Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel International Advanced Level

Centre Number	Candidate Number

Thursday 22 October 2020

Afternoon (Time: 1 hour 30 minutes) | Paper Reference **WST01/01**

Mathematics

International Advanced Subsidiary/Advanced Level
Statistics S1

You must have:
Mathematical Formulae and Statistical Tables (Blue), calculator

Total Marks

Candidates may use any calculator permitted by Pearson regulations. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulae stored in them.

Instructions
- Use **black** ink or ball-point pen.
- If pencil is used for diagrams/sketches/graphs it must be dark (HB or B).
- **Fill in the boxes** at the top of this page with your name, centre number and candidate number.
- Answer **all** questions and ensure that your answers to parts of questions are clearly labelled.
- Answer the questions in the spaces provided – *there may be more space than you need.*
- You should show sufficient working to make your methods clear. Answers without working may not gain full credit.
- Values from the statistical tables should be quoted in full. If a calculator is used instead of the tables, the value should be given to an equivalent degree of accuracy.
- Inexact answers should be given to three significant figures unless otherwise stated.

Information
- A booklet 'Mathematical Formulae and Statistical Tables' is provided.
- There are 6 questions in this question paper. The total mark for this paper is 75.
- The marks for **each** question are shown in brackets – *use this as a guide as to how much time to spend on each question.*

Advice
- Read each question carefully before you start to answer it.
- Try to answer every question.
- Check your answers if you have time at the end.
- If you change your mind about an answer, cross it out and put your new answer and any working underneath.

Turn over

1/1/1/1/1/1/1/

Leave blank

1. The discrete random variable X takes the values $-1, 2, 3, 4$ and 7 only.

Given that

$$P(X = x) = \frac{8 - x}{k} \quad \text{for } x = -1, 2, 3, 4 \text{ and } 7$$

find the value of E(X)

(5)

(a) a disrete random variable X is one that only takes a certain number of specific values - in the question, they're given as -1, 2, 3, 4 and 7, and P(X = x) represents the likelihood of each occurring

↳ subbing into the given expression:

...for $X = -1$,

$$P(X=-1) = \frac{8-(-1)}{k}$$

$$= 9/k$$

...for $X = 2$,

$$P(X=2) = \frac{8-(2)}{k}$$

$$= 6/k$$

...for $X = 3$,

$$P(X=3) = \frac{8-(3)}{k}$$

$$= 5/k$$

...for $X = 4$,

$$P(X=4) = \frac{8-(4)}{k}$$

$$= 4/k$$

...for $X = 7$,

$$P(X=7) = \frac{8-7}{k}$$

$$= 1/k$$

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

Leave blank

Question 1 continued

hence putting it all in a table:

x	-1	2	3	4	7
$P(X=x)$	$9/k$	$6/k$	$5/k$	$4/k$	$1/k$

and using the fact that the probabilities in a table sum to 1:

formula: $\Sigma P(X=x)=1$

$$9/k+6/k+5/k+4/k+1/k=1$$

collect like terms

$$\frac{25}{k}=1$$

$$\boxed{\Rightarrow k=25}$$

and we're given in the formula booklet that:

formula: $E(X)=\Sigma x_i P(X=x_i)$

sub into the above:

$$=-1(9/k)+2(6/k)+3(5/k)+4(4/k)+7(1/k)$$

expand the above

$$=-9/k+12/k+15/k+16/k+7/k$$

$$=41/k$$

subin $k=25$

$$\boxed{=41/25}$$

Q1

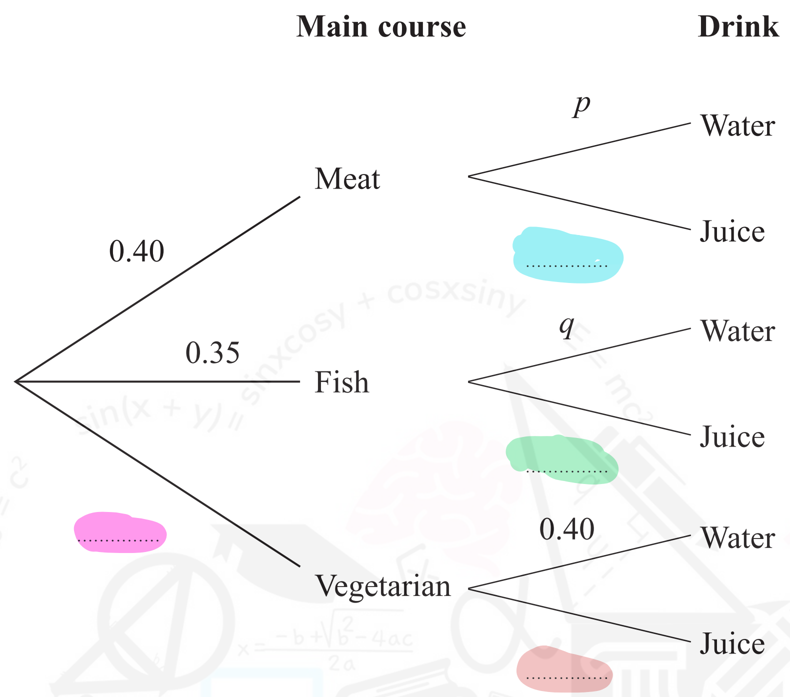
(Total 5 marks)

P 6 5 7 6 1 A 0 3 2 4

Turn over

2. In a school canteen, students can choose from a main course of meat (M), fish (F) or vegetarian (V). They can then choose a drink of either water (W) or juice (J).

The partially completed tree diagram, where p and q are probabilities, shows the probabilities of these choices for a randomly selected student.



(a) Complete the tree diagram, giving your answers in terms of p and q where appropriate.
(2)

(b) Find an expression, in terms of p and q, for the probability that a randomly selected student chooses water to drink.
(1)

The events "choosing a vegetarian main course" and "choosing water to drink" are independent.

