Chapter 2 Descriptive statistics

Measures of Central tendency
مقاييس النزعة المركزية

- Mean
- Median
- Mode

3 M's

① Mean (Arithmetic mean)

$$\bar{X} = \frac{\sum X}{n} = \frac{\text{مجموع البيانات}}{\text{عددها}}$$

∴

Example Find the mean for:

(2, 7, 5, 11, 5)

$$\bar{X} = \frac{2+7+5+11+5}{5} = 6$$

Example If the mean of $(a, a, 7, 11, 2)$ is 6, find the (a)?

$$\frac{6}{1} = \frac{a+a+7+11+2}{5}$$

$$30 = 2a + 20$$

$$2a = 30 - 20$$

$$2a = 10$$

$$\boxed{a = 5}$$

Example If the mean of $(x, y, 12)$ is 10, find the mean of (x, y)?

$$\frac{10}{1} = \frac{x+y+12}{3}$$

$$30 = x + y + 12$$

$$\boxed{x + y = 18}$$

$$\frac{x+y}{2} = \frac{18}{2} = 9$$

Example find the mean of $(-9, -7, -11, 8, 2)$

$$\bar{X} = \frac{-9 + -7 + -11 + 8 + 2}{5}$$

$$= -3.4$$

NOTE $\bar{X}$ could be negative

$$\frac{\bar{X}}{1} = \frac{\sum X}{n} \quad \therefore \quad \boxed{\sum X = \bar{X} \cdot n} \quad \therefore$$

عندما يحدث تغيير على مجموع قيم X

طالب يدخل على الصف / يطلع

معلم يغير صنف / يشيل

Example If the mean mark of 9 students is 15. Ahmad with mark 20 joined the class, find the new mean?

الحل

$n = 9$

$\bar{X} = 15$

$X = 20$ ← Ahmad

$n = 10$

$\bar{X} = ?$

$\sum X_{\text{قبل}} = \bar{X} \cdot n$

$= 15 * 9 = 135$

$\sum X_{\text{جديد}} = 135 + 20 = 155$

$\bar{X}_{new} = \frac{\sum X_{new}}{n_{new}}$

$\bar{X}_{new} = \frac{\sum X_{new}}{10}$

$= \frac{155}{10}$

$\bar{X}_{new} = 15.5$

Example If the mean mark of 10 Boys is 12, and the mean mark of 12 Girls is 10 Find the mean mark of students altogether?

الحل

$n = 10$

$\bar{X} = 12$

Boys

$\sum X_B = \bar{X} \cdot n$

$= 12 * 10$

$= 120$

$n = 12$

$\bar{X} = 10$

Girls

$\sum X_G = \bar{X} \cdot n$

$= 10 * 12$

$= 120$

$$\bar{X}_{total} = \frac{\Sigma X_B + \Sigma X_G}{22} = \frac{120+120}{22}$$

$$= 10.9$$

---

② The mode "المنوال"

⇐ العنصر الأكثر تكراراً

Example find the mode for:

1) $(2, 7, 5, 11, 5)$

⇒ The mode is 5

2) $(2, 7, 5, 11, 5, 2)$

The mode is 5, 2

3) $(2, 7, 5, 11, 5, 2, 5)$

The mode is 5

NOTE

1 mode ⇒ unimodal

2 modes ⇒ Bimodal

3 modes ⇒ Trimodal

---

③ The median (Q_2)

① الترتيب

② عد البيانات

زوجي → نأخذ القيمتين بالمنتصف ÷ 2

$$\frac{\text{نأخذ القيمتين بالمنتصف}}{2}$$

فردي → نأخذ القيمة التي بالمنتصف

Example 5, 5, 8, 10, 12, 12, 13, 13, 15

Find the median ?

⇒ The median is 12

Example 2, 9, 11, 5, 6, 27

Find Q_2 ?

التَرتيب 2, 5, 6, 9, 11, 27

The median is $\frac{6+9}{2} = 7.5$

Example 2, 6, 16, 9, 3, 8, 11

التَرتيب 2, 3, 6, 8, 9, 11, 16

The median is 8

Example 5, 7, 7, 9, 10, 10, 12, 14

The median is $\frac{9+10}{2}$

$= 9.5$

Example For the following ORDERED DATA (a, 2, 3, 5, b, 9, 10, c). IF the mean is 7, the median is 6 and the mode is 2 Find the a, b, c?

