2014-06

1 The stem-and-leaf diagram shows the heights, in metres to the nearest 0.1 m, of a random sample of trees of species A.

$1+2+3+4+6+3+1=20$

5		
5	9	(1)
6	1 4	(2)
6	5 5 9	(3)
7	2 3 3 4	(4)
7	5 6 6 6 7 8	(6)
8	0 3 4	(3)
8	5	(1)

Key: 6 | 4 means 6.4 m

(i) Find the median and interquartile range of the heights. **[3]**

Median position $= \frac{n+1}{2} = \frac{20+1}{2} = 10.5$

$\therefore$ find 10^{th} and 11^{th} values

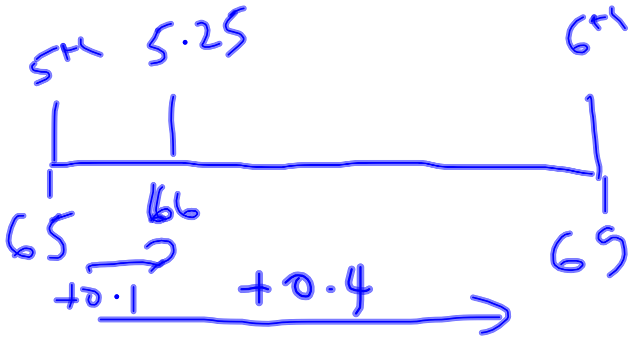
10^{th} value $= 7.4$ 11^{th} value $= 7.5$

$\therefore$ median $= \frac{7.4+7.5}{2} = \underline{\underline{7.45\text{ m}}}$ ✓

LQ position $= \frac{n+1}{4} = \frac{21}{4} = 5.25$

5^{th} value $= 6.5$

6^{th} value $= 6.9$

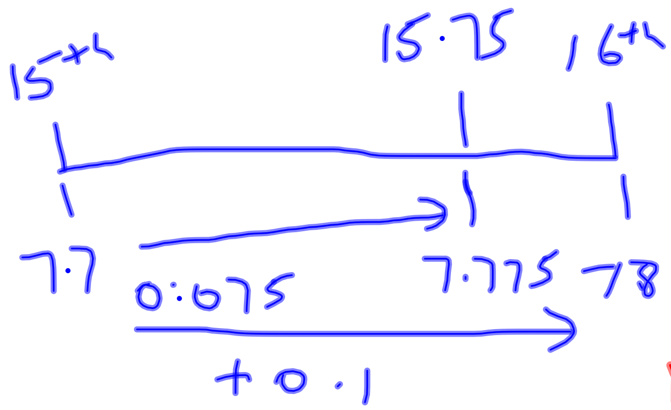


LOWER QUARTILE = 6.6

UQ position $= \frac{3(n+1)}{4} = \frac{63}{4} = 15.75$

15^{th} value $= 7.7$

16^{th} value $= 7.8$



UQ $= \underline{\underline{7.775}}$

IQR = UQ − LQ = 7.775 − 6.6

$\underline{\underline{\text{IQR} = 1.175\text{m}}}$

NOTE: MS allows IQR = 7.75 − 6.7 = 1.05 as this is both QUICKER and EASIER you should do it using simple MIDPOINTS

2014-06

(ii) The heights, in metres to the nearest 0.1 m, of a random sample of trees of species *B* are given below.

~~7.6~~ ~~5.2~~ ~~8.5~~ ~~5.2~~ ~~6.3~~ ~~6.3~~ ~~6.8~~ ~~7.2~~ ~~6.7~~ ~~7.3~~ ~~5.4~~ ~~7.5~~ ~~7.4~~ ~~6.0~~ ~~6.7~~

In the answer book, complete the back-to-back stem-and-leaf diagram. **[2]**

UNORDERED STEM AND LEAF

4	2	2	5
			5
0	3	3	6
7	7	8	6
4	3	2	7
	5	6	7
			8
		5	8

ORDERED STEM AND LEAF

4	2	2	5	
			5	9
3	3	0	6	1 4
8	8	7	6	5 5 9
4	3	2	7	2 3 3 4
	6	5	7	5 6 6 6 7 8
			8	0 3 4
		5	8	5

must have

Key: 6 | 4 means 6.4 m

MAKE SURE YOU
1. state a key
and
2. align your columns

2014-06

(iii) Make two comparisons between the heights of the two species of tree. **[2]**

median = 6.8 m

LQ position = $\frac{16}{4}$ = 4^{th} LQ = 6.0

UQ position = $\frac{48}{4}$ = 12^{th} UQ = 7.4

IQR = 7.4 − 6.0 = 1.4 m

SPECIES B has a LOWER MEDIAN ✓

and a WIDER IQR ✓

2014-06

2 **(a)** The probability distribution of a random variable W is shown in the table.