V W

(c) Find a linear equation in terms of p and q.
(2)

condition

A student who has chosen juice to drink is selected at random. The probability that they chose fish for their main course is $\frac{7}{30}$

(d) Find the value of p and the value of q.
(5)

condition

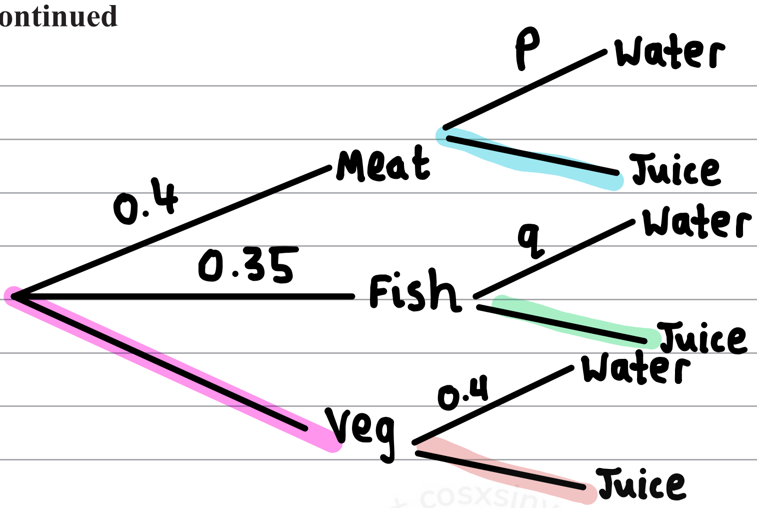
The canteen manager claims that students who choose water to drink are most likely to choose a fish main course.

(e) State, showing your working clearly, whether or not the manager's claim is correct.
(3)

Question 2 continued

Leave blank

(a)



we know that for tree diagrams, the probability for each set of branches must equal to 1

$\Rightarrow P(Veg) = 1 - 0.4 - 0.35$

$= 0.25$

$\Rightarrow P(Juice) = 1 - p$

$= 1 - p$

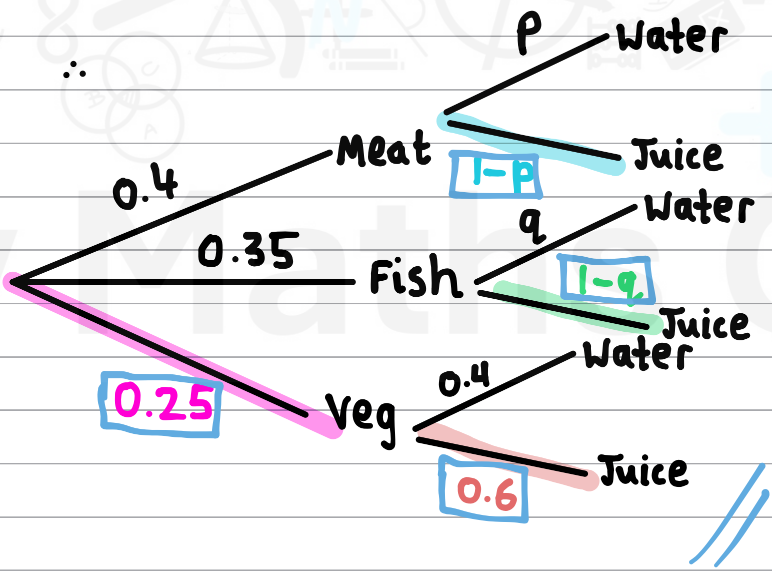
$\Rightarrow P(Juice) = 1 - q$

$= 1 - q$

$\Rightarrow P(Juice) = 1 - 0.4$

$= 0.6$

$\therefore$

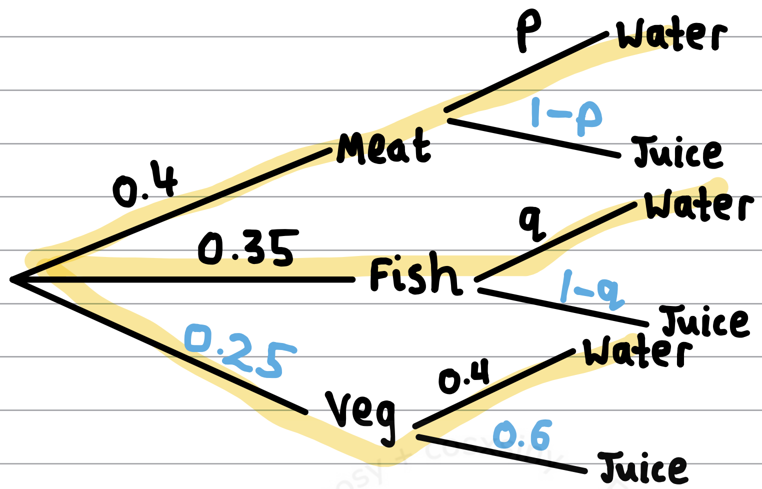


(b) the question is asking for $P(W)$:

↳ this involves the branches highlighted below:

Turn over

P65761A0524



⇒ P(W) = P(M∩W) + P(F∩W) + P(V∩W)

↳ these are independent events, so can multiply them

↳ the plus signs represent 'or'

= 0.4(p) + 0.35(q) + 0.25(0.4)

= $\boxed{0.4p + 0.35q + 0.1}$

(c) for independent events,

formula: P(V∩W) = P(V) × P(W)

reading off the prob. tree diagram

P(V) = 0.25 ← answer to (a)

P(W) = 0.4p + 0.35q + 0.1 ← answer to (b)

P(V∩W) = 0.1 ← see from (b)

subbing into above formula:

0.1 = 0.25(0.4p + 0.35q + 0.1)

0.1 = 0.1p + 0.0875q + 0.025

⇒ $\boxed{0.1p + 0.0875q = 0.075}$

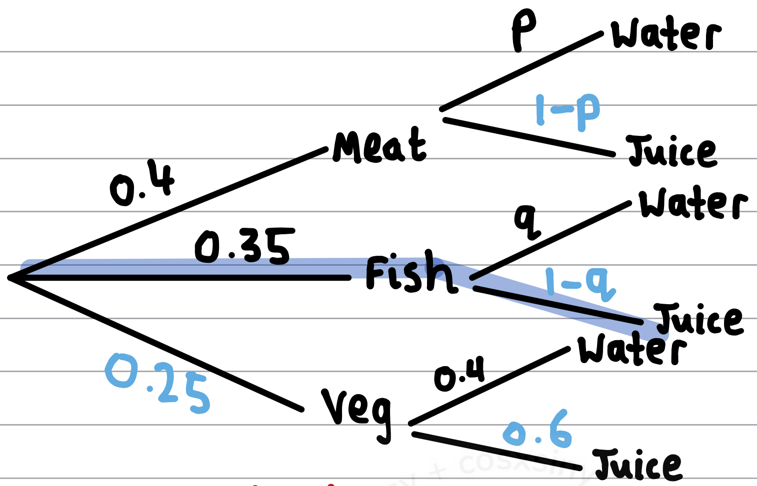
(d) we are asked to find a conditional probability, precisely: P(F | J)

