ww} = 289.6$$

(a) Find S_{ll} and S_{ww} (3)

(b) Calculate, to 3 significant figures, the product moment correlation coefficient between l and w . (2)

(c) Give an interpretation of your coefficient. (1)

$$a) S_{ll} = \sum l^2 - \frac{(\sum l)^2}{n} = 327754.5 - \frac{4027^2}{50} = \underline{3419.92}$$

$$S_{ww} = \sum w^2 - \frac{(\sum w)^2}{n} = 29330.5 - \frac{357.1^2}{50} = \underline{569.666}$$

$$b) r = \frac{S_{lw}}{\sqrt{S_{ll} \times S_{ww}}} = \frac{569.666}{\sqrt{3419.92 \times 289.6}} = \underline{0.572}$$

c) Slight evidence to suggest positive correlation \Rightarrow longer the length, bigger they weigh.

2. Keith records the amount of rainfall, in mm, at his school, each day for a week. The results are given below.

2.8 5.6 2.3 9.4 0.0 0.5 1.8

Jenny then records the amount of rainfall, x mm, at the school each day for the following 21 days. The results for the 21 days are summarised below.

$$\sum x = 84.6$$

(a) Calculate the mean amount of rainfall during the whole 28 days. (2)

Keith realises that he has transposed two of his figures. The number 9.4 should have been 4.9 and the number 0.5 should have been 5.0
Keith corrects these figures.

(b) State, giving your reason, the effect this will have on the mean. (2)

$$a) \text{ Keith total} = 22.4 \quad \text{mean} = \frac{22.4 + 84.6}{28}$$

$$\text{ Jenny total} = 84.6$$

$$\text{mean} = \underline{3.82 \text{ mm (3sf)}}$$

$$b) \underline{\text{no effect}} \quad 4.9 - 9.4 = 4.5$$

$$5.0 - 0.5 = 4.5$$

One was 4.5 too high and the other was 4.5 too low.

3. Over a long period of time a small company recorded the amount it received in sales per month. The results are summarised below.

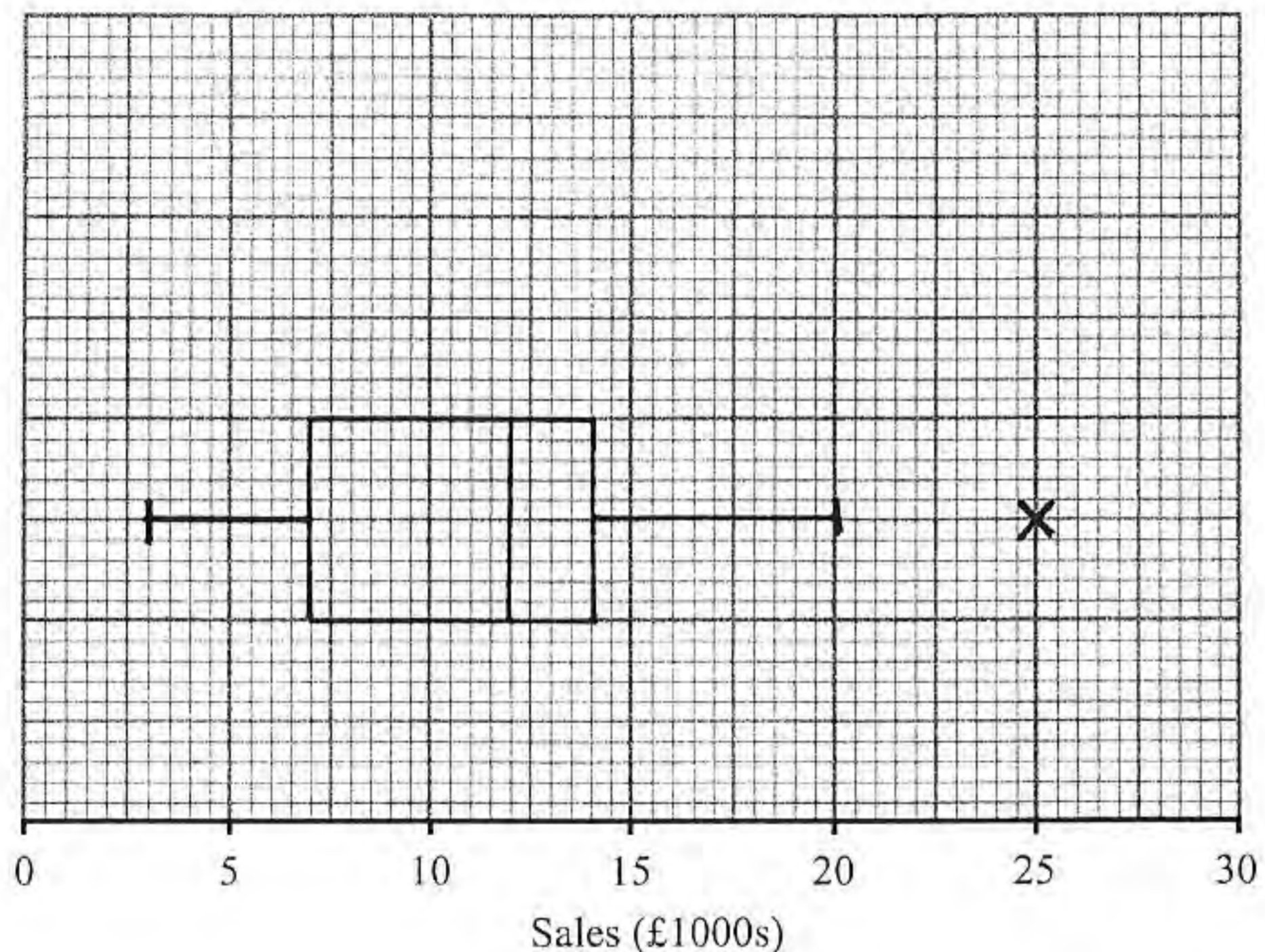
	Amount received in sales (£1000s)
Two lowest values	3, 4
Lower quartile	7
Median	12
Upper quartile	14
Two highest values	20, 25

