Date 12.02.24
Page

S.1

Probability and Statistics - 1

Representation of Data

Ex 1. Solution (Revision)

SP-20	M-20	M-22	S-20	S-22	W-20	W-22
—	M-21	M-23	S-21	S-23	W-21	—

Suresh Goel
(Former Director)
Alliance World School,
Noida - Delhi. NCR.
INDIA.

(+91 9810444804)

Representation of Data – Revision

Example 1: The following back-to-back stem-and-leaf diagram shows the annual salaries of a group of 39 females and 39 males.

	Females		Males	
(4)	5 2 0 0	20	3	(1)
(9)	9 8 8 7 6 4 0 0 0	21	0 0 7	(3)
(8)	8 7 5 3 3 1 0 0	22	0 0 4 5 6 6	(6)
(6)	6 4 2 1 0 0	23	0 0 2 3 3 5 6 7 7	(9)
(6)	7 5 4 0 0 0	24	0 1 1 2 5 5 6 8 8 9	(10)
(4)	9 5 0 0	25	3 4 5 7 7 8 9	(7)
(2)	5 0	26	0 4 6	(3)

Key: 2|20|3 mean \$20 200 for females and \$20300 for males.

(a) Find the median and the quartiles of the female's salaries. -- [2]

You are given that the median salary of the males is \$24,000, the lower quartile is \$22600 and the upper quartile is \$25300.

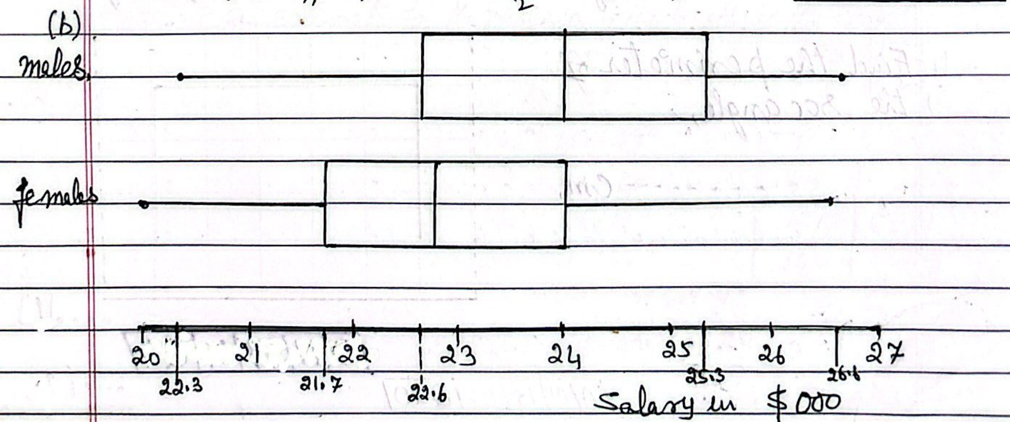
(b) Draw a pair of box-and-whisker plots in a single diagram on the grid below to represent the data. [SP-20/05/Q1] -- [3]

Solution (a) Median salary of females $= \frac{39+1}{2} = 20_{th}$ data $=$ \$22700 ✓

(Remove 20_{th} data and we are left with 19 data), Lower quartile $= \frac{19+1}{2} = 10_{th}$ data $=$ \$21700 ✓

Similarly the upper quartile $= \frac{19+1}{2} = 10_{th}$ data from median $=$ \$24,000 ✓

(b)



Example 2: A summary of the speeds, x km per hour, of 22 cars passing a certain point gave the following information:

$\Sigma(x-50) = 81.4$ and $\Sigma(x-50)^2 = 671.0$

Find the variance of the speeds and hence find the value of Σx^2 --[4]

[SP-20/05/Q2]

Solution: $\text{Coded mean} = \frac{\Sigma(x-50)}{n} = \frac{81.4}{22} = 3.7$

$$\text{Variance} = \frac{\Sigma(x-50)^2}{n} - \left(\frac{\Sigma(x-50)}{n}\right)^2 = \frac{671}{22} - (3.7)^2 = 16.81$$

$$\bar{x}, \text{ Mean} = a + \frac{\Sigma(x-a)}{n} = 50 + 3.7 = 53.7$$

$$\text{Variance} = \frac{\Sigma x^2}{n} - (\bar{x})^2 = \frac{\Sigma x^2}{22} - (53.7)^2 = 16.81$$

$$\Rightarrow \Sigma x^2 = 63811$$

Example 3: Helen measures the lengths of 150 fish of a certain species in a large pond. These lengths, correct to the nearest centimetre, are summarised in the following table.

Length (Cm)	0-9	10-14	15-19	20-30
Frequency	15	48	66	21

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

(b) 40% of these fish have a length of d cm or more. Use your graph to estimate the value of d. --[2]

The mean length of these 150 fish is 15.295 cm.

(c) Calculate an estimate for the variance of the lengths of the fish. --[3]

[M-20/52/Q7]

Solution(c)

Length (Cm)	f	Class boundries	Mid point x	fx^2	C.f
0-9	15	0-9.5	4.75	15×4.75^2	15
10-14	48	9.5-15.5	12	48×12^2	63
15-19	66	15.5-19.5	17	66×17^2	129
20-30	21	20.5-30.5	25	21×25^2	150

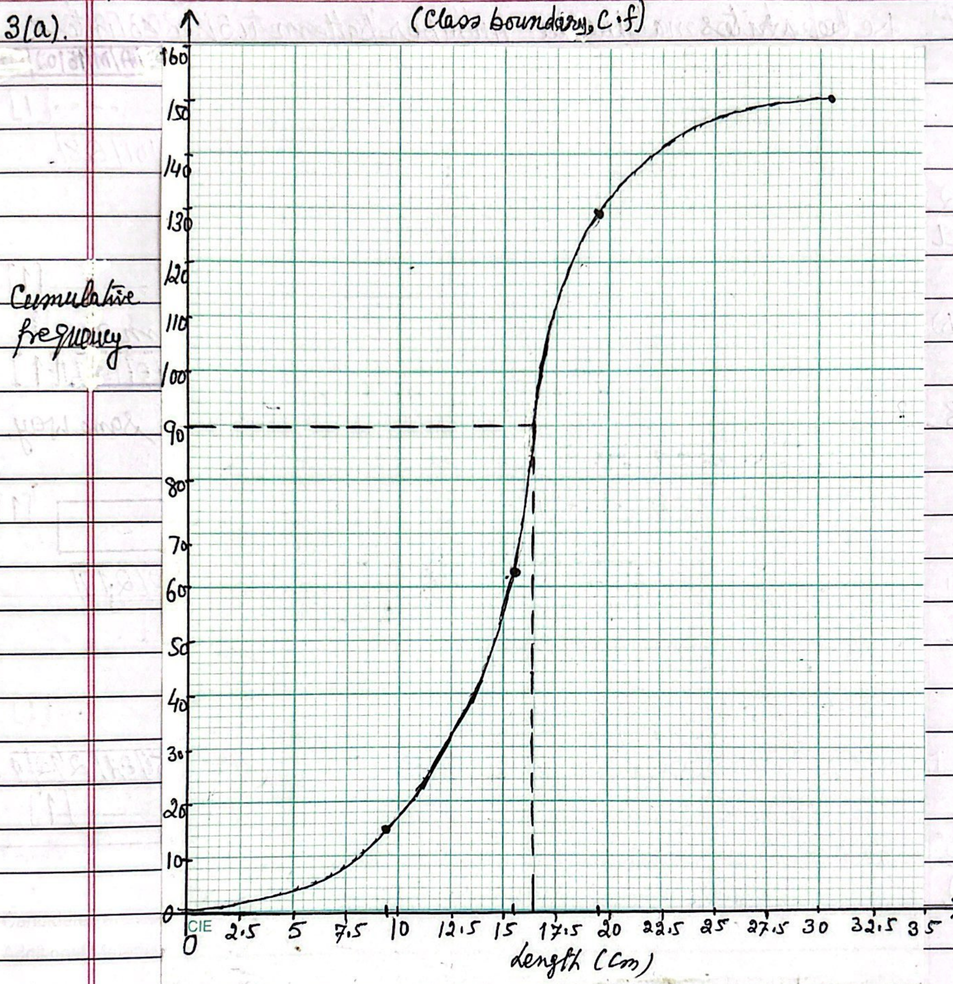
$N = \Sigma f = 150$. $\Sigma fx^2 = 39449.4$

$$Var = \frac{\Sigma fx^2}{N} - (\bar{x})^2 = \frac{39449.4}{150} - (15.295)^2 = 29.1$$

(Continued →)

(Continued →)

3(a).



(b) 60% of 150 = 90

Approx. 16.5 cm.

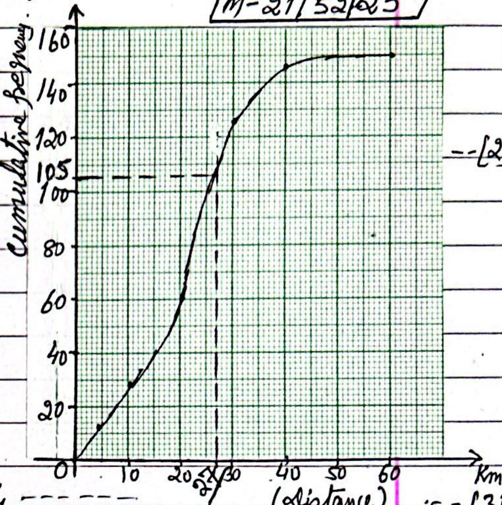
| M-20 | 52 | Q7 |

4 A driver records the distance travelled in each of 150 journeys. These distances correct to the nearest km, are summarized in the table:

Distance (km)	0-4	5-10	11-20	21-30	31-40	41-60
Frequency	12	16	32	66	20	4

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

[M-21/52/Q5]



(b) For 30% of these journeys the distance travelled is d km or more. Use your graph to estimate the value of 'd'. --[2]

30% are above 'd'

70% are below 'd'

70% of $150 = \frac{70}{100} \times 150 = 105$

$\therefore$ $d = 27$ km (approx).