↳ formula: $P(F|J) = \dfrac{P(F \cap J)}{P(J)}$

... for P(F∩J):

↳ see highlighted

Question 2 continued



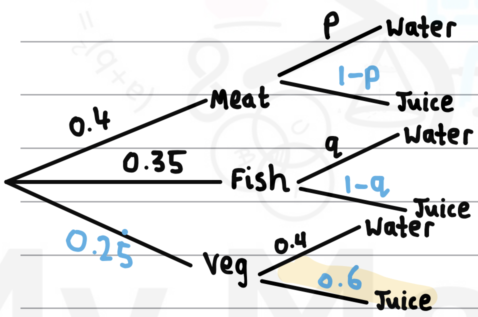
$\Rightarrow 0.35(1-q)$
$= 0.35 - 0.35q$

... for $P(J)$ – there are two ways:

WAY 1: use fact that V and W are independent, so V and W' (J) are independent, hence:

formula: $P(J|V) = P(J)$

reading off the prob. tree:



$P(J|V) = 0.6$
$\Rightarrow P(J) = 0.6$

sub into the conditional probability:

$P(F|J) = \dfrac{0.35-0.35q}{0.6}$

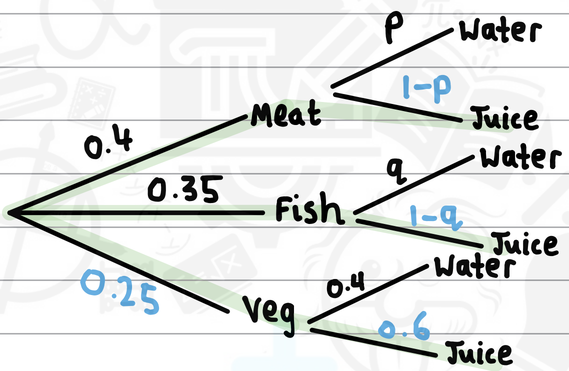
we're given that $P(F|J) = 7/30 = 0.233...$

$7/30 = \dfrac{0.35-0.35q}{0.6}$

$\Rightarrow 0.35 - 0.35q = 0.14$
$\Rightarrow 0.35q = 0.21$
$\boxed{q = 0.6}$

sub into equtn ① (part (c))

WAY 2: similar as for $P(W)$



$P(J) = P(M\cap J) + P(F\cap J) + P(V\cap J)$
$= 0.4(1-p) + 0.35(1-q) + 0.25(0.6)$
$= 0.4 - 0.4p + 0.35 - 0.35q + 0.15$
$= 0.9 - 0.4p - 0.35q$

sub into the conditional probability:

$P(F|J) = \dfrac{0.35-0.35q}{0.9-0.4p-0.35q}$

we're given that $P(F|J) = 7/30 = 0.233...$

$7/30 = \dfrac{0.35-0.35q}{0.9-0.4p-0.35q}$

trying to get this into a linear equation in 'p' and 'q':

$7/30(0.9-0.4p-0.35q) = 0.35-0.35q$

$0.21 - 7/75\,p - \frac{49}{600}q = 0.35 - 0.35q$

Q2

(Total 13 marks)

Leave blank

P 6 5 7 6 1 A 0 7 2 4

$$0.1p + 0.0875q = 0.075$$
$$0.1p + 0.0875(0.6) = 0.075$$
$$0.1p = 0.0225$$
$$\boxed{p = 0.225}$$

collect like terms:

$$\frac{161}{600}q - \frac{7}{75}p = \frac{7}{50}$$

solve simultaneously - the above with part (c)

$$0.1p + 0.0875q = 0.075 \quad \text{— ①}$$
$$\frac{161}{600}q - \frac{7}{75}p = \frac{7}{50} \quad \text{— ②}$$

on Classwiz

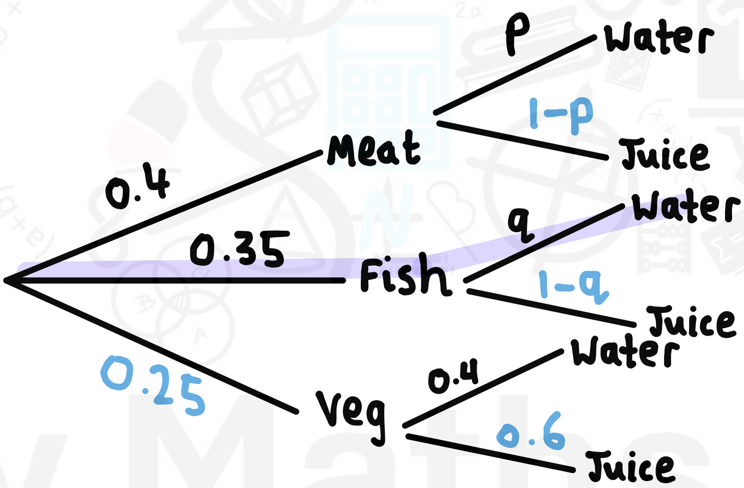
$$\Rightarrow \boxed{\begin{array}{l} p = 0.225 \text{ or } 9/40 \\ q = 0.6 \text{ or } 3/5 \end{array}}$$

(e) the manager's claim is another conditional probability - this time $P(F|W)$

↳ sub into the conditional probability formula:

formula: $P(F|W) = \dfrac{P(F \cap W)}{P(W)}$

...for $P(F \cap W)$:



$= 0.35q$ ← from (d)

$= 0.35(0.6)$

$= 0.21$

...$P(W)$:

$0.4p + 0.35q + 0.1$ ← from (b)

from (d) → sub $p = 0.225, q = 0.6$ into above

$= 0.4(0.225) + 0.35(0.6) + 0.1$

$= 0.4$

sub into the **cond. prob. formula**:

$$P(F|W) = \frac{0.21}{0.4}$$
$$= 0.525$$

now we have to **compare** this with:

- $P(V|W)$

↳ can do this in two ways:

WAY 1: $P(V|W) = P(V|W') = P(V)$

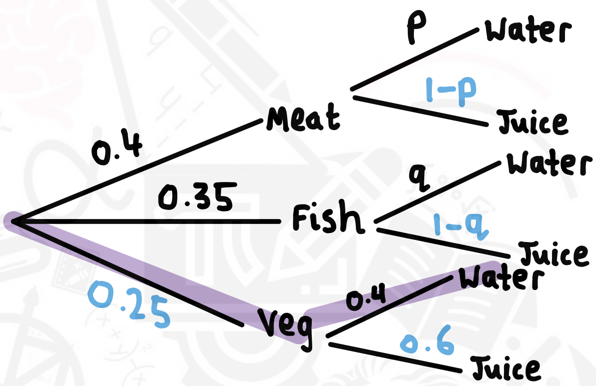
↳ this is true for **independent events** - which is given for part (c):

$P(V) = 0.25$ ← from (a)

$\Rightarrow P(V|W) = 0.25$

WAY 2: $P(V|W) = P(V \cap W)/P(W)$

...for $P(V \cap W)$:



$= 0.25(0.4)$

$= 0.1$

...$P(W)$:

$= 0.4$

sub into the **cond. prob. formula**:

$$\frac{0.1}{0.4} = 0.25$$

$$0.25 < 0.525$$

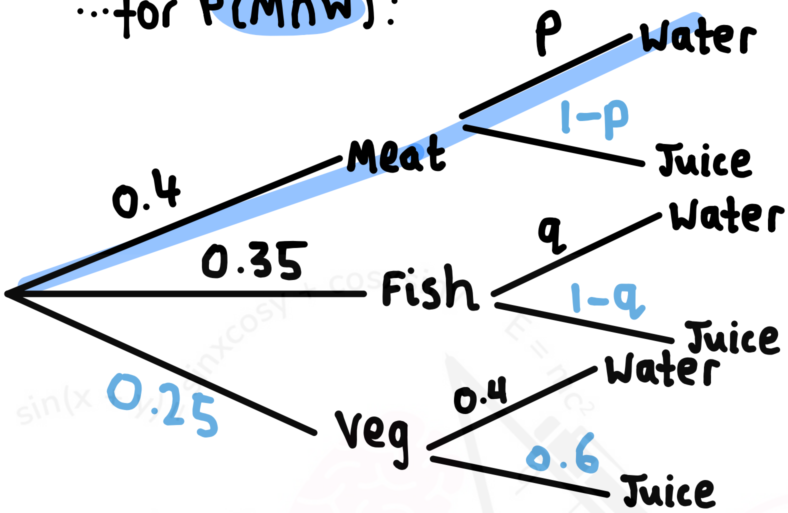
$\therefore$ so far, manager's claim is correct

· $P(M|W)$

↳ use conditional probability formula:

formula: $P(M|W) = \dfrac{P(M \cap W)}{P(W)}$

…for $P(M \cap W)$:



part (d)

$= 0.4p$

$= 0.4(0.225)$

$= 0.09$

…for $P(W)$:

$= 0.4$

$\Rightarrow P(M|W) = \dfrac{0.09}{0.4}$

$= 0.225$

$0.225 < 0.525$

$\boxed{\therefore \text{manager's claim is true}}$

3. The distance achieved in a long jump competition by students at a school is normally distributed with mean 3.8 metres and standard deviation 0.9 metres.

Students who achieve a distance greater than 4.3 metres receive a medal.

(a) Find the proportion of students who receive medals.

(3)

The school wishes to give a certificate of achievement or a medal to the 80% of students who achieve a distance of at least *d* metres.

(b) Find the value of *d*.

condition

(3)

Of those who received medals, the $\frac{1}{3}$ who jump the furthest will receive gold medals.

(c) Find the shortest distance, *g* metres, that must be achieved to receive a gold medal.

(4)

A journalist from the local newspaper interviews a randomly selected group of 3 medal winners.

(d) Find the exact probability that there is at least one gold medal winner in the group.

(3)

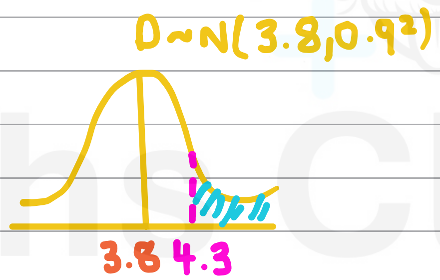
(a) the question is basically asking us to recognise that:

$D \sim N(3.8, 0.9^2)$

and find:

$P(D > 4.3)$

↳ let's visualise this:



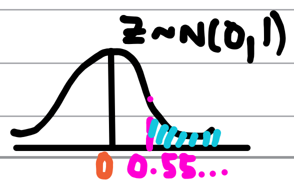
in order to use probability tables we need to standardise:

formula: $z = \frac{x - \mu}{\sigma}$

sub into the above:

$P\left(z > \frac{4.3 - 3.8}{0.9}\right)$

$= P(z > 0.555...)$

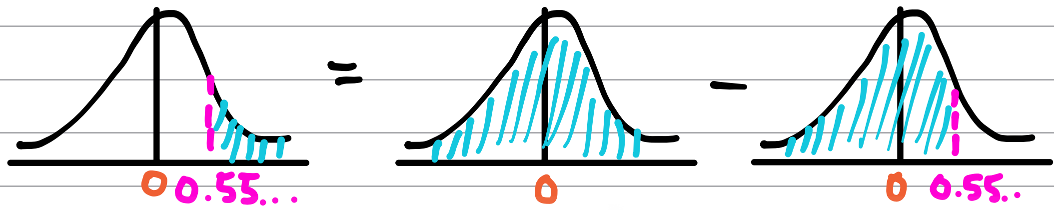


P65761A0824

Question 3 continued

but notice how our prob. tables only give results for the +ve tail, and where $P(Z<z)$ - hence, we need to find:

$$P(Z>0.55\ldots) = 1 - P(Z<0.55\ldots)$$



$= 1 - 0.7123$

↳ prob. tables for $P(Z<0.56)$

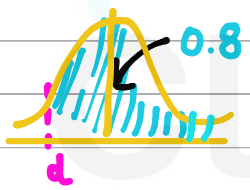
$= 0.2877 = \boxed{0.288 \text{ (3 s.f)}}$

