

# Chapter 3: Descriptive Statistics

## Numerical Measure

---

3.1  $\Rightarrow$  Measure of Location

3.2  $\Rightarrow$  Measure of Variation

3.3  $\Rightarrow$  Measure of ~~test~~ distribution shape

3.4  $\Rightarrow$  Exploratory Data Analysis

3.5  $\Rightarrow$  Measure of Association Between two Variables

3.6  $\Rightarrow$  The Weighted Mean and Working with Grouped Data

### 3.1 $\Rightarrow$ Measure of Location :-

(Central Tendency) الهدف من هذا الفصل هو تحديد المركز



البيانات التي يتم من خلالها تحديد ال Center

① Mean  
الوسط الحسابي

② Median  
الوسط الكمي

③ Mode  
البيانات

① The mean: (Average) الوسط الحسابي

انظر الى  
\* في الصفحة

$\mu \Rightarrow$  The Population mean (تقرأ  $\mu$ )

$\bar{X} \Rightarrow$  The Sample mean (تقرأ  $\bar{X}$ )

$$\text{Average} = \frac{\text{Sum of Values}}{\text{number of values}}$$

$$\mu = \frac{\sum X}{N} \quad , \quad \bar{X} = \frac{\sum X}{n}$$

$\downarrow$  Population size                       $\downarrow$  Sample size

Ex: for the following Data find the Mean (Average)

10, 6, 13, 12, 8, 14

$$\bar{X} = \frac{\sum X}{n} = \frac{10 + 6 + \dots}{6} = \frac{63}{6} = 10.5$$

هذا القانون في اكله الاسوي وهو عند استار Raw Data

## \*\* The Weighted Mean : 3.6

هذا النوع عندما يكون هناك اختلاف في تأثير ال Data  
 مثل علامات الجامعة حيث المادة اى  $\frac{1}{2}$  ساعات يكون تأثير  
 اكثر من المادة ~~ال~~ اى على ساعة واحدة.

Value	frequency
$x_1$	$f_1$
$x_2$	$f_2$
$\vdots$	$\vdots$
$x_n$	$f_n$

$$\bar{X} = \frac{\sum Xf}{\sum f}$$

ex: find the mean for the following frequency Distribution

$x$	$f$
6	10
8	5
5	6
7	9

$$\begin{aligned} \bar{X} &= \frac{\sum fx}{\sum f} \\ &= \frac{(6 \times 10) + (8 \times 5) + (5 \times 6) + (7 \times 9)}{10 + 5 + 6 + 9} \\ &= \frac{193}{30} = 6.43 \end{aligned}$$

Ex:	Units $x$	Price $f$
	100	10
	150	30
	⋮	⋮

$$\text{Average Price per unit } (\bar{x}) = \frac{\sum xf}{\sum f}$$

Weighted

Mean

\*\*\* The mean for a grouped Data

Grouped Data :-

→ It is a set of grouped frequency distribution

Class	freq
A → B	$f_1$
C → D	$f_2$
E → F	$f_3$
⋮	⋮
	$f_n$

→ grouped frequency distribution

Grouped Data

# Group Data

جواباً

~~$\bar{X} = \frac{\sum Mf}{\sum f}$~~

$$\bar{X} = \frac{\sum Mf}{\sum f}$$

$M \Rightarrow$  Mid point  $\Rightarrow \frac{\text{Upper} + \text{Lower}}{2}$   
 Class Mark

Ex: find the mean for the following distribution:

Class	frequency (f)
5-11	5
12-18	4
19-25	3
26-32	2
Total	14

∴

$$\bar{X} = \frac{\sum Mf}{\sum f}$$

$$8 = \frac{5+11}{2}$$

$$15 = \frac{12+18}{2}$$

∴



(M) Mid point

اوسط کلاس کے اعداد

Class	f	M	Mf
5-11	5	8	40 (8×5)
12-18	4	15	60 (4×15)
19-25	3	22	66 (22×3)
26-32	2	29	58 (29×2)
			<u>224 (Total)</u>

8

$$\bar{X} = \frac{\sum Mf}{\sum f} = \frac{224}{14} = \boxed{16}$$

Mean يوجد هناك خلافة في ترتيب حساب ال

الطريقة

$$\bar{X} = \frac{\sum X}{n} \Leftrightarrow \text{Raw Data} \quad \text{عند الحفظ ①}$$

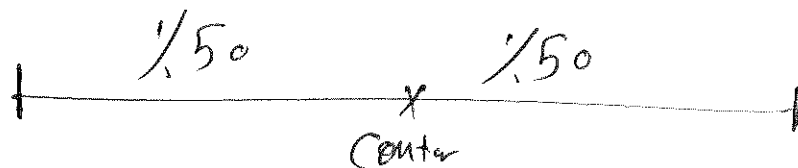
$$\bar{X} = \frac{\sum Xf}{\sum f} \Leftrightarrow \text{Data وكل واحد له تأثير مختلف} \quad \text{عند الحفظ ②}$$

$$\bar{X} = \frac{\sum Mf}{\sum f} \Leftrightarrow \text{Group Data} \quad \text{عند الحفظ ③}$$

## ② The Median الوسيط

is the value in the middle of Arranged

Data values



# # Step to find the Median

① Arranged the Data from the Smallest to Larger

ترتیب ال Data از کوچکترین

② find the median position

ابجاد موقع ال Median و یو در داند

A] If the number of values (Data) (n) is odd

اذا كان عدد ال Data فردي يتم ال  
ال position

$$\text{Position} = \frac{n+1}{2}$$

B] If the number of value (n) is even (اذا كان عدد ال (n) عدد زوجي)

توجد ال position

$$\text{Position} = \frac{n}{2} \text{ و } \frac{n}{2} + 1$$



### 3 Median

A إذا كان عدد الـ Data فردي  $\Rightarrow$  Median  $\Rightarrow$  قيمة الرقم الناتج من الموقع

B إذا كان عدد الـ Data زوجي  $\Rightarrow$  Median =  $\frac{\text{قيمة}_1 + \text{قيمة}_2}{2}$

يجب ترتيب هذه الأرقام بالترتيب

Example:

find the median for the following Data

2, 3, 5, 2, 9, 15, 8, 7, 19

خطوات الحل :-  
① ترتيب الأرقام تصاعدياً

2, 2, 3, 5, 7, 8, 9, 15, 19  
 $\underbrace{\quad}_1 \quad \underbrace{\quad}_2 \quad \underbrace{\quad}_3 \quad \underbrace{\quad}_4 \quad \underbrace{\quad}_5 \quad \underbrace{\quad}_6 \quad \underbrace{\quad}_7 \quad \underbrace{\quad}_8 \quad \underbrace{\quad}_9$

② هل عدد الـ Data زوجي أم فردي  
 عدد 4 إذا فردي

وبعد ذلك نجد الـ position حسب قانونه الفردي

$$\text{Position} = \frac{n+1}{2} = \frac{9+1}{2} = \frac{10}{2} = \boxed{5}$$

هذا رقم 5 يعني انه القيمة التي نرتبها 5 هي ال Median

وليس رقم 5 هو ال Median وإنما رقم 5 هو عبارة عن

Median = 7

Position = 5  
Median = 7

Example: find the median?

8, 4, 10, 3, 4, 1, 12 و 2

خطوات الحل

① ⇒ 1, 2, 3, 4, 4, 8, 10, 12

② هل ال رقم زوجي ام فردي (بعد العد نجد ان ال زوجي)  $n = 8$

③ نجد الموقع حسب قاعدة الزوجي

Position =  $\frac{n}{2}$  ,  $\frac{n}{2} + 1$

$\frac{8}{2}$  ,  $\frac{8}{2} + 1$

4, 5

Average      Median      (11)

$$\text{Median} = \frac{\text{فردی ترتیب 4} + \text{فردی ترتیب 5}}{2}$$
$$= \frac{4 + 4}{2} = \frac{8}{2} = \boxed{4}$$

Example find the Median

8, 10, 12, 6, 5, 1

① 1, 5, 6, 8, 10, 12

$\underbrace{\quad}_3$        $\underbrace{\quad}_4$

②  $\Rightarrow$  Position =  $\frac{6}{2}$  ,  $\frac{6}{2} + 1$

3, 4

$$\text{Median} = \frac{6 + 8}{2} = \frac{14}{2} = \boxed{7}$$

### 3 The Mode المنوال

\* The Value that Occurs Most frequently

\* هي القيمة التي تكرر في ال Data

Ex. 1, 1, 2, 2, 3, 3, 4, 4, 4

Mode = 4 { لانواكثر وتكرر 3 مرات }  
4

Ex: 1, 2, 3, 4, 5, 6

No mode

Ex: 11, 11, 12, 12, 6, 6, 3

Mode = 11, 12, 6

Ex<sup>o</sup>. for the following Salaries (1000 ILS)

3, 6, 7, 4, 10, 35

find the Mean, Median, Mode if found!

① Mean  $\Rightarrow \bar{x} = \frac{\sum X}{n} = \frac{3+6+7+4+10+35}{6} = \frac{65}{6} = 10.83$

② Median  $\Rightarrow$

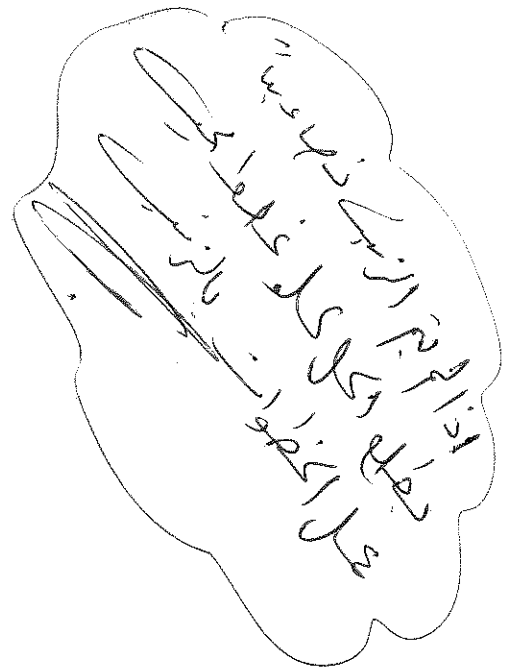
الترتيب حسب الصغر  $\Rightarrow$  3, 4, 6, 7, 10, 35

③ Data Points  $\Rightarrow$  (n=6) عدد

Position  
موقع  
الترتيب  $= \frac{n}{2}, \frac{n}{2} + 1$   
 $\frac{6}{2}, \frac{6}{2} + 1 \Rightarrow 3, 4$

Median  $= \frac{6+7}{2} = \frac{13}{2} = 6.5$

③ Mode = No mode.