w	0	2	4
$P(W = w)$	0.3	0.4	0.3

Calculate Var(W). **[3]**

pw 0 0.8 1.2

w^2 0 4 16

pw^2 0 1.6 4.8

$\mu = E(w) = 0 + 0.8 + 1.2 = 2$

$\sum pw^2 = 0 + 1.6 + 4.8 = 6.4$

$\sigma^2 = Var(w) = \sum pw^2 - \mu^2$

$= 6.4 - 2^2$

$= 6.4 - 4$

$Var(w) = \underline{\underline{2.4}}$ ✓

2014-06

(b) The random variable X has probability distribution given by

$$P(X = x) = k(x + 1) \quad \text{for } x = 1, 2, 3, 4.$$

(i) Show that $k = \frac{1}{14}$. **[1]**

x	1	2	3	4
$P(X=x)$	$2k$	$3k$	$4k$	$5k$

$$\sum P(X=x) = 1 = 2k + 3k + 4k + 5k$$

$$14k = 1$$

$$k = \frac{1}{14}$$ QED

2014-06

(ii) Calculate E(X). **[3]**

x	1	2	3	4
$P(X=x)$	$\frac{2}{14}$	$\frac{3}{14}$	$\frac{4}{14}$	$\frac{5}{14}$
px	$\frac{2}{14}$	$\frac{6}{14}$	$\frac{12}{12}$	$\frac{20}{14}$

$$E(x) = \sum px = \frac{2}{14} + \frac{6}{14} + \frac{12}{12} + \frac{20}{14}$$

$$= \frac{40}{14} = 2\frac{12}{14} = 2\frac{6}{7}$$

$$\underline{\underline{E(x) = 2}}\frac{6}{7}$$

2014-06

3 The table shows information about the numbers of people per household in 280 900 households in the north-west of England in 2001.

Number of people	1	2	3	4	5 or more
Number of households	86 900	92 500	45 000	37 100	19 400

(i) Taking '5 or more' to mean '5 or 6', calculate estimates of the mean and standard deviation of the number of people per household. **[5]**

No. people	1	2	3	4	5.5
No. household	86900	92500	45000	37100	19400
people × households	86900	185000	135000	148400	106700

$\sum$ people $= 86900 + 185000 + 135000 + 148400 + 106700$
$= 662000$

$\sum$ households $= 86900 + 92500 + 45000 + 37100 + 19400$
$= 280900$

estimate of mean $= \dfrac{662000}{280900} = 2.36$ ✓ people per household (3sf)

Variance formula for a FREQUENCY DISTRIBUTION (not in tables) is very similar to variance formula for a PROBABILITY DISTRIBUTION (in tables!)

$$Var(X) = \sum x^2 p - \mu^2 \quad (P.D.)$$

$$Var(X) = \frac{\sum x^2 f}{\sum f} - \bar{x}^2 \quad (F.D.)$$

Here 'x' is 'number of people' watch out here

x	1	2	3	4	5.5
x^2	1	4	9	16	30.25
f	86900	92500	45000	37100	19400
x^2f	86900	370000	405000	593600	586850

$$\sum x^2f = 86900 + 370000 + 405000 + 593600 + 586850$$

$$= 2042350$$

$$Var(X) = \frac{2042350}{280900} - 2.36^2$$

$$= 1.70 \ (3sf)$$

$$sd = \sqrt{1.70} = 1.30 \ (3sf)$$

2014-06

(ii) State the values of the median and upper quartile of the number of people per household. **[2]**

There are 280900 household in the survey.

The median position is 140450th

The UQ position is 210675th

Use CUMULATIVE FREQUENCY

x	1	2	3	4	5.5
f	86900	92500	45000	37100	19400
cf	86900	179400	224400	261500	280900

MEDIAN is between 86900th and 179400th

So MEDIAN is 2 ✓

UQ is between 179400th and 224400th so UQ is 3 ✓

2014-06

4 Each time Ben attempts to complete a crossword in his daily newspaper, the probability that he succeeds is $\frac{2}{3}$. The random variable X denotes the number of times that Ben succeeds in 9 attempts.

(i) Find

(a) $P(X=6)$, **[3]**

$X \sim \text{Binomial}(9, \frac{2}{3})$

$P(X=6) = 9C6 \times \left(\frac{2}{3}\right)^6 \times \left(\frac{1}{3}\right)^3$

$= \frac{1792}{6561} = 0.273$ (3sf)

2014-06

(b) $P(X < 6)$, **[1]**

Use tables if possible!

