Ex:- Grouped Frequency tables,

(1) Continuous data.

Given class boundaries	F	proved C.F
		0
$0 \le x < 5$	27	27
5 to 10	36	63
10 to 20	54	117
20 to 30	49	166
30 to 60	24	190
60 to 100	12	202

$= n$

order

L.Q $= 50.5$

$Q_2 = 101$

$Q_3 = 151.5$

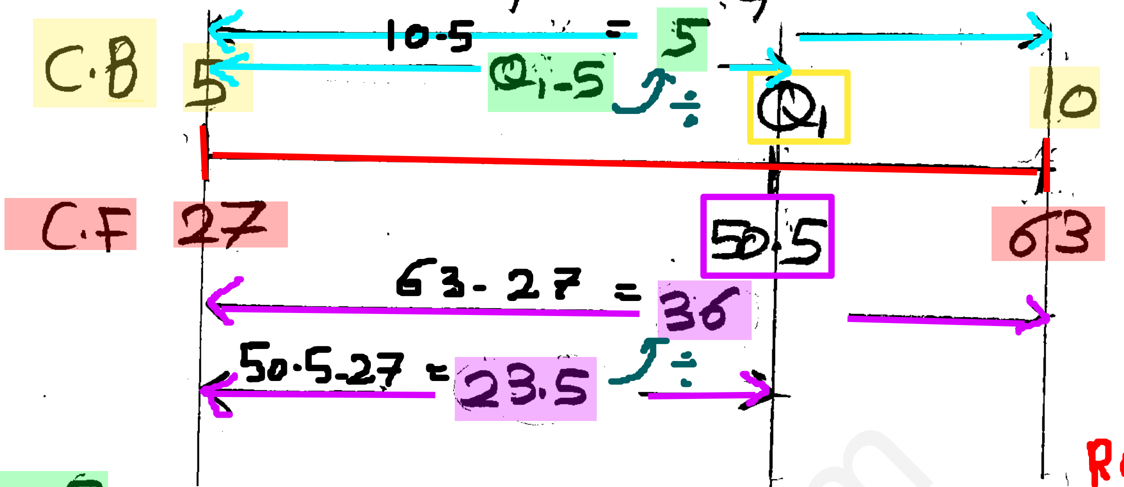
Note:- with continuous data, the end of 1-interval is the same as the start of the next. [No gaps]

Find:- 1) L.Q , 2) Median, 3) U.Q

solution

To find Q_1 (L.Q) :-

$*$ order of $Q_1 = \frac{1}{4} \times n = \frac{1}{4} \times 202 = 50.5$

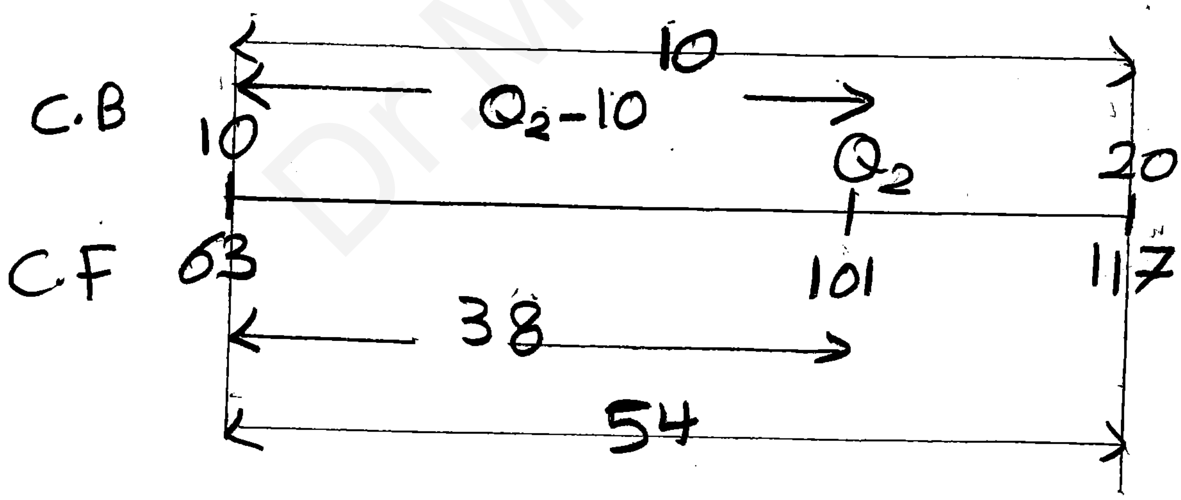


$$\frac{Q_1 - 5}{5} = \frac{23.5}{36} \rightarrow Q_1 = \boxed{\frac{23.5 \times 5}{36} + 5}$$