$\bar{X} = 7$

$Q_2 = 6$

mode = 2

from mode $\Rightarrow$ $a = 2$

from median $\Rightarrow$ $\frac{5+b}{2} = \frac{6}{1}$

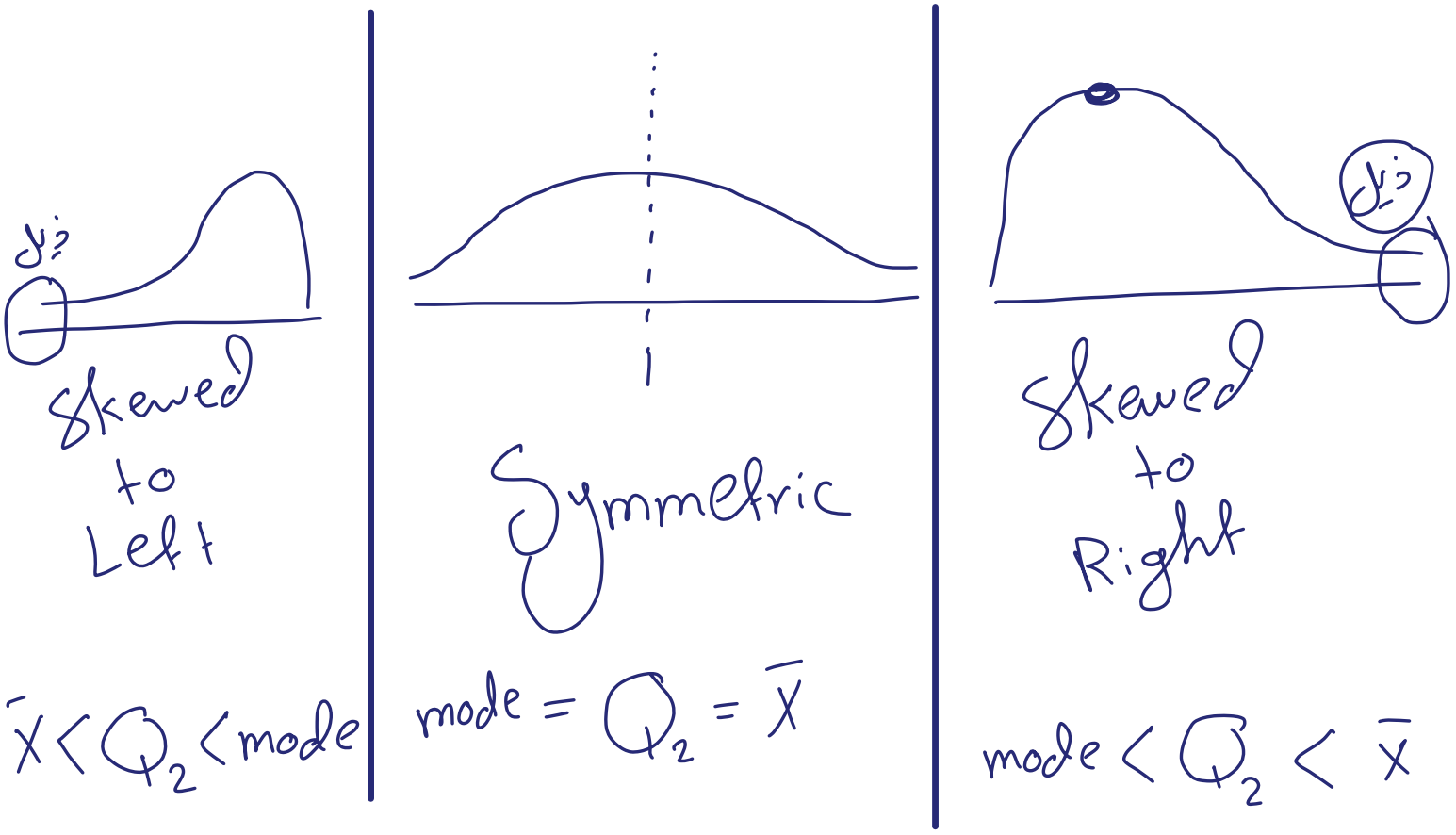
$b = 7$

from mean $\Rightarrow$

$$7 = \frac{2+2+3+5+7+9+10+C}{8}$$

$C = 18$

Skewness



$\bar{X} < Q_2 < \text{mode}$ | $\text{mode} = Q_2 = \bar{X}$ | $\text{mode} < Q_2 < \bar{X}$

Example Determine the shape of the following:

1) $\text{mode} = 16$, $\bar{X} = 14$

$\text{mode} > \bar{X}$ skewed to left

2) $\text{mode} = 7630$, $\bar{X} = 7630$

mode = mean Symmetric

3) mode = 2 , $\bar{x} = 4$

mean > mode skewed to Right

Characteristics of mean

① mean is affected by an outlier

outlier القيم الشاذة

2, 2, 2, 2, (20) outlier

2, 2, 2, 2 $\bar{x} = 2$

outlier (2), 19, 19, 18, 20

2, 2, 2, 2, 20

$$\bar{x} = \frac{2+2+2+2+20}{5}$$

$$= 5.6$$

2) (3) 2, 2, 2, 2 $\bar{x} = 2$

5, 5, 5, 5 $\bar{y} = 5$ (+3)

$$\bar{y} = \bar{x} + b$$

③ 2, 2, 2, 2 $\quad \bar{x} = 2$

*3 ↓ 6, 6, 6, 6 $\quad \bar{y} = 6$ ↓ *3

★ Measures of variability (Spread)

مقاييس التشتت

مستحيل يجونا سوالب

① The Range = Max - Min

المدى

Example find the range for:

1) (2, 7, 5, 11, 2)

The Range is = 11 - 2 = 9

② The variance and standard deviation

التباين والانحراف المعياري

① $S^2 = \frac{\Sigma (x - \bar{x})^2}{n-1}$

② $S^2 = \frac{\Sigma x^2}{n-1} - \frac{(\Sigma(x))^2}{n(n-1)}$

$S = \sqrt{\text{variance}} = \sqrt{\frac{\Sigma x^2}{n-1} - \frac{(\Sigma x)^2}{n(n-1)}}$

Σx^2 vs $(\Sigma(x))^2$

Σx^2: تربيع القيم ثم تجمع

2, 2, 2, 2

4, 4, 4, 4 $\Sigma x^2 = 16$

$(\Sigma x)^2$: تجمع القيم ثم تربع

2, 2, 2, 2

$(8)^2 = 64$

Example find variance for (2, 7, 5, 11, 5)

$\bar{x} = \frac{2+7+5+11+5}{5} = 6$

(1) طريقة $S^2 = \frac{\Sigma (x - \bar{x})^2}{n-1}$

x	2	7	5	11	5
$(x-\bar{x})$	-4	1	-1	5	-1
$(x-\bar{x})^2$	16	1	1	25	1

Always $\Sigma (x-\bar{x}) = 0$

$\Sigma (x-\bar{x})^2 = 44$

$$S^2 = \frac{44}{5-1} = \frac{44}{4} = 11$$

$$S = \sqrt{11} = 3.31$$

طريقة (2)

$$S^2 = \frac{\sum X^2}{n-1} - \frac{(\sum X)^2}{n(n-1)}$$

X	2	7	5	11	5	$\sum X = 30$
X^2	4	49	25	121	25	$\sum X^2 = 224$

$$S^2 = \frac{224}{5-1} - \frac{(30)^2}{5(5-1)} = 11$$

$$S = \sqrt{11} = 3.31$$

Example find the variance for (5, 7, 1, 2, 4)

طريقة (1)

$$S^2 = \frac{\sum (X-\bar{X})^2}{n-1} \qquad \bar{X} = 3.8$$

X	5	7	1	2	4	
$(X-\bar{X})$	1.2	3.2	-2.8	-1.8	0.2	
$(X-\bar{X})^2$	1.44	10.24	7.84	3.24	0.04	$\sum (X-\bar{X})^2 = 22.8$