$n = 9 \quad p = 2/3$

$P(X<6) = P(X \leq 5)$

From tables $P(X \leq 5) = 0.3497$

2014-06

(c) E(X) and Var(X). → in tables for binomial distribution **[2]**

$$E(X) = np = 9 \times \frac{2}{3} = 6$$

$$Var(X) = np(1-p) = npq = 9 \times \frac{2}{3} \times \frac{1}{3} = 2$$

$$\underline{\underline{E(X) = 6}} \quad \underline{\underline{Var(X) = 2}}$$

2014-06

Ben notes three values, X_1, X_2 and X_3, of X.

(ii) State the total number of attempts to complete a crossword that are needed to obtain three values of X. Hence find $P(X_1 + X_2 + X_3 = 18)$. **[4]**

X is defined as the number of successes from 9 attempts

so to obtain 3 values of X requires 27 attempts.

$(X+X+X) \sim B(27, 2/3)$

$P(X_1+X_2+X_3=18) = 27C18 \times \left(\frac{2}{3}\right)^{18} \times \left(\frac{1}{3}\right)^{9}$

$= 0.161182672$

$= 0.161$ (3sf)

2014-06

5 Tariq collected information about typical prices, £y million, of four-bedroomed houses at varying distances, x miles, from a large city. He chose houses at 10-mile intervals from the city. His results are shown below.

x	10	20	30	40	50	60	70	80
y	1.2	1.4	1.2	0.9	0.8	0.5	0.5	0.3

$n = 8$ $\Sigma x = 360$ $\Sigma x^2 = 20\,400$ $\Sigma y = 6.8$ $\Sigma y^2 = 6.88$ $\Sigma xy = 241$

(i) Use an appropriate formula to calculate the product moment correlation coefficient, r, showing that $-1.0 < r < -0.9$. **[3]**

Here we have been given summary statistics so we just need to substitute into the Pmcc equation from the tables.

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$

$$S_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 241 - \frac{(360 \times 6.8)}{8} = -65 \checkmark$$

$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 20400 - \frac{360^2}{8} = 4200 \checkmark$$

$$S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} = 6.88 - \frac{6.8^2}{8} = 1.1 \checkmark$$

$$r = \frac{-65}{\sqrt{4200 \times 1.1}} = -0.9562960653$$

$$= \underline{\underline{-0.956}} \ (3\text{sf}) \checkmark$$

2014-06

(ii) State what this value of *r* shows in this context. **[1]**

This high negative value of r shows that there is a very strong link between house prices reducing as distance from the city centre increases. ✓

2014-06

(iii) Tariq decides to recalculate the value of r with the house prices measured in hundreds of thousands of pounds, instead of millions of pounds. State what effect, if any, this will have on the value of r. **[1]**

This will have no effect on the value of r as the product moment correlation coefficient is INDEPENDENT OF SCALE.

2014-06

(iv) Calculate the equation of the regression line of y on x. **[3]**

Use the formula from the book of tables (S_{xy} and S_{xx} already calculated)

regression coefficient $b = \frac{S_{xy}}{S_{xx}} = \frac{-65}{4200} = \frac{-13}{840}$

$= -0.015476190$

$= -0.0155$ (3sf)

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

$\bar{y} = \frac{\sum y}{n} = \frac{6.8}{8} = \frac{17}{20} = 0.85$

$\bar{x} = \frac{\sum x}{n} = \frac{360}{8} = 45$

$a = 0.85 - \left(\frac{-13}{840} \times 45\right)$

$= \frac{433}{280} = 1.546428571$

$= 1.55$ (3sf)

$y = a + bx$

$\underline{\underline{y = 1.55 - 0.0155x}}$

2014-06

(v) Explain why the regression line of y on x, rather than x on y, should be used for estimating a value of x from a given value of y. **[1]**

In this case it is the PRICE OF HOUSES which is affected by the DISTANCE FROM THE CITY CENTRE. It wouldn't make sense to think about the DISTANCE FROM THE CITY CENTRE being dependent on the PRICE OF HOUSES.

2014-06

6 Fiona and James collected the results for six hockey teams at the end of the season. They then carried out various calculations using Spearman's rank correlation coefficient, r_s.

(i) Fiona calculated the value of r_s between the number of goals scored FOR each team and the number of goals scored AGAINST each team. She found that $r_s = -1$. Complete the table in the answer book showing the ranks.