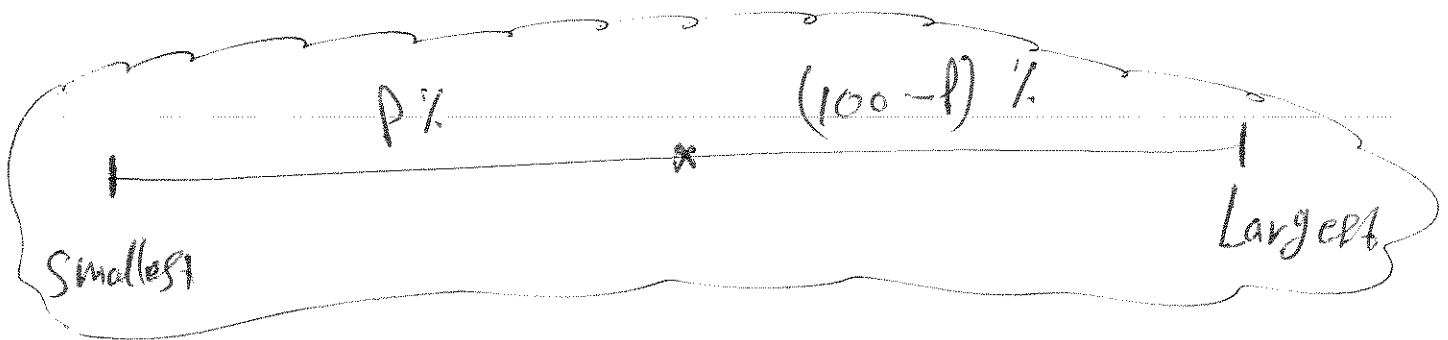


# # The Percentiles

Measure of Position  $\equiv$  The Percentile Measure  
Position or the rank of Data Value

Data القياس المسمى الرتبة القدرية

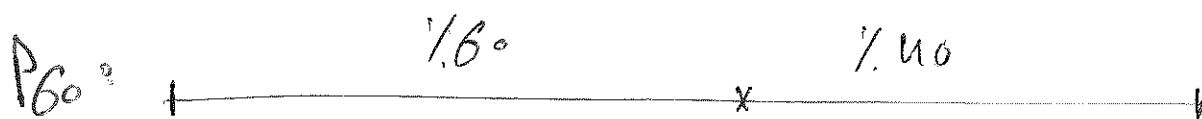
The  $P^{\text{th}}$  Percentile divides the data into two parts, is a value such that at least  $P\%$  of the data value are Less or equal to this value and at least  $(100 - P)\%$  of the data value are greater than or equal to this value



\*  
 $25^{\text{th}}$  Percen  $\equiv P_{25}$



$P_{25}$  : ۲۵٪ از داده ها کمتر از  $x$  است و ۷۵٪ از داده ها بیشتر از  $x$  است.  
۲۵٪ از داده ها کمتر از  $x$  است و ۷۵٪ از داده ها بیشتر از  $x$  است.



$P_{50} = \text{Median}$       وسطی

# How to find the  $p^{\text{th}}$  Percentile :

خطوات اكل: كيف بالترتيب التالي

1 Arrange the Data from smallest  $\rightarrow$  Largest

2 Compute the index ( $i$ )

$$i = \left( \frac{p}{100} \right) \times n$$

$n$ : # of values

3 if ( $i$ ) is not integer then round up

بعد ان نحسب ال ( $i$ ) اذا كان الجواب عدد غير صحيح نقربه  
ال اعلى، الصحيح اول منه مثلا

$$i = 9.3 \Rightarrow i = 10$$

$$i = 8.1 \Rightarrow i = 9$$

$$i = 8.8 \Rightarrow i = 9$$

$$i = 20.9 \Rightarrow i = 21$$

⋮

هذا ال عيب، ممكن

Position

$p^{\text{th}}$  Percentile is the Value of this Position



Q if  $i$  is an integer then the  $P^{\text{th}}$  Percentile is  
 is the Average of the two Values in position

$$(i, i+1) \Rightarrow \text{Position}$$

Ex: for the following Data find  $P_{25}, P_{50}, P_{40}, P_{75}$

4, 10, 13, 2, 17, 18, 20, 35

خطوات الحل:

① الترتيب تصاعدياً

2, 4, 10, 13, 17, 18, 20, 35

$P_{25} \Rightarrow$

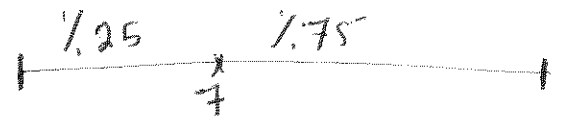
$$i = \left(\frac{P}{100}\right) * n = \left(\frac{25}{100}\right) * 8 = \frac{2}{1} \text{ (integer)}$$

Position =  $i, i+1$

$$(2), (2+1) \Rightarrow 2, 3$$

$$P_{25} = \frac{\text{صفحة الترتيب 3} + \text{صفحة الترتيب 2}}{2}$$

$$= \frac{4 + 10}{2} = 7$$



P50 :

$$i = \left(\frac{P}{100}\right) \times n$$

$$= \left(\frac{50}{100}\right) \times 8 = \frac{4}{1} \text{ (is integer)}$$

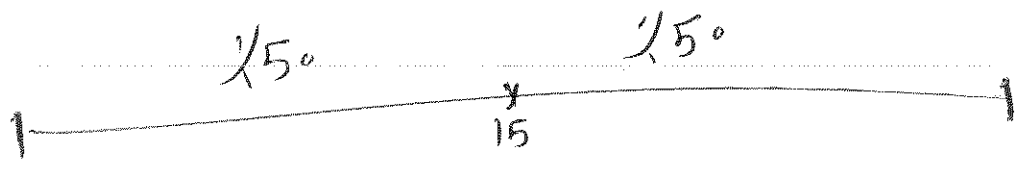
Position =  $i, i+1$

$$4, 4+1 \Rightarrow 4, 5$$

$$P_{50} = \frac{13 + 14}{2} = 13.5$$

الترتيب 5  
الترتيب 4  
الترتيب 3  
الترتيب 2  
الترتيب 1  
الترتيب 0  
الترتيب -1  
الترتيب -2  
الترتيب -3  
الترتيب -4  
الترتيب -5

$$P_{50} = 15$$



$$P_{40\%}$$

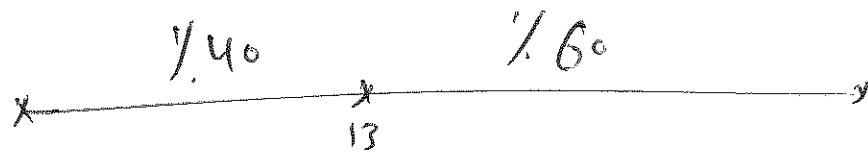
$$i = \left(\frac{P}{100}\right) \times n$$

$$= \left(\frac{40}{100}\right) \times 8 = 3.2 \text{ (is Not integer)}$$

$$i = 3.2 \approx \boxed{4} \quad \boxed{i=4} \text{ position}$$

$P_{40\%}$  فيه الزبب الرابع

$$\underline{\underline{P_{40\%} = 13}}$$



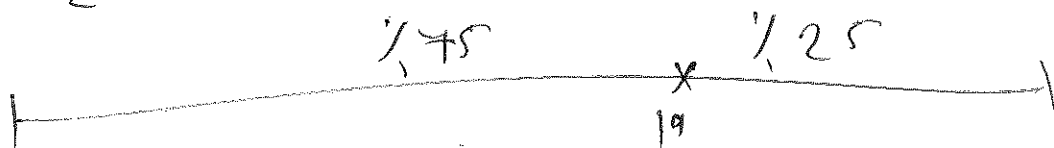
$$P_{75}$$

$$i = \left(\frac{75}{100}\right) \times 8 = 6$$

$$i = 6, 6+1$$

$$\boxed{6, 7}$$

$$P_{75} = \frac{18 + 20}{2} = \frac{38}{2} = 19$$



# # Special Percentile ((Quartiles))

في هذه الأقسام، سنناقش تعريفات خاصة بالنسبة لـ

1<sup>st</sup> quartile  $\equiv Q_1 \equiv P_{25} \equiv$  Lower Quartile

2<sup>nd</sup> quartile  $\equiv Q_2 \equiv P_{50} \equiv$  Median

3<sup>rd</sup> quartile  $\equiv Q_3 \equiv P_{75} \equiv$  Upper Quartile

~~##~~ ~~##~~ Five number summary:

وهذا هو التلخيص الخماسي

Smallest,  $Q_1$ ,  $Q_2$ ,  $Q_3$ , Largest

# حساب الكتل السابقة ( 2, 4, 10, 13, 17, 18, 20, 35 )  
 17 في الترتيب

five number summary

إذا كان أصب ( )

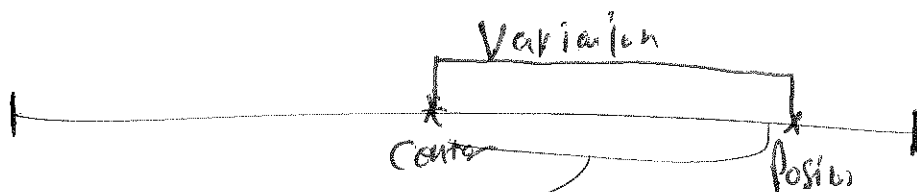
2, 7, 15, 19, 35  
 أصغر رقمي Data  
 $Q_1$   $Q_2$   $Q_3$  أكبر رقمي Data

### 3.2 : The Measure of Variation (( Dispersion ))

# التشتت

تغير في هذا الكمية هو انحراف القيمة عن

Center ( )



منه يقيسه التشتت

يوجد أكثر من طريقة (متغير) من خلال حساب

هذا التشتت :

① The Range ( المدى )

$$\text{Range} = \text{Largest} - \text{Smallest}$$

Ex: 2, 1, 7, 12, 18, 40, 35

$$\text{Range} = 40 - 1 = 39$$

Ex: 2, 8, 9, 12, 15

$$\text{Range} = 15 - 2 = 13$$

معلومة :-

ولكن هذا المناسبات تكون غير دقيق في حال وجود Outliers تشرح لقدام

لكنو وجود رقم كبير (يعرف) خلا في الحساب.