An outlier is an observation that falls either $1.5 \times$ interquartile range above the upper quartile or $1.5 \times$ interquartile range below the lower quartile.

$1.5 \times IQR = 10.5$
 $Q_1 - 1.5 IQR = -3.5$
 no lower outliers

- (a) On the graph paper below, draw a box plot to represent these data, indicating clearly any outliers.

$Q_1 + 1.5 IQR = 24.5$
 (5)
 25 is an outlier.



- (b) State the skewness of the distribution of the amount of sales received. Justify your answer.

negative skew $Q_2 - Q_1 > Q_3 - Q_2$ (2)
 (5) (2)

- (c) The company claims that for 75% of the months, the amount received per month is greater than £10 000. Comment on this claim, giving a reason for your answer.

Incorrect. 75% of months are greater than £7000 (2)

4. A farmer collected data on the annual rainfall, x cm, and the annual yield of peas, p tonnes per acre.

The data for annual rainfall was coded using $v = \frac{x-5}{10}$ and the following statistics were found.

$S_{vv} = 5.753$ $S_{pv} = 1.688$ $S_{pp} = 1.168$ $\bar{p} = 3.22$ $\bar{v} = 4.42$

- (a) Find the equation of the regression line of p on v in the form $p = a + bv$. (4)

- (b) Using your regression line estimate the annual yield of peas per acre when the annual rainfall is 85 cm. (2)

a) $x \rightarrow v$ $b = \frac{S_{vp}}{S_{vv}} = \frac{1.688}{5.753} = 0.293$
 $y \rightarrow p$

$a = \bar{p} - b\bar{v} = 3.22 - 0.293 \dots \times 4.42 = 1.92$

$p = 1.92 + 0.293v$

b) $x = 85 \Rightarrow v = \frac{80}{10} = 8$ $p = 1.92 + 0.293 \times 8$
 $p = 4.26t$ (3sf)

5. On a randomly chosen day, each of the 32 students in a class recorded the time, t minutes to the nearest minute, they spent on their homework. The data for the class is summarised in the following table.

