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)

$84.6 + 2.8 + 5.6 + 2.3 + 9.4 + 0.5 + 1.8 = \frac{107}{28} = 3.82$

b) no change, as Σx remains constant!

Cotinine is a chemical that is made by the body from nicotine which is found in cigarette smoke. A doctor tested the blood of 12 patients, who claimed to smoke a packet of cigarettes a day, for cotinine. The results, in appropriate units, are shown below.

Patient	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*
Cotinine level, x	160	390	169	175	125	420	171	250	210	258	186	243

[You may use $\sum x^2 = 724\,961$]

(a) Find the mean and standard deviation of the level of cotinine in a patient's blood.
(4)

(b) Find the median, upper and lower quartiles of these data.
(3)

A doctor suspects that some of his patients have been smoking more than a packet of cigarettes per day. He decides to use $Q_3+1.5(Q_3-Q_1)$ to determine if any of the cotinine results are far enough away from the upper quartile to be outliers.

(c) Identify which patient(s) may have been smoking more than a packet of cigarettes a day. Show your working clearly.
(4)

Research suggests that cotinine levels in the blood form a skewed distribution.

One measure of skewness is found using $\dfrac{(Q_1 - 2Q_2 + Q_3)}{(Q_3 - Q_1)}$.

(d) Evaluate this measure and describe the skewness of these data.
(3)

a) $\dfrac{2757}{12} = \underline{229.75}$ mean

$\sqrt{\dfrac{\Sigma x^2}{12} - \text{mea}^2} = \sqrt{\dfrac{724961}{12} - 229.75^2}$

$= 87.34$

A survey of 100 households gave the following results for weekly income £y.

Income y (£)	Mid-point	Frequency f	
$0 \leqslant y < 200$	100	12	1200
$200 \leqslant y < 240$	220	28	6160
$240 \leqslant y < 320$	280	22	6160
$320 \leqslant y < 400$	360	18	6480
$400 \leqslant y < 600$	500	12	600
$600 \leqslant y < 800$	700	8	5600
		100	31600

(You may use $\sum fy^2 = 12\,452\,800$)

A histogram was drawn and the class $200 \leqslant y < 240$ was represented by a rectangle of width 2 cm and height 7 cm.

(a) Calculate the width and the height of the rectangle representing the class $320 \leqslant y < 400$ **(3)**

(b) Use linear interpolation to estimate the median weekly income to the nearest pound. **(2)**

(c) Estimate the mean and the standard deviation of the weekly income for these data. **(4)**

One measure of skewness is $\dfrac{3(\text{mean} - \text{median})}{\text{standard deviation}}$.

(d) Use this measure to calculate the skewness for these data and describe its value. **(2)**

Katie suggests using the random variable X which has a normal distribution with mean 320 and standard deviation 150 to model the weekly income for these data.

(e) Find $P(240 < X < 400)$. **(2)**

(f) With reference to your calculations in parts (d) and (e) and the data in the table, comment on Katie's suggestion. **(2)**

a) $\frac{28}{40} = 0.7 \quad \therefore \quad f.d \times 10$

$1\text{cm} = £20,$

so $320 \leq y < 400$

$\frac{18}{80} = 0.225, \quad \therefore$ Height $= 2.25$cm and width $= 4$cm

b) Q_2

240 — Q_2 — 320

40 — 50 — 62

$$\frac{Q_2 - 240}{50 - 40} = \frac{320 - 240}{62 - 40}$$

$Q_2 = 276.3636 \approx £276$

c) $\frac{31600}{100} = 316 = \bar{x}$

$\sqrt{124528 - 316^2} \approx £157$

d) skewness $= \dfrac{3(316 - 276.3636)}{157.073} = 0.7570$

$\therefore$ positive skew

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) $84.6 + 22.4 = 107$

$\frac{107}{28} = 3.82$

b) $9.4 \to 4.9 = 4.5 \downarrow$

$0.5 \to 5.0 = 4.5 \uparrow$

No change, sum stays the same.

The following table summarises the times, t minutes to the nearest minute, recorded for a group of students to complete an exam.