Team	A	B	C	D	E	F
Number of goals FOR (rank)	1	2	3	4	5	6
Number of goals AGAINST (rank)	6	5	4	3	2	1

[1]

2014-06

(ii) James calculated the value of r_s between the number of goals scored and the number of points gained by the 6 teams. He found the value of r_s to be 1. He then decided to include the results of another two teams in the calculation of r_s. The table shows the ranks for these two teams.

Team	G	H
Number of goals scored (rank)	7	8
Number of points gained (rank)	8	7
DIFFERENCE	−1	1
d^2	1	1

Calculate the value of r_s for all 8 teams. **[4]**

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

For the first 6 teams

$r_s = 1$ so

$$1 = 1 - \frac{6\sum d^2}{6 \times 35}$$

$$= 1 - \frac{6\sum d^2}{210}$$

so $\frac{6\sum d^2}{210} = 0$

$\therefore \sum d^2 = 0$

For all 8 teams

$\sum d^2 = 0 + 1 + 1$
$= 2$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 2}{8 \times 63}$$

$$= 1 - \frac{12}{504}$$

$$r_s = \frac{41}{42}$$ ✓

2014-06

7 The table shows the numbers of members of a swimming club in certain categories.

	Male	Female	TOTAL
Adults	78	45	123
Children	52	n	
TOTAL	130		

It is given that $\frac{5}{8}$ of the female members are children.

(i) Find the value of n. **[2]**

If $\frac{5}{8}$ of female members are children then $\frac{3}{8}$ of female members are adults.

So 45 is $\frac{3}{8}$ of all female members

$\therefore$ $\frac{1}{8}$ is 15 members

So $n = 5 \times 15 = 75$

$\underline{\underline{n = 75}}$ ✓

2014-06

(ii) Find the probability that a member chosen at random is either female or a child (or both). **[2]**

	M	F	TOTAL
A	78	45	123
C	52	75	127
TOTAL	130	120	250

Total of females and children = 52+75+45
= 172

P(female, children or both) = $\frac{172}{250}$ = 0.688 (3sf)

2014-06

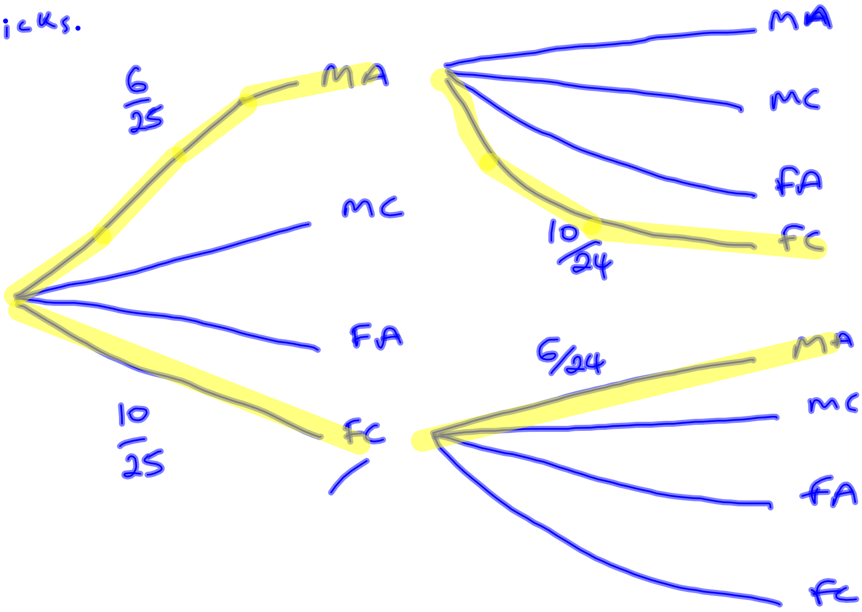
The table below shows the corresponding numbers for an athletics club.

	Male	Female	
Adults	6	4	10
Children	5	10	15
	11	14	25

(iii) Two members of the athletics club are chosen at random for a photograph.

(a) Find the probability that one of these members is a female child and the other is an adult male. **[2]**

You have to model the fact that you are making <u>Two</u> picks.



$\frac{6}{25} \times \frac{10}{24} = \frac{60}{600} = \frac{1}{10}$

$\frac{10}{25} \times \frac{6}{24} = \frac{60}{60} = \frac{1}{10}$

so $P(FC \text{ and } MA) = P(FC, MA) + P(MA, FC) = \frac{1}{10} + \frac{1}{10} = \frac{2}{10} = \frac{1}{5}$ ✓

(b) Find the probability that exactly one of these members is female and exactly one is a child. **[2]**

Start by thinking about the outcomes which satisfy this condition

<u>Outcome 1</u> FC and MA $P(FC \text{ and } MA) = \frac{1}{5}$

<u>Outcome 2</u> FA and MC

$P(FA \text{ and } MC) = P(FA, MC) + P(MC, FA) = \left(\frac{4}{25} \times \frac{5}{24}\right) + \left(\frac{5}{25} \times \frac{4}{24}\right)$

$= \frac{20}{600} + \frac{20}{600} = \frac{40}{600} = \frac{1}{15}$

If the child is female the adult has to be male

and if the adult is female the child has to be male.