Time, t	Number of students
10 - 19	2
20 - 29	4
30 - 39	8
40 - 49	11
50 - 59	5
70 - 79	2

0
2
6
14
25 → 16M
30
32

(a) Use interpolation to estimate the value of the median.

(2)

Given that

$$\sum t = 1414 \quad \text{and} \quad \sum t^2 = 69378$$

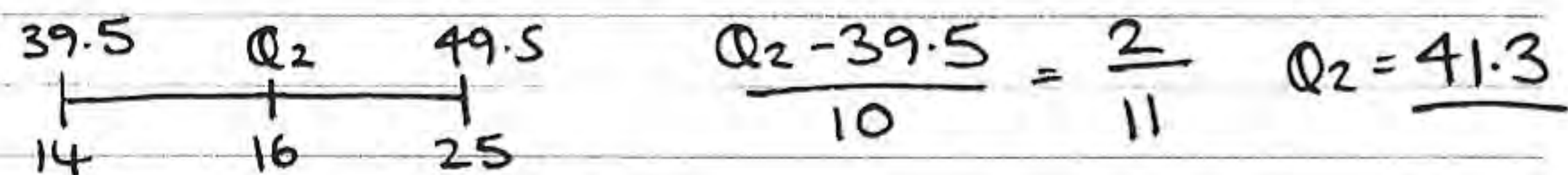
(b) find the mean and the standard deviation of the times spent by the students on their homework.

(3)

(c) Comment on the skewness of the distribution of the times spent by the students on their homework. Give a reason for your answer.

(2)

a) $\frac{1}{2}n = 16^{\text{th}}$ time, continuous.



b) $\bar{t} = \frac{\sum t}{n} = \frac{1414}{32} = 44.2$

s.d. = $\sqrt{\frac{\sum t^2}{n} - \bar{t}^2} = \sqrt{\frac{69378}{32} - 44.2^2} = 14.7$

c) positive skew
median < mean

6. The discrete random variable X has the probability distribution

x	1	2	3	4
$P(X=x)$	k	$2k$	$3k$	$4k$

$\sum P = 1$
 $10k = 1$
 $k = \frac{1}{10}$

(a) Show that $k = 0.1$

$\frac{1}{10} \quad \frac{2}{10} \quad \frac{3}{10} \quad \frac{4}{10}$

(1)

Find

(b) $E(X) = \frac{1}{10} + \frac{4}{10} + \frac{9}{10} + \frac{16}{10} = \frac{30}{10} = 3$

(2)

(c) $E(X^2) = \frac{1}{10} + \frac{8}{10} + \frac{27}{10} + \frac{64}{10} = \frac{100}{10} = 10$

(2)

(d) $\text{Var}(2-5X)$
 $V(X) = E(X^2) - E(X)^2 = 10 - 3^2 = 1$
 $V(2-5X) = (-5)^2 \times V(X) = 25$

(3)

Two independent observations X_1 and X_2 are made of X .

(e) Show that $P(X_1 + X_2 = 4) = 0.1$ (1,3) or (3,1) or (2,2)
only possibilities $\Rightarrow 4$

(2)

(f) Complete the probability distribution table for $X_1 + X_2$

(2)

y	2	3	4	5	6	7	8
$P(X_1 + X_2 = y)$	0.01	0.04	0.10	0.20	0.25	0.24	0.16

(g) Find $P(1.5 < X_1 + X_2 \leq 3.5)$

(2)

e) $P(1,3) = \frac{3}{100} \Rightarrow P(3,1) = \frac{3}{100}$

$P(2,2) = \frac{4}{100} \Rightarrow P(X_1 + X_2 = 4) = \frac{3+3+4}{100} = \frac{10}{100} = 0.1$