(c) Calculate the estimate of mean distance travelled for 150 Journeys. --[3]

Distance (Km)	Frequency f_i	Class Boundaries	Mid point x_i of Classes.	$f_i \cdot x_i$	C. f.
0-4	12	0 - 4.5	2.25	12 × 2.25	12
5-10	16	4.5-10.5	7.5	16 × 7.5	28
11-20	32	10.5-20.5	15.5	32 × 15.5	60
21-30	66	20.5-30.5	25.5	66 × 25.5	126
31-40	20	30.5-40.5	35.5	20 × 35.5	146
41-60	4	40.5-60.5	50.5	4 × 50.5	150

$n = \Sigma f = 150$ $\qquad$ $\Sigma f_i x_i = 3238$

(c) $\text{Mean} = \frac{\Sigma f_i x_i}{n} = \frac{3238}{150} = 21.6$ (21.5866)

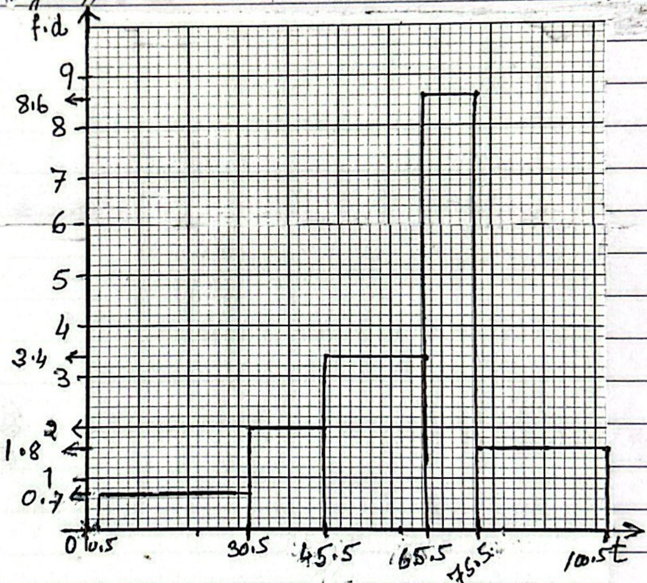
5 At a summer camp an arithmetic test is taken by 250 children. The times taken, to the nearest minute, to complete the test were recorded. The results are summarised as:

Time taken, in minutes	1-30	31-45	46-65	66-75	76-100
Frequency	21	30	68	86	45

(a) Draw a histogram to represent this information. ---[4]

(b) Which class interval contains the median ---[1]

(c) Given that an estimate of the mean time is 61.05 minutes, state what feature of the distribution accounts for the median and mean being different. [M-22/52/Q3] ---[1]



Continuous distribution	Classes	frequency	Class width (C.W)	Fre. Density = $\frac{f}{C.W}$
0.5-30.5	1-30	21	30	$21/30 = 0.7$
30.5-45.5	31-45	30	15	$30/15 = 2$
45.5-65.5	46-65	68	20	$68/20 = 3.4$
65.5-75.5	66-75	86	10	$86/10 = 8.6$
75.5-100.5	76-100	45	25	$45/25 = 1.8$

(b) median $= \frac{250}{2} = 125$th Value lies in the interval → 66-75

(c) mean = 61.05; median is in 66-75 interval: Mean and Median are different as the distribution is not symmetrical.

6 Each year the total number of hours, x, of sunshine in Kintoo is recorded during the month of June. The result for the last 60 years is given:

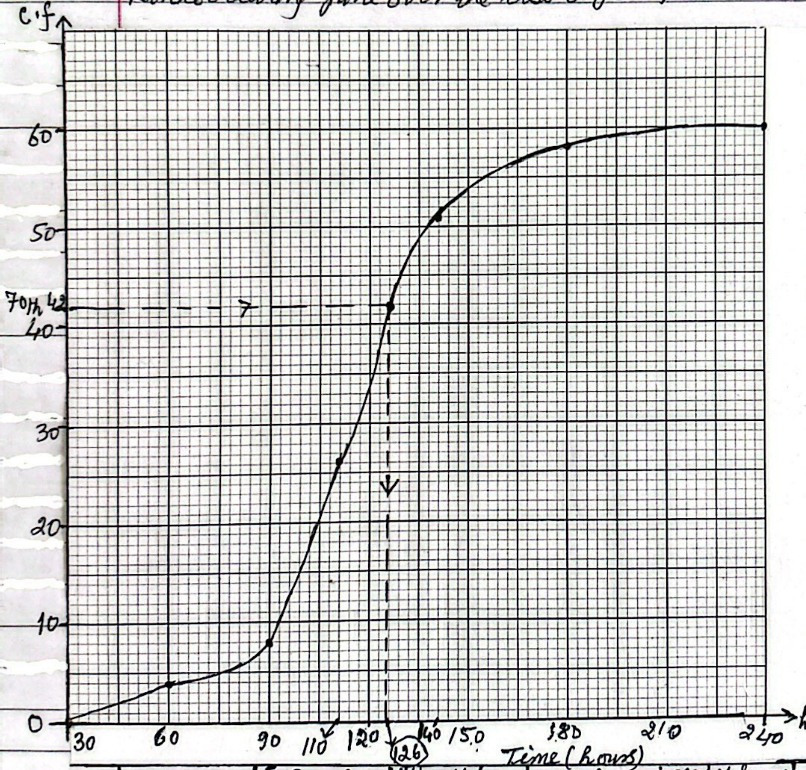
x	$30 \leq x < 60$	$60 \leq x < 90$	$90 \leq x < 110$	$110 < x \leq 140$	$140 \leq x < 180$	$180 \leq x < 240$
Number of Years	4	8	14	25	7	2

(a) Draw a cumulative frequency graph to illustrate the data. ---[3]

(b) use your graph to estimate the 70th percentile of the data. ---[2]

(c) Calculate an estimate for the mean number of hours of sunshine in Kintoo during June over the last 60 years. --[3]

[M-23/52/Q1]



(b) 70th percentile

$= \frac{70}{100} \times 60 = 42^{nd}$ Value

$= \underline{126}$ ✓

(c) Mean $= \frac{\Sigma f \cdot x_1}{\Sigma f}$

$= \frac{6845}{60}$

$= \underline{114.08}$ ✓

Classes	Frequency f	Upper Value	C.f	Mid Value x_1	$f \cdot x_1$
30 – 60	4	$x < 60$	4	45	$4 \times 45 = 180$
60 – 90	8	$x < 90$	12	75	$8 \times 75 = 600$
90 – 110	14	$x < 110$	26	100	$14 \times 100 = 1400$
110 – 140	25	$x < 140$	51	125	$25 \times 125 = 3125$
140 – 180	7	$x < 180$	58	160	$7 \times 160 = 1120$
180 – 240	2	$x < 240$	60	210	$2 \times 210 = 420$
	$\Sigma f = 60$				$\Sigma f \cdot x_1 = 6845$

Example 7: The numbers of chocolate bars sold per day in a cinema over a period of 100 days are summarised in the following table:

No of Chocolate bars sold	1-10	11-15	16-30	31-50	51-60
Number of days	18	24	30	20	8

(a) Draw a histogram to represent the information. --[5]

(b) What is the greatest possible value of the interquartile range for the data -[2]

(c) Calculate estimate of the mean and standard deviation of the number of chocolate bars sold. [S-20/51/Q7] --[4]

Solution:

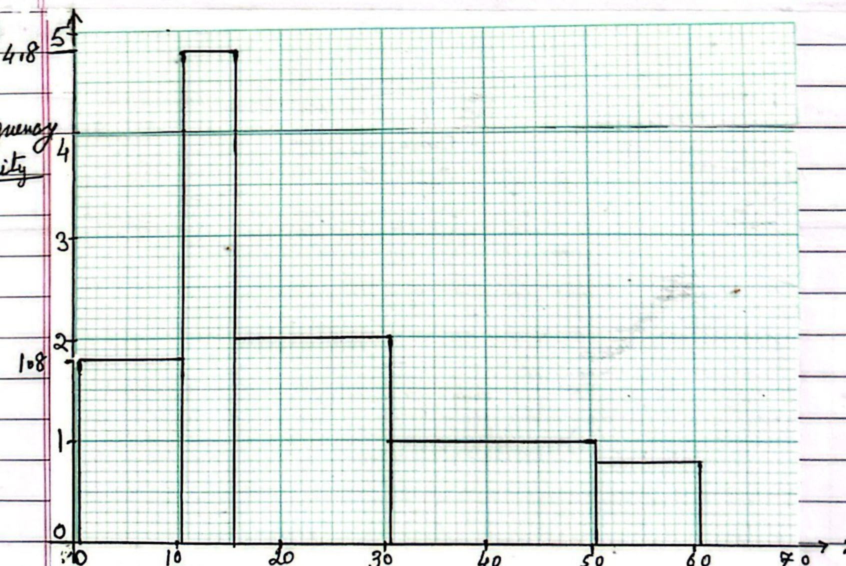
Classes	Class Boundaries	frequency	midpoint x	fx	fx^2	Frequency density = Frequency / Class width
1-10	0.5-10.5	18	5.5	18×5.5	18×5.5^2	$\frac{18}{10}=1.8$
11-15	10.5-15.5	24	13	24×13	24×13^2	$\frac{24}{5}=4.8$
16-30	15.5-30.5	30	23	30×23	30×23^2	$\frac{30}{15}=2$
31-50	30.5-50.5	20	40.5	20×40.5	20×40.5^2	$\frac{20}{20}=1$
51-60	50.5-60.5	8	55.5	8×55.5	8×55.5^2	$\frac{8}{10}=0.8$
		$\Sigma f=100$ $n=$	$\Sigma fx=2355$		$\Sigma fx^2=77917.5$	

(C) $\text{Mean}=\Sigma fx/n = 2355/100 = 23.6$ ✓ (Continued →)

$\text{Variance} = \frac{\Sigma fx^2}{N} - \text{mean}^2 = \frac{77917}{100} - (23.6)^2 = 224.57$ ✓

$\therefore$ Standard deviation $=\sqrt{224.57} = 15.0$

(a)



(b)

LQ - 11-15, UQ = 31-50

Greatest IQR = 50 - 11 = 39 ✓

Example 8: For n values of the variable x, it is given that
$$\Sigma(x-50)=144 \quad \text{and } \Sigma x=944$$
Find the value of n. [S-20/52/Q1] ---[3]

Solution: $\Sigma(x-50)=144 \Rightarrow \Sigma x-50n=144$
$\Rightarrow 944-50n=144$ $\quad [\because \Sigma x=944]$
$\Rightarrow 50n=800 \Rightarrow n=16$ ✓

Example 9: Two machines, A and B, produce metal rods of a certain type. The lengths, in metres, of 19 rods produced by machine A and 19 rods produced by machine B are shown in the following back-to-back stem-and-leaf diagram. [S-20/52/Q3]

A		B
	21	1 2 4
7 6 3 0	22	2 4 5 5 6
8 7 4 3 1 1	23	0 2 6 8 9 9
5 5 5 3 2	24	3 3 4 6
4 3 1 0	25	6

Key: 7/22/4 means 0.227 m for machine A and 0.224 for machine B.

(a) Find the median and the interquartile range for machine A. ---[3]

It is given that for machine B the median is 0.232 m, the lower quartile is 0.224 m and the upper quartile is 0.243 m.