NOTE: on CASIO CLASSWIZ, you'd use the NORMAL C.D function, then have lower value as 4.3

(b) the question is basically asking us for the 'd' needed for the following to be true:

$P(D>d) = 0.8$ (inverse normal)

↳ illustrate this: $D \sim N(3.8, 0.9^2)$

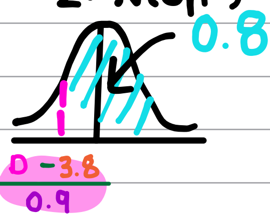


in order to use probability tables we need to standardise

formula: $z = \frac{x-\mu}{\sigma}$

sub into the above:

$P\left(Z > \frac{D-3.8}{0.9}\right) = 0.8$ $\quad Z \sim N(0, 1^2)$



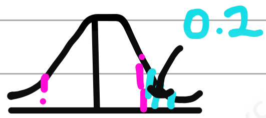
Question 3 continued

Leave blank

but notice how this is a round integer probability, hence we should utilise the % points table in the formula booklet (where probs given for $P(Z>z)$ and only for the +ve tail

↳ hence we can work out the z for the following:

$Z \sim N(0,1^2)$



$-z = \frac{D-3.8}{0.9}$

$P(Z>z) = 0.2$

↳ using $P(Z>0.8416) = 0.2$ from formula booklet

and negate it to get the -ve tail $\frac{D-3.8}{0.9}$ value (by symmetry)

$$\frac{D-3.8}{0.9} = -0.8416$$

$\times 0.9$ $\qquad \times 0.9$

$$D-3.8 = -0.75744$$

$$\Rightarrow D = 3.04256$$

$$\boxed{= 3.04\text{km } (3 \text{ s.f})}$$

NOTE: on CASSIO CLASSWIZ, use the INVERSE NORMAL function, then $p=0.2$ to get 3.04km

(c) here we need to recognise that we're looking at a conditional probability, with the condition being that the student wins a medal i.e $D>4.2$

↳ and we're asked for the shortest distance, g, for a student to receive a gold medal - prob. of which is ⅓

↳ i.e $P(D>g \mid D>4.2) = \frac{1}{3}$

subbing into the conditional probability formula

formula: $\frac{P(D>g \cap D>4.2)}{P(D>4.2)} = \frac{1}{3}$

part (a)

for this, always pick the stricter inequality i.e $P(D>g)$ as $g>4.2$

P 6 5 7 6 1 A 0 1 0 2 4

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

Leave blank

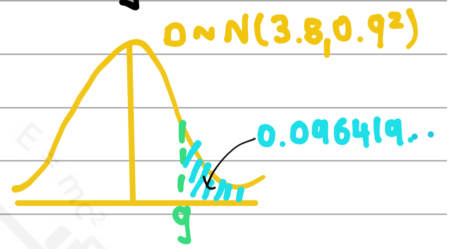
Question 3 continued

$$\Rightarrow \frac{P(D>g)}{0.2877} = \frac{1}{3}$$

$\times 0.2877$ $\times 0.2877$

$$P(D>g) = 0.096419\ldots$$

... illustrate diagrammatically:

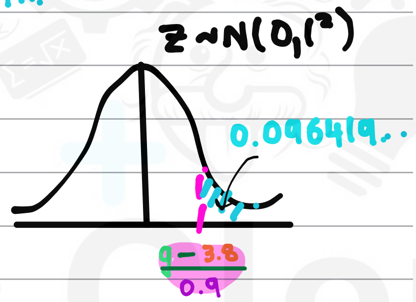


in order to use probability tables we need to standardise

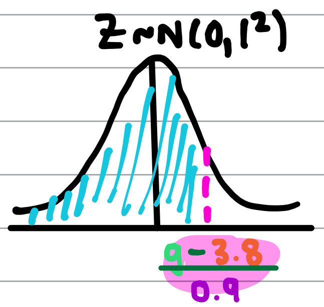
formula: $z = \frac{x-\mu}{\sigma}$

sub into the above:

$$P\left(D > \frac{g-3.8}{0.9}\right) = 0.096419\ldots$$



because the above probability isn't a round integer, we need to to use our prob. tables – these only give results for the +ve tail, and for results $P(Z<z)$, hence we can work out the following:



$P(Z<z) = 1 - 0.096419\ldots$
$= 0.9035\ldots$

we use from the prob. tables:

$P(Z<1.30) = 0.9032$

Q3

(Total 13 marks)

equating above:

$$\frac{g-3.8}{0.9} = 1.30$$

$\times 0.9$ $\qquad \times 0.9$

$$g - 3.8 = 1.17$$

$$\boxed{g = 4.97}$$

NOTE: on CASSIO CLASSWIZ, use the INVERSE NORMAL function, then p=0.9035... to get 4.97

(d) we know from part (c), that the amount of gold medallist winners is $1/3$

$\Rightarrow P(\text{not gold medallist}) = 1 - 1/3$
$= 2/3$

there are two ways to find the prob(at least one gold medallist)

WAY 1: using P(no gold)

P(at least one) = 1 − (no gold)
$= 1 - 2/3$

↳ but because there was a group of 3 and they are independent we need to do:

$= 1 - (2/3)^3$
$= 1 - 8/27$
$= \boxed{19/27}$

WAY 2: combinations

let G = gold medallist

...and we're asked for the probability of at least one gold winner:

... one:

$G\,G'G' = 1/3\,(2/3)(2/3)$

↳ and there are 3C1 combinations of the above = 3 (Pascal triangle or calculator)

$= \cancel{3} \times \frac{1}{\cancel{3}}(2/3)^2$
$= 4/9$

... two:

$G\,G\,G' = 1/3\,(1/3)(2/3)$

↳ and there are 3C2 combinations of the above = 3 (Pascal triangle or calculator)

$= \cancel{3} \times (1/3)^2 (2/\cancel{3})$
$= 2/9$

... three:

$G\,G\,G = 1/3 \times 1/3 \times 1/3$

↳ and there are 3C3 combinations of the above = 1 (Pascal triangle or calculator)

$$\left(\frac{1}{3}\right)^2 = \frac{1}{27}$$

$$\therefore \frac{4}{9} + \frac{2}{9} + \frac{1}{27} = \frac{19}{27}$$

Data Representations - box plots, outliers, measures of central tendency

4. A group of students took some tests. A teacher is analysing the average mark for each student. Each student obtained a different average mark.