There are no other outcomes which satisfy the condition.

So

$P(\text{one female, one child}) = P(FC \text{ and } MA) + P(FA \text{ and } MC)$

$= \frac{1}{5} + \frac{1}{15} = \frac{3}{15} + \frac{1}{15}$

$= \frac{4}{15}$ ✓

$= 0.267$ (3sf)

2014-06

8 A group of 8 people, including Kathy, David and Harpreet, are planning a theatre trip.

(i) Four of the group are chosen at random, without regard to order, to carry the refreshments. Find the probability that these 4 people include Kathy and David but not Harpreet. **[3]**

O O O O x x x x $^8C_4 = 70$

There are 70 ways of choosing 4 from 8.

If Kate and David are picked then there are 2 more slots to fill from 5 others (Harpreet is excluded)

K D O O x x x $^5C_2 = 10$

not H

There are 10 ways of choosing 2 from 5.

So probability of having a group of 4 from 8 including Kate and David but not Harpreet is

$$\frac{10}{70} = \frac{1}{7}$$ ✓

2014-06

(ii) The 8 people sit in a row. Kathy and David sit next to each other and Harpreet sits at the left-hand end of the row. How many different arrangements of the 8 people are possible? **[3]**

stuck together

H   K D   X   X   X   X   X

fixed   effectively 6 slots

So there are 6! ways of arranging the 8 people with these conditions

$6! = 720$

but Kate and David can swap places

so $2 \times 6! = 1440$ different combinations.

2014-06

(iii) The 8 people stand in a line to queue for the exit. Kathy and David stand next to each other and Harpreet stands next to them. How many different arrangements of the 8 people are possible? **[3]**

KDH is a block

$\underline{KDH}$ $\underline{x}$ $\underline{x}$ $\underline{x}$ $\underline{x}$ $\underline{x}$

so $6! = 720$ ways in which the people can be arranged

but KD can be in either order

so $720 \times 2 = 1440$ combinations

and Harpreet can be either side of Kate and David

so $1440 \times 2 = \underline{\underline{2880}}$ combinations ✓

2014-06

9 Each day Harry makes repeated attempts to light his gas fire. If the fire lights he makes no more attempts. On each attempt, the probability that the fire will light is 0.3 independent of all other attempts. Find the probability that

(i) the fire lights on the 5th attempt, **[2]**

This is a Geometric distribution

$$X \sim G(0.3)$$

$$P(X=5) = 0.7^4 \times 0.3$$
$$= 0.07203$$
$$= 0.0720 \text{ (3sf)}$$

2014-06

(ii) Harry needs more than 1 attempt but fewer than 5 attempts to light the fire. **[3]**

so the fire lights on the 2nd, 3rd or 4th attempt

$P(X=2) = 0.7 \times 0.3 = 0.21$

$P(X=3) = 0.7^2 \times 0.3 = 0.147$

$P(X=4) = 0.7^3 \times 0.3 = 0.1029$

$P(1<X<5) = 0.21 + 0.147 + 0.1029$

$= 0.4599$

$= 0.460$ (3sf)

2014-06

If the fire does not light on the 6th attempt, Harry stops and the fire remains unlit.

(iii) Find the probability that, on a particular day, the fire lights. **[3]**

$P(X>6) = 0.7^6 = 0.117649$

probability of 6 failures

so $P(X \leq 6) = 1 - P(X>6)$

$= 1 - 0.117649$

$= 0.882351$

$= \underline{\underline{0.882}}$ (3sf)

2014-06

(iv) Harry's week starts on Monday. Find the probability that, during a certain week, the first day on which the fire lights is Wednesday. **[2]**

We now need a new Geometric distribution modelling success each day

$$Y \sim Geo(0.882)$$

$$P(Y=3) = 0.118^2 \times 0.882$$

$$= 0.0122809068$$

$$= \underline{\underline{0.0123}} \text{ (3sf)}$$

Mark Scheme allows 0.0123 but correct answer is 0.0122. This is obtained by using the unrounded answer to part (iii) in the calculation for part (iv) (Use calculator memory)