(b) Draw box-and-whisker plots for A and B. --[3]

(c) Hence make two comparisons between the lengths of the rods produced by machine A and those produced by machine B. ---[2]

[S-20/52/Q3]

Solution (a) $n=19$, Median of machine A $=\frac{19+1}{2}$ th $=10$ th data $=0.238$ ✓

Now in lower half No $=9$, $LQ=\frac{9+1}{2}=5$ th data $=0.231$ ✓

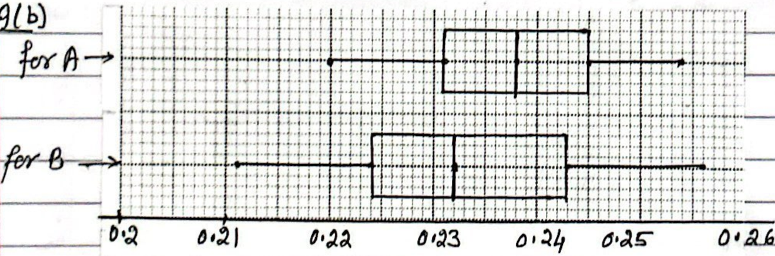
and in upper half No $=9$, $UQ=\frac{9+1}{2}=5$ th from median $=0.245$ ✓

$\therefore$ IQR $=0.245-0.231=\underline{0.014}$ ✓

(Continued →)

(Continued→)

Example 9(b)



	Lower End point	LQ	M	UQ	Upper End Pt	IQR
A	0·220	0·231	0·238	0·245	0·254	0·014
B	0·211	0·224	0·232	0·243	0·256	0·019

(C) (i) Lengths of rods produced by machine A are longer (Comparision of medians)

(ii) Lengths of rods produced by machine A are less spread out (Comparision of IQR)

Example 10: The annual salaries, in thousands of dollars, for 11 employees at each of two companies A and B are shown below.

Company A	30	32	35	41	41	42	47	49	52	53	64
Company B	26	47	30	52	41	38	35	42	49	31	42

(a) Represent the data by drawing a back-to-back stem-and-leaf diagram with company A on the left hand side of the diagram. --[4]

(b) Find the median and the interquartile range of the salaries of the employees in company A. --[3]

A new employee joins company B. The mean salary of the 12 employees is now \$ 38500.

(c) Find the salary of the new employee. [S-20/53/Q6] --[3]

Solution:

(a)

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

Key 1|4|2 means \$ 41000 for A and \$ 42000 for B. ✓

(b) Median of A, $\frac{11+1}{2}$th $= 6_{th}$ data $=$ \$ 42000

below Med → LQ = lower half 5 data $= \frac{5+1}{2} =$ 3rd data = \$ 35000 ✓

Above Med → UQ = upper half has 5 data $= \frac{5+1}{2} =$ 3rd data = \$ 52000 ✓

IQR = UQ − LQ = 52000 − 35000 = \$ 17000 ✓

(c) for B, Sum of 11 number = 433000

Sum of 12 numbers including the new = 38500 × 12 = 462000

∴ Salary of the new employee = 462000 − 433000 = \$ 29000 ✓

11. The time taken by 200 players to solve a computer puzzle are summarised in the following table:

Time (t seconds)	$0 \leqslant t < 10$	$10 \leqslant t < 20$	$20 \leqslant t < 40$	$40 \leqslant t < 60$	$60 \leqslant t < 100$
Number of players	16	54	78	32	20

(a) Draw a histogram to represent this information. --[4]

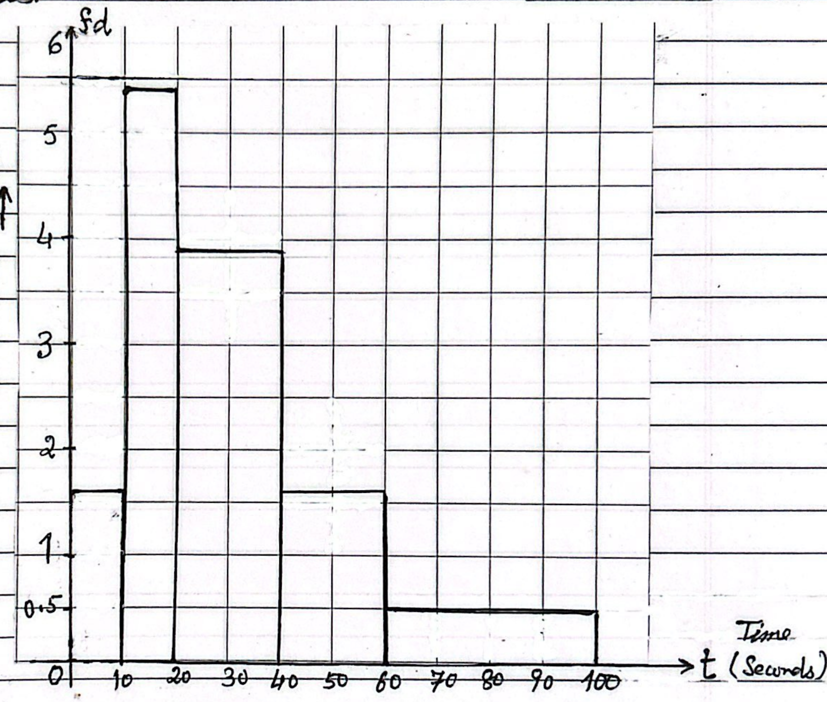
(b) Calculate an estimate of mean time by these 200 players. ---[2]

* (C) Find the greatest possible value of interquartile range of these times. [S-21/51/Q5] --[2]

Solution (a)

Frequency density:

Class Width	frequency density
10	1.6
10	5.4
20	3.9
20	1.6
40	0.5



(b)

Classes	f	Mid Value x	$f x_i$
0 - 10	16	5	80
10 - 20	54	15	810
20 - 40	78	30	2340
40 - 60	32	50	1600
60 - 100	20	80	1600

$\Sigma f = 200$ | $\Sigma fx = 6430$

$$\text{Mean} = \frac{\Sigma fx}{\Sigma f} = \frac{6430}{200} = \underline{32.15} \checkmark$$

(C) $UQ = 200 \times \frac{3}{4} = 150_{th} = (40-60)$ Interval

$L.Q = 200 \times \frac{1}{4} = 50_{th} = (10-20)$ Interval

$\therefore$ Greatest possible value of interquartile range $= 60 - 10$
$= \underline{50} \checkmark$

12. The heights, in cm, of the 11 basketball players in each of two clubs, the Amazons and the Giants, are shown below.

Amazons	205	198	181	182	190	215	201	178	202	196	184
Giants	175	182	184	187	189	192	193	195	195	195	204

(a) State an advantage of using a stem-and-leaf diagram compared to a box-and-whisker plot to illustrate this information. --[1]

(b) Represent the data by drawing a back-to-back stem-and-leaf diagram with Amazons on the left-hand side of the diagram. --[4]

(c) Find the interquartile range of the heights of the players in the Amazons [2]

Four new players join the Amazons. The mean height of the 15 players in the Amazons is now 191·2 cm. The heights of three of the new players are 180 cm, 185 cm and 190 cm.

(d) Find the height of the fourth new player. --[3]

S-21/52/Q7

Solution (a) Includes all data.

(b)

Amazons		Giants
8	17	5
4 2 1	18	2 4 7 9
8 6 0	19	2 3 5 5 5
5 2 1	20	4
5	21	

Key: 1/18/2 means 181 cm for Amazons and 182 cm for Giants.

(C) $Me = \frac{11+1}{2} = 6_{th}$ item.

Now for L.Q. consider, Lower half: 178, 181, 182, 184, 190

$\frac{5+1}{2}th = 3rd \Rightarrow LQ = 182$

Upper half: 198, 201, 202, 205, 215,

$\therefore$ UQ = 3rd = 202

$\therefore$ Interquartile range for Amazons = UQ − LQ = 202 − 182

$\therefore$ IQR = 20 (cm)

(d) Mean of 15 players $\bar{x} = 191\cdot2$

$\therefore$ Total height of 15 players $= n\bar{x} = 15 \times 191\cdot2 = 2868$ ---(i)

Let the height of 4_{th} new player = h cm.

Sum of 15 heights $= \Sigma_{11} h_i + (180+185+190) + h = 2132 + 555 + h$
$= 2687 + h$ ---(ii)

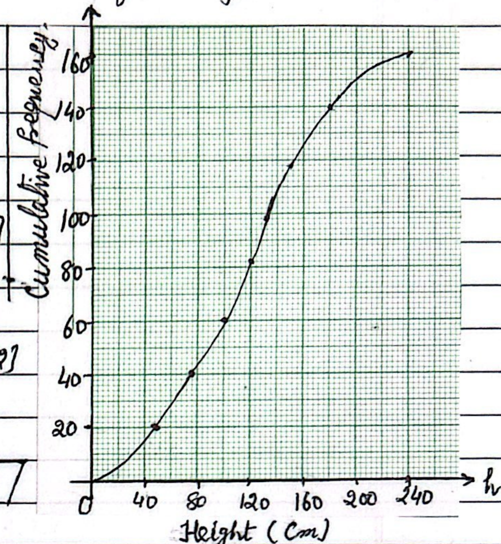
from (i) and (ii) $2687 + h = 2868 \Rightarrow h = 181$ cm

13. The heights in cm of 160 sunflower plants were measured. The results are summarised on the following cumulative frequency curve.

(a) Use the graph to estimate the number of plants with heights less than 100cm. ---[1]

(b) Use the graph to estimate the 65th percentile of the distribution. ---[2]

(c) Use graph to estimate the interquartile range of the heights of these plants. ---[2]



| S-21 | 53 | Q 1 |

Solution (a) Number of plants with heights less than 100cm = <u>60</u> ✓

(b) 65_{th} percentile $= 65\%\ 140 = \frac{65}{160} \times 140 = 104_{th} = \underline{136\ cm}$ ✓