Required in M.S

$$= \boxed{8.26} \text{ (to 3 S.F)}$$

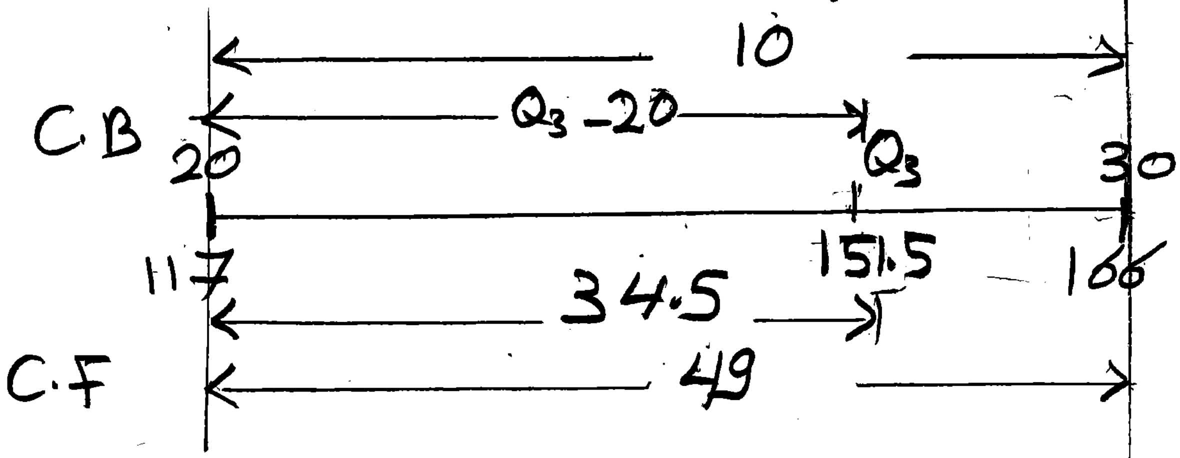
$*$ order of $Q_2 = \frac{1}{2} \times n = \frac{1}{2} \times 202 = 101$



$$\frac{Q_2 - 10}{10} = \frac{38}{54} \rightarrow Q_2 = \frac{38 \times 10}{54} + 10 = 17.0$$

(to 3 S.F)

* order of $Q_3 = \frac{3}{4} \times n = \frac{3}{4} \times 202 = 151.5$



$$\frac{Q_3 - 20}{10} = \frac{34.5}{49} \rightarrow Q_3 = \frac{10 \times 34.5}{49} + 20$$

$$= 27.0 \text{ (to 3 s.f)}$$

Ex: Grouped Frequency tables:

· Discrete data

$\frac{4+5}{2}$

0 - 0.5 rejected

Given		Proved	
class interval	Frequency	class boundaries	Cumulative Frequency
0-4	25	0 to 4.5	25
5-9	32	4.5 to 9.5	57
10-19	51	9.5 to 19.5	108
20-29	47	19.5 to 29.5	155
30-59	20	29.5 to 59.5	175
60-99	8	59.5 to 99.5	183

+0.5

Note: The discrete data in grouped Frequency tables is treated as continous.