150

Sum: 4837.5

Time (minutes) t	11 – 20	21 – 25	26 – 30	31 – 35	36 – 45	46 – 60
Number of students f	62	88	16	13	11	10
xf	961	2024	448	429	495.5	530
x^2f	14895.5	46552	12544	14157	18042.75	28090

$\to 200$

Sum 134281.25

[You may use $\sum ft^2 = 134281.25$]

(a) Estimate the mean and standard deviation of these data. **(5)**

(b) Use linear interpolation to estimate the value of the median. **(2)**

(c) Show that the estimated value of the lower quartile is 18.6 to 3 significant figures. **(1)**

(d) Estimate the interquartile range of this distribution. **(2)**

(e) Give a reason why the mean and standard deviation are not the most appropriate summary statistics to use with these data. **(1)**

The person timing the exam made an error and each student actually took 5 minutes less than the times recorded above. The table below summarises the actual times.

Time (minutes) t	6 – 15	16 – 20	21 – 25	26 – 30	31 – 40	41 – 55
Number of students f	62	88	16	13	11	10

(f) Without further calculations, explain the effect this would have on each of the estimates found in parts (a), (b), (c) and (d). **(3)**

a) Mean $= \frac{4837.5}{200} = 24.1875$

$\sigma = \sqrt{\frac{134281.25^2}{200} - 24.1875^2}$

$= 9.2936 \approx 9.30$

b)
21 Q_2 25
62 100 150

$\frac{Q_2 - 21}{100 - 62} = \frac{25 - 21}{150 - 62}$

$\frac{Q_2 - 21}{38} = \frac{4}{22}$

$Q_2 = 22.7272$

≈ 22.73

c)
11 Q_1 20
0 50 62

$$\frac{Q_1 - 11}{50} = \frac{9}{62}$$

$Q_1 = 18.258$

d) $Q_3 = 25$

$Q_1 = 18.258$

$IQR = 25 - 18.258 = 6.742$

e) data is positively skewed

f)

a) would decrease by 5, SD is same

b) would decrease by 5

c) " " " "

d) same

The birth weights, in kg, of 1500 babies are summarised in the table below.

Weight (kg)	Midpoint, x kg	Frequency, f
0.0 – 1.0	0.50	1
1.0 – 2.0	1.50	6
2.0 – 2.5	2.25	60
2.5 – 3.0		280
3.0 – 3.5	3.25	820
3.5 – 4.0	3.75	320
4.0 – 5.0	4.50	10
5.0 – 6.0		3

[You may use $\sum fx = 4841$ and $\sum fx^2 = 15\,889.5$]

(a) Write down the missing midpoints in the table above. **(2)**

(b) Calculate an estimate of the mean birth weight. **(2)**

(c) Calculate an estimate of the standard deviation of the birth weight. **(3)**

(d) Use interpolation to estimate the median birth weight. **(2)**

(e) Describe the skewness of the distribution. Give a reason for your answer. **(2)**

a) 2.75 , 5.5

b)

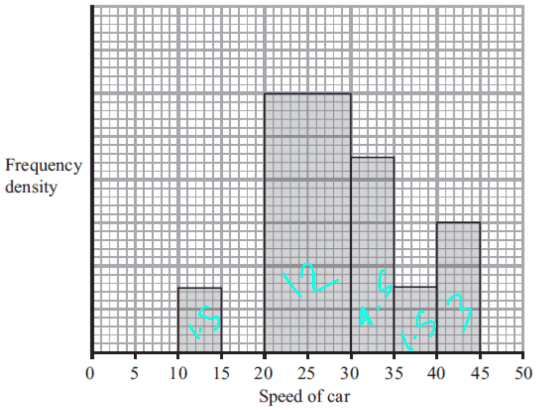


Figure 2

A policeman records the speed of the traffic on a busy road with a 30 mph speed limit. He records the speeds of a sample of 450 cars. The histogram in Figure 2 represents the results.

(a) Calculate the number of cars that were exceeding the speed limit by at least 5 mph in the sample. **(4)**

(b) Estimate the value of the mean speed of the cars in the sample. **(3)**

(c) Estimate, to 1 decimal place, the value of the median speed of the cars in the sample. **(2)**

(d) Comment on the shape of the distribution. Give a reason for your answer. **(2)**

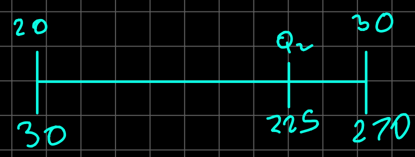
(e) State, with a reason, whether the estimate of the mean or the median is a better representation of the average speed of the traffic on the road. **(2)**

a) $450 \div 22.5$ units

$\cancel{90} \div 4.5$

b) $28.8\dot{3}$

c) $\frac{450}{2} = 225$



$$\frac{Q_2 - 20}{225 - 30} = \frac{10}{240}$$

$Q_2 = 28.125$

≈ 28.13

d) positive skew, mean > median

c)

The histogram in Figure 1 shows the time taken, to the nearest minute, for 140 runners to complete a fun run.