(c) $UQ = \frac{75}{100} \times 160 = 120_{th} = 150$

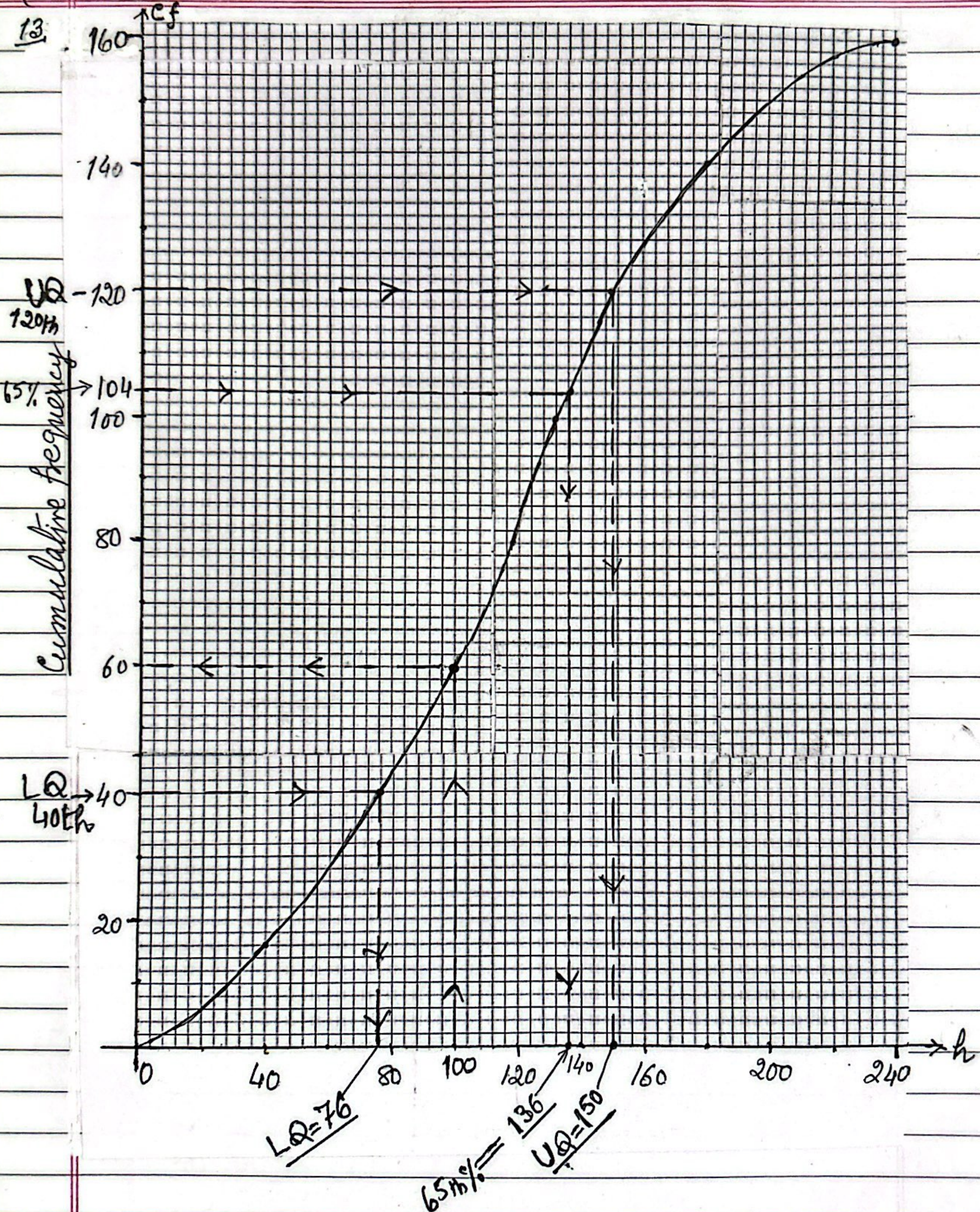
$LQ = \frac{25}{100} \times 160 = 40_{th} = 76$

$\therefore\ IQR = UQ - LQ = 150 - 76 = \underline{74\ cm.}$

(Graph on the next page)

(Continued→)

13.



14 A sports club has a volleyball team and a hockey team. The heights of the 6 members of the volleyball team are summarised by $\Sigma x = 1050$ and $\Sigma x^2 = 193700$, where x is the height of a member in cm. The heights of the 11 members of the hockey team are summarised by $\Sigma y = 1991$ and $\Sigma y^2 = 366400$, where y is the height of a member in cm.

(a) Find the mean height of all 17 members of the club. ---[2]

(b) Find the standard deviation of the heights of all 17 members of the club. ---[3]

S-21/53/Q3

Solution (a) Mean height $= \dfrac{\Sigma x + \Sigma y}{6+11}$

$= \dfrac{1050 + 1991}{17}$

$= \dfrac{3041}{17} = 178.9$

$\therefore$ Mean height $= 178.9$ cm ---(i)

(b) Variance of heights of all 17 members

$= \dfrac{\Sigma x^2 + \Sigma y^2}{6+11} - (\text{Mean }fall)^2$ ---(ii)

Now $\dfrac{\Sigma x^2 + \Sigma y^2}{6+11} = \dfrac{193700 + 366400}{17} = \dfrac{560100}{17}$

$\therefore$ Variance $= \dfrac{560100}{17} - (178.9)^2$ [from (i) & (ii)]

$= 948.289$

$\therefore$ Standard deviation $= \sqrt{Var} = \sqrt{948.289}$

$= 30.794$

$\therefore$ S.D $= 30.8$

Scanned with CamScanner

15. The times taken to travel to college by 2500 students are summarised:

Time taken (t minutes)	$0 \leq t < 20$	$20 \leq t < 30$	$30 \leq t < 40$	$40 \leq t < 60$	$60 \leq t < 90$
Frequency	440	720	920	300	120

(a) Draw a histogram to represent this information. ---[4]

From the data, the estimate of the mean value of t is 31.44.

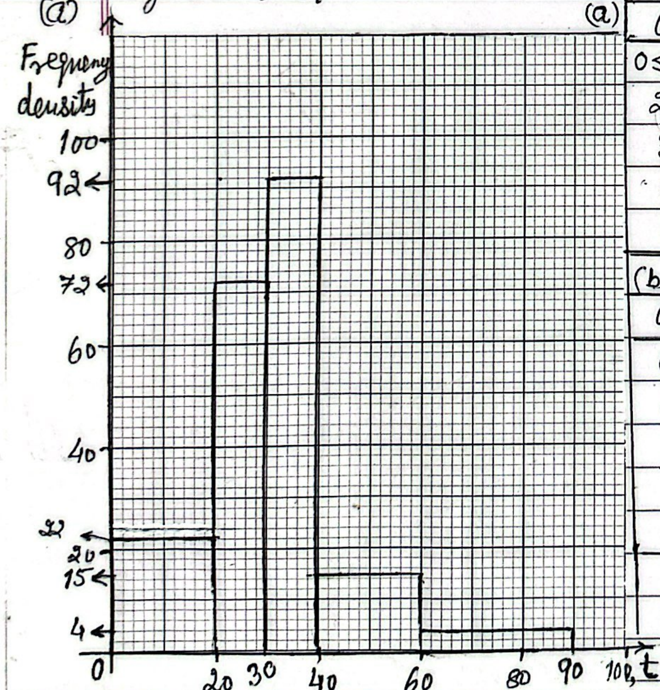
(b) Calculate the estimate of the standard deviation of the times taken to travel to college. ---[3]

(c) In which class interval does the upper quartile lie. ---[1]

It was later discovered that the times taken to travel to college by two students were incorrectly recorded. One student's time was recorded 15 instead of 5 and the other student time was recorded as 65 intead of 75.

(d) Without doing any further calculations, state whether the estimate of the standard deviation part (b) would be increased, decreased or stay the same. [S-22/51/Q3] ---[1]

(a)



(a)

Classes	Frequency	Class width	Frequency density $= f/c.w$
$0 \leq t < 20$	440	20	$\frac{440}{20} = 22$
20-30	720	10	$\frac{720}{10} = 72$
30-40	920	10	$920/10 = 92$
40-60	300	20	$300/20 = 15$
60-90	120	30	$120/30 = 4$

(b)

Classes	Frequency	Mid Value x	fx^2
0-20	440	10	440×10^2
20-30	720	25	720×25^2
30-40	920	35	920×35^2
40-60	300	50	300×50^2
60-90	120	75	120×75^2

$n = 2500$; $\Sigma fx^2 = 304,6000$

$$\text{Variance} = \frac{\Sigma fx^2}{n} - (\text{mean})^2$$

$$= \frac{3046000}{2500} - (31.44)^2$$

$$= 229.9264$$

$$S.D = \sqrt{229.9264} = \underline{15.2}$$

(c) upper quartile $\frac{3}{4} \times 2500 = 1875$ lies in $\boxed{30-40}$

(d) Stays the same; as the data still in the same interval.

16. For n values of the variabe x, It is given that:

$\Sigma(x-200)=446$ and $\Sigma x = 6846$; Find the value of n.

[S-22/52/Q1] ---[3]

Solution: $\Sigma(x-200)=446$

$\Rightarrow \Sigma x - \Sigma 200 = 446$

$\Rightarrow 6846 - 200n = 446$ [Given $\Sigma x = 6846$]

$\Rightarrow 200n = 6846 - 446 = 6400$

$n = 6400/200 = \underline{32}$ ✓

Scanned with CamScanner

17 The back-to-back stem-and-leaf diagram shows the diameters, in cm, of 19 cylindrical pipes produced by each of two companies A & B.

Company A		Company B	
4	33	1 2 8	Key: 1/35/3 means the pipe diameter of Company A is 0.351 cm and from Company B is 0.353 cm.
9 8 3 2 0	34	1 6 8 9 9	
8 7 5 4 1 1	35	1 2 2 3	
9 6 5 2	36	5 6	
4 3 1	37	0 3 4	
	38	2 8	

(a) Find the median and interquartile range of the pipes produced by company A. ---[3]

It is given that for the pipes produced by company B the lower quartile, median and the upper quartile are 0.346 cm, 0.352 cm and 0.370 cm respectively.

(b) Draw box-and-whisker plots for companies A and B on the grid. ---[3]

(c) Make one comparision between the diameters of the pipes produced by companies A and B. [S-22/52/Q3] --[1]

Solution (a) for company A.

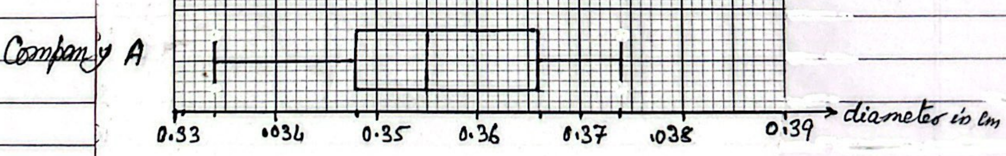
Median $= \frac{19+1}{2}$ th $= 10$th value $= \underline{0.355}$ ✓

L.Q is the mid value of lower half, 9 items $= \frac{9+1}{2}$ th $= 5$th value $= 0.348$ ✓

U.Q is the mid value of upper half 9 items $= \frac{9+1}{2}$ th $= 5$th value of upper half $= 0.366$

$\therefore$ IQR $=$ UQ $-$ LQ $= 0.366 - 0.348 = \underline{0.018\text{ cm}}$. ✓

(b)



(C) Company B has a higher spread than company A.

18 The time taken, t minutes, to complete a puzzle was recorded for each of 150 students. These times summarised in the table.