## ② The inter quartile range (IQR)

$$IQR = Q_3 - Q_1$$

وحد هذا الجيب من حيث Largest and smallest

(outliers) = (Extreme Value) التي تخرج ال

Ex: If the five number summary is

20, 35, 50, 70, 100 find the Range and IQR

السؤال هو إيجاد ال five number sum (Largest,  $Q_2$ ,  $Q_1$ , smallest)

(الحد الأدنى - الحد الأعلى) = الحد الأعلى - الحد الأدنى

$$\star \text{Range} = \text{Largest} - \text{smallest} = 100 - 20 = \underline{80}$$

$$\star IQR = Q_3 - Q_1 = 70 - 35 = \underline{35}$$

### 3 The Variance

التباين

قوانين الحساب

$$\sqrt{\text{Variance}} = \text{Standard deviation (S.d)}$$

$$(\text{Standard deviation})^2 = \text{Variance}$$

Notation:

$\sigma$  = small sigma

$\Sigma$  = ~~small~~ capital sigma

$\sigma^2$  = Population Variance } Parameter  
 $\sigma$  = Population Standard deviation }

$s^2$  = Sample Variance } Statistics  
 $s$  = Sample Standard deviation }

$s^2$  = Sample Variance  
 $s$  = Sample Standard deviation



# formulas:

$$\sigma^2 = \frac{\sum (X - \mu)^2}{N}$$

$$\sigma = \sqrt{\frac{\sum (X - \mu)^2}{N}}$$

$\mu$ : Population mean  $\Rightarrow \frac{\sum X}{N}$

$N$ : Population size

$X - \mu$ : The deviation of the value  $X$  from the mean ( $\mu$ )

$$S^2 = \frac{\sum (X - \bar{X})^2}{n-1} \quad \text{و} \quad \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1} \quad \text{أو} \quad \frac{\sum X^2 - n(\bar{X})^2}{n-1}$$

$\bar{X}$  = Sample ~~size~~ mean

$n$ : // size

هذه التلخيص  
قد أتينا له  
نفس الجواب بتعريف  
احفظوا اي و ام بين  
اي

$$\frac{\sum^2}{\text{Variance}} = \frac{\sum (X - \bar{X})^2}{n-1}$$

$$\frac{\sigma^2}{\text{Variance}} = \frac{\sum (X - \mu)^2}{N}$$

$$\text{Standard deviation} = \sqrt{\text{Variance}}$$

$$(\text{S.d.})^2 = \text{Variance}$$

Example :

for the following Data find the Standard Deviation

5, 10, 4, 6, 15

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

إذاً للزمن على هذا الشكل ولا حتم لكل شيء

X	$\bar{X}$	$X - \bar{X}$	$(X - \bar{X})^2$
5	8	-3	9
10	8	2	4
4	8	-2	4
6	8	-4	16
15	8	7	49
Total 40		0	82

$$\bar{X} = \frac{\sum X}{n} = \frac{40}{5} = 8$$

ليجب ان يكون

$$\text{Total} = 0$$

$$\sum X - \bar{X} = 0$$

$$S = \sqrt{\frac{82}{5-1}} = \sqrt{\frac{82}{4}} = \sqrt{20.5}$$

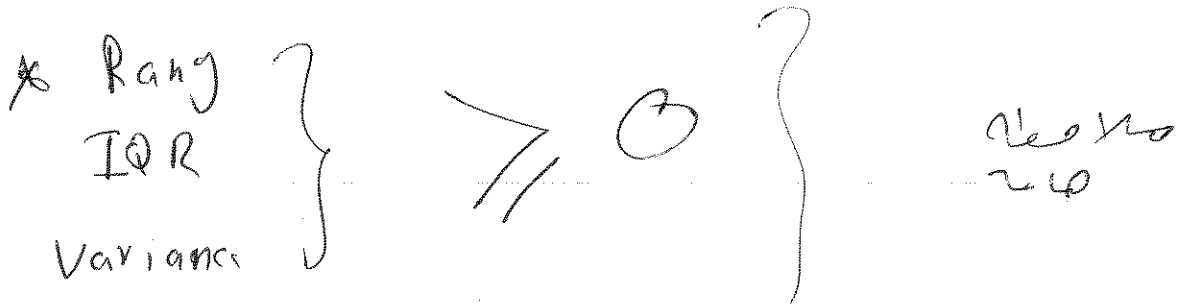
find the Sampl. Variance

$$\text{Sampl Variance} = S^2 = (\sqrt{20.5})^2 = \underline{\underline{20.5}}$$

# # The measure of Variation القياس

قياس تباين البيانات

- ① Range      ② IQR      ③ Variance



طريقة استخدام آلة الحاسبة العلمية في اكل

~~الكل~~

الخطوات

مقارنة استخدام الآلة الحاسبة في اكل

في هذه الخطوات بالترتيب

من الممكن اكل

1

هي مسح وتنظيف الذاكرة من المعلومات  
وتسمى كالتالي

Shift +  $\rightarrow$  Mode

يظهر في الذاكرة الحاسب بعد ذلك

SC1	Mode	All
1	2	3

اختار الرقم 3 واضغط عليه بهذا ليقتر مسح كل ما في

الذاكرة الحاسب وبعد ذلك اضغط على الزر AC

2

اضغط على الزر Mode بالتالي يظهر كالتالي

Comp	SD	REG
1	2	3

اختار الرقم 2 (أكسس 2)

3

بعد ذلك يتم ادخال ال Data في فراغات

ال Data

10, 8, 7, 6, 5

كيف بعد ذلك ان يفرغ (البيانات)

M +	10	ادخل الرقم
M +	8	" "
M +	7	" "
M +	6	" "
M +	5	" "

$N = 5$   
عدد ال Data ال دخلت

4

بعد هذه الخطوات نبدأ بكل وهو كالتالي

Shift + 1

نفس كالتالي

$\sum x^2$   
1

$\sum x$   
2

n  
3

حسب الاعداد السابقة

$\sum x^2 = 274$

$\sum x = 36$

$n = 5$

الخطوة 2 بعد ذلك

$n = 3 = =$

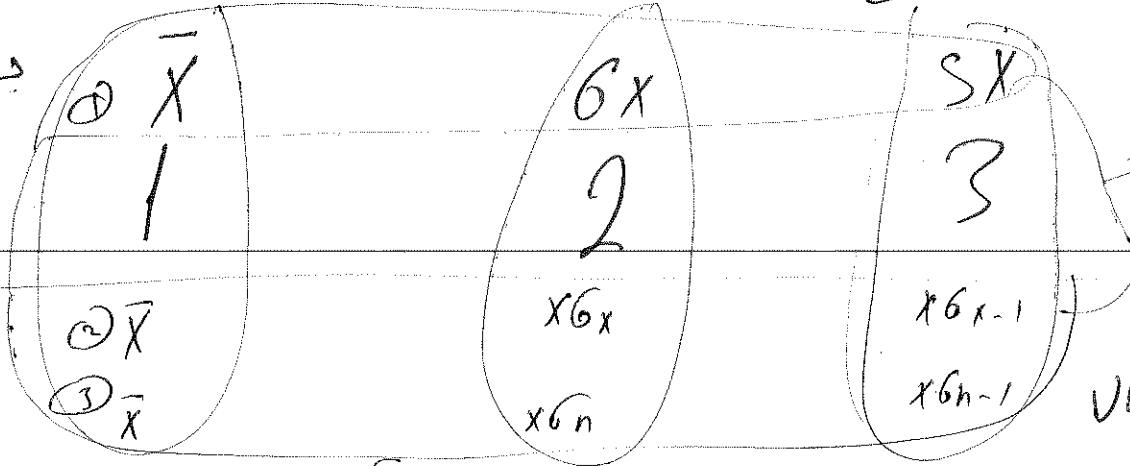
$\sum x^2 = 1 = =$

$\bar{X}$

Shift + 2  $\implies$

دفعه ۵ و ۷  
 $\downarrow$

دفعه اول  
دفعه دوم  
دفعه سوم  
دفعه چهارم  
دفعه پنجم



تفاوت بین دفعه اول و دوم  
و تفاوت بین دفعه دوم و سوم

حساب اکتال

$\bar{X} = 7.2$

~~XXXXXXXXXX~~

$S_x = 1.92$

از آنجا که میانگین و Variance ربع اول S.d

$Variance = (S.d)^2$  و  $S.d = \sqrt{Variance}$

از آنجا که او بعد از  $S_x$  و بعد از  $6x$  است و این  
این عبارت را Sample و این

Population

\* Standard deviation  $\uparrow$  = Risk  $\uparrow$

لأنه كلما زاد التباين كلما زاد الخطر

لأنه كلما زاد التباين كلما زاد الخطر

Var 1  $\Rightarrow$  S.D. 1

two Variable

Var 2  $\Rightarrow$  S.D. 2

لذلك فإننا نستخدم التباين لقياس الخطر

Variation من حيث التباين

Coefficient of Variation (C.V)

$$C.V = \frac{\text{Standard deviation (S.D)}}{\text{Mean}} * 100\%$$

Population  $= \left( \frac{\sigma}{\mu} * 100 \right)$  و  $\left( \frac{s}{\bar{x}} * 100\% \right)$  Sample



Exo

for the following ~~the~~ Table Which Exam has more Variation?