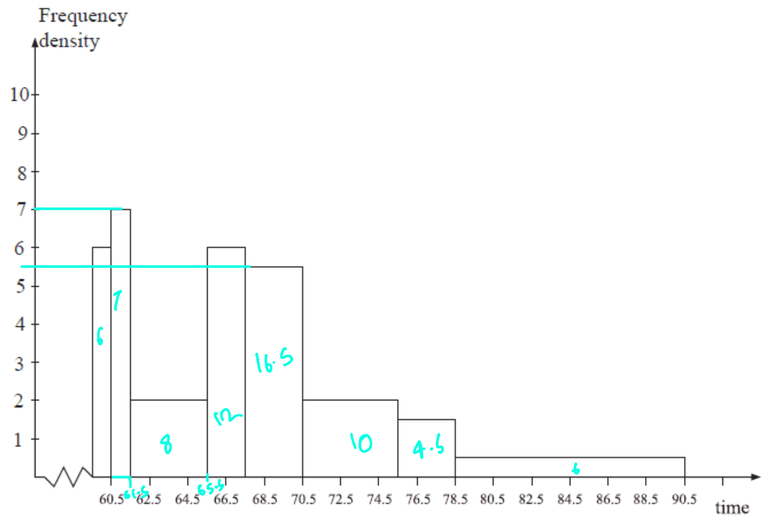


Figure 1

Use the histogram to calculate the number of runners who took between 78.5 and 90.5 minutes to complete the fun run.

(5)

a) $\frac{6}{70} \times 140 = 12$

1 The following back-to-back stem-and-leaf diagram shows the times to load an application on 61 smartphones of type A and 43 smartphones of type B.

	Type A		Type B	
(7)	9 7 6 6 4 3 3	2	1 3 5 8	(4)
(7)	5 5 4 4 2 2 2	3	0 4 4 5 6 6 6 6 7 8 8 9	(12)
(13)	9 9 8 8 8 7 6 6 4 3 2 2 0	4	0 1 1 2 3 6 8 8 9 9	(10)
(9)	6 5 5 4 3 2 1 1 0	5	2 5 6 6 9	(5)
(4)	9 7 3 0	6	1 3 8 9	(4)
(6)	8 7 4 4 1 0	7	5 7	(2)
(10)	7 6 6 6 5 3 3 2 1 0	8	1 2 4 4	(4)
(5)	8 6 5 5 5	9	0 6	(2)

31

Key: 3 | 2 | 1 means 0.23 seconds for type A and 0.21 seconds for type B.

(i) Find the median and quartiles for smartphones of type A. [3]

You are given that the median, lower quartile and upper quartile for smartphones of type B are 0.46 seconds, 0.36 seconds and 0.63 seconds respectively.

(ii) Represent the data by drawing a pair of box-and-whisker plots in a single diagram on graph paper. [3]

(iii) Compare the loading times for these two types of smartphone. [1]

