

3.4 Exploratory data analysis

* Five number summary

"minimum, Q_1 , Q_2 , Q_3 , Maximum"

Exp:- Given the sample 12, 10, 18, 13, 25, 18

Find the Five number summary.

Sol: Sort Data 10, 12, 13, 18, 18, 25.

① Min = 10

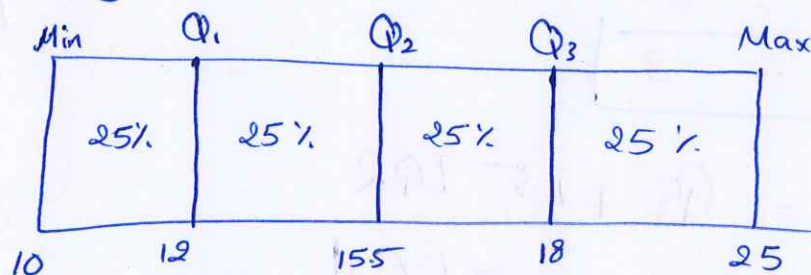
② $Q_1 = P_{25} \longrightarrow i = \frac{25}{100} * 6 = 1.5 \longrightarrow Q_1 = X_2 = 12$

$Q_2 = P_{50} \longrightarrow i = \frac{50}{100} * 6 = 3 \longrightarrow Q_2 = \frac{X_3 + X_4}{2} = 15.5$

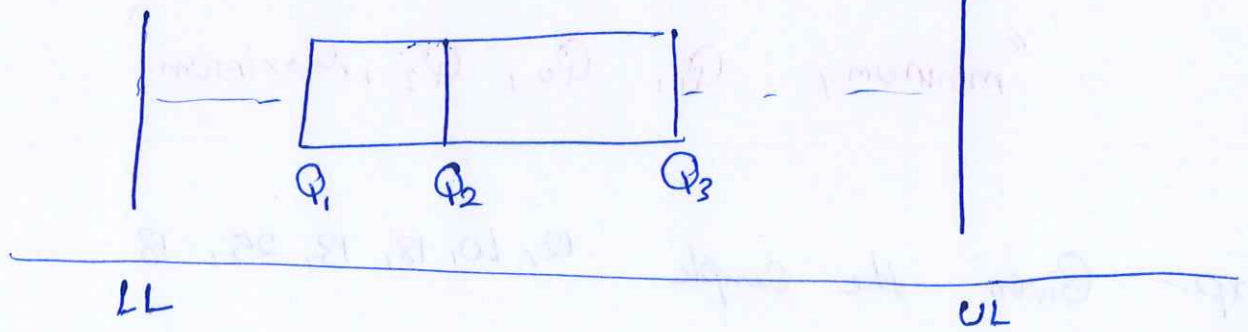
$Q_3 = P_{75} \longrightarrow i = \frac{75}{100} * 6 = 4.5 \longrightarrow Q_3 = X_5 = 18$

Max = 25

Five number summary : 10, 12, 15.5, 18, 25



* Box plot:- Graphical Summary of data based on the Five number summary



Where: $LL = Q_1 - 1.5 IQR$ "Lower Limit"

$UL = Q_3 + 1.5 IQR$ "Upper Limit"

Exp: Construct a boxplot for 12, 10, 18, 13, 25, 18

Sol. we have already found $Q_1 = 12$
 $Q_2 = 15.5$
 $Q_3 = 18$

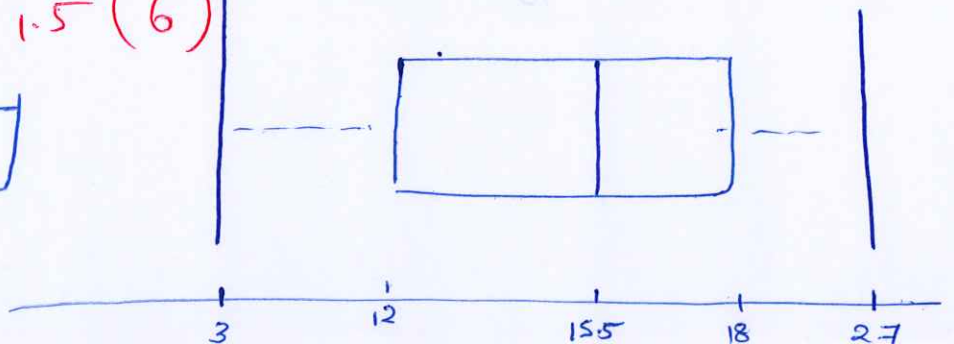
Now: $LL = Q_1 - 1.5 IQR$
 $= 12 - 1.5 (6)$