	Exam 1 (x)	Exam 2 (y)
$\bar{x}$	60	82
Sid	12	20

$$CV_1 = \frac{12}{60} \times 100\% = 20\%$$

$$CV_2 = \frac{20}{81} \times 100\% = 24.93\%$$

\* Exam 2 has More Variation

has more Variation

لما اختلف في السؤال

لما اختلف C.V

وعند 2 Variance

(C.V Variab. )

وكونه (C.V)

# # The Standard deviation for Grouped Data

$$S = \sqrt{\frac{\sum (M.P - \bar{X})^2 * f}{n-1}}$$

$$n = \sum f$$

f = frequency

M.P = Mid point

$\bar{X}$  = Mean

S: Sample Standard deviation

Q 53/122

Class	frequency
3-7	4
8-12	7
13-17	9
18-22	5
Total	25

find the Mean and  
Standard deviation and  
Variance

$$S = \sqrt{\frac{\sum (M.P - \bar{x})^2 \times f}{n-1}}$$

$$\bar{x} = \frac{\sum (M.P \times f)}{\sum f}$$

من كل هذا السؤال نلزمنا ان نكتب

M.P,  $\bar{x}$ ,  $(M.P - \bar{x})^2$ ,  $(M.P \times f)$

السؤال

$\left(\frac{U+L}{2}\right)$ M.P	M.P * f	M.P - $\bar{x}$	$(M.P - \bar{x})^2$	$(M.P - \bar{x})^2 \times f$
3-7	5 5x4 = 20	-8	64	64x4 = 256
8-12	10 10x7 = 70	-3	9	9x7 = 63
13-17	15 135	+2	4	4x9 = 36
18-22	20 100	+7	49	49x5 = 245
Total	325			600

$$\bar{x} = \frac{\sum M.P \times f}{\sum f} = \frac{325}{25} = 13$$

$$S = \sqrt{\frac{600}{25-1}} = 5$$

Variance =  $S^2 = 25$

35

### 3.3 The Standardized Value $z$ , ~~is called~~

## The Z-score

Def

~~The Z-score~~: Consider a data set with the mean ( $\mu$ ) and standard deviation ( $\sigma$ ), The Z-score for a value ( $X$ ) is given by

$$Z_x = \frac{X - \mu}{\sigma}$$

Population

$$Z_x = \frac{X - \bar{X}}{s}$$

Sample

Ex: if  $\mu = 60$ ,  $\sigma = 10$  for the following Data

$x = 80, 70, 60, 40$

$$Z_{80} = \frac{80 - 60}{10} = +2$$

$$Z_{60} = \frac{60 - 60}{10} = 0$$

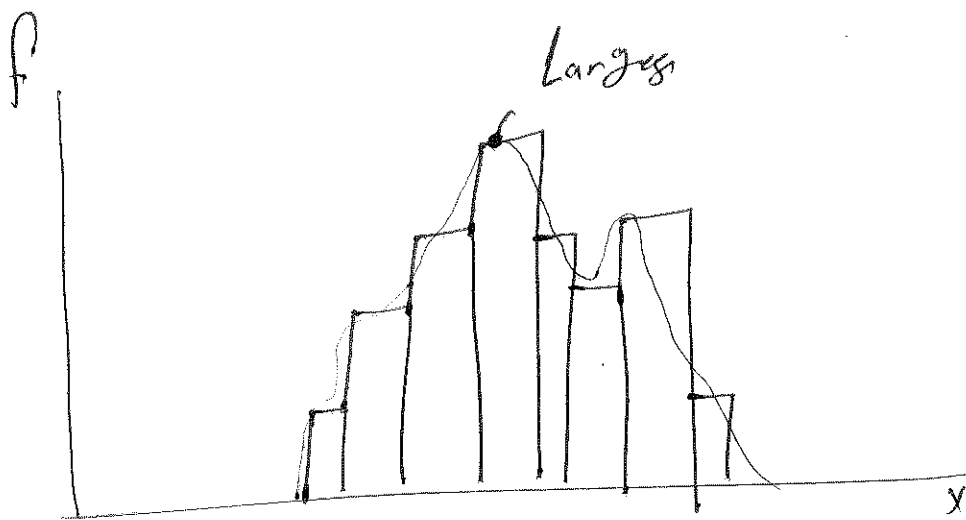
$$Z_{70} = \frac{70 - 60}{10} = +1$$

$$Z_{40} = \frac{40 - 60}{10} = -2$$

## \* Empirical Rule :

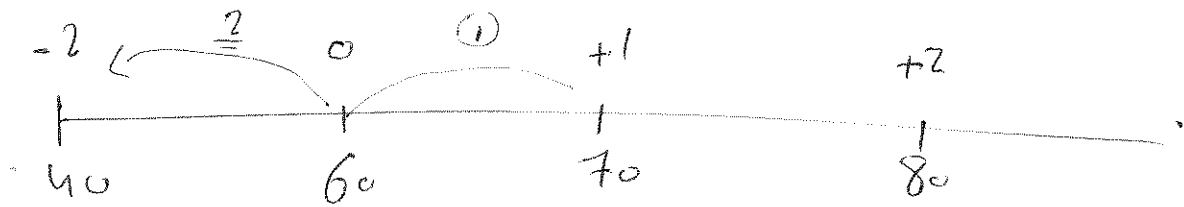
نکد من خلال نسبة Data الى Range من  
(توضیح مع اہل آہن)

Shape of distributio  $\Rightarrow$  Histogram



Largest = Mode

لکھو اکثریتی



So: the Z-score measures the Variation

Mean کی نسبت سے انحراف کی پیمائش

Z-score: the number of S.D for value X

From the mean

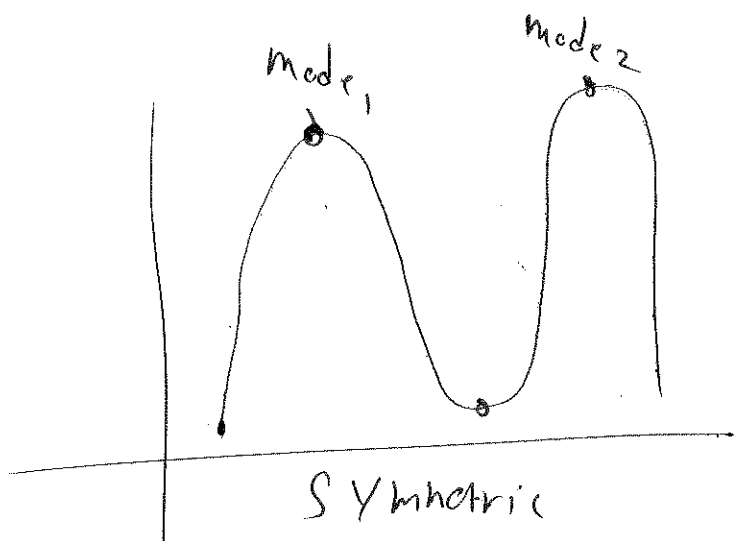
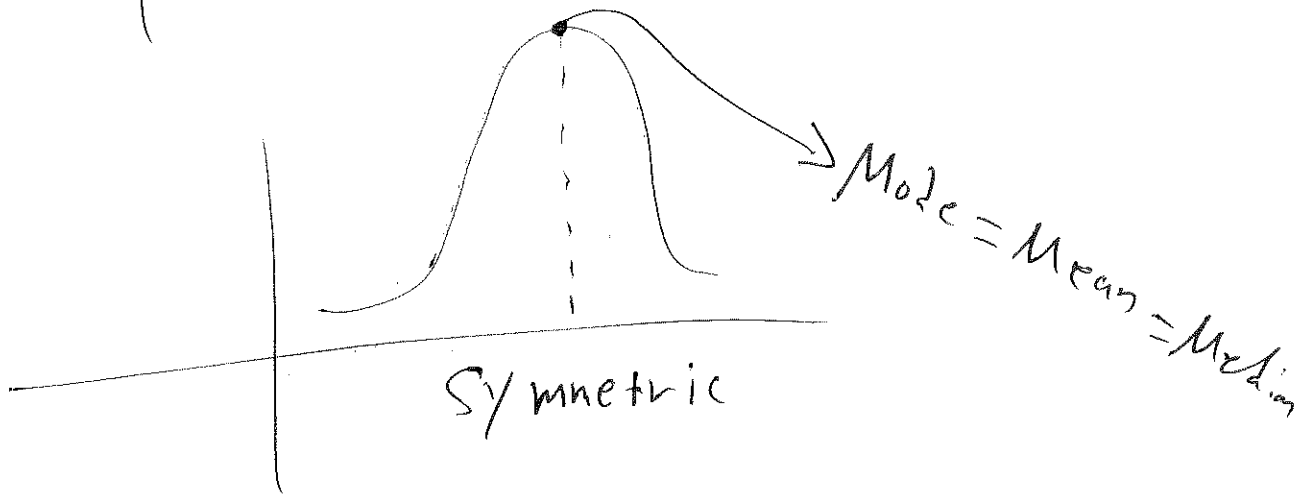
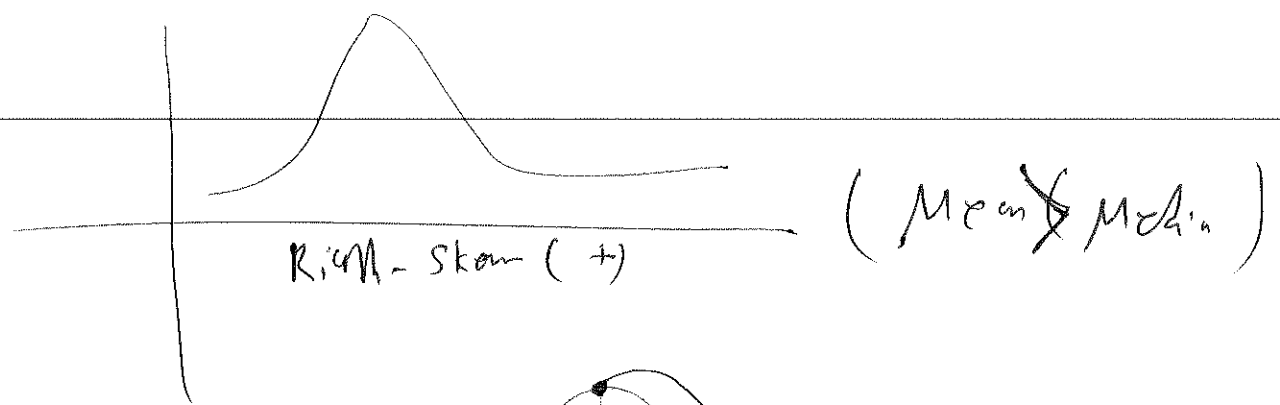
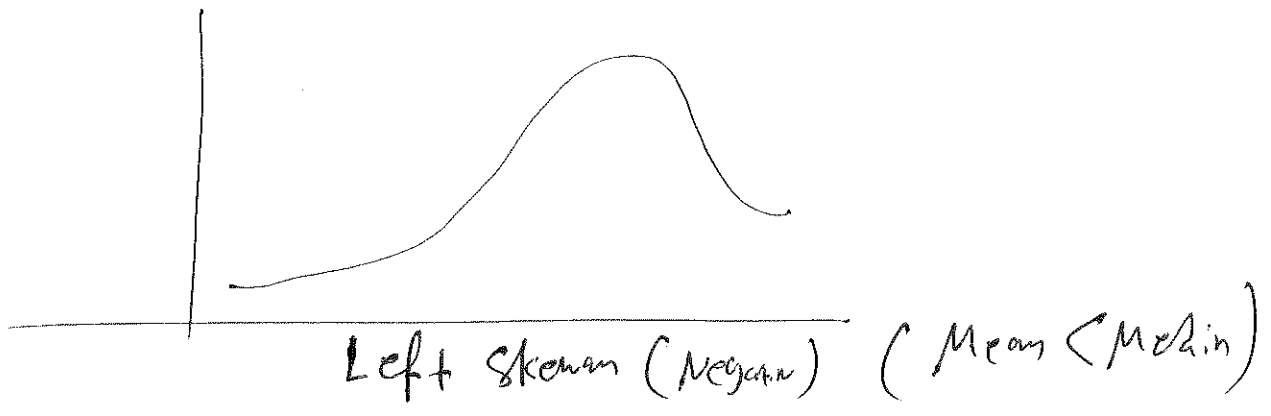
مقدار X سے انحراف کی پیمائش