f) $5 \Rightarrow P(2,3), P(3,2), P(1,4), P(4,1) = \frac{6+6+4+4}{100}$

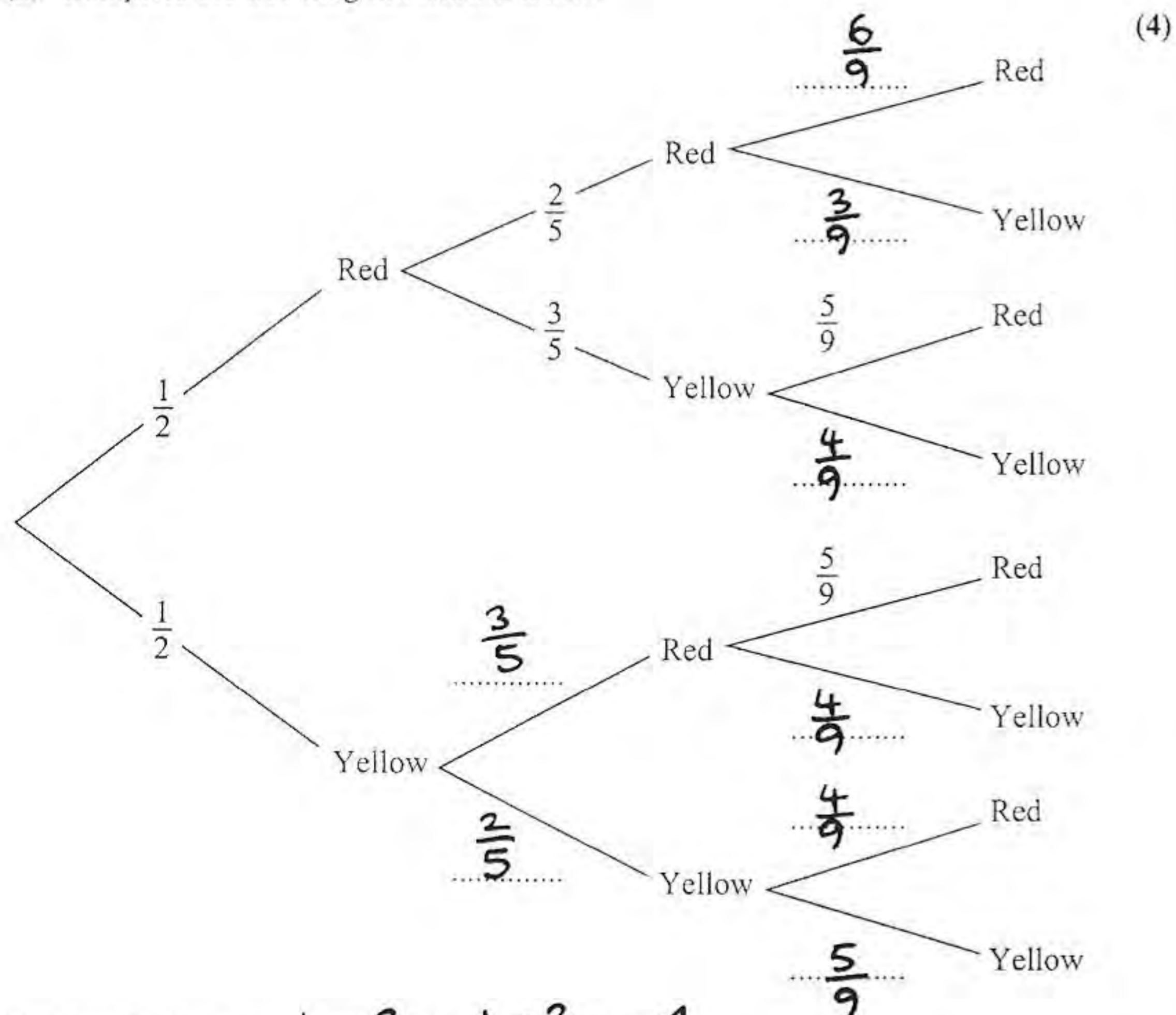
$\Rightarrow P(4,4) = \frac{16}{100} = 0.16$

g) $P(2) + P(3) = 0.05$

7. The bag P contains 6 balls of which 3 are red and 3 are yellow.
 The bag Q contains 7 balls of which 4 are red and 3 are yellow.
 A ball is drawn at random from bag P and placed in bag Q . A second ball is drawn at random from bag P and placed in bag Q .
 A third ball is then drawn at random from the 9 balls in bag Q .