For these average marks, the lower quartile is 24, the median is 30 and the interquartile range (IQR) is 10
The three lowest average marks are 8, 10 and 15.5 and the three highest average marks are 45, 52.5 and 56

The teacher defines an outlier to be a value that is either

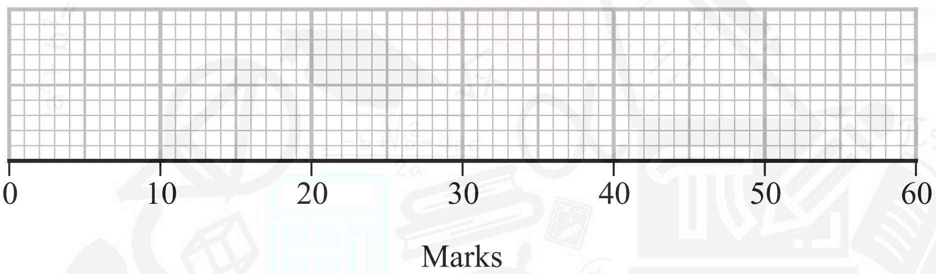
more than $1.5 \times$ IQR below the lower quartile or
more than $1.5 \times$ IQR above the upper quartile

(a) Determine any outliers in these data.

(4)

(b) On the grid below draw a box plot for these data, indicating clearly any outliers.

(3)



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

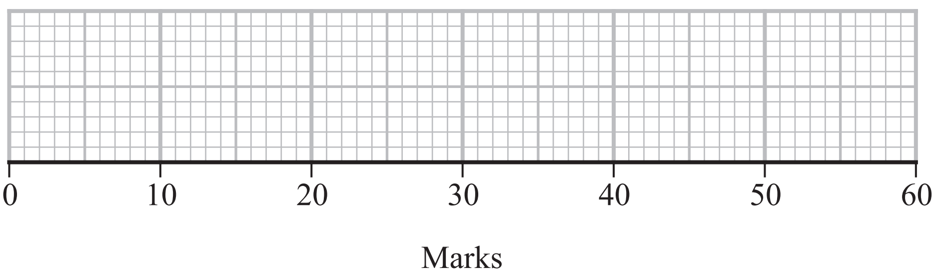
(2)

Two more students also took the tests. Their average marks, which were both less than 45, are added to the data and the box plot redrawn.

The median and the upper quartile are the same but the lower quartile is now 26

(d) Redraw the box plot on the grid below.

(3)



(e) Give ranges of values within which each of these students' average marks must lie.

(2)

Question 4 continued

(a) we're checking if any of 8, 10, 15.5, 45, 52.5, 56 can be outliers

...given that an outlier $= Q_1 - 1.5(IQR)$:

$= 24 - 15$

$= 9$

$\therefore$ we can see that 10 is an outlier

outlier $= Q_3 + 1.5(IQR)$

$IQR = Q_3 - Q_1$

sub into above

$Q_3 = 10 + 24$

$= 34$

$= 34 + 15$

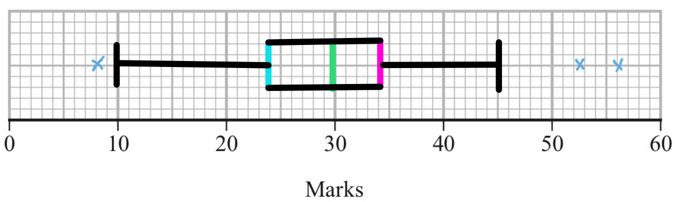
$= 49$

$\therefore$ we can see that 52.5, 56 are outliers

$\boxed{\therefore\ 8,\ 52.5 \text{ and } 56 \text{ are outliers}}$

(b) remembering the features of a box plot:

- outliers – 8 (mark with a cross)
- lower value – 10
- lower quartile – 24
- median – 30
- upper quartile – 34
- upper value – 45
- outliers – 52.5, 56



Turn over for spare grids if you need to redraw your answer for part (b) or part (d).

Question 4 continued

(c) we know that, when evaluating the skewness of data :

- if $Q_2 - Q_1 > Q_3 - Q_2$, then -vely skewed
- if $Q_2 - Q_1 < Q_3 - Q_2$, then +vely skewed

here, $30 - 24 = 6 > 34 - 30 = 4$

$\boxed{\therefore \text{-vely skewed}}$

(d) remembering the features of a box plot:

if we're given a new $Q_1 = 26$, then we need to recalculate the outliers and the lowest values

$IQR_{new} = 34 - 26$
$= 8$

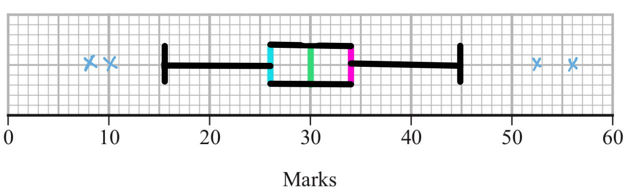
$\Rightarrow$ new outliers (subbing into the given formulae):

outlier $= Q_1 - 1.5(IQR_{new})$
$= 26 - 1.5(8)$
$= 14$

outlier $= Q_3 + 1.5(IQR_{new})$
$= 34 + 1.5(8)$
$= 46$

$\Rightarrow$ outliers : 8, 10 (<14) and 52.5, 56

- outliers – 8, 10 (mark with a cross)
- lower value – 15.5
- lower quartile – ~~24~~ 26
- median – 30
- upper quartile – 34
- upper value – 45
- outliers – 52.5, 56



(e) if the Q_1 has changed, then this means both values have had to have been above the prev. 24, but below or at Q_2, since the median has remained unchanged

Leave blank

Question 4 continued

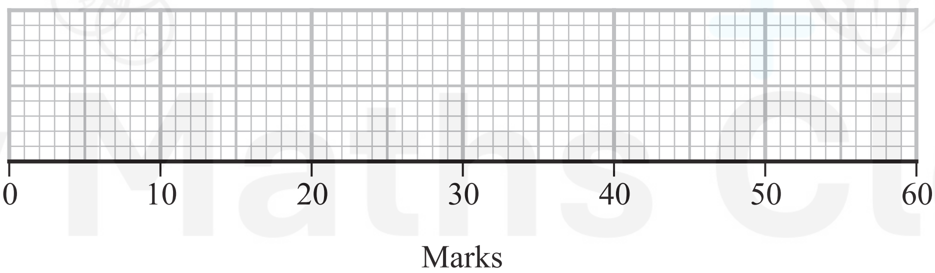
$\boxed{\Rightarrow 26 \le \text{average} \le 30}$

if Q_3 has remained unchanged, then the average must be either at the Q_3 or higher (but below the upper value)