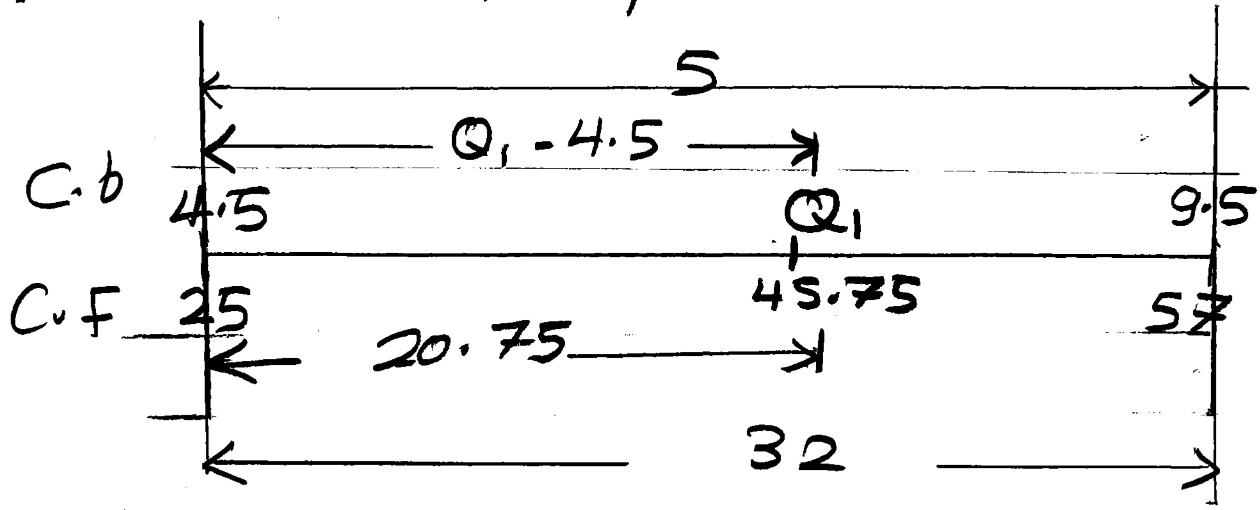
step 1:- change the class interval to class boundaris

0-4 ⟶ 0 to 4.5

step 2:- Continue like continuos data to find Q_1, Q_2, Q_3

solution

* Order of $Q_1 = \frac{1}{4} \times 183 = 45.75$



$$\frac{Q_1 - 4.5}{5} = \frac{20.75}{32} \rightarrow Q_1 = \frac{5 \times 20.75}{32} + 4.5$$
$$= 7.24$$

* percentiles:-

percentiles are calculated in exactly the same way as quartiles.

Ex:- For the 90^{th} percentile

sol:

Find $\frac{90}{100} \times n$ & proceed as above.

order L.Q $= \frac{1}{4} \times n$

$= 25\% \times n$

$Q_1 = L.Q = 25^{th}$ percentile

$Q_2 = 50^{th}$ percentile

$Q_3 = 75^{th}$ percentile

* Box plots:-

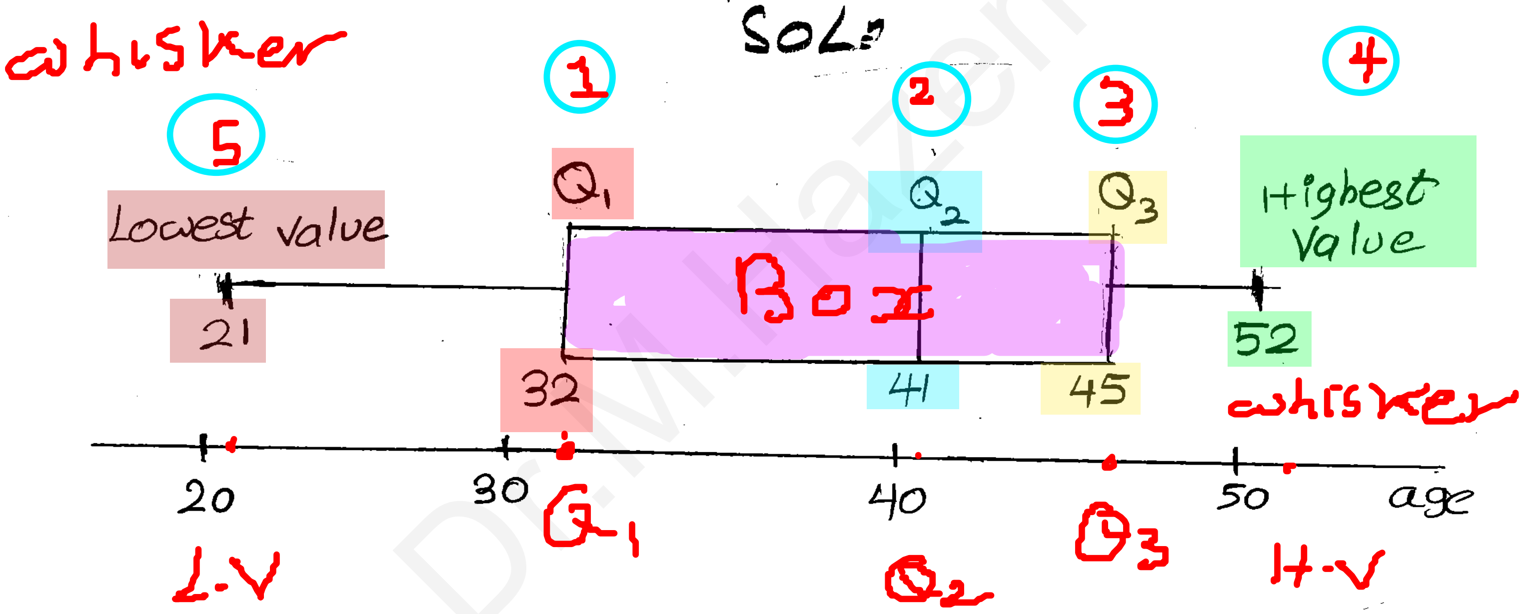
Ex:- In a group of people the youngest is 21 & the oldest is 52.