$$Z = +2$$

$$\begin{aligned} 2 \text{ S.D} \\ = 2 \times 10 = 20 \end{aligned}$$

$$\left\{ \begin{aligned} Z_{80} &= 2 \\ 80 &\Rightarrow 20 \text{ سے } 20 \text{ زیادہ} \\ &\text{کی } 1 \text{ } \\ (Z=0) &\text{ Center} \end{aligned} \right.$$

(37)

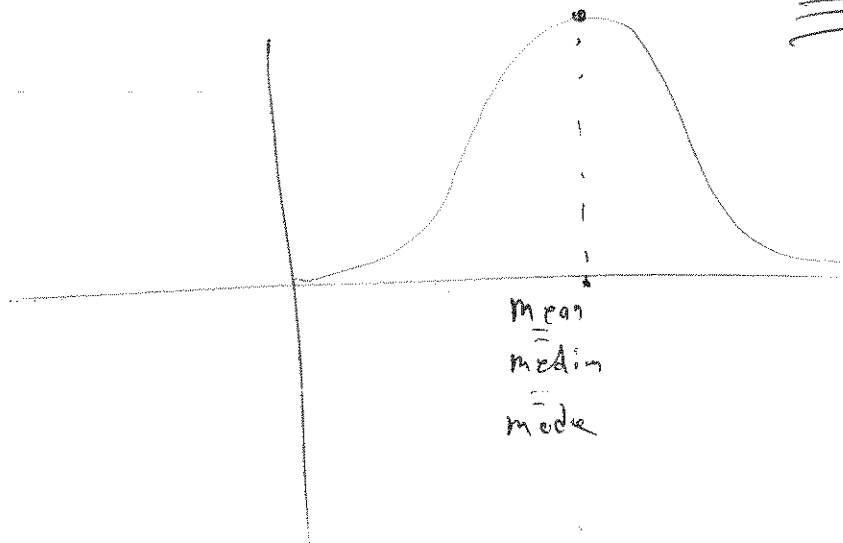


2 mod. سوزن دو کوهه

\* Empirical Rule

شروط هذا القانون

1 for a bell-shaped data (just)



2

Z score

سوف يكتب ال

$Z = 1$  or  $2$  or  $3$

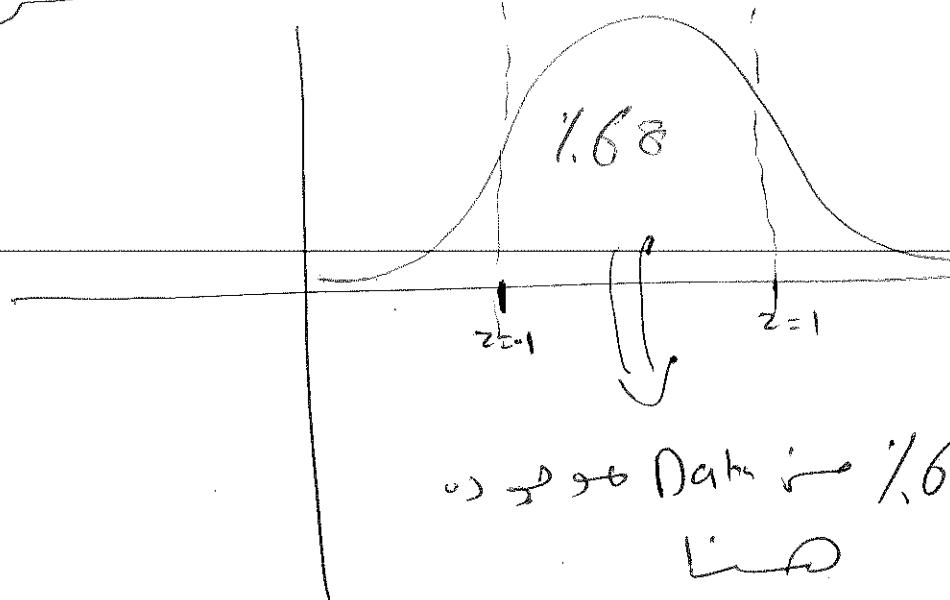
كتب

\*  $Z = 1.6, 1.7 \Rightarrow$  Chapter 6

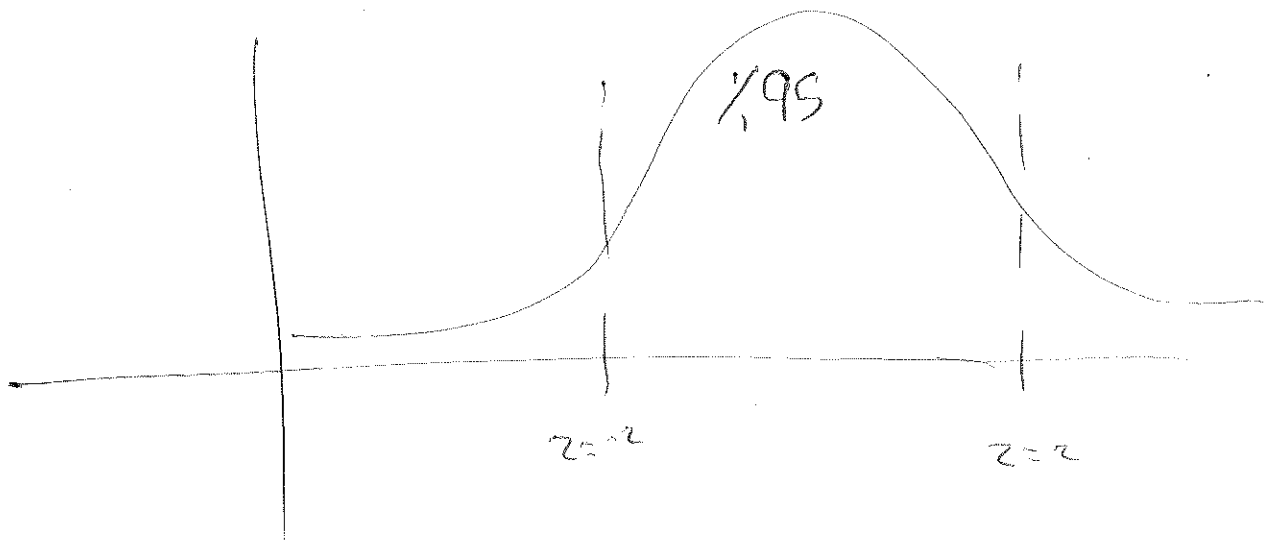


A If  $Z = 1$  ~~to~~ to  $Z = -1$  Approximately

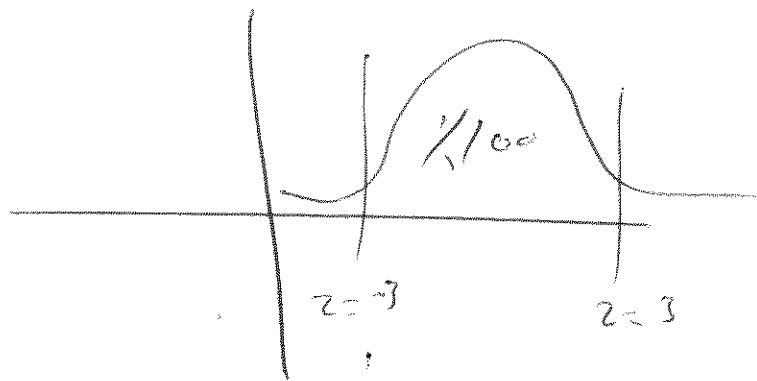
68% of the Data Values S.d of the mean



B Approximately 95% of Data Value  
are within  $(\pm) 2$  S.d of the mean



☐ Almost All of Data (100%) Values are within  $(\pm) 3$  S.d of the mean



∴ 99.99% of data

① Bell-shaped

② Range of Value (400-600), (200-300) ...