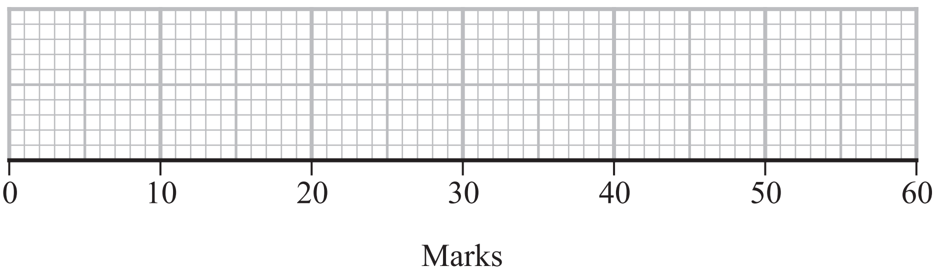
$\boxed{\Rightarrow 34 \le \text{average} \le 45}$

Only use these grids if you need to redraw your answer for part (b) or part (d).

Copy of grid for part (b)



Copy of grid for part (d)



Q4

(Total 14 marks)

Leave blank

5. A large company rents shops in different parts of the country. A random sample of 10 shops was taken and the floor area, x in $10\,\text{m}^2$, and the annual rent, y in thousands of dollars, were recorded.
The data are summarised by the following statistics

$$\sum x = 900 \quad \sum x^2 = 84\,818 \quad \sum y = 183 \quad \sum y^2 = 3434$$

and the regression line of y on x has equation $y = 6.066 + 0.136x$

(a) Use the regression line to estimate the annual rent in dollars for a shop with a floor area of $800\,\text{m}^2$
(2)

(b) Find S_{yy} and S_{xx}
(3)

(c) Find the product moment correlation coefficient between y and x.
(4)

An 11th shop is added to the sample. The floor area is $900\,\text{m}^2$ and the annual rent is 15 000 dollars.

(d) Use the formula $S_{xy} = \sum(x - \bar{x})(y - \bar{y})$ to show that the value of S_{xy} for the 11 shops will be the same as it was for the original 10 shops.
(2)

(e) Find the new equation of the regression line of y on x for the 11 shops.
(3)

The company is considering renting a larger shop with area of $3000\,\text{m}^2$

(f) Comment on the suitability of using the new regression line to estimate the annual rent. Give a reason for your answer.
(1)

(a) the question is basically asking us to sub $x=800$ into the given regression line – however careful of the units – we're given the area in 10s $\therefore$ sub in $x=80$:

$y = 6.066 + 0.136(80)$
$= 16.946$ thousands of dollars
$\boxed{= \$16{,}946}$

(b) S_{yy} is the covariance between the y variable

formula: $S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}$

P 6 5 7 6 1 A 0 1 6 2 4

Leave blank

Question 5 continued

sub into the above:

$$S_{yy} = 3434 - \frac{(183)^2}{10}$$

$$\boxed{= 85.1}$$

S_{xx} is the covariance between the x variable

formula: $S_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$

sub into the above:

$$S_{xx} = 84818 - \frac{(900)^2}{10}$$

$$\boxed{= 3818}$$

(c) now we're asked for the PMCC between y and x - this measures the strength and type of relationship between the two variables

formula: $r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$ part (a) part (a)

however, we don't have S_{xy}:

we can use the fact that:

the regression coeff. of y on x is $b = \frac{S_{xy}}{S_{xx}}$, for $y = a + bx$

sub into the above:

$$0.136 = \frac{S_{xy}}{3818}$$

$\times 3818$ $\times 3818$

$$\Rightarrow S_{xy} = 519.248$$

sub back into the PMCC formula:

$$r = \frac{519.248}{\sqrt{(3818)(85.1)}}$$

Question 5 continued

$= 0.910944...$

$\boxed{= 0.911 \text{ (3 s.f)}}$

(d) sub $x = 90$ (remembering the 10s units) and $y = 15$ (remembering the thousands units) into the given formula for S_{xy}

formula: $S_{xy} = \Sigma(x-\bar{x})(y-\bar{y})$

the new $\bar{x} = \dfrac{\Sigma x + x}{n+1}$

$= \dfrac{900+90}{11}$

$= 990/11$

$= 90$

so just from the first $(x-\bar{x})$ this term is 0 $\therefore$ the 11th shop makes no difference to S_{xy} //

(e) we are trying to find a regression line in the form $y = a + bx$ where:

formula: $b = \dfrac{S_{xy}}{S_{xx}}$ } no change to either of these $\therefore b$ is the same $= 0.136$ (as given in q)

$a = \bar{y} - b\bar{x}$

where $\bar{y} = \dfrac{\Sigma y + y}{n+1}$

$= \dfrac{183+15}{11}$

$= 18$

sub into formula for 'a':

$a = 18 - (0.136)(90)$

$= 5.76$

$\Rightarrow \boxed{y = 5.76 + 0.136x}$

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

Leave blank

Question 5 continued

(f) we know that our range for the linear regression line
above is $0 \leq x \leq 90$ (the 11th shop)
$\therefore x = 300$ (remembering the 10s units)
is beyond this range $\therefore$ inappropriate
as would be extrapolating

Q5

(Total 15 marks)

P 6 5 7 6 1 A 0 1 9 2 4

Leave blank

6. The random variable A represents the score when a spinner is spun. The probability distribution for A is given in the following table.

a	1	4	5	7
$\mathrm{P}(A = a)$	0.40	0.20	0.25	0.15

(a) Show that $\mathrm{E}(A) = 3.5$
(2)

(b) Find $\mathrm{Var}(A)$
(3)

The random variable B represents the score on a 4-sided die. The probability distribution for B is given in the following table where k is a positive integer.

b	1	3	4	k
$\mathrm{P}(B = b)$	0.25	0.25	0.25	0.25

(c) Write down the name of the probability distribution of B.
(1)

(d) Given that $\mathrm{E}(B) = \mathrm{E}(A)$ state, giving a reason, the value of k.
(1)

The random variable $X \sim \mathrm{N}(\mu, \sigma^2)$

Sam and Tim are playing a game with the spinner and the die.