The quartiles are 32 & 45.

& the median age is 41.

Draw box plot:-

Sol:-



* Outliers:-

An outlier is an extreme value. You are not required to remember how to find an outlier - you always be given a rule.

Ex:- The ages of 11 children are given below.

outlier   outlier

Age	3	6	12	12	13	14	14	15	17	21	26

$Q_1 = 12$, $Q_2 = 14$ & $Q_3 = 17$

Outliers are values outside the range:-

L.B $\boxed{Q_1 - 1.5 \times (Q_3 - Q_1)}$ to $\boxed{Q_3 + 1.5 \times (Q_3 - Q_1)}$ U.B

Find any outliers, & draw a box plot.

Soln

Lower boundary for outlier is :-

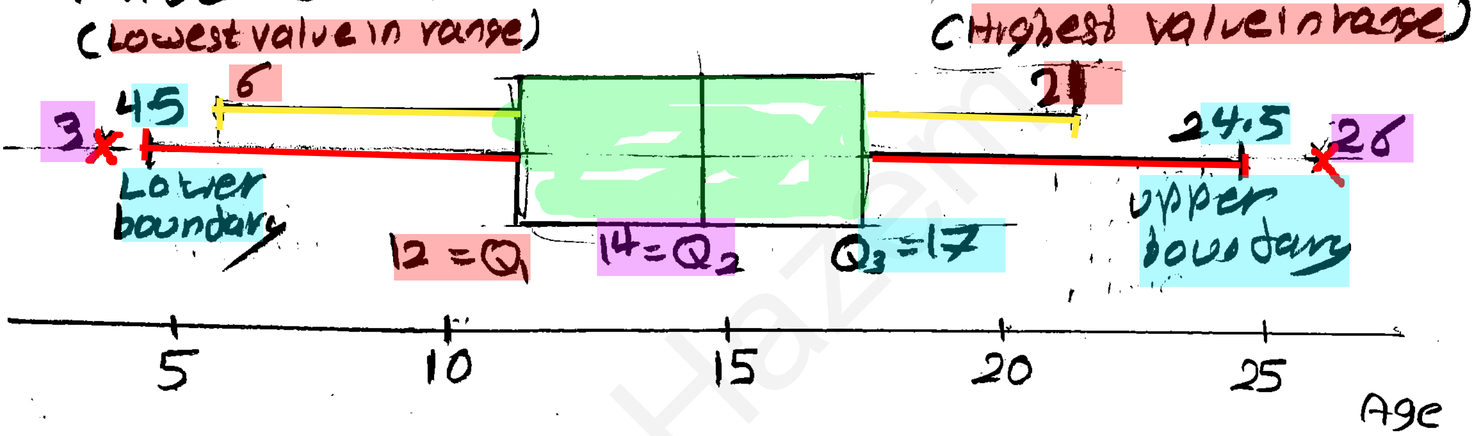
$12 - 1.5(17 - 12) = 4.5$

Upper boundary for outlier is:-

$17 + 1.5(17 - 12) = 24.5$

3 & 26 are the only outliers.

