Introduction to Statistics

1. Find the mean, median, mode and range of the following dataset : 10, 7, 14, 23, 15, 7, 32

Ans: Mean $(\bar{x}) = \frac{10+7+14+23+15+7+32}{7}$

$= \frac{108}{7} = 15.43$

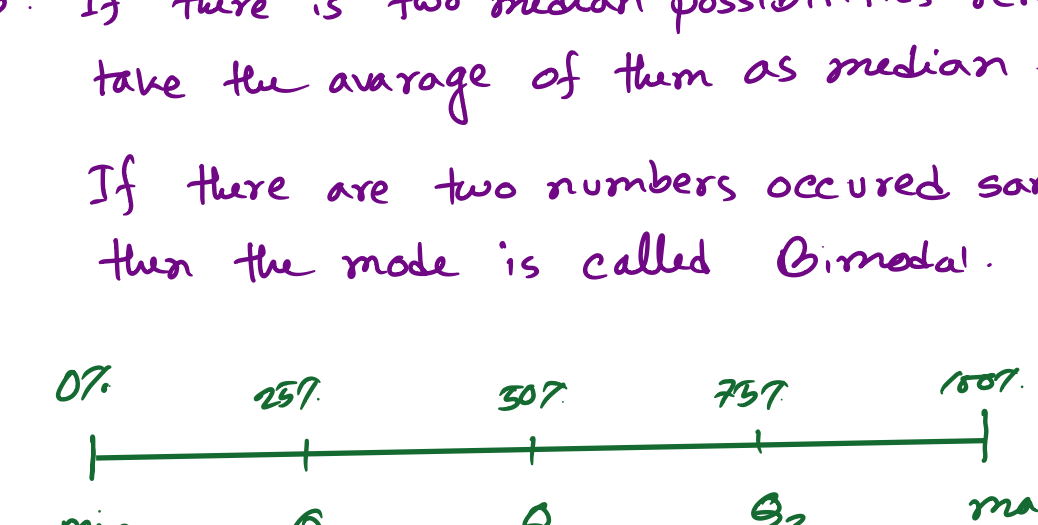
Arranging the numbers in increasing format
7, 7, 10, (14), 15, 23, 32

Median is basically the middle number.
So, here 14 is the median.

Mode: The most "occured" numbers
Here, '7' occures 2 times.

Range: The difference between the highest and lowest number
Here, $32 - 7 = 25$

N.B: If there is two median possibilities remain then take the average of them as median.
If there are two numbers occured same times then the mode is called Bimodal.



Here, Q_2 is the median of whole range.
Q_1 " " " " min & Q_2 range
Q_3 " " " " Q_3 & max range.

IQR → Inter Quartile range
$Q_3 - Q_1$

A number $[Q_1 - 1.5\,IQR,\ Q_3 + 1.5\,IQR]$ within this range is not an outlier.

Example:
7, 11, 14, 5, 8, 27, 16, 10, 13, 17, 16

rearrange → 5, 7, (8), 10, 11, (13), 14, 16, (16), 17, 27
median (Q_1) — median (Q_2) — median (Q_3)