They each spin the spinner once to obtain their value of A and each roll the die once to obtain their value of B.
Their value of A is taken as their value of μ and their value of B is taken as their value of σ.
The person with the larger value of $\mathrm{P}(X > 3.5)$ is the winner.

(e) Given that Sam obtained values of $a = 4$ and $b = 3$ and Tim obtained $b = 4$ find, giving a reason, the probability that Tim wins.
(2)

(f) Find the largest value of $\mathrm{P}(X > 3.5)$ achievable in this game.
(4)

(g) Find the probability of achieving this value.
(2)

Leave blank

Question 6 continued

(a) we're given in the formula booklet that:

formula: $E(A) = \sum a\,P(A=a)$

↳ i.e multiply each a with its probability and find its sum

$$E(A) = 1(0.4) + 4(0.20) + 5(0.25) + 7(0.15)$$

$$\boxed{= 3.5}$$

(b) we're given in the formula booklet that:

formula: $Var(A) = \sum a^2 P(A=a) - \mu^2$

↳ this is part (a)'s answer

... we need to do $\sum a^2 P(A=a)$:

$$= (1)^2(0.4) + (4)^2(0.20) + (5)^2(0.25) + (7)^2(0.15)$$

$$= 17.2$$

sub into the variance formula:

$$Var(A) = 17.2 - (3.5)^2$$

$$\boxed{= 4.95}$$

(c) we can see that we are given the prob. distribution for the discrete random variable, A in the form of a table, with the 'a' representing options for the variable A to take and the 'P(A=a)' the likelihood of each option to be taken

↳ because the P(A=a)s are all equal - we're looking at a $\boxed{\text{discrete uniform distribution}}$

(d) because this is a 1 mark question, best not to waste time finding E(B) - notice the symmetry in the prob. distributio

+2 +1 +2

b	1	3	4	k
$P(B = b)$	0.25	0.25	0.25	0.25

$$\boxed{\Rightarrow k = 6}$$

Leave blank

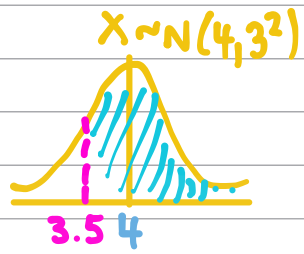
Question 6 continued

(e) $X \sim N(\mu, \sigma^2)$

because we're given more info on Sam, let's first try to find her $P(X > 3.5)$

for Sam, $X \sim N(4, 3^2)$

and we need $P(X > 3.5)$



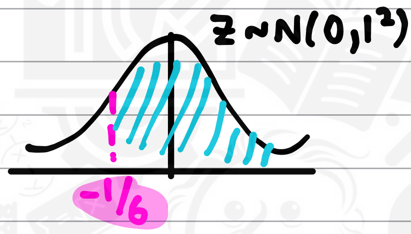
let's standardise to use the probability tables:

$$Z = \frac{X - \mu}{\sigma}$$

sub into the above:

$$P\left(Z > \frac{3.5 - 4}{3}\right)$$

$$= P(Z > -1/6)$$



this therefore implies that for Tim $(X \sim N(\mu, 4^2))$ to win, his standardised:

$$P\left(Z > \frac{3.5 - \mu}{4}\right) > P(Z > -1/6)$$

$$\Rightarrow \frac{3.5 - \mu}{4} < -1/6 \Big\} \text{(i.e to the left of } -1/6\text{)}$$

$\times 4$ $\quad\quad$ $\times 4$

$$3.5 - \mu < -2/3$$

$$\Rightarrow \mu > 25/6$$

$$\Rightarrow \mu > 4.166\ldots$$

i.e $a > 4.166\ldots$

looking at our table - this is $P(A = 5)$ and $P(A = 7)$

Question 6 continued

a	1	4	5	7
$P(A = a)$	0.40	0.20	0.25	0.15

$\therefore P(\text{Tim wins}) = 0.25 + 0.15$
$\boxed{= 0.4}$

(f) in order to get the largest possible value of $P(X > 3.5)$, we need to think of the best possible combo of μ, so a and σ, so b out of :

a	1	4	5	7
$P(A = a)$	0.40	0.20	0.25	0.15

and

from part (d)

b	1	3	4	6
$P(B = b)$	0.25	0.25	0.25	0.25

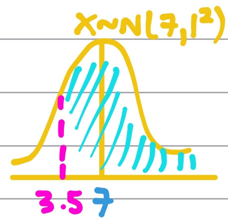
considering the standardizing formula

$z = \frac{X - \mu}{\sigma}$ } we want the most -ve z to get the biggest $P(X > 3.5)$

$\therefore$ μ biggest and σ smallest

$\therefore X \sim N(7, 1^2)$

we want $P(X > 3.5)$



standardize to use the probability tables:

Question 6 continued

$z = \frac{x - \mu}{\sigma}$

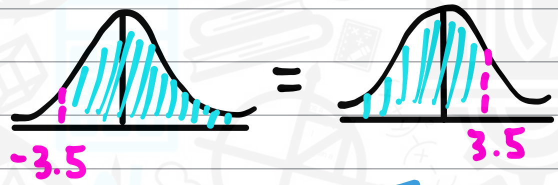
sub into the above:

$P\left(Z > \frac{3.5 - 7}{1}\right)$

$= P(Z > -3.5)$

but the formula booklet only gives us values for $P(Z > z)$ and the +ve tail

$\therefore$ let's use $P(Z > -3.5) = P(Z < 3.5)$



because of symmetry

$\boxed{= 0.9998}$

NOTE: could've used NORMAL C.D on calc, and have had the lower value as -3.5

(g) now asked for the prob. of $a = 7$ (0.15) or $b = 1$ (0.25) ← from prob. distrib. table
↳ these are independent so can multiply

$\Rightarrow 0.15 \times 0.25$

$\boxed{= 0.0375}$

Leave blank

Q6

(Total 15 marks)

TOTAL FOR PAPER: 75 MARKS

END