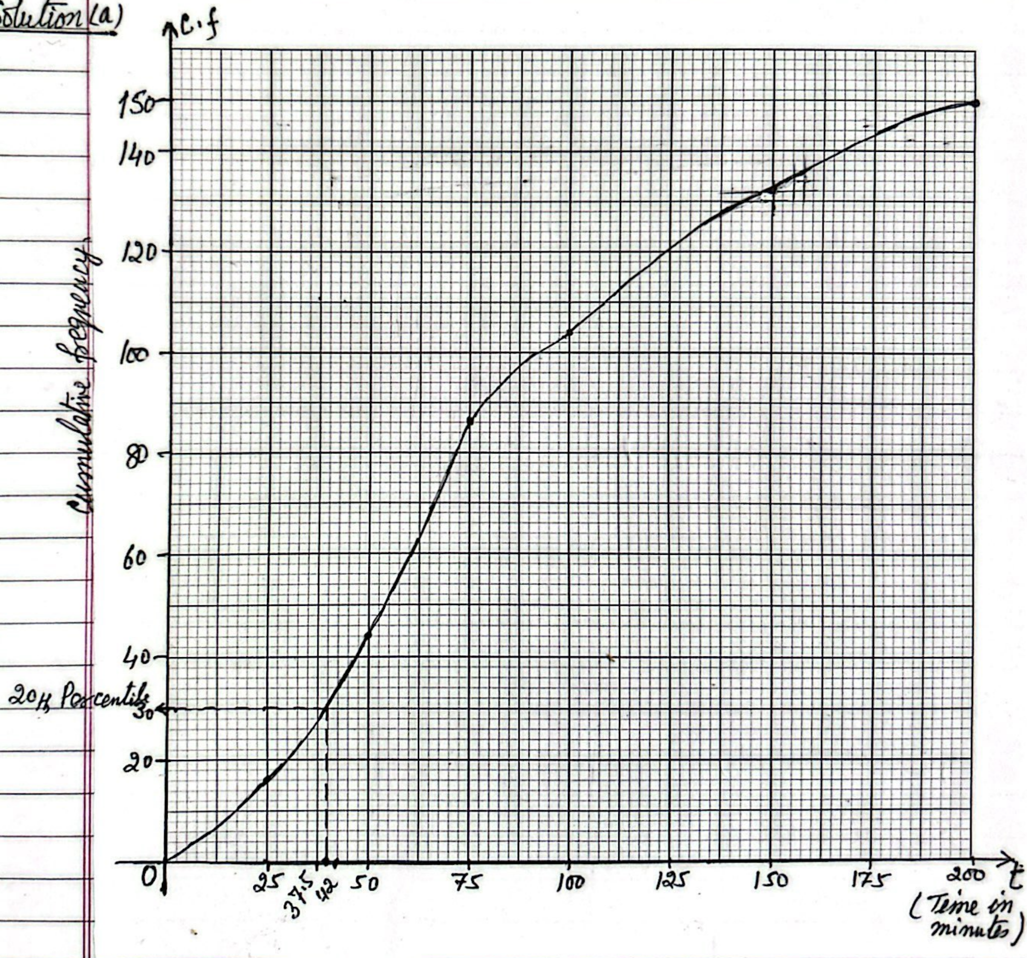
Time taken (t minutes)	$t \leq 25$	$t \leq 50$	$t \leq 75$	$t \leq 100$	$t \leq 150$	$t \leq 200$
Cumulative frequency	16	44	86	104	132	150

(a) Draw a cumulative frequency graph to illustrate the data. ---[2]

(b) Use your graph to estimate the 20th percentile of the data. ---[1]

S-22/53/Q1

Solution (a)



(b) 20th percentile $= \frac{20}{100} \times 150_{th} = 30_{th}$ item $=$ lies between $t = 37.5$ and 42 minutes.

19. Twenty children were asked to estimate the height of a particular tree. Their estimates, in metres, were as follows:

4.1	4.2	4.4	4.5	4.6	4.8	5.0	5.2	5.3	5.4
5.5	5.8	6.0	6.2	6.3	6.4	6.6	6.8	6.9	19.4

(a) Find the mean of the estimated heights. ---[1]

(b) Find the median of the estimated heights. ---[1]

(c) Give a reason why the median is likely to be more suitable than the mean as a measure of the central tendency for this information.

[S-22/53/Q2] ---[1]

Solution (a) Mean $= \frac{\Sigma x}{n} = \frac{123.4}{20} = \underline{6.17}$ ✓

(b) Median $= \frac{10th + 11th \text{ value}}{2} = \frac{5.4+5.5}{2} = \underline{5.45}$ (m) ✓

(c) The mean is unduly influenced by an extreme value, 19.4 ✓

Scanned with CamScanner

20 A summary of 50 values of x gives: $\Sigma(x-q)=700$, $\Sigma(x-q)^2=14235$ where q is a constant.

(a) Find the standard deviation of these values of x. ---[2]

(b) Given that $\Sigma x = 2865$, find the value of q. ---[2]

S-23/51/Q1

Solution: $n=50$, $\Sigma(x-q)=700$ and $\Sigma(x-q)^2=14235$

(a) $\text{Var } x = \left[\frac{\Sigma(x-q)^2}{n} - \left(\frac{\Sigma(x-q)}{n}\right)^2\right] = \frac{14235}{50} - \left(\frac{700}{50}\right)^2 = 284.7 - 196 = 88.7$

$\therefore$ $s.d = \sqrt{\text{Var } x} = \sqrt{88.7} = \underline{9.42}$

(b) Given $\Sigma(x-q)=700 \Rightarrow \Sigma x - nq = 700 \Rightarrow 2865 - 50q = 700 \Rightarrow q = 43.3\ \left(43\frac{3}{10}\right)$

21. The population of 150 villages in the U.K. to the nearest hundred, are:

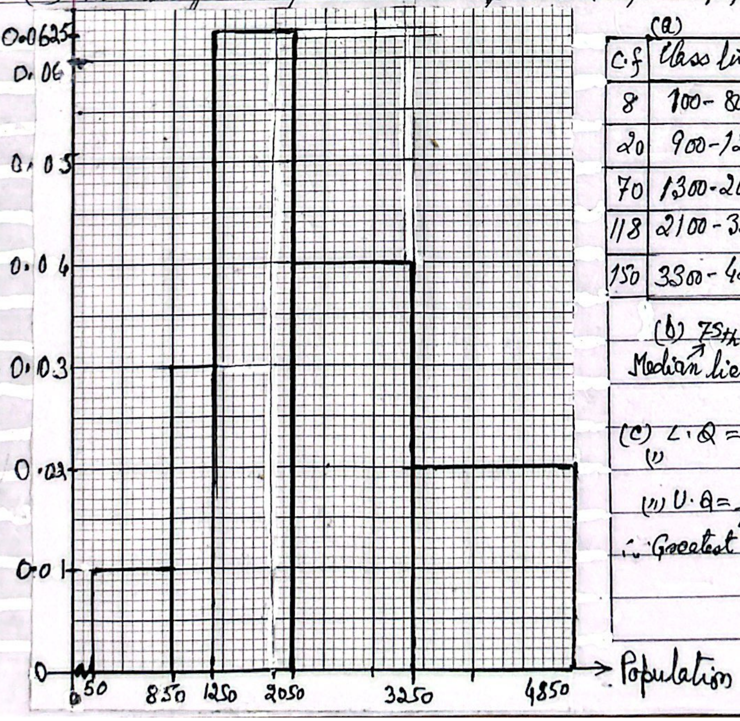
Population	100 - 800	900 - 1200	1300 - 2000	2100 - 3200	3300 - 4800
Number of villages	8	12	50	48	32

(a) Draw a histogram to represent this information; ---[4]

(b) Write down the class interval which contains the median for this information. ---[1]

(c) Find the greatest possible value of the I.Q.R for the population of 150 villages. ---[2]

S-23/51/Q5



(a)

C.f	Class limit	Class Boundaries	f	$f.d = \frac{fre}{C.W}$
8	100 - 800	050 - 850	8	$\frac{8}{800} = 0.01$
20	900 - 1200	850 - 1250	12	$\frac{12}{400} = 0.03$
70	1300 - 2000	1250 - 2050	50	$\frac{50}{800} = 0.0625$
118	2100 - 3200	2050 - 3250	48	$\frac{48}{1200} = 0.04$
150	3300 - 4800	3250 - 4850	32	$\frac{32}{1600} = 0.02$

$\Sigma f = 150$

(b) 75_{th} value

Median lies in the Class (2100 - 3200) or (2050 - 3250)

(c) (i) $L.Q = \frac{150}{4} = 37.5_{th}$ values lies in (1250 - 2050)

(ii) $U.Q = \frac{3}{4} \times 150 = 112.5_{th}$ (2050 - 3250)

$\therefore$ Greatest value of I.Q.R

$= (3249 - 1250)$

$= \underline{1999}$

22. The following back-to-back step-and-leaf diagram represents the monthly salaries, in dollars, of 27 employees at each of two companies, A and B.

Company A		Company B
5 4 1 1 0	25	4 4 5 6 6 7
9 9 8 7 2 1 0	26	0 1 3 5 5 7 9 9
8 6 4 2 1 0	27	1 3 4 6 6 8 8
6 5 4 2 0	28	0 1 2 2 2
9 8 5	29	
1	30	9

Key 1|27|6 means \$2710 for Company A, and \$2760 for Company B.

(a) Find the median and interquartile range of the monthly salries of employees in company A. ...[3]

The lower quartile, median and upper quartile for company B are \$2600, \$2690 and \$2780 respectively.

(b) Draw two box-and whisker plots in a single diagram to represent the information for the salaries of the employees at companies A and B. ...[3]

(c) Comment on whether the mean would be a more appropriate measure than the median for comparing the given information for the two companies. ...[1]

S-23/52/Q3

For Company A:

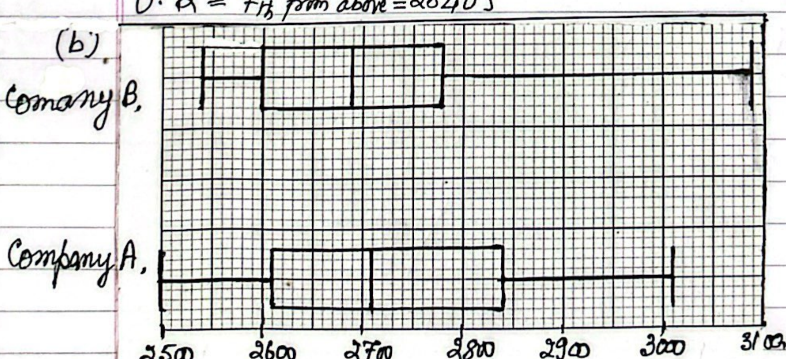
Solution (a) $Med = \frac{27+1}{2} = 14_{th}$ Value $= \$2710$.

$L.Q = \frac{13+1}{2} = 7_{th} = 2610$

$U.Q = 7_{th}$ from above $= 2840$

$\Rightarrow I.Q.R = 2840 - 2610 = \underline{230}$ ✓

(b)



(c) Mean is less appropriate than median because of the extreme value for company B (at \$3090).

23 The times taken, in minutes, to complete a cycle race by 19 cyclists from each of two clubs, the cheetahs and the Panthers, are represented in the following back-to-back stem-and-leaf diagram:

Cheetahs		Panthers
9 8	7	4
8 7 3 2 0	8	6 8
9 8 7	9	1 7 8 9 9
6 5 3 3 1	10	2 3 4 4 5 6
9 8 2	11	1 2 8
4	12	0 6

Key 7|9|1 means 97 minutes for Cheetahs and 91 minutes for Panthers.

--- [3]

(a) Find the median and the interquartile range of times of the Cheetahs.

The median and the interquartile change for the Panthers are 103 minutes and 14 minutes respectively.

(b) Make two comparisons between the times taken by the Cheetahs and the times taken by the Panthers. --- [2]

Another cyclist, Kenny, from cheetahs also took part in the race. The mean time taken by 20 cyclists from the Cheetahs was 99 minutes.

(c) Find the time taken by Kenny to complete the race. --- [3]

S-23/53/Q4

Solution: For Cheetahs, Median $= \frac{19+1}{2}^{th} = 10^{th}$ Value $= \underline{99}$ (minutes)

(a) (Leave Med) $LQ = \frac{9+1}{2}^{th} = 5^{th}$ Value $= 83$ } Hence IQR $= 106 - 83$
$UQ = \frac{9+1}{2}^{th} = 5^{th}$ from above $= 106$ } $= \underline{23}$ (minutes)

(b) for Panthers Me $= 103$ min and IQR $= 14$ (min) Given

(less time → faster)

(i) The times for Cheetahs are faster than the times for the Panthers,

(ii) The times for the Cheetahs are more spread than times for Panthers.