③ fin Z-score

④ Rule

- $\pm 1 = 1/68$
- $\pm 2 = 1/95$
- $\pm 3 = 1/100$

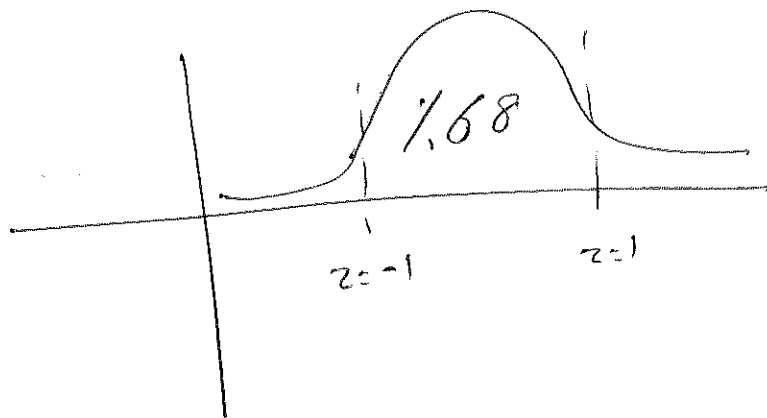
Ex:

The STAT Scores have a bell-shaped distribution  
with  $\mu = 500$ ,  $\sigma = 100$

① find the Perceny of score between 400 and 600

$$Z_{400} = \frac{x - \mu}{\sigma}$$
$$= \frac{400 - 500}{100} = -1$$

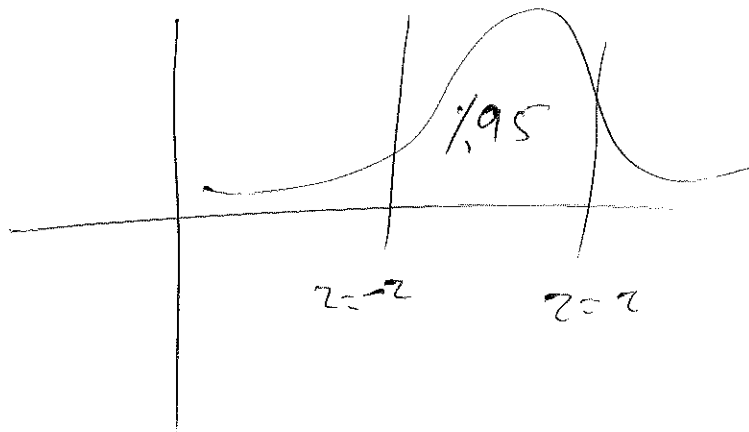
$$Z_{600} = \frac{600 - 500}{100} = +1$$



② between 300 and 700

$$Z_{300} = \frac{300 - 500}{100} = -2$$

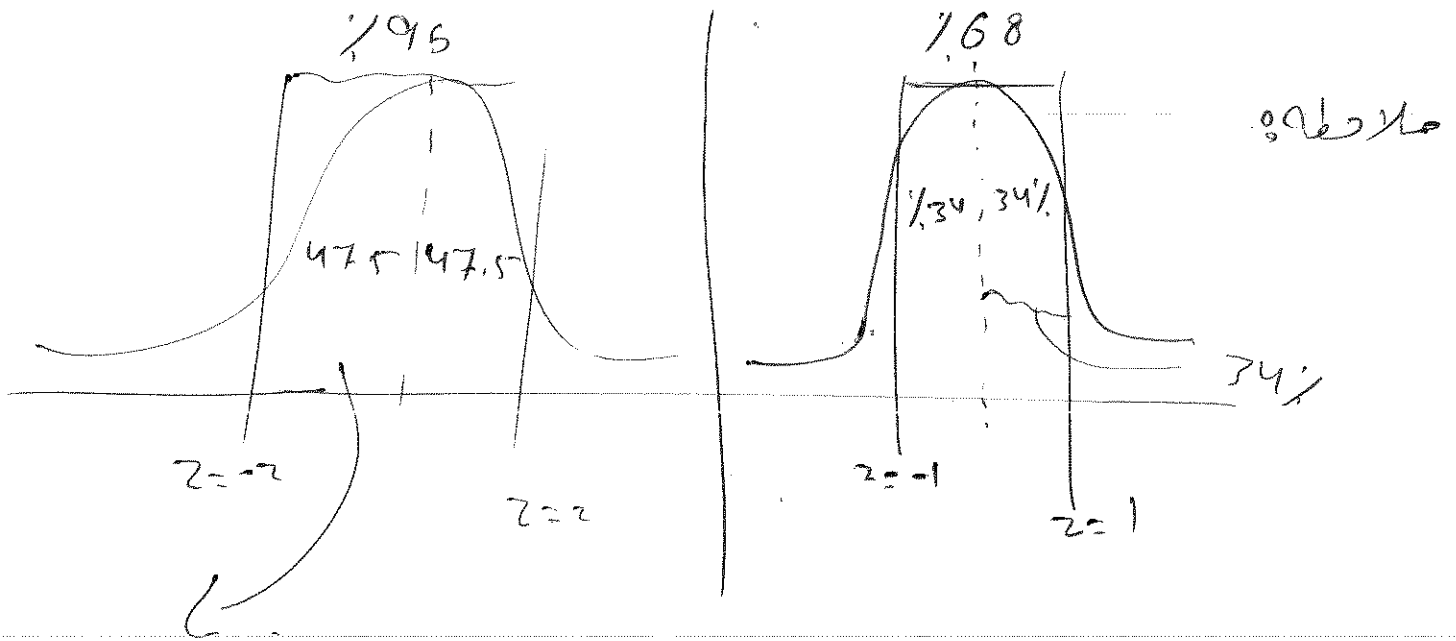
$$Z_{700} = \frac{700 - 500}{100} = +2$$