The event A occurs when the 2 balls drawn from bag P are of the same colour.
 The event B occurs when the ball drawn from bag Q is red.

- (a) Complete the tree diagram shown below.



(b) Find $P(A) = \frac{1}{2} \times \frac{2}{5} + \frac{1}{2} \times \frac{2}{5} = \frac{4}{10} = \underline{0.4}$ (3)

(c) Show that $P(B) = \frac{5}{9}$ (3)

(d) Show that $P(A \cap B) = \frac{2}{9}$ (2)

(e) Hence find $P(A \cup B)$ (2)

(f) Given that all three balls drawn are the same colour, find the probability that they are all red. (3)

Question 7 continued

c) $P(B) = \frac{1}{2} \times \frac{2}{5} \times \frac{6}{9} + \frac{1}{2} \times \frac{3}{5} \times \frac{5}{9} + \frac{1}{2} \times \frac{3}{5} \times \frac{5}{9} + \frac{1}{2} \times \frac{2}{5} \times \frac{4}{9}$
 $= \frac{12}{90} + \frac{15}{90} + \frac{15}{90} + \frac{8}{90} = \frac{50}{90} = \frac{5}{9} \#$

c) $P(A \cap B) = P(R,R,R) + P(Y,Y,R) = \frac{12}{90} + \frac{8}{90} = \frac{20}{90}$
 $= \frac{2}{9} \#$

d) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
 $= \frac{2}{5} + \frac{5}{9} - \frac{2}{9} = \underline{\frac{11}{15}}$

f) $P(R,R,R) = \frac{12}{90}$ $P(Y,Y,Y) = \frac{10}{90}$

$\Rightarrow P(\text{Same colour}) = \frac{22}{90}$

$\Rightarrow P(R,R,R | \text{Same colour}) = \frac{\frac{12}{90}}{\frac{22}{90}} = \frac{12}{22} = \underline{\frac{6}{11}}$

8. The weight, X grams, of soup put in a tin by machine A is normally distributed with a mean of 160 g and a standard deviation of 5 g.

A tin is selected at random.

(a) Find the probability that this tin contains more than 168 g.

(3)

The weight stated on the tin is w grams.

(b) Find w such that $P(X < w) = 0.01$

(3)

The weight, Y grams, of soup put into a carton by machine B is normally distributed with mean μ grams and standard deviation σ grams.

(c) Given that $P(Y < 160) = 0.99$ and $P(Y > 152) = 0.90$ find the value of μ and the value of σ .

(6)

$$a) X \sim N(160, 5^2)$$

$$P(X > 168) \Rightarrow P\left(Z > \frac{168-160}{5}\right) = P(Z > 1.6) = 1 - \Phi(1.6) \\ = \underline{0.0548}$$

$$b) P(X < w) = 0.01 \Rightarrow P\left(Z < \frac{w-160}{5}\right) = 0.01$$

$$\Rightarrow P\left(Z > \frac{160-w}{5}\right) = 0.01 \Rightarrow \frac{160-w}{5} = 2.3263$$

$$w = \underline{148.4g}$$

$$c) P\left(Z < \frac{160-\mu}{\sigma}\right) = 0.99 \Rightarrow P\left(Z > \frac{160-\mu}{\sigma}\right) = 0.01$$

$$\Rightarrow \frac{160-\mu}{\sigma} = 2.3263 \Rightarrow 160-\mu = 2.3263\sigma$$

$$P\left(Z > \frac{152-\mu}{\sigma}\right) = 0.90 \Rightarrow P\left(Z < \frac{152-\mu}{\sigma}\right) = 0.10$$

$$\Rightarrow P\left(Z > \frac{\mu-152}{\sigma}\right) = 0.10 \Rightarrow \frac{\mu-152}{\sigma} = 1.2816$$

$$160-\mu = 2.3263\sigma$$

$$\sigma = \underline{2.22} \text{ (3sf)}$$

$$\mu - 152 = 1.2816\sigma$$

$$\underline{8} = 3.6079\sigma$$

$$\underline{\mu = 154.8g}$$