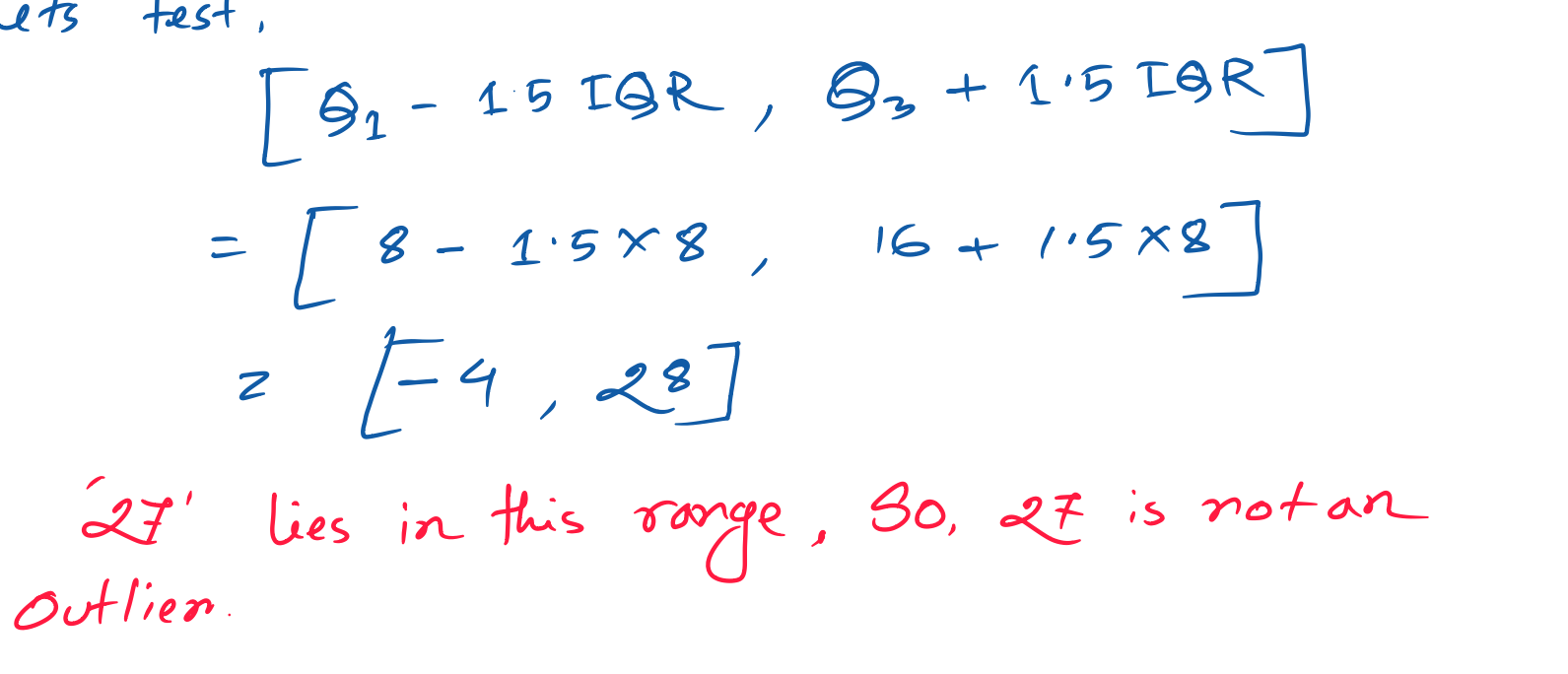
Then the $IQR = Q_3 - Q_1$
$= 16 - 8$
$= 8$

Is '27' is an outlier?

Lets test,
$[Q_1 - 1.5\,IQR,\ Q_3 + 1.5\,IQR]$
$= [8 - 1.5 \times 8,\ 16 + 1.5 \times 8]$
$= [-4, 28]$

'27' lies in this range, So, 27 is not an outlier.

More example.



Ex. 13, 16, 18 | 18, 22, 23 | 25, 28, 29 | 31, 38, 50
(18) Q_1 — (24) Q_2 median $\frac{23+25}{2}$ — (30) Q_3