(c) Total time including Kenny $= 99 \times 20 = 1980$

and time for 19 cheetahs $= 1862$

$\therefore$ Kenny's time $= 1980 - 1862 = \underline{118}$ (minutes)

24. The time, t minutes, taken by 150 students to complete a particular challenge are summarised in the following cumulative frequency table:

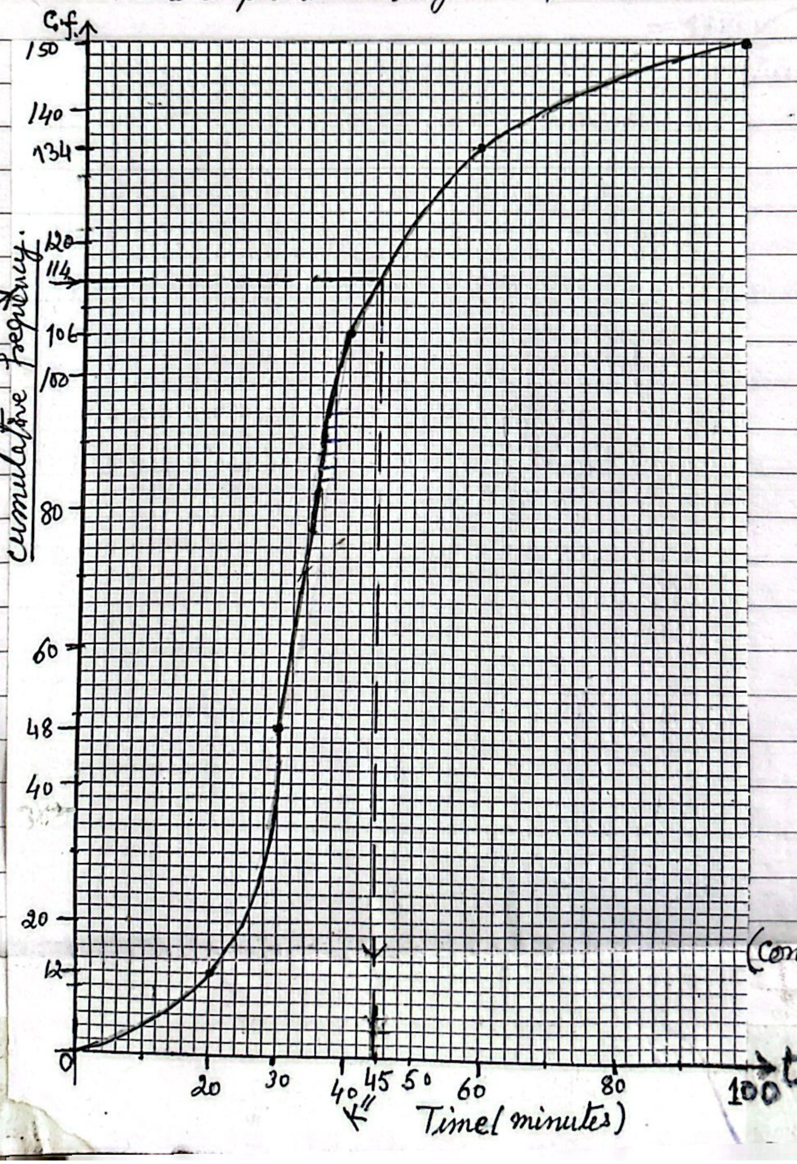
Time taken (t minutes)	$t \leq 20$	$t \leq 30$	$t \leq 40$	$t \leq 60$	$t \leq 100$
Cumulative frequency	12	48	106	134	150

(a) Draw a cumulative frequency graph to illustrate the data. ---[2]

(b) 24% of the students take k minutes or longer to complete the challenge. Use your graph to estimate the value of k. ---[2]

(c) Calculate the estimate of the mean and the standard deviation of the time taken to complete the challenge. [W-20/51/Q6] --[6]

Solution (a)



(b) $(100-24)\% = 76\%$

76% of $150 = 114$

$K = 45$ min

(Continued)

(continued→)

24(c)

Classes	frequency f	Mid Value x	$f\times x$	$f\times x^2$
0-20	12	10	120	1200
20-30	36	25	900	22500
30-40	58	35	2030	71050
40-60	28	50	1400	70000
60-100	16	80	1280	102400
	$\Sigma f = 150$		$\Sigma fx = 5730$	$\Sigma fx^2 = 267150$

$$\text{Mean} = \frac{\Sigma fx}{\Sigma f} = \frac{5730}{150}$$

$$\bar{x} = 38.2 ✓$$

$$\text{Var } x = \frac{\Sigma fx^2}{n} - (\bar{x})^2$$

$$= \frac{267150}{150} - (38.2)^2$$

$$= 1781 - 1459.24$$

$$\sigma^2 = 321.76$$

$\therefore$ Standard deviation $\sigma = \sqrt{321.76}$

$$= 17.9 ✓$$

25 The following table gives the weekly snowfall, in centimetres, for 11 weeks in 2018 at two ski resorts, Dados and Linva.

Dados	6	8	12	15	10	36	42	28	10	22	16
Linva	2	11	15	16	0	32	36	40	10	12	9

(a) Represent the information in a back to back stem and leaf diagram. --[4]

(b) Find the median and the interquartile range for the weekly snowfall in Dados. --[3]

(c) The median, lower quartile and upper quartile of the weekly snowfall for Linva are 12, 9 and 32cm. respectively. Use this information and your answers to part (b) to compare the central tendency and the spread of the weekly snowfall in Dados and Linva. [W-20/52/Q5] --[2]

Solution(a)

Dados		Linva
8 6	0	0 2 9
6 5 2 0 0	1	0 1 2 5 6
8 2	2	
6	3	2 6
2	4	0

Key 6/3/2 means 36cm (snow) in Dados and 32cm (snow) in Linva

(b) Median (or Q_2) =

$\frac{11+1}{2}$th = 6th = 15 (cm)

UQ (or Q_3) = 28 (cm) ✓

(upper half: $\frac{5+1}{2}$th = 3rd)

LQ (or Q_1) = 10 ✓

IQR = $Q_3 - Q_1$

= 28 − 10

= 18 (cm) ✓

(c) On the average the snow fall in Dados is higher

spread → The amount of snowfall in Linva varies more than in Dados

26. A particular piece of music was played by 91 pianist and for each pianist, the number of incorrect notes was recorded. The results are summerised in the table.

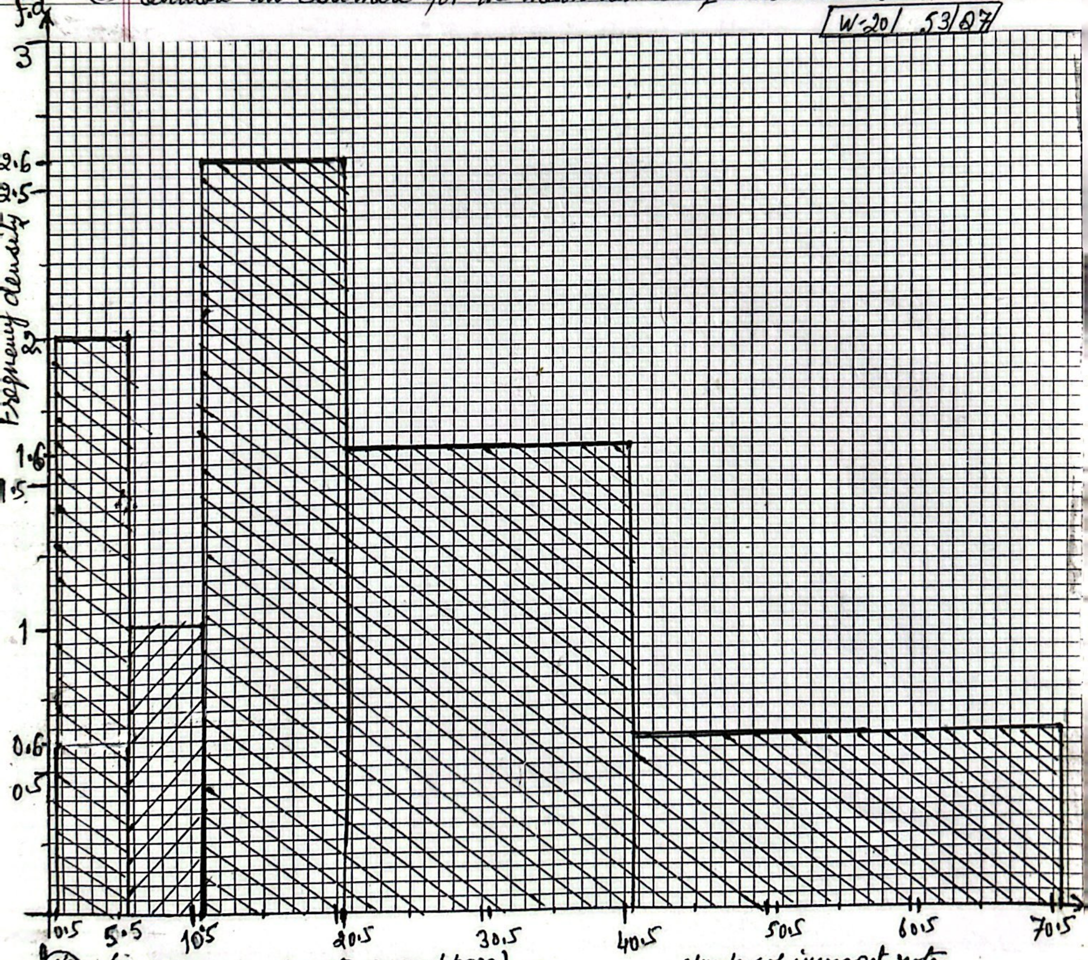
Number of incorrect Notes	1-5	6-10	11-20	21-40	41-70
Frequency	10	5	26	32	18

(a) Draw a histogram to represent this information ---[5]

(b) State which class interval contains the lower quartile and which class interval contains the upper quartile.
Hence find the greatest possible value of the interquartile range. ---[2]

(c) Calculate an estimate for the mean number of incorrect notes. ---[3]

[W-20/ 53/Q7]



(Find frequency density on next page) (continued)

Continued→

26 (a)

	Classes	Frequency f	Frequency density $= \frac{f}{\text{Class size}}$	Mid Value x	$f.x$
(C)	1-5	10	$\frac{10}{5} = 2$	3	30
	6-10	5	$\frac{5}{5} = 1$	8	40
	11-20	26	$\frac{26}{10} = 2.6$	15.5	403
	21-40	32	$\frac{32}{20} = 1.6$	30.5	976
	41-70	18	$\frac{18}{30} = 0.6$	55.5	999
		$\Sigma f = 91$			$\Sigma fx = 2448$

$$\text{Mean} = \frac{\Sigma fx}{\Sigma f} = \frac{2448}{91} = \underline{26.9}$$

(b) $L.Q = \frac{45+1}{2} = 23^{th}$ = lies in the class 11-20 (in the lower half)

$U.Q = \frac{45+1}{2} = 23^{th}$ in upper half = lies in the class = 21-40

$\therefore$ Greatest IQR = $40 - 11 = \underline{29}$

27. The summary of 40 values of x gives the information:

$$\Sigma(x-k) = 520 \quad , \Sigma(x-k)^2 = 9640$$

where k is a constant.