i) $Q_2 = 0.52$

$Q_1 = 0.25 \times 62 = 15.5 \quad \therefore 0.41$

$Q_3 = 0.75 \times 62 = 46.5 \quad \therefore 0.79$

ii) $Q_1 = 0.36$

$Q_2 = 0.46$

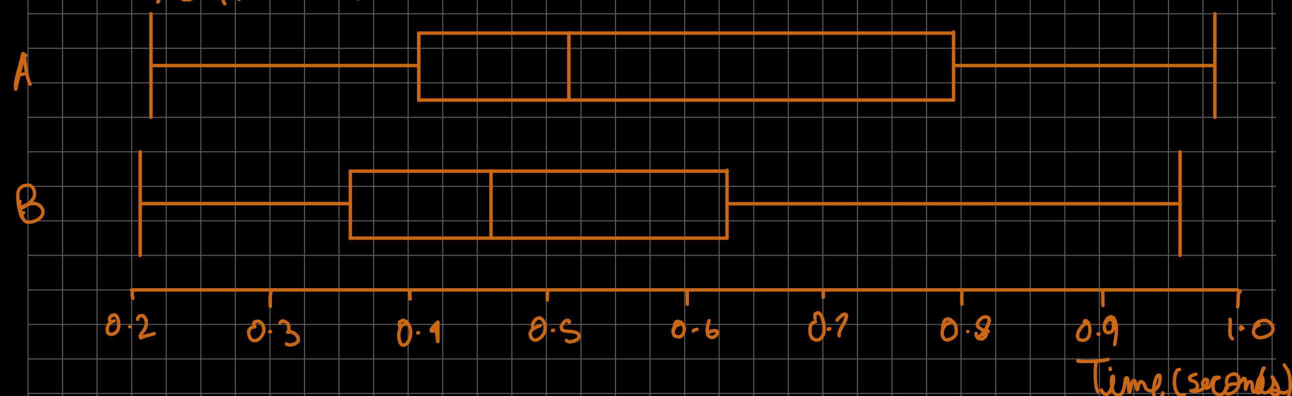
$Q_3 = 0.63$

$IQR = 0.27$

$1.5 \times IQR = 0.405$

Lower bound, ~~-0.045~~, 0.21

Upper bound, ~~1.035~~, 0.96



2 On a certain day in spring, the heights of 200 daffodils are measured, correct to the nearest centimetre. The frequency distribution is given below.

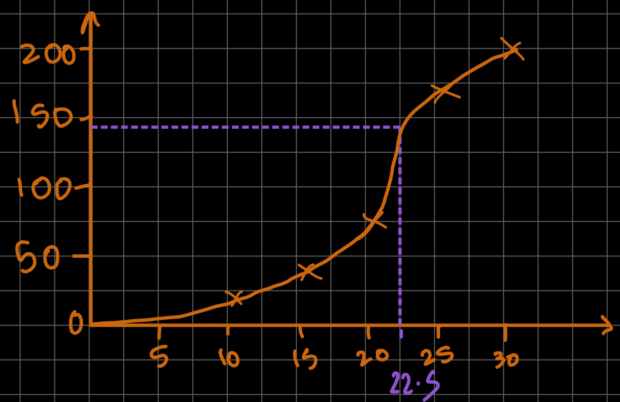
Height (cm)	4 – 10	11 – 15	16 – 20	21 – 25	26 – 30
Frequency	22	32	78	40	28
CF	22	54	132	172	200

(i) Draw a cumulative frequency graph to illustrate the data. [4]

(ii) 28% of these daffodils are of height h cm or more. Estimate h. [2]

(iii) You are given that the estimate of the mean height of these daffodils, calculated from the table, is 18.39 cm. Calculate an estimate of the standard deviation of the heights of these daffodils. [3]

i)



ii) $0.28 \times 200 = 56$

$200 - 56 = 144$

22.5

iii) $(7^2 \times 22) + (13^2 \times 32) + (18^2 \times 78) + (23^2 \times 40) + (28^2 \times 28)$

$$\sqrt{\frac{74810}{200} - 18.39^2}$$

$\sigma = 6.01$

5 The following are the house prices in thousands of dollars, arranged in ascending order, for 51 houses from a certain area.

253	270	310	354	386	428	433	468	472	477	485	520	520	524	526	531	535
536	538	541	543	546	548	549	551	554	572	583	590	605	614	638	649	652
666	670	682	684	690	710	725	726	731	734	745	760	800	854	863	957	986

(i) Draw a box-and-whisker plot to represent the data. [4]

An expensive house is defined as a house which has a price that is more than 1.5 times the interquartile range above the upper quartile.

(ii) For the above data, give the prices of the expensive houses. [2]

(iii) Give one disadvantage of using a box-and-whisker plot rather than a stem-and-leaf diagram to represent this set of data. [1]