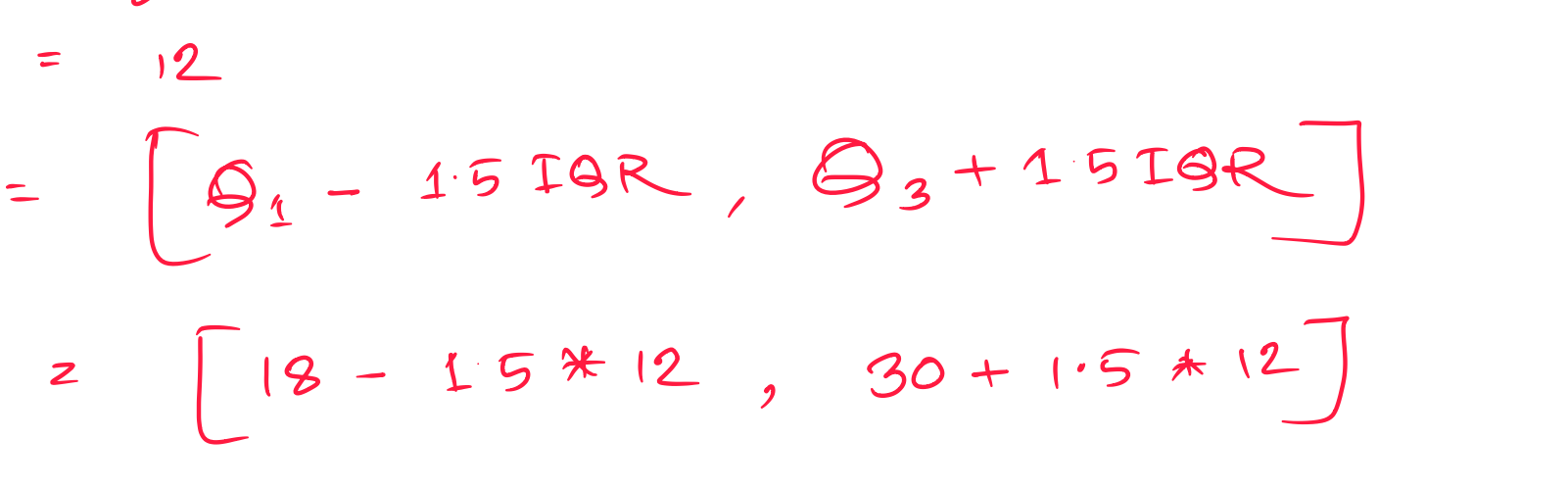
$IQR = Q_3 - Q_1$
$= 30 - 18$
$= 12$

Range $= [Q_1 - 1.5\,IQR,\ Q_3 + 1.5\,IQR]$
$= [18 - 1.5 * 12,\ 30 + 1.5 * 12]$
$= [0, 48]$

So, is '50' an outlier?

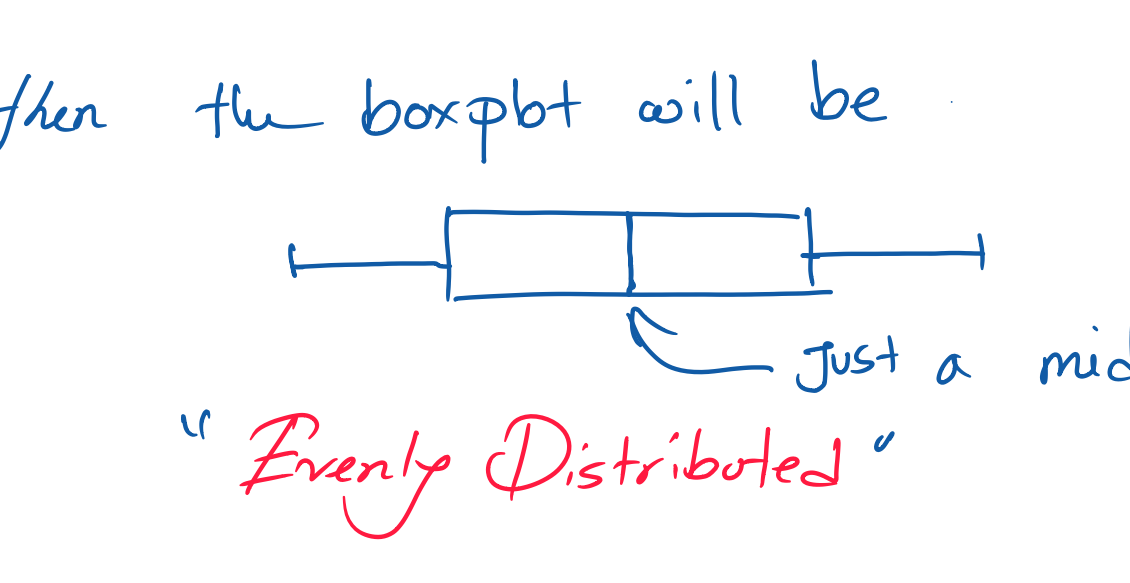
Ans: yes, it is. Because it is not lie on the range

Lets plot,



Skewness:

Lets assume a "symmetrical" representation as belows

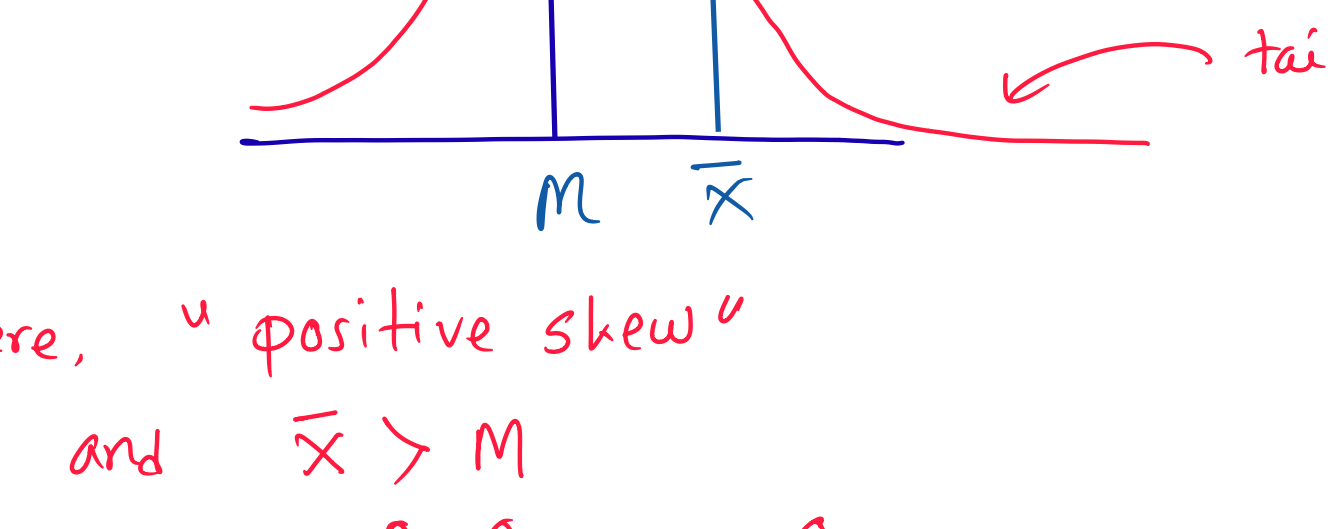


(*) if symmetric then mean $(\bar{x})$ = Median

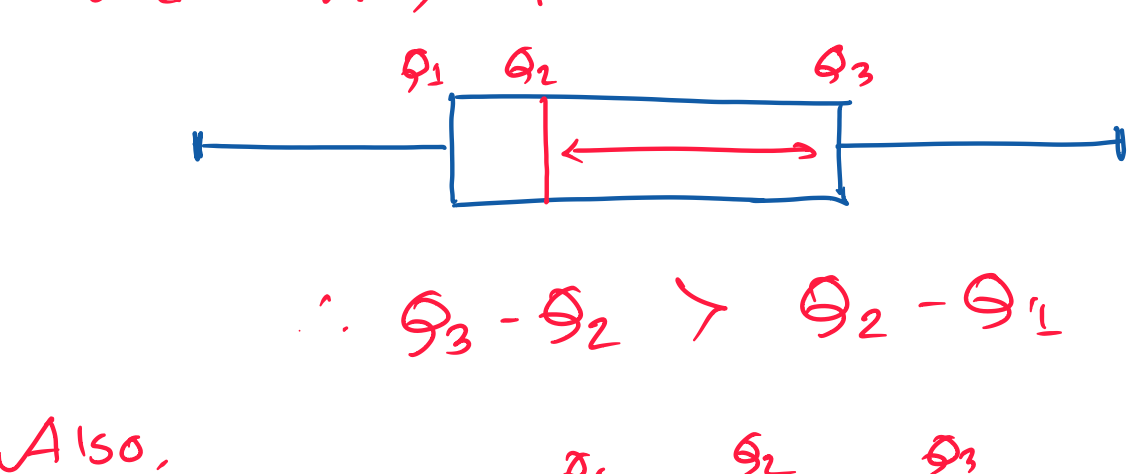
Then the boxplot will be

" Evenly Distributed "

(**) Lets take "Right skewed"