(a) Given that the mean of 40 values of x is 34, find the value of k. ---[2]

(b) Find the variance of these 40 values of x. ---[2]

[W-21/51/Q2]

Solution: $n=40$, Given mean $=34$

(a) $\text{Mean} = k + \frac{\Sigma(x-k)}{40} = 34$ Given

$\Rightarrow k + \frac{520}{40} = 34$ [$\because \Sigma(x-k)=520$]

$\Rightarrow k = 34 - 13 = 21$ ✓

(b) $\text{Var} = \frac{\Sigma(x-k)^2}{n} - \left(\frac{\Sigma(x-k)}{n}\right)^2$

$= \frac{9640}{40} - \left(\frac{520}{40}\right)^2$

$= 241 - 13^2 = 72$ ✓

28. The weights, in kg, of 15 rugby players in the Rebels club and 15 soccer players in the Sharks club are:

Rebels	75	78	79	80	82	82	83	84	85	86	89	93	95	99	102
Sharks	66	68	71	72	74	75	75	76	78	83	83	84	85	86	92

(a) Represent the data by drawing a back-to-back stem-and-leaf diagram with Rebels on the left-hand side of the diagram. ---[4]

(b) Find the median and inter-quartile range for the Rebels. ---[3]

[W-21/51/Q6]

Solution:

(a)

Rebels		Sharks
	6	6 8
9 8 5	7	1 2 4 5 5 6 8
9 6 5 4 3 2 2 0	8	3 3 4 5 6
9 5 3	9	2
2	10	

Key 8/7/2 mean 78 kg for Rebels and 72 kg for Sharks.

(b) Now for Rebels:

$n = 15$, Median $= \frac{15+1}{2} = 8_{th}$ Value $= 84$ (kg) ✓

Lower quartile: Mid Value of lower half: 75, 78, 79, 80, 82, 82, 83, $n_1 = 7$

$LQ = \frac{7+1}{2}{}_{th} = 4_{th}$ Value $= 80$

Upper quartile: Mid value of upper half: 85, 86, 89, 93, 95, 99, 102

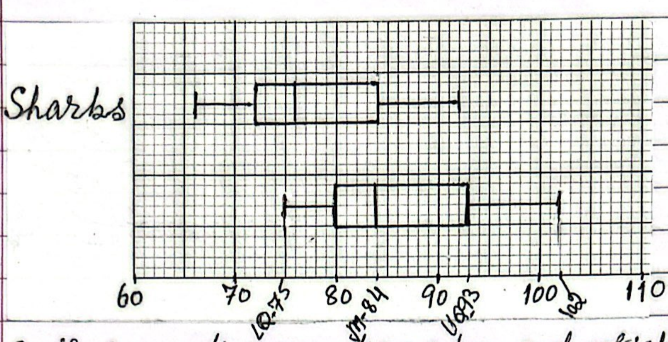
$UQ = \frac{7+1}{2}{}_{th} = 4_{th}$ value $= 93$ ✓

$\therefore IQR = UQ - LQ = 93 - 80 = 13$ (kg)

(Continued →)

Continued:

28. A box-and-whisker plot for the Sharks is shown below:



(c) On the same diagram, draw a box-and-whisker plot for the Rebels. ---[2]

(d) Make one comparision between the weight of the players in the Rebels club and the weights of the players in the Sharks club. ---[1]

Ans: Average weight of Rebels is higher than average weight of Sharks.

29. The distances, x m, travelled to school by 140 children were recorded. The results are summarised in the table below:

Distance, x m	$x \leq 200$	$x \leq 300$	$x \leq 500$	$x \leq 900$	$x \leq 1200$	$x \leq 1600$
Cumulative frequency	16	46	88	122	134	140

(a) On the grid, draw a cumulative frequency graph to represent the results. [2]

(b) Use your graph to estimate the interquartile range of the distances. [2]

(c) Calculate estimates of mean and standard deviation of the distances. --[6]

[W-21/52/Q7]

Solution: (c)

Distances	C.f	Classes	Frequency f	Mid Value x	$f.x$	$f.x^2$
$x \leq 200$	16	0-200	16	100	$16\times100=1600$	$16\times100^2=1600000$
$x \leq 300$	46	200-300	30	250	$30\times250=7500$	$30\times250^2=1875000$
$x \leq 500$	88	300-500	42	400	$42\times400=16800$	$42\times400^2=6720000$
$x \leq 900$	122	500-900	34	700	$34\times700=23800$	$34\times700^2=16660000$
$x \leq 1200$	134	900-1200	12	1050	$12\times1050=12600$	$12\times1050^2=13230000$
$x \leq 1600$	140	1200-1600	6	1400	$6\times1400=8400$	$6\times1400^2=11760000$
			$n=140$		$\Sigma f.x=70700$	$\Sigma fx^2=50405000$

$$\text{Mean} = \frac{\Sigma fx}{n} = \frac{70700}{140} = \underline{505}$$

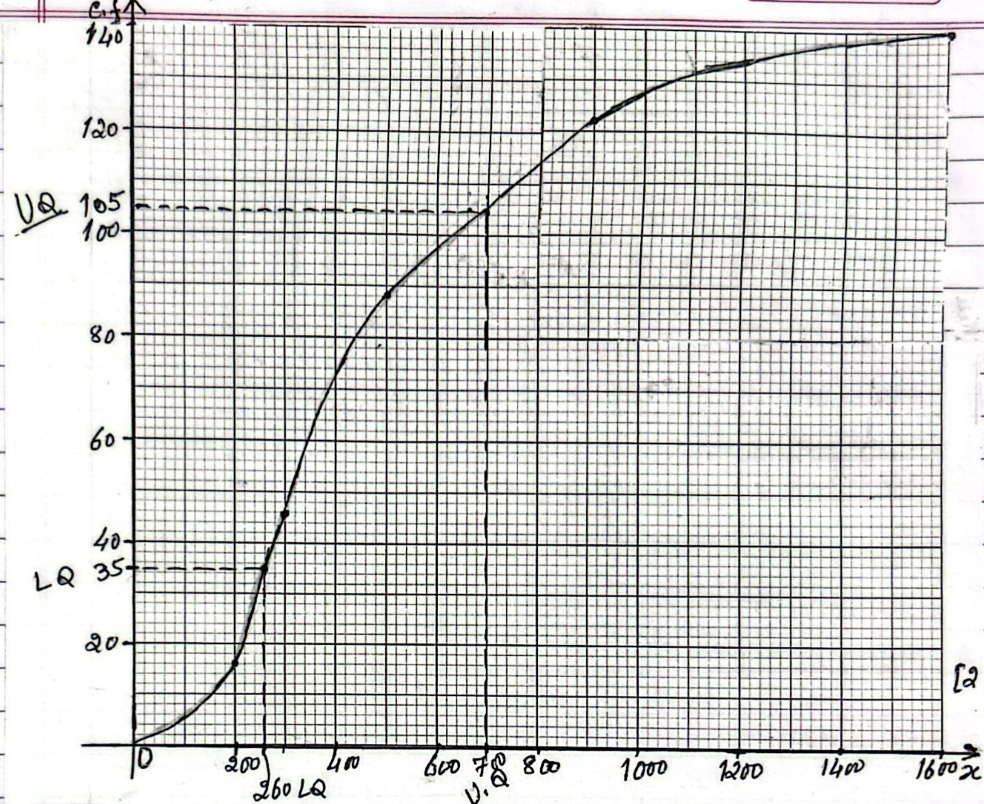
$$\text{Variance} = \frac{\Sigma fx^2}{n} - (\text{mean})^2$$

$$\text{Var} = \frac{50405000}{140} - (505)^2 = 105010.7$$

$$\therefore \text{S.D} = \sqrt{105010.7} = \underline{324}$$

(Continued →)

29 → (continued →)



[2]

(b) $UQ = 75\%$ of $140 = 105_{th}$ Value $= 700$ ✓

$LQ = 25\%$ of $140 = 35_{th}$ Value $= 260$

$\therefore$ IQR $= 700 - 260 = \underline{440}$ ✓

30. Lakeview and Riverside are two schools. The pupils at both schools took part in a competition to see how far they could throw a ball. The distances thrown, to the nearest metre, by 11 pupils from each school are shown as:

Lakeview	10	14	19	22	26	27	28	30	32	33	41
Riverside	23	36	21	18	37	25	18	20	24	30	25

(a) Draw a back-to-back stem-and-leaf diagram to represent this information, with Lakeview on the left-hand side. --[4]

(b) Find the interquartile range of the distances thrown by 11 pupils at Lakeview school. --[2]

[W-21/53/Q2]

Solution (a)

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

Key 6/2/3 means 26 m for Lakeview and 23 for Riverside.

"For Lakeview" → $\boxed{\text{Median } 6^{th} = 27}$

(b) LQ is the mid value of Lower half: 10, 14, 19, 22, 26,

$LQ = \frac{5+1}{2}^{th} = 3^{rd}\text{ Value} = 19$

UQ is the mid value of Upper half: 28, 30, 32, 33, 41

$UQ = \frac{5+1}{2}^{th} = 3^{rd} = 32$

$\therefore IQR = UQ - LQ = 32 - 19 = 13$

31 The times taken, in minutes, by 360 employees at a large company to travel from home to work are summarised as: [W-21/53/Q3]

Time, t minute	$0 \leq t < 5$	$5 \leq t < 10$	$10 \leq t < 20$	$20 \leq t < 30$	$30 \leq t < 50$
Frequency	23	102	135	76	24

[2]

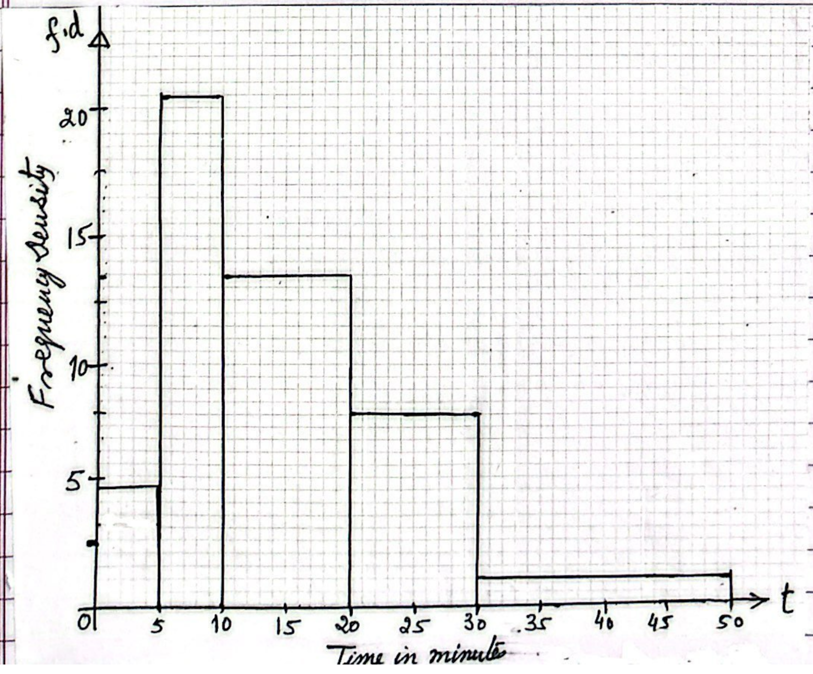
(b) Calculate an estimate of the mean time taken by an emplyee to travel to work.