$LL = 3$

$UL = Q_3 + 1.5 IQR$
 $= 18 + 1.5 (6)$

$UL = 27$

$IQR = Q_3 - Q_1$
 $= 6$

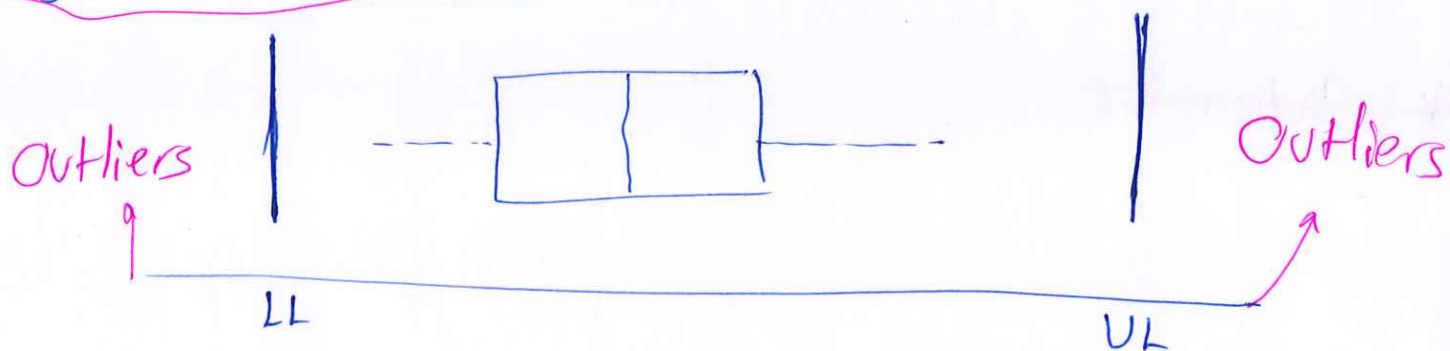


* Detecting Outliers using the Boxplot.

* We can use box plot (LL and UL) to detect outliers, if any

Outliers:-

Any Value $> UL$ or Any value $< LL$



Exp: Consider the sample 110, 150, 210, 180, 210, 70, 400
Use the box plot to find outliers, if any

Sol: Sort data : 70, 110, 150, 180, 210, 210, 400

$$Q_1 \longrightarrow i = \frac{25}{100} * 7 = 1.75 \longrightarrow \boxed{Q_1 = X_2 = 110}$$

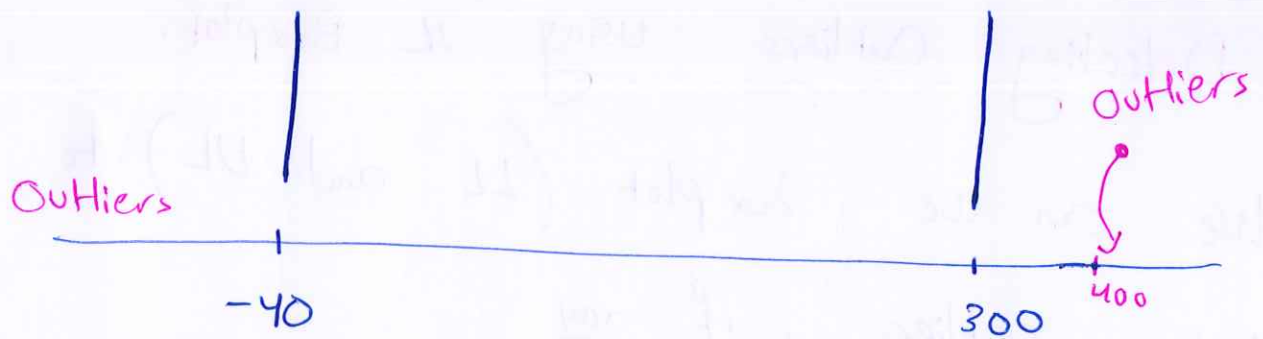
$$Q_2 \longrightarrow i = \frac{50}{100} * 7 = 3.5 \longrightarrow \boxed{Q_2 = X_4 = 180}$$

$$Q_3 \longrightarrow i = \frac{75}{100} * 7 = 5.25 \longrightarrow \boxed{Q_3 = X_6 = 210}$$

$$* IQR = 210 - 110 = 100$$

$$* LL = Q_1 - 1.5 (IQR) = -40$$

$$* UL = Q_3 + 1.5 (IQR) = 360$$



So 400 is an outliers because $400 > UL$

~~* Section 3.6~~

$$Q1 = X = 10$$

$$Q3 = X = 180$$

$$Q2 = X = 30$$

Section 3.6

The weighted Mean of grouped Data.

$$\text{Mean} = \frac{\sum (x_i w_i)}{\sum w_i}$$

where x_i = data value

w_i = weights

Exp: The grades of four courses were

	x_i	w_i
Course 1:	95	3 hours
Course 2:	80	2 hours
Course 3:	77	4 hours
Course 4:	85	2 hours.

Find the Mean of the grades

Sol:- By using SD Mode

$$\text{Mean} = \bar{x} = 83.91$$

* Grouped Data : we deal with a Mid point as x_i
and Frequency as w_i

Exp: The following are the weights of students

Weights	Frequency
50 — 56	10
57 — 63	15
64 — 70	7
71 — 77	9
78 — 84	18

Find by using SD Mode

① The sample mean $\bar{x} = 68.19$

② The sample standard deviation $S = \underline{10.10} \quad 10.6519$

③ The sample variance $s^2 = \underline{101.99} \quad 113.46$

④ The population std deviation $\sigma = \underline{10.05} \quad 10.5653$

⑤ The Pop. Variance $\sigma^2 = \underline{101.05} \quad 111.54$