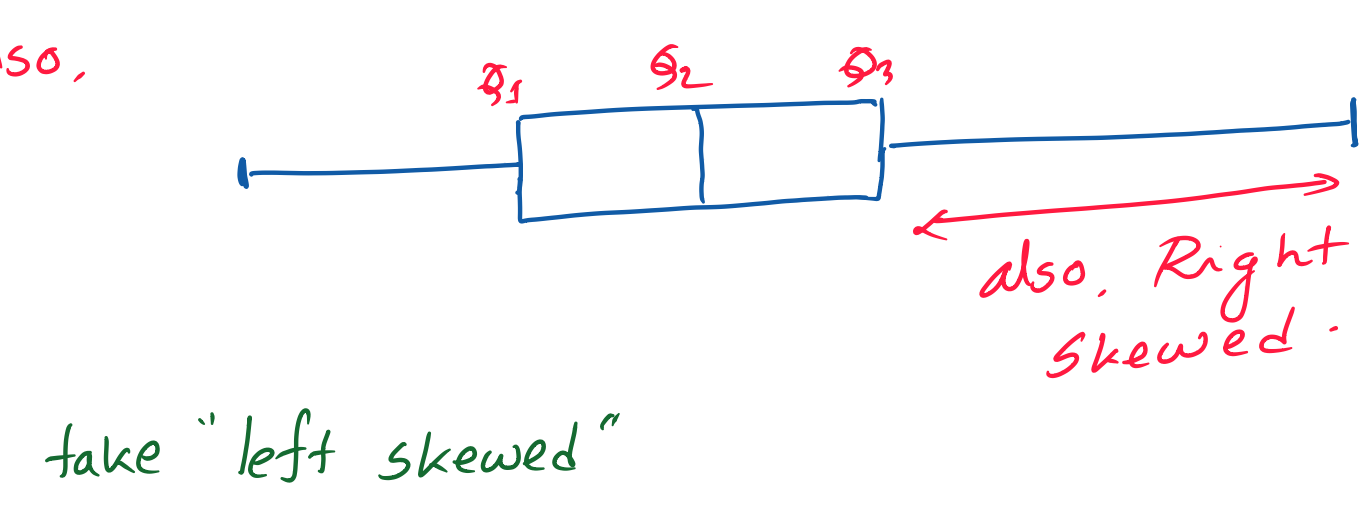


Here, " positive skew "
and $\bar{X} > M$

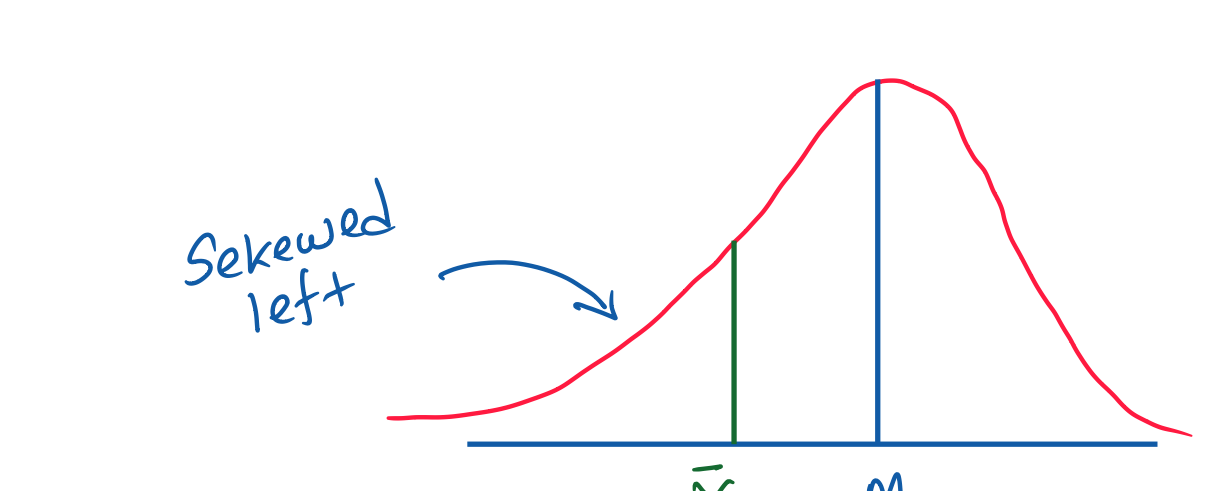


$\therefore Q_3 - Q_2 > Q_2 - Q_1$

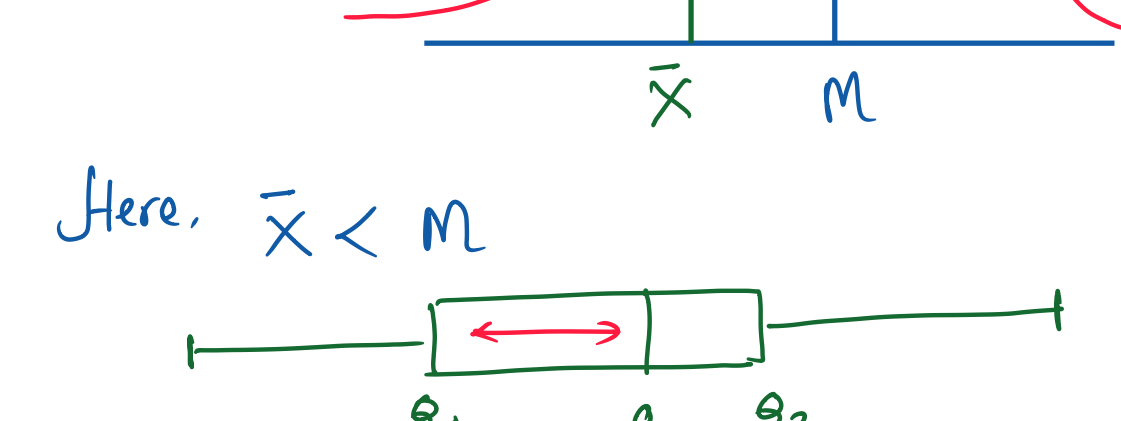
Also,