Classes	f	Mid Value x	fx
$0 \leq t < 5$	23	2.5	$23 \times 2.5 = 57.5$
$5 \leq t < 10$	102	7.5	$102 \times 7.5 = 765$
$10 \leq t < 20$	135	15	$135 \times 15 = 2025$
$20 \leq t < 30$	76	25	$76 \times 25 = 1900$
$30 \leq t < 50$	24	40	$24 \times 40 = 960$
$n = \Sigma f = 360$			$\Sigma fx = 5707.5$

$$\text{Mean} = \frac{\Sigma fx}{n} = \frac{5707.5}{360} = 15.854 = 15.9 \text{ min.}$$

(a) Draw a histogram to represent the given information. --[4]

Classes	frequency f	frequency density = $\frac{\text{frequency}}{\text{Class size}}$
$0 \leq t < 5$	23	$fd = 23/5 = 4.6$
$5 \leq t < 10$	102	$102/5 = 20.4$
$10 \leq t < 20$	135	$135/10 = 13.5$
$20 \leq t < 30$	76	$76/10 = 7.6$
$30 \leq t < 50$	24	$24/20 = 1.2$



32 The Lions and the Tigers are two basketball clubs. The heights, in cm, of the 11 players in each of the squads are given in the table.

Lions	178	186	181	187	179	190	189	190	180	169	196
Tigers	194	179	187	190	183	201	184	180	195	191	197

(a) Draw a back-to-back stem-and-leaf diagram to represent this information. ---[4]

(b) Find the median and I.Q.R of the height of the Lions squad. ---[3]

It is given for the Tigers, the L.Q = 183 cm, the Me = 190 cm and UQ = 195 cm.

(c) Make two comparisions between the heights of the players of the two teems.

[W-22/51/Q3] ---[2]

Solution (a)

Lions		Tigers
9	16	
9 8	17	9
9 7 6 1 0	18	0 3 4 7
6 0 0	19	0 1 4 5 7
	20	1

Key 1/18/3 means 181 cm for Lions and 183 cm for Tigers.

(b) For Lions Squad;

Median $= \frac{11+1}{2}$th $= 6$th item $= \underline{186}$ ✓

Lower half (169, 178, 179, 180, 181) LQ $\frac{5+1}{2}$th $= 3$rd $\Rightarrow LQ = \underline{179\text{ cm}}$ ✓

Upper half (187, 189, 190, 190, 196) $UQ = \frac{5+1}{2} = 3$rd $= \underline{190}$ ✓

$\therefore IQR = UQ - LR = 190 - 179 = \underline{11}$ ✓

(c) (i) Tigers are (generally) taller.

(ii) Heights of Tigers are slightly less consistent than the heights of Lions.

[IQR of Tigers $195 - 183 = 12 >$ IQR of Lions $= \underline{11}$.]

33 The time taken, in minutes, to complete a word processing task by 250 employees are summarised in the table:

Time taken (t minutes)	$0 \leq t < 20$	$20 \leq t < 40$	$40 \leq t < 50$	$50 \leq t < 60$	$60 \leq t < 100$
Frequency	32	46	96	52	24

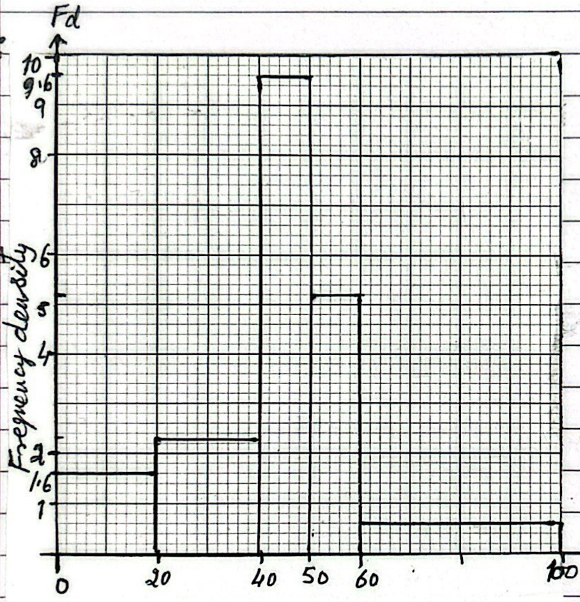
(a) Draw a histogram to represent this information. ---[4]

The estimate of mean time taken by 250 employees is 43.2 minutes.

(b) Calculate an estimate for the standard deviation of these times. ---[3]

W-22/52/Q4

Solution:



Classes	Frq.	Frequency density = $\frac{\text{Frequency}}{\text{C. Width}}$
0-20	32	$32/20 = 1.6$
20-40	46	$46/20 = 2.3$
40-50	96	$96/10 = 9.6$
50-60	52	$52/10 = 5.2$
60-100	24	$24/40 = 0.6$

(b)

Classes	f	Mid point x	$f \cdot x^2$
0-20	32	10	$32 \times 10^2 = 3,200$
20-40	46	30	$46 \times 30^2 = 41,400$
40-50	96	45	$96 \times 45^2 = 194,400$
50-60	52	55	$52 \times 55^2 = 157,300$
60-100	24	80	$24 \times 80^2 = 153,600$
	$\Sigma f = 250$		$\Sigma f x^2 = 549,900$

Given mean = 43.2

$$\text{Variance } \sigma^2 = \frac{\Sigma f x^2}{n} - (\text{mean})^2$$

$$= \frac{549900}{250} - (43.2)^2$$

$$= 2199.6 - 1866.24$$

$$\sigma^2 = 333.36$$

$$\therefore \text{Standard deviation } \sigma = \sqrt{333.36}$$

$$= 18.258$$

$$= \underline{18.3}$$

Scanned with CamScanner

34 50 values of variable x are summarised by.

$$\Sigma(x-20) = 35 \quad \text{and} \quad \Sigma x^2 = 25036$$

Find the variance of 50 values. [W-22/53/Q1] --- [3]

Solution: $\Sigma(x-20) = 35 \Rightarrow \Sigma x - 50\times 20 = 35$; $n = 50$

$$\Rightarrow \Sigma x = 1035$$

$$\text{Variance} = \frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2 = \frac{25036}{50} - \left(\frac{1035}{50}\right)^2 = 500.72 - (20.7)^2$$

$$Var(x) = 72.23$$

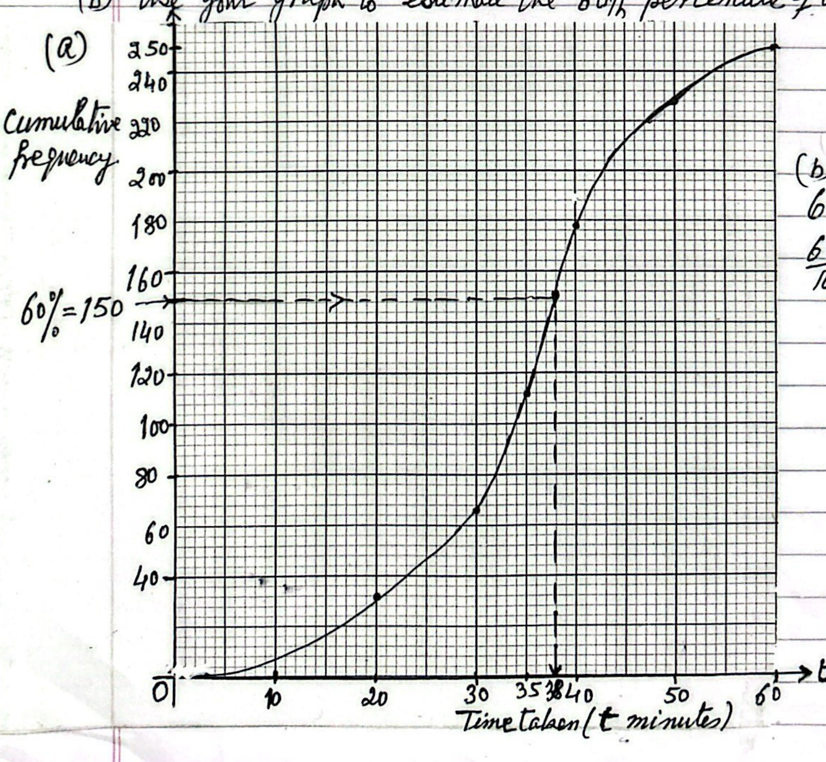
35 The time, t minutes, taken by 250 members of a club to take a walking challenge is summarised in the table:

Time taken (t minutes)	$t \le 20$	$t \le 30$	$t \le 35$	$t \le 40$	$t \le 50$	$t \le 60$
Cumulative frequency	32	66	112	178	228	250

(a) Draw a cumulative frequency graph to illustrate the data --[2]

(b) Use your graph to estimate the 60th percentile of the data. --[1]

(a)



[W-22/53/Q3]

(b)

60th Percentile

$\frac{60}{100} \times 250 = 150$th Value

$= 38$ minute (approx)

(Continued →)

(Continued→)

It is given that an estimate for the mean time taken to complete the challenge by these 250 members is 34.4 minutes.

35(c) Calculate an estimate for the standard deviation of the time taken to complete the challenge by these 250 members. --- [4]

W-22/53/Q3

Solution:

Time	Cumulati- frequen-	Classes	Fre. f	Mid point x	$f \cdot x^2$
$t \le 20$	32	0 - 20	32	10	$32 \times 10^2 = 3{,}200.0$
$t \le 30$	66	20 - 30	34	25	$34 \times 25^2 = 21{,}250.0$
$t \le 35$	112	30 - 35	46	32.5	$46 \times 32.5^2 = 48{,}587.5$
$t \le 40$	178	35 - 40	66	37.5	$66 \times 37.5^2 = 92{,}812.5$
$t \le 50$	228	40 - 50	50	45	$50 \times 45^2 = 101{,}250.0$
$t \le 60$	250	50 - 60	22	55	$22 \times 55^2 = 66{,}550.0$

$\Sigma f = 250$ $\Sigma f x^2 = 333650$

Variance $\sigma^2 = \dfrac{\Sigma f x^2}{\Sigma f} - (\bar{x})^2$

$= \dfrac{333650}{250} - (34.4)^2$ $(\bar{x} = 34.4 \text{ Given})$

$= 1334.6 - 1183.36$

$\sigma^2 = 151.24$

$\therefore$ Standard deviation $\sigma = \sqrt{151.24}$

$= 12.3$