$$S^2 = \frac{22.8}{5-1} = 5.7$$

$$S = \sqrt{5.7} = 2.38$$

(2) طريقة

$$S^2 = \frac{\Sigma X^2}{n-1} - \frac{(\Sigma X)^2}{n(n-1)}$$

X	5	7	1	2	4	$\Sigma X = 19$
X^2	25	49	1	4	16	$\Sigma X^2 = 95$

$$S^2 = \frac{95}{4} - \frac{(19)^2}{5(4)} = 5.7$$

$$S = \sqrt{5.7} = 2.38$$

✱ Characteristics of variance

① مستحيل يكون سالب

② 2, 2, 2, 2 $\quad S_x^2 = a$

$\xrightarrow[-2]{+3}$ 5, 5, 5, 5 $\quad S_y^2 = a$

لا يتأثر بالجمع والطرح

③ 2, 2, 2, 2 $S_x^2 = a$

S_x^2: variance الأصلي

S_y^2: variance الجديد

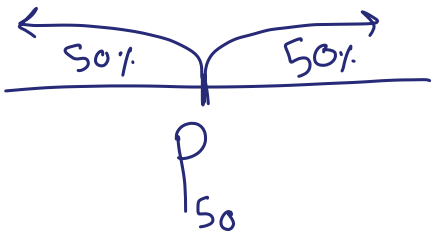
*3 ↓

6, 6, 6, 6 $S_y^2 = C^2 * S_x^2$

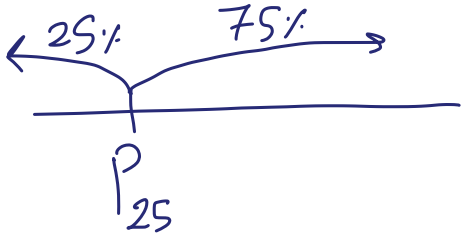
$S_y^2 = 9 * a$

✡ Quartiles and percentiles

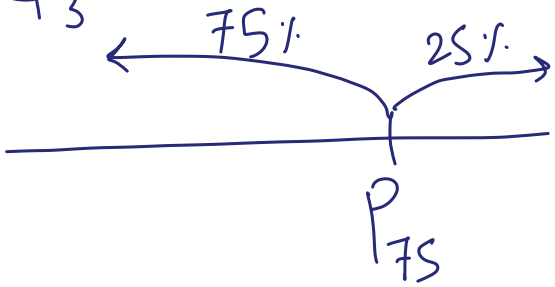
Q_2: median

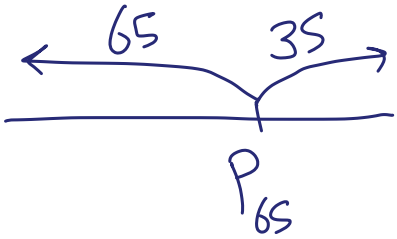


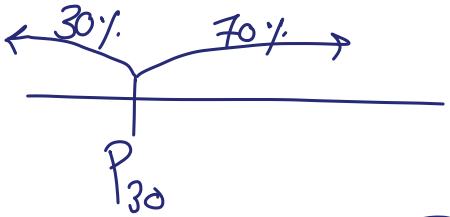
Q_1: Lower quartile



Q_3: upper quartile







Percentile

$$P_p = \frac{n \cdot p}{100}$$

n: Sample Size

p: percentile

→ next integer

$2.2 \Rightarrow (3)^{th}$

$3.9 \Rightarrow (4)^{th}$

$$\frac{\text{العدد} + (\text{العدد} + 1)}{2}$$

$$\frac{(3)^{th} + (4)^{th}}{2}$$

$$\frac{(5)^{th} + (6)^{th}}{2}$$

Inter-quartile Range (IQR)

$$IQR = Q_3 - Q_1$$

Example 2, 7, 5, 11, 3, 8, 10 Find IQR:

الترتيب ← 2, 3, 5, 7, 8, 10, 11

$$Q_1 = P_{25} = \frac{n \cdot P}{100} = \frac{7 * 25}{100} = \frac{175}{100} = (1.75)^{th}$$

$$= (2)^{th}$$

$$Q_1 = 3$$

$$Q_3 = P_{75} = \frac{7 * 75}{100} = \frac{525}{100} = (5.25)^{th}$$

$$= (6)^{th}$$

$Q_3 = 10$

$IQR = 10 - 3 = 7$

Example 260, 290, 300, 320, 330, 340, 340, 550

find IQR?

$$Q_1 = P_{25} = \frac{8 * 25}{100} = \frac{200}{100} = \frac{(2)^{th} + (3)^{th}}{2}$$

$$\frac{290 + 300}{2} = 295$$

$$Q_3 = P_{75} = \frac{8 * 75}{100} = \frac{600}{100} = \frac{(6)^{th} + (7)^{th}}{2}$$

$$\frac{340 + 340}{2} = 340$$

$IQR = 340 - 295 = 45$

Example 2, 3, 5, 7, 9, 10, 11, 12 find the value that 65% of data lie below it?

$$P_{65} = \frac{8 * 65}{100} = \frac{520}{100} = (5.2)^{th} = (6)^{th}$$

$$\boxed{P_{65} = 10}$$

Example 3, 5, 5, 6, 8, 8, 9, 10, 11, 12

find the value that 30% of data lie above it?