i) $Q_2 = \frac{52}{2} = 26^{th}$ term $=$ 554

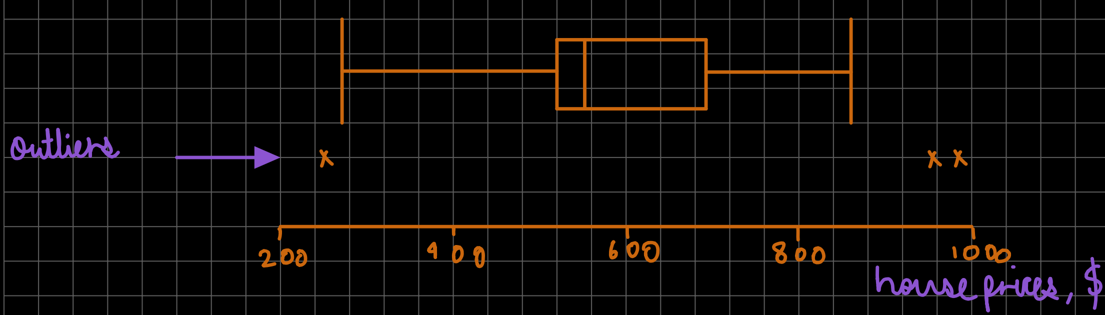
$Q_1 = \frac{52}{4} = 13^{th}$ term $=$ 520

$Q_3 = 52 \times 0.75 = 39^{th}$ term $=$ 690

IQR $= 170$

L.B $= 520 - (1.5 \times 170) = 265$

U.B $= 690 + (1.5 \times 170) = 945$



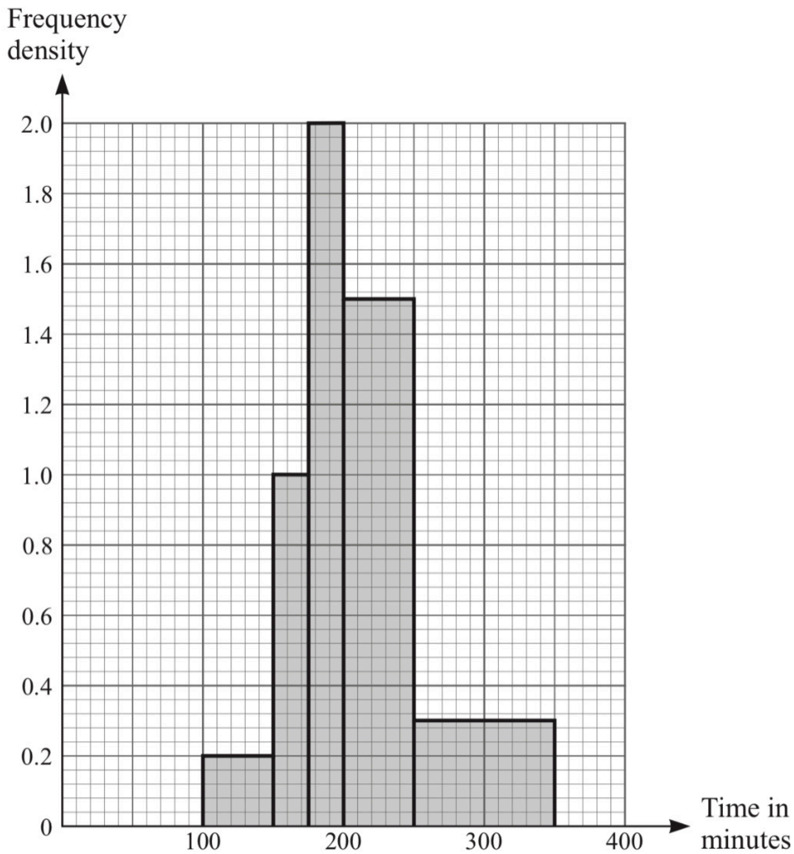
5 □ = 200

1 □ = 40

ii) \$957 and \$986

iii) not all data can be seen

6 The following histogram summarises the times, in minutes, taken by 190 people to complete a race.



(i) Show that 75 people took between 200 and 250 minutes to complete the race. [1]

(ii) Calculate estimates of the mean and standard deviation of the times of the 190 people. [6]

(iii) Explain why your answers to part **(ii)** are estimates. [1]

i) $250 - 200 = 50$
$50 \times 1.5 = 75$

ii)