(**) Lets take "left skewed"

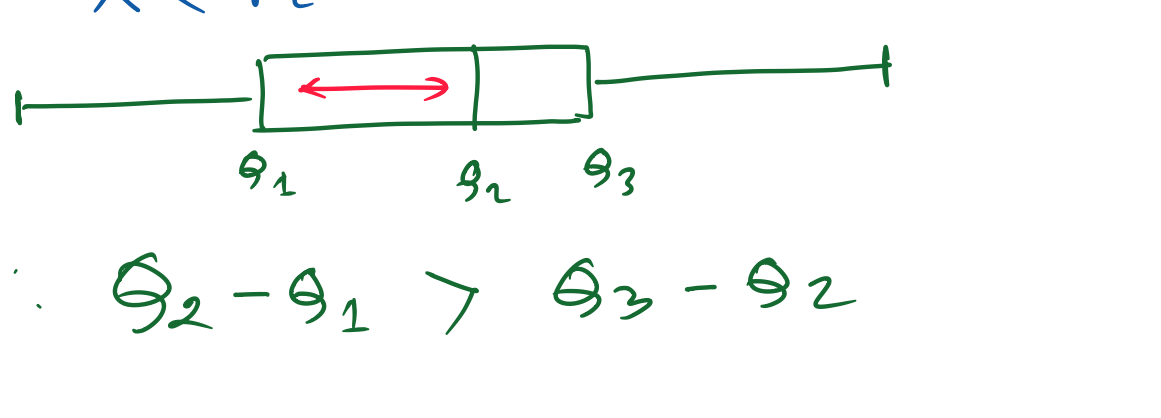


Here, $\bar{X} < M$



$\therefore Q_2 - Q_1 > Q_3 - Q_2$

Also,



Stem & Leaf:

11, 12, 13, 21, 24, 31, 32

Stem	leaf
1	1 2 3
2	1 4
3	1 2

Frequency Table:

5, 9, 8, 7, 8, 12, 9, 8, 10, 8, 9, 7

#	frequency	sum
5	1	5
7	2	14
8	4	32
9	3	27
10	1	10
12	1	12
	12	100

frequency table to mean

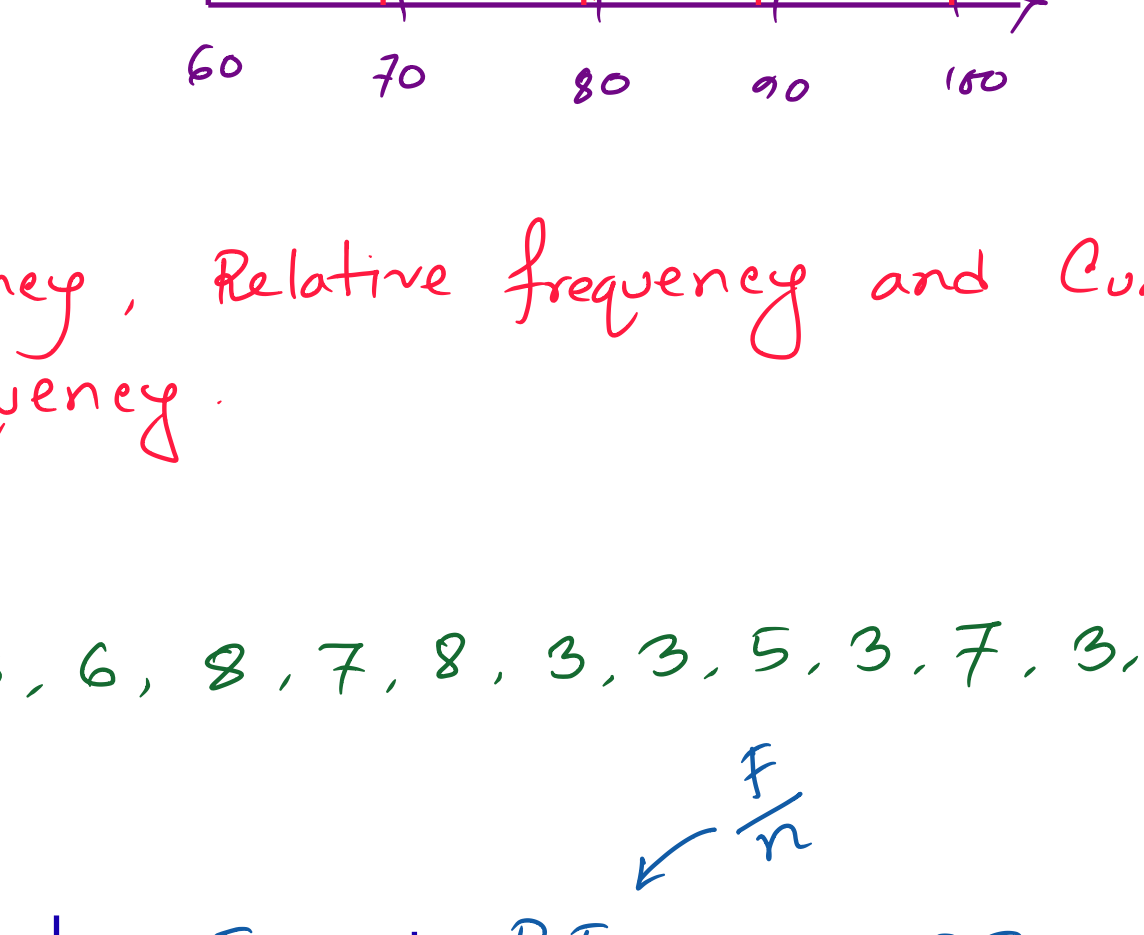
$\therefore \bar{x} = \frac{100}{12} = 8.3$

(**) Histogram:

Example:
65, 72, 93, 68, 76, 98, 84, 84, 79, 88, 90, 82, 83, 87, 78

Consider these are the grades lets measure the frequency

Grades	frequency
60 – 69	2
70 – 79	4
80 – 89	6
90 – 100	3



(**) Now, frequency, Relative frequency and Cumulative relative frequency

Lets have,
2, 3, 5, 3, 6, 8, 7, 8, 3, 3, 5, 3, 7, 3, 8, 5, 2, 7, 8, 3

V	F	RF	CRF
2	2	0.10	0.10
3	7	0.35	0.45
5	3	0.15	0.60
6	1	0.05	0.65
7	3	0.15	0.80
8	4	0.20	1.00
	n = 20	1.00	

RF = $\frac{F}{n}$

Should/must be 1.00

* What is the 60th percentile?

Ans: we need to add the 'V'
at 5 → it is 60
So, $(5+6)/2 = 5.5$

Then, 80th per...
$(7+8)/2 = 7.5$

if 20th per.?
there is no exact 20%, so it lies between 0.10 – 0.45
So, it will be → 3

↳ lets understand visually.

2 2 | 3 3 | 3 3 | 3 3 | 3 5 | 5 5 | 6 7 | 7 7 | 8 8 | 8 8
10% 20% 30% 40% 50% 60% 70% 80% 90%

↑ 3 — ↑ 5.5 — ↑ 7 — ↑ 7.5