بدل

$$P_{70} = \frac{10 * 70}{100} = \frac{(7)^{th} + (8)^{th}}{2}$$

$$= \frac{9 + 10}{2} = \boxed{9.5}$$

Outliers

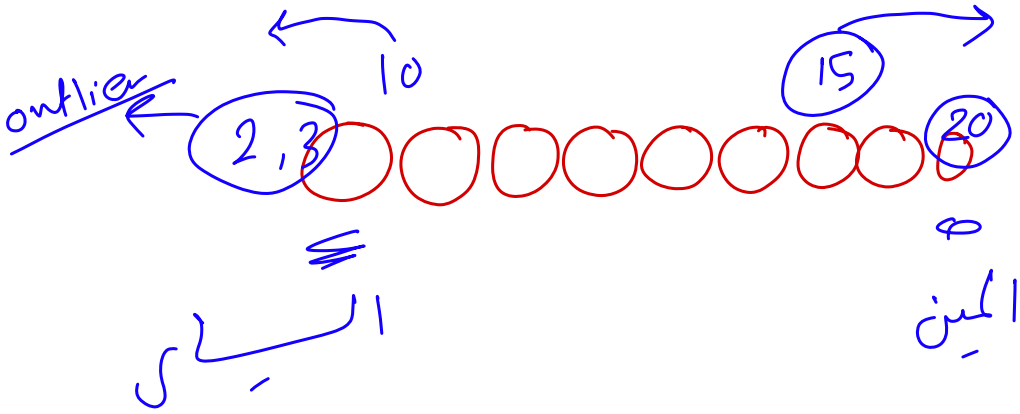
القيم الشاذة

من اليمين

$$Q_3 + 1.5 * IQR$$

حد الشمال

$$Q_1 - 1.5 * IQR$$



132 160

Example (123) 140, 145, 146, 147, 149, 150, (172)

find an outliers (if there was)

حد اليمين

$$Q_3 + 1.5 * IQR$$

$$149.5 + 1.5 * 7 = 160$$

so 172 is an outlier

حد الشمال

$$Q_1 - 1.5 * IQR$$

$$142.5 - 1.5 * 7 = 132$$

so 123 is an outlier

$$Q_1 : P_{25} = \frac{8 * 25}{100} = \frac{(2)^{th} + (3)^{th}}{2}$$

$$= \frac{140 + 145}{}$$

$$= \frac{142}{2}$$

$$= 142.5$$

$$Q_3 : P_{75} = \frac{8 * 75}{100} = \frac{(6)^{th} + (7)^{th}}{2}$$

$$= \frac{149 + 150}{2}$$

$$= 149.5$$

$$IQR = 149.5 - 142.5 = 7$$

$\therefore$

Example 340, 300, 520, 340, 320, 290, 260, 330

find outliers (if there was) ?

حد العليا

$$Q_3 + 1.5 * IQR$$

$$340 + 1.5 * 45 = 407.5$$

∴ 520 is an outlier

حد السفلى

$$Q_1 - 1.5 * IQR$$

$$295 - 1.5 * 45 = 227.5$$

×

الترتيب 227.5 407.5

260, 290, 300, 320, 330, 340, 340, 520

$$Q_1 : P_{25} = \frac{8*25}{100} = \frac{(2)^{th} + (3)^{th}}{2}$$

$$= \frac{290 + 300}{2}$$

$$= 295$$

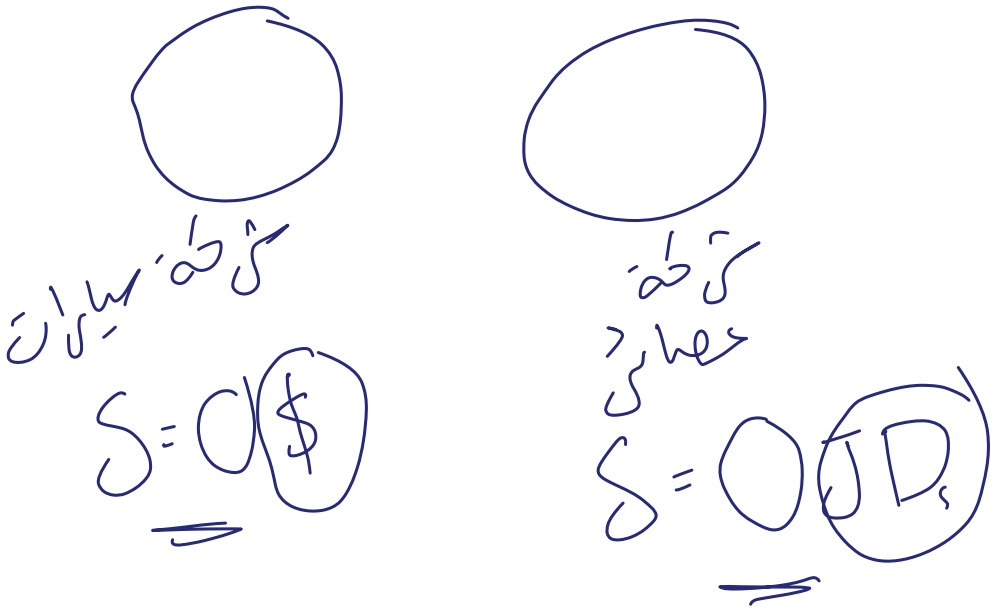
$$Q_3 : P_{75} = \frac{8*75}{100} = \frac{(6)^{th} + (7)^{th}}{2}$$

$$= \frac{340 + 340}{2}$$

$$= 340$$

$$IQR = 340 - 295 = 45$$

✱ Coefficient of variation (CV)



$$CV = \frac{S}{\bar{X}} * 100\%$$

Example In a class, if the mean is 30 and standard deviation is 2, find the Coefficient of variation for this class?