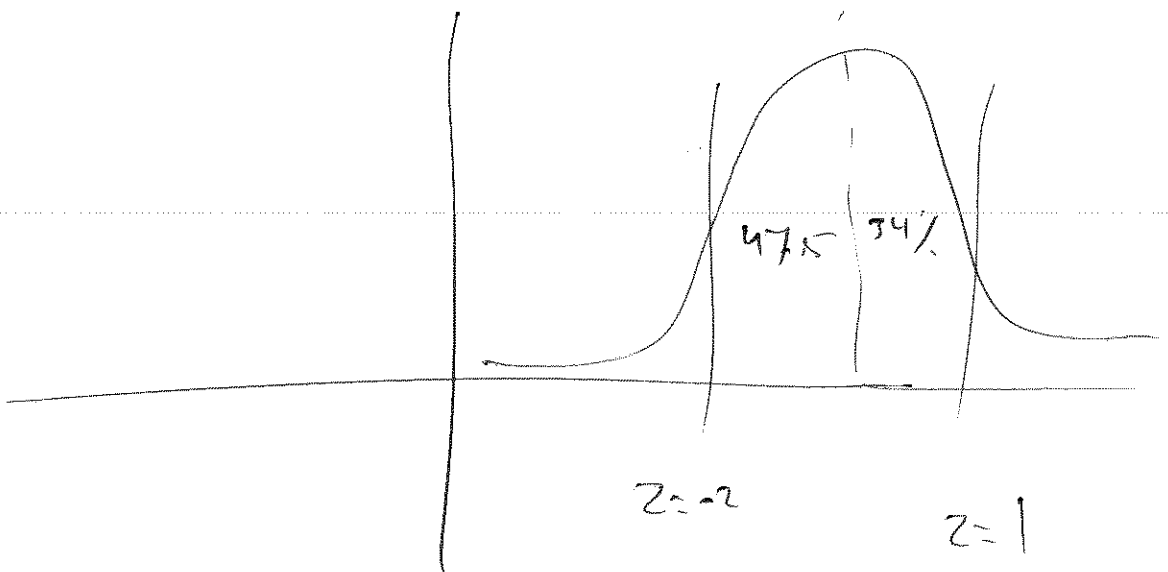
③ between 300 and 600

$$Z_{300} = -2$$

$$Z_{600} = +1$$



47.5



$$47.5 + 34\% = ~~181.5~~$$

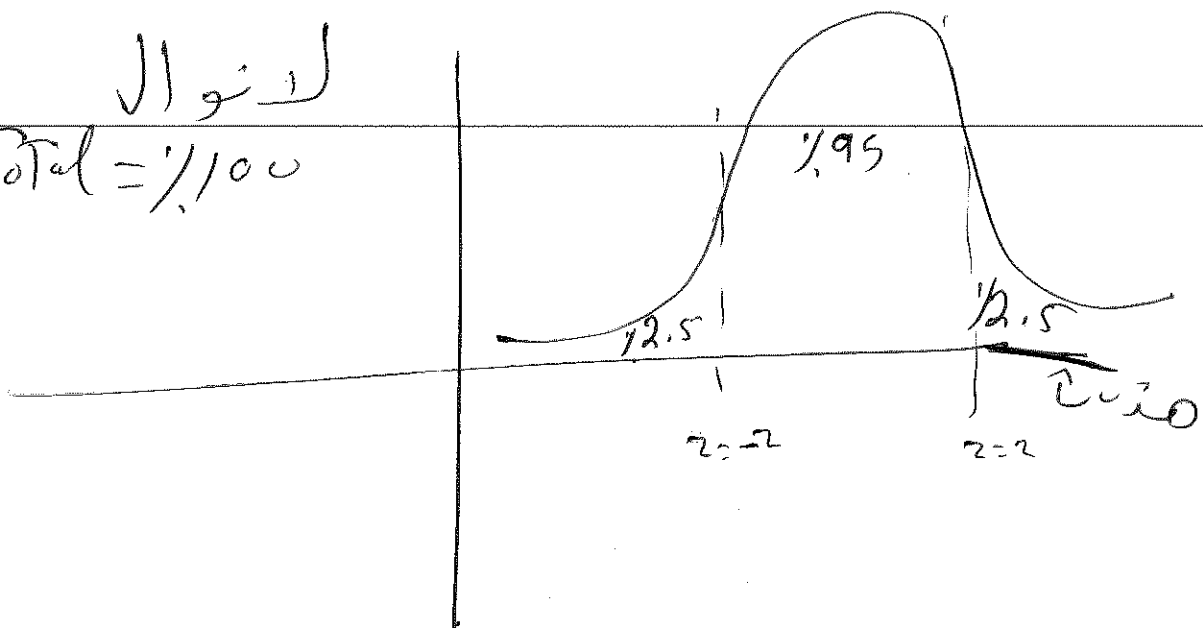
81.5%

44

4) % greater than 700

$$Z_{700} = \frac{700 - 500}{100} = +2$$

الاجابة  
Total = 1/100

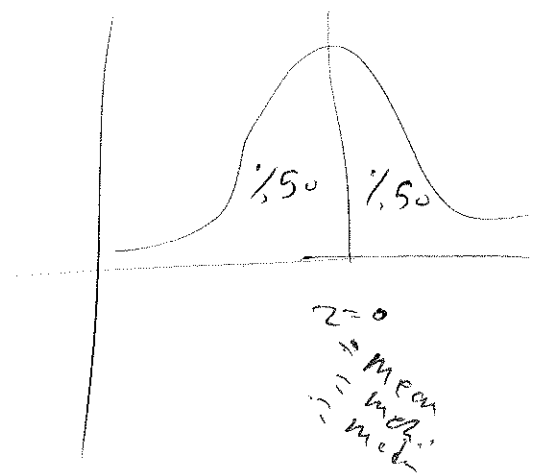


# Answer = 2.5%

5) % Less than 500

$$Z_{500} = \frac{500 - 500}{100} = 0$$

Answer = 1/50

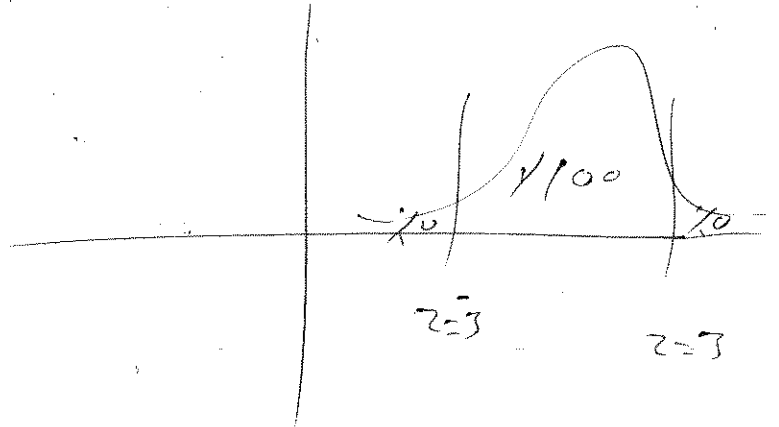


45

⑥ % greater than 800

$$Z_{800} = \frac{800 - 800}{100} = 0$$

$$= 0\%$$



⑦ Less than 800

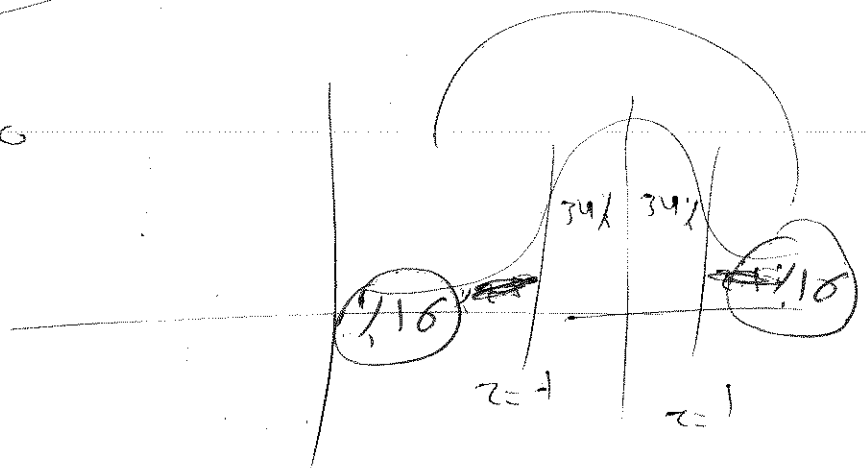
$$Z_{800} = 0$$

Answer = 100%

%/100

⑧ % Less than 600

$$Z_{600} = -1$$



$$= \underbrace{34\% + 34\%}_{168} + 18\% = 84\%$$

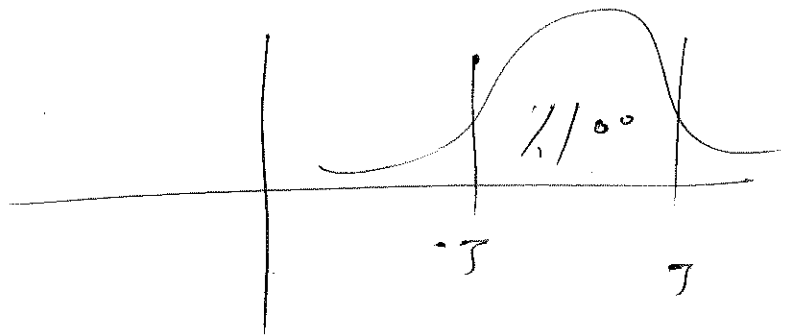
46

Q) is possible to have a score greater than 850

اگر

$$Z_{850} = 3.5$$

فقرتاً لا يوجد  $\Rightarrow 3.5 > 3$



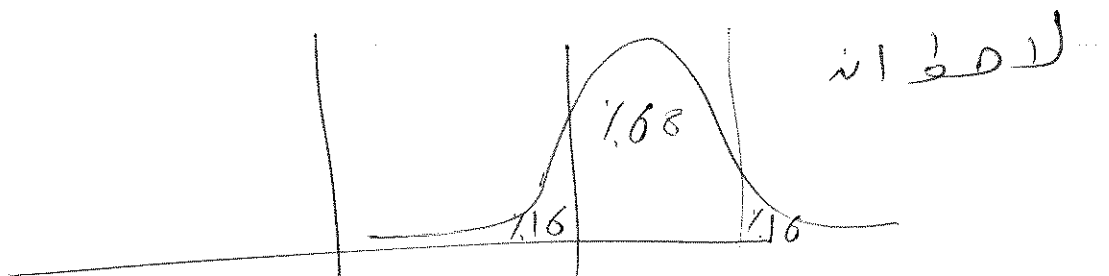
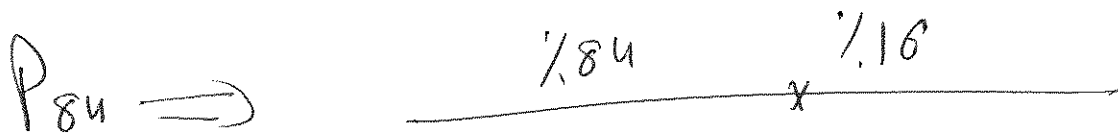
ولكن هناك احتمال وجود 850 وهذه

تكون نادرة و نسي

Extrem  $\equiv$  Outliers

١٥

# حساب المئوية من الذود



إذا نفس  $P_{84}$

$$\left\{ \begin{array}{l} 84 = 68 + 16 \\ \%16 \end{array} \right.$$

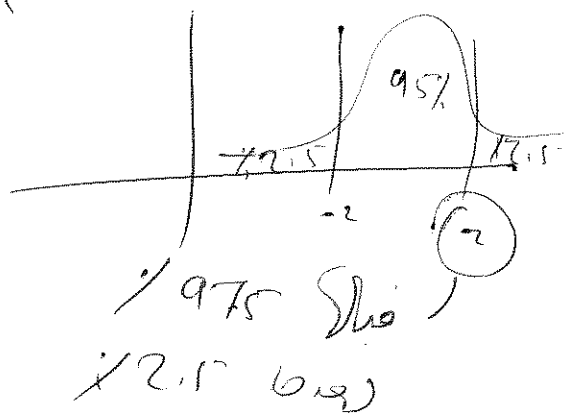
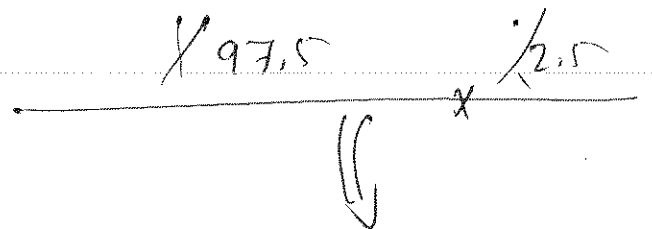
يوجد في  $z=1$   
 يوجد بعد  $z=1$

$P_{84} \Rightarrow z = 1$

$P_{16} \Rightarrow z = 1$

$P_{2.5} \Rightarrow z = -2$

$P_{97.5} \Rightarrow z = 2$





# # Outliers & Extrem Value

Very  
Large

Very  
small

Why outliers? ببعض أسباب

1] In Correctly Recorded Data

في بعض الأحيان معلومات خطأ مثل ذلك عند ادخال  
حساب ميزت وجود خطأ في 91 بدل 14  
عوضه اكله اذا اكتسبت القليل ثم زحفه

2] In Correctly included data value

عند فوجده اكله ان خطأ ب ميزت ثم بالخطأ  
سؤال ان حكاك وقال مرة 45 فكله اكله  
كثرت هذه المعلومات لان ليس من  
لا نوانا بدره بالخطأ (45)

③ Unusual data ٧٤١

عنه قال الطلاب جاوب آخدم ٧٥ المر وهذا  
صحيح لانه انقطع عن التدريب و فر عدا  
فر هذه الاماكن نيقرا ال ٧٥ وتوقع في اكل  
ولكن يجب ان توقع في Note

# If the Data bell shape

$z > 3$   
 $z < -3$  }  $\Rightarrow$  extrem = outliers

قفنا لـ Bell shape

لانه هذه الارقام فقط لـ Bell shape حتى توقع وجود

outliers هذا غير كانه لانه هناك طريقة

تسمى Boxplot حيث ان outliers

كانت اقل

# # BOX-Plot

حصان جسی

1] Given any Data

2] Arrange the dat. (Smallest  $\rightarrow$  Largest)

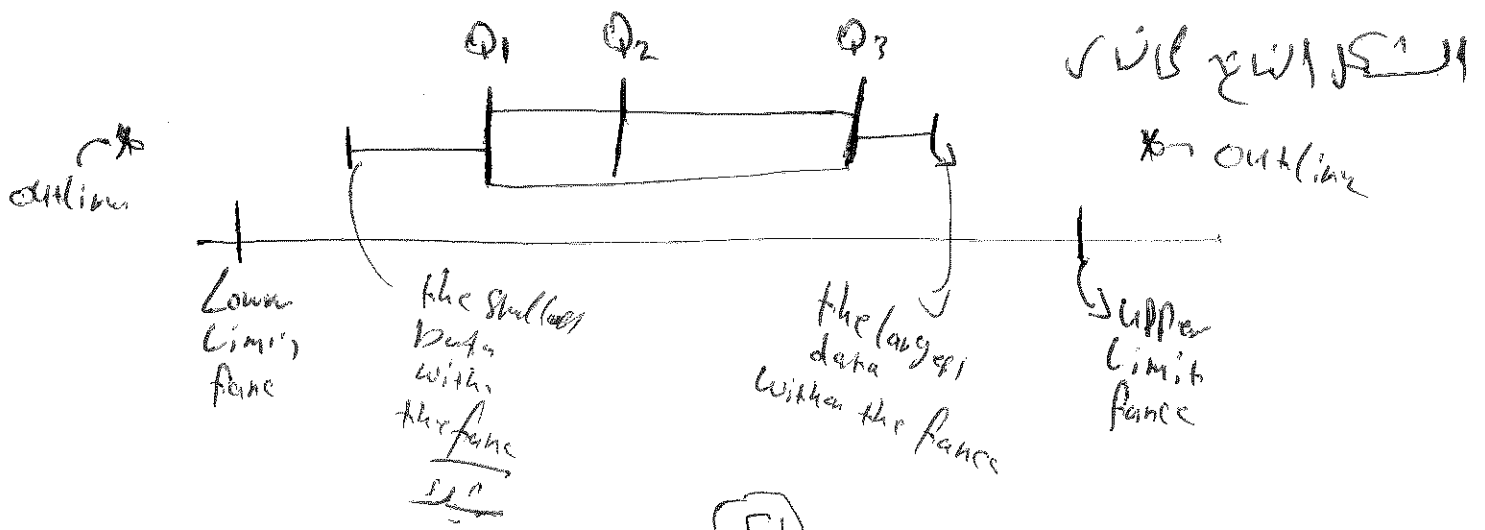
3] find  $Q_1, Q_2, Q_3$

4] find IQR ( $Q_3 - Q_1$ )