125	162.5	187.5	225	300
100-150	150-175	175-200	200-250	250-350
10	25	50	75	30

$\Sigma x^2 f = 9071093.75$

$\Sigma xf = 40562.5$

$\Sigma f = 190$

$\bar{x} = \frac{40562.5}{190} = 213.48$

$\sigma = \sqrt{\frac{9071093.75}{190} - 213.48^2}$

$= 46.57 \approx 46.6$

iii) Mid-points used, not raw data

May/June 2002

2 The manager of a company noted the times spent in 80 meetings. The results were as follows.

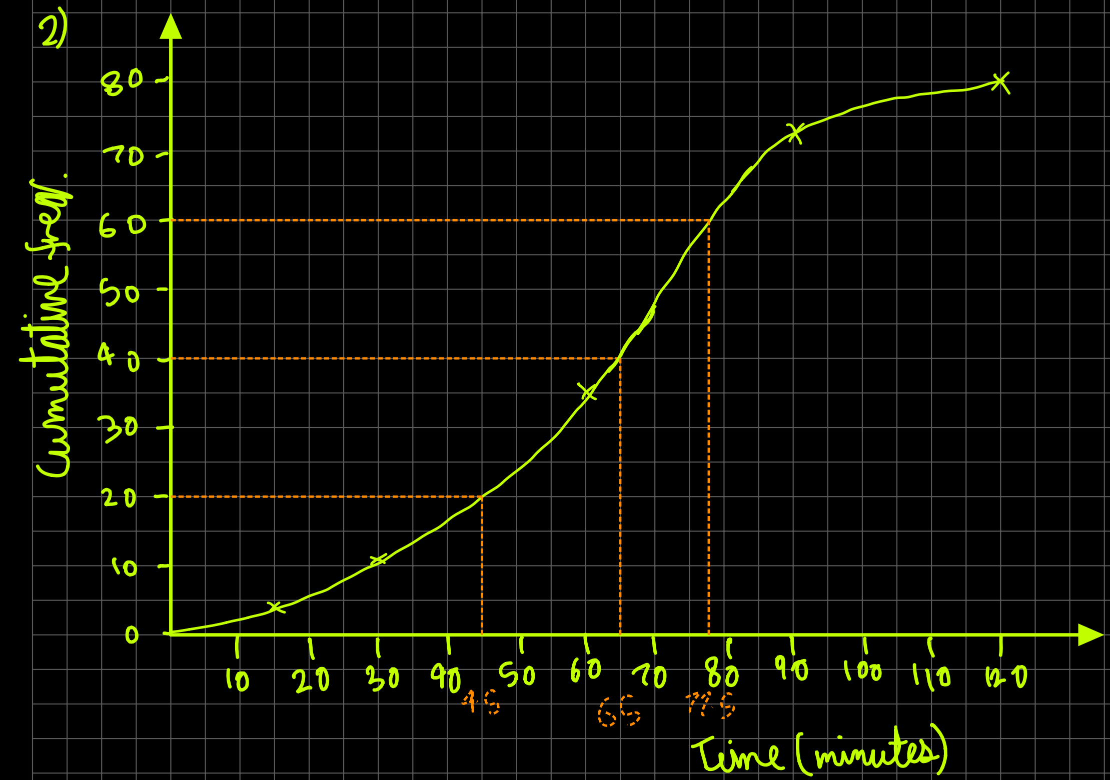
Time (t minutes)	$0 < t \leqslant 15$	$15 < t \leqslant 30$	$30 < t \leqslant 60$	$60 < t \leqslant 90$	$90 < t \leqslant 120$
Number of meetings	4	7	24	38	7
CF	4	11	35	73	80

Draw a cumulative frequency graph and use this to estimate the median time and the interquartile range. [6]

4 (i) In a spot check of the speeds $x\,\mathrm{km\,h^{-1}}$ of 30 cars on a motorway, the data were summarised by $\Sigma(x - 110) = -47.2$ and $\Sigma(x - 110)^2 = 5460$. Calculate the mean and standard deviation of these speeds. [4]

(ii) On another day the mean speed of cars on the motorway was found to be $107.6\,\mathrm{km\,h^{-1}}$ and the standard deviation was $13.8\,\mathrm{km\,h^{-1}}$. Assuming these speeds follow a normal distribution and that the speed limit is $110\,\mathrm{km\,h^{-1}}$, find what proportion of cars exceed the speed limit. [3]

2)



Median: 65

IQR: 77.5 − 45 = 32.5

4) i) $110 \times 30 = 3300$

$\Sigma x = -47.2 + 3300 = 3252.8$

$\bar{x} = \dfrac{3252.8}{30} = 108.43$

$\Sigma(x-110)^2 = 5460$

$\sigma = \sqrt{\dfrac{5460}{30} - (108.43 - 110)^2}$

$\sigma = 13.4$

ii) X Different chap.