$$CV = \frac{S}{\bar{X}} * 100\% = \frac{2}{30} * 100\%$$

$$= 6.67\%$$

Example If the mean and Coefficient of variation of a data are 15 and 48 Respectively, then find the value of stand. deviation

$\bar{X} = 15$

$CV = 48$

$$CV = \frac{S}{\bar{X}} * 100$$

$$48 = \frac{S}{15} * 100$$

$$S = 48 * 0.15 = 7.2$$

Example If $n = 5$, $\bar{X} = 6$, $\Sigma X^2 = 765$, then Find the Coefficient of variation ?

$$CV = \frac{S}{\bar{X}} * 100\% \Rightarrow \frac{12.09}{6} * 100\%$$

$$= 201.55\%$$

$$S^2 = \frac{\Sigma X^2}{n-1} - \frac{(\Sigma X)^2}{n(n-1)}$$

$$= \frac{765}{5-1} - \frac{(30)^2}{5(4)}$$

$$= 146.25$$

$$\Sigma X = \bar{X} \cdot n$$

$$= 6 * 5$$

$$= 30$$

$$S = \sqrt{146.25}$$

$$= 12.09$$

Example Find the Coefficient of Variation for (24, 26, 33, 37, 29, 31) ?

Sol $CV = \frac{S}{\bar{X}} * 100\%$

$(x-\bar{x})$	-6	-4	3	7	-1	1
$(x-\bar{x})^2$	36	16	9	49	1	1

$\Sigma (X-\bar{x})^2 = 112$

$$\bar{X} = \frac{\Sigma X}{n} = \frac{180}{6} = 30$$

$$S^2 = \frac{\Sigma (X-\bar{X})^2}{n-1} = \frac{112}{6-1} = 22.4$$

$$S = 4.73$$

$$CV = \frac{4.73}{30} * 100\% = 15.77\%$$

Example Two plants C and D of a factory show the following Results about the number of workers and the wages paid to them.

	C	D
no. of workers	5000	6000
Average monthly wages	2500	2500
Standard deviation	9	10

Using Coefficient of variation, find in which plant C or D, is there greater variability in individual wages

$$CV_C = \frac{9}{2500} * 100$$

$$= 0.36$$

$$CV_D = \frac{10}{2500} * 100 = 0.40$$

$CV_D > CV_C$ so D has greater variability

CV	SD
Relative measure of variabity	absolute measure of variability
used to campare variability between 2 samples	used to measure the dispersion of data in a single set

Example The mean and SD of marks obtained by 40 students of a class in three subjects maths, science and social science are given

	Mean	SD
Maths	56	12
Science	65	14
Social science	60	10

which of three subjects shows the highest variation?

$$CV_{maths} = \frac{12}{56} * 100\% = 21.43$$

$$CV_{sci.} = \frac{14}{65} * 100\% = 21.54$$

$$CV_{\substack{soc \\ sci.}} = \frac{10}{60} * 100\% = 16.67$$

∴ ∴

✯ Graphical methods

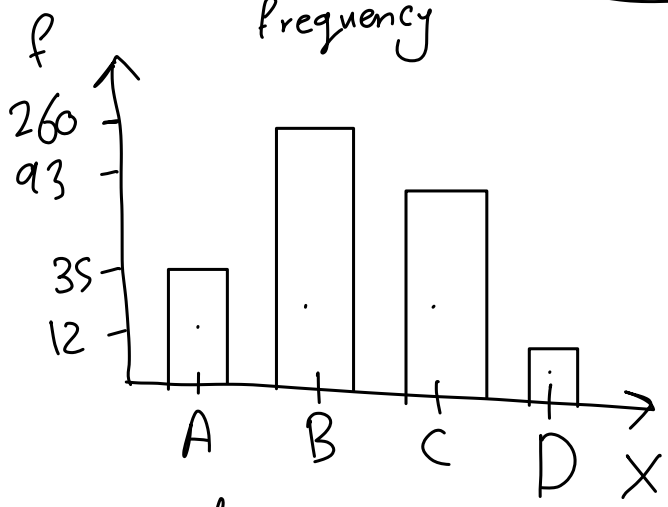
① Bar Graph ② Stem and leaf plot

③ Box plot

---

① Bar Graph

X	A	B	C	D
f	35	260	93	12



- difficult to construct
- identify sample points is lost

---

② Stem and leaf plot

ساق وورقة

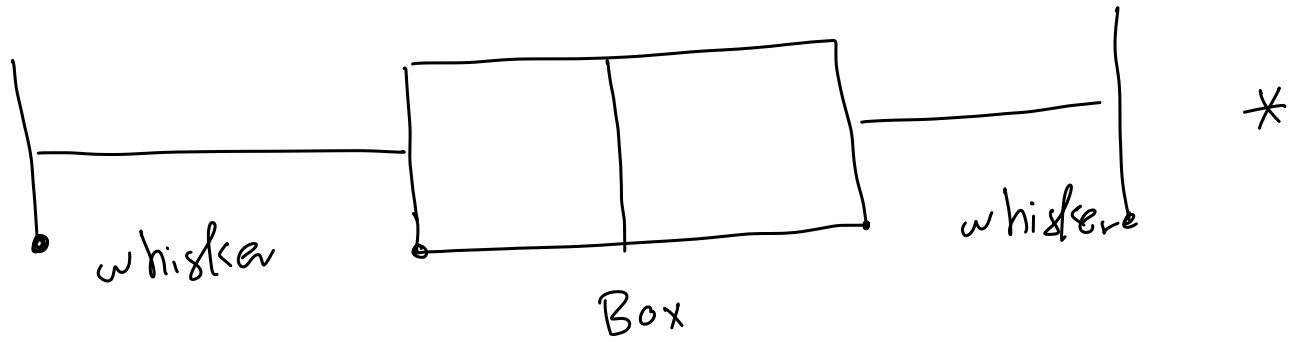
Example 23 71
58 71

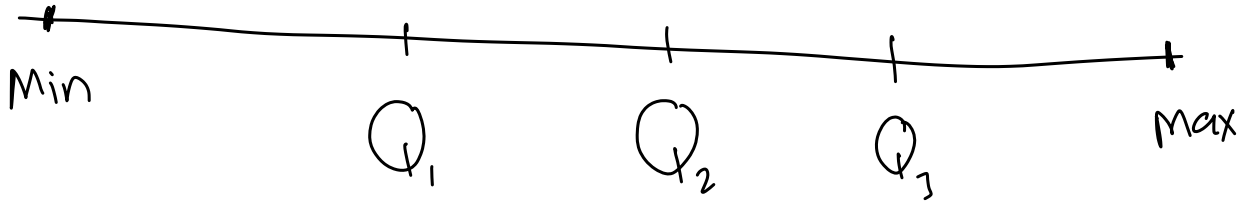
S	L
2	3
3	

62	72
62	80
63	82
65	82
67	82

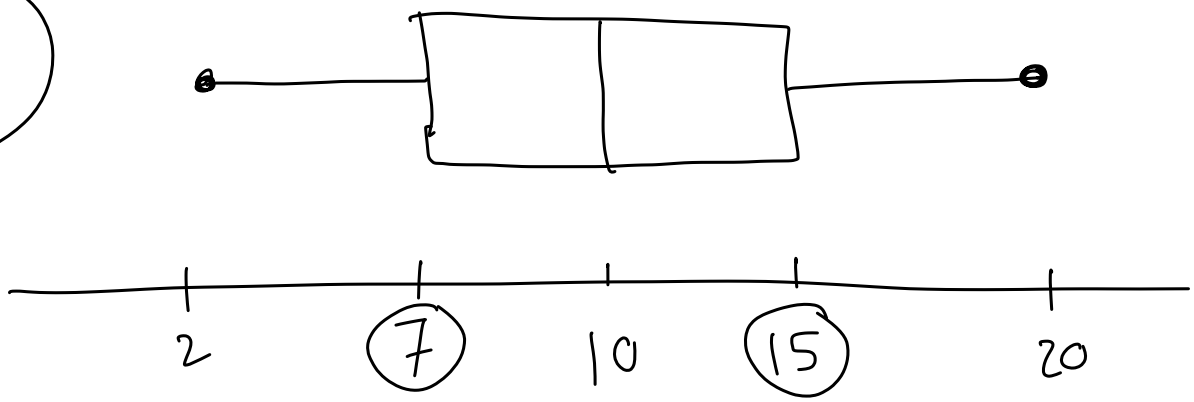
4	
5	8
6	2 2 3 5 7
7	1 1 2
8	0 2 2

(3) Box plot "Box and whiskers"





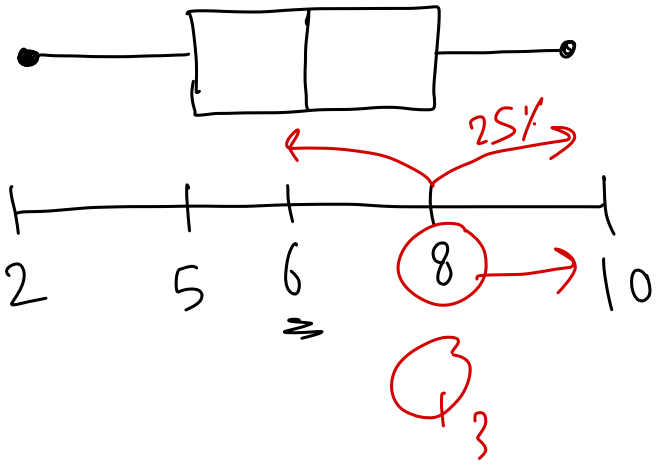
Example



find IQR ?

$$IQR = Q_3 - Q_1 = 15 - 7 = 8$$

Example If there are 32 students in the class, find:



1. $Q_2 = 6$
2. Range $= 10 - 2 = 8$
3. $IQR = 8 - 5 = 3$
4. The number of students achieved more than 8 ?

$0.25 * 32 = 8$

✯ Central tendency = Location

→ useful to define the center or middle of sample

→ Could be negative

→ Mean (generally not part of data set)

1. oversensitive to extreme values
2. easy to calculate
3. each sample has only one sample mean

→ Median (Maybe part of data set)

⇒ 1. less affected by outliers

2. less efficient than mean

→ mode (always part of data set)

Measures of variation (dispersion) (spread)

1. Range
   - Simplest MOV
   - quick summary of variation
   - extremely affected by outlier

2. IQR
   not affected by outlier

---

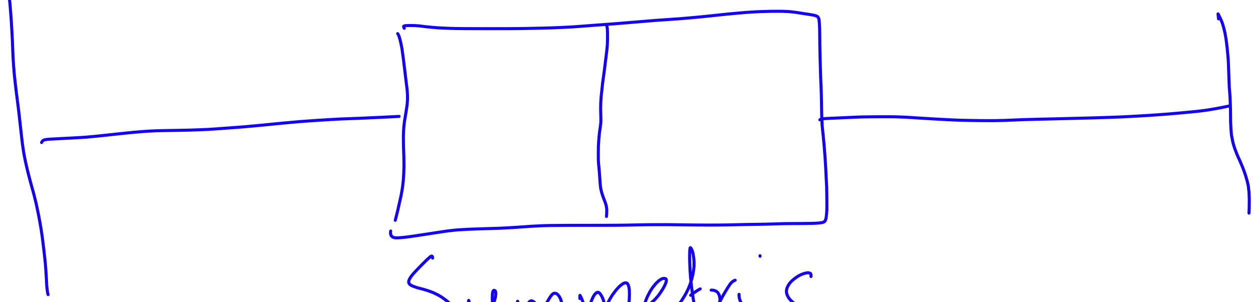
skewed to Right (+ skewness)

skewed to left (- skewness)

Variance & SD

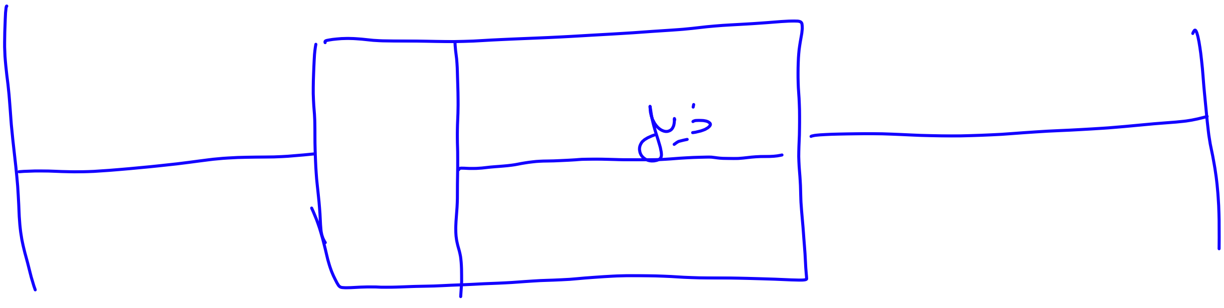
2, 2, 2, 2, 2 متشابهة

$S^2 = 0$



Symmetric

$Q_1 \rightarrow Q_2 = Q_2 \rightarrow Q_3$



skewed to Right

$Q_1 \rightarrow Q_2 < Q_2 \rightarrow Q_3$

skewed to left

$Q_1 \rightarrow Q_2 > Q_2 \rightarrow Q_3$