5] find  $1.5 * IQR$

6] find Lower Limit  $\Rightarrow Q_1 - (1.5 * IQR)$

7] find Upper Limit  $\Rightarrow Q_3 + (1.5 * IQR)$



Ex: Construct Box Plot for the following Data

10, 15, 1, 12, 8, 17, 24, 13, 20, 40

درجات اولى بجانب (51 في الترتيب)

1, 8, 10, 12, 13, 15, 17, 20, 24, 40

2  $Q_1 = P_{25}$

$$i = \frac{p}{100} \times n = \frac{25}{100} \times 10 = 2.5 \approx 3$$

$$Q_1 = 10$$

$Q_2 = P_{50}$

$$i = \frac{50}{100} \times 10 = 5 \Rightarrow P = 5, 6$$

$$Q_2 = \frac{13 + 15}{2} = 14$$

$$Q_2 = 14$$

$Q_3 = P_{75}$

$$\frac{75}{100} \times 10 = 7.5 \approx 8$$

$$Q_3 = 20$$

$$\textcircled{3} \text{ IQR} = Q_3 - Q_1 \\ = 20 - 10 = 10$$

$$\text{IQR} = 10$$

$$\textcircled{4} 1.5 \text{ IQR}$$

$$= 1.5 \times 10 = 15$$

$$1.5 \text{ IQR} = 15$$

$$\textcircled{5} \text{ Lower Limit} = Q_1 - 1.5 \text{ IQR} \\ = 10 - 15 = -5$$

$$\text{L. Limit} = -5$$

$$\textcircled{6} \text{ Upper Limit} = Q_3 + 1.5 \text{ IQR} \\ = 20 + 15 = 35$$

$$\text{U. Limit} = 35$$

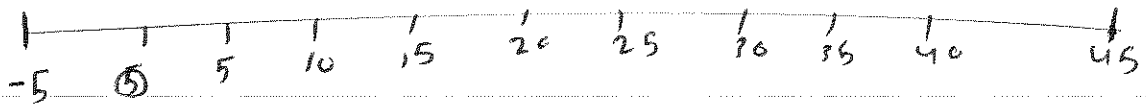
# خطوات الرسم : بحسب الترتيب

ملاحظة :-

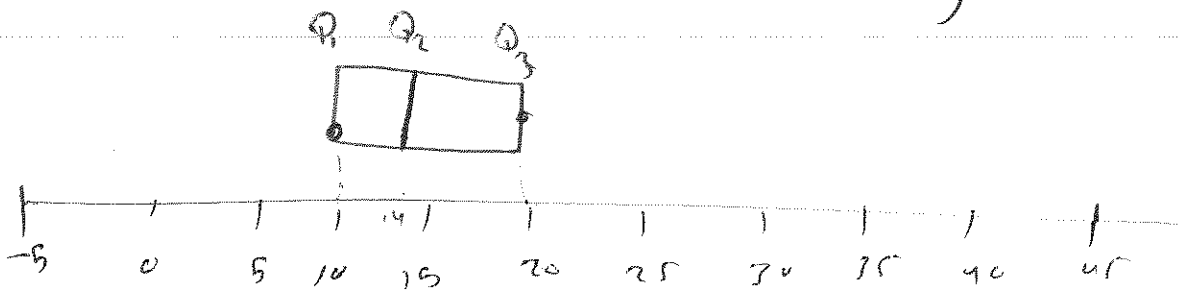
هي عبارة عن  $Q_1$   $\Rightarrow$  Upper Limit, Lower Limit  
 وبالنسبة الي قيمة  $Q_1$  من ال Upper او اقل  
 من Lower بحيث Outlier

## خطوات الرسم ( بالترتيب )

① نرسم خط الاعداد يكون مناسباً



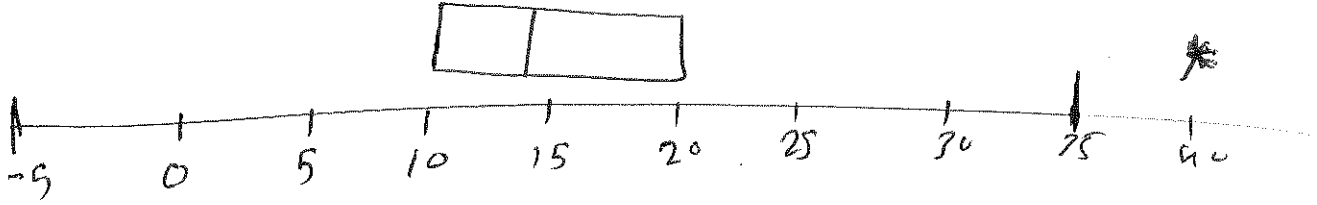
② نحدد  $Q_1$  اول اثنى ثم  $Q_3$  ثم نرسم بيضاً توصلنا  $Q_1$  و  $Q_3$   
 من اجل ثم نحدد  $Q_2$



③ نحدد ال Upper و Lower  
 Lower  
 Upper  
 54

4] اي قوه قوه ال  $U_{pp}$  اركت  $Lower$

يعر عن  $(*)$  Start كتر  $OUT(i)$



5] ~~نبت~~ عن ان ~~نبت~~  $Lower$  اوتادي

ان  $U_{pp}$  و اقل قوه  $Lower$  اوتادي

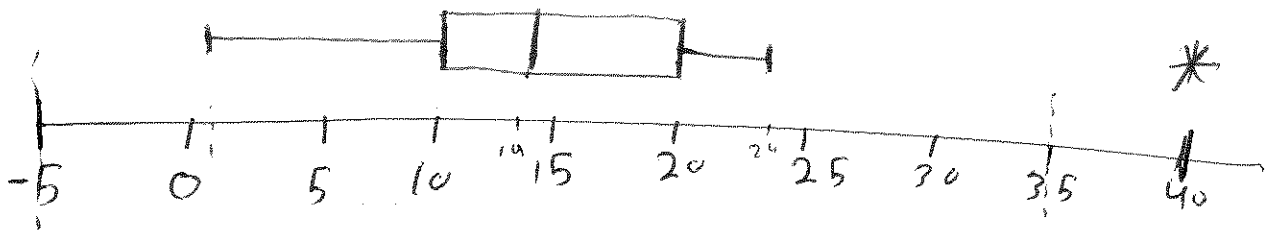
من ان  $Data$  من هذا المثال  $(1, 24)$

كقول  $U_{pp} = 35 = Lim$   
 اركت عن ان قوه و  $35$   
 بشرط تكون  $35$  ان  
 اول اقل من  $35$  ففوقه  
 ان  $35$

$Lower = 5$   
 اركت عن قوه بشرط  
 اقل من  $5$  وانما ناريه او  
 اول و حده ان عنده

و بعد  $5$  من  $5$  الى  $5$  ان  $5$

بعد ذلك نرسم المربع النهائي  
 (بقررة كذا سنأخذها) وكلنا قبل نوضح



هذا الخطان هما  
 Whiskers



3.5

# Measurement of Association Between 2 Variables

\* 2 Variables  $\Rightarrow$  Quant  $\Rightarrow$  Numerical

X  $\Rightarrow$  Independent Variable

Y = Dependent variable

المتغيران هما

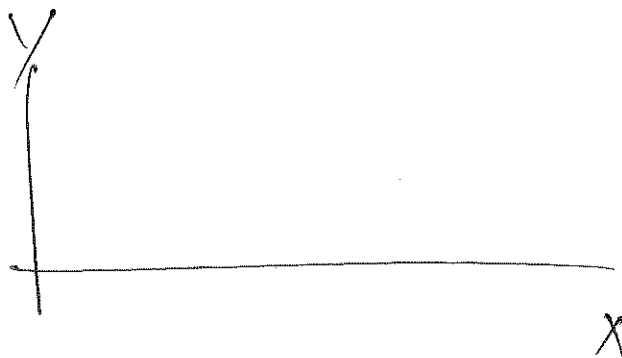
## Find the Relation $\rho$ : $\rho$ Between 2 Variables

العلاقة بينهما

1

### Scatter diagram

نقطة



② Co-Variance

③ Correlation

④ Regression

---

$X, y \Rightarrow$  two numerical variables

Notation:

$\sigma_{xy}$  = Population Co-Variance

$S_{xy}$  = sample //

formulas:

$$S_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{n-1}$$

$$\sigma_{xy} = \frac{\sum (x - \mu_x)(y - \mu_y)}{N}$$

$S_{xy}$ :

$> 0 \Rightarrow$  Positive Relationship

$< 0 =$  Negative Relationship

$= 0 \Rightarrow$  No Relationship

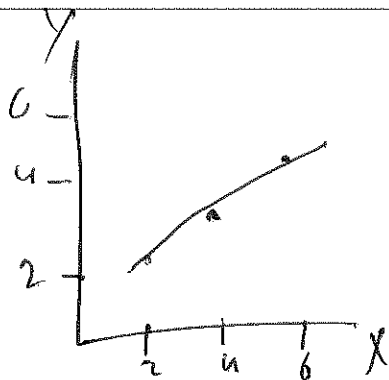
$\Sigma X$ :

X	Y
4	3
6	4
2	2

find the Relation ship  
between 2 variables (x, y)

جس

### 1 Scatter Diagram



Positive  
Covariance

ب ايجابية  
والتغير

2

$$S_{xy} = \frac{\Sigma (x - \bar{x})(y - \bar{y})}{n - 1}$$

$$= \frac{4}{3 - 1} = \frac{4}{2} = 2$$

حساب

$$\bar