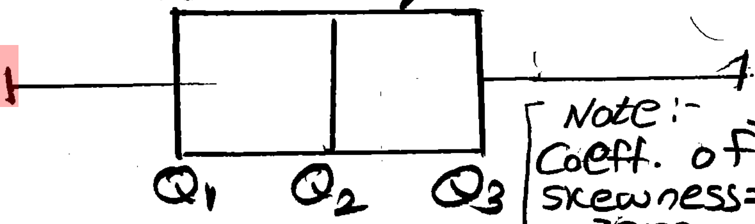
* To draw a box plot, put crosses at 3 & 26, & draw the lines to 6 (the lowest value which is not an outlier), & to 21 (the highest value which is not an outlier).



1- Q_1
2- Q_2
3- Q_3
4- L.V
5- H.V
6- outliers
7- L.B
8- U.B

* Skewness :-

A distribution which is symmetrical is not skewed :

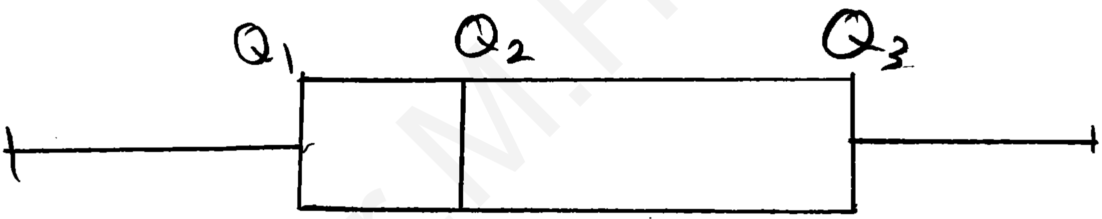


$$\boxed{Q_3 - Q_2 = Q_2 - Q_1}$$

Note :- Coeff. of skewness = zero

* Positive skew :-

If a symmetrical box plot is stretched in the direction of the positive x-axis, then the resulting distribution has positive skew :-



$$\boxed{Q_3 - Q_2 > Q_2 - Q_1}$$

Note :- Coefficient of skewness = + ve

The same ideas apply for a continous distribution, & a little bit of thought should show that for:-

Positive skew: Mean > Median > Mode

No skew

+ve skew

$\boxed{\text{mean} > \text{median} > \text{mode}}$

Negative skew:

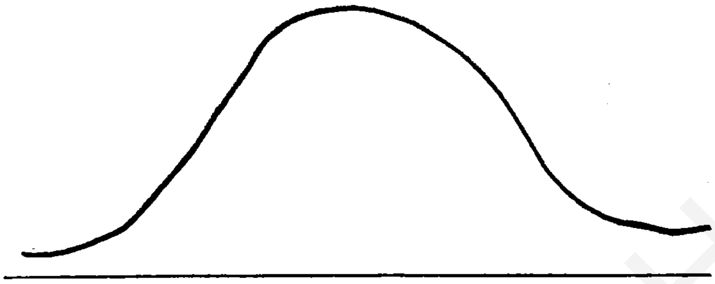
If a symmetrical box plot is stetched in the direction of the negative x-axis, then the resulting distribution has negative skew.

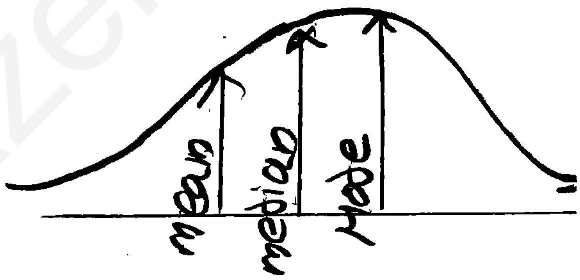
$$\boxed{Q_2 - Q_1 > Q_3 - Q_2}$$

Note:- coefficint of skewness = -ve

The same ideas apply for a continuous distribution, & a little bit of thought should show that for :-

Negative skew :- $\boxed{\text{Mean} < \text{Median} < \text{Mode}}$





_ve skew
Mean < Median < Mode

Note :- For skewed data, we use median better than mean because it is less affected by outliers

Measures of spread

Range & Interquartile range:-

- Range:-

  largest value - Smallest value

- Interquartile range:-

  Upper Quartile - lower Quartile

  $Q_3 - Q_1$

- Variance & standard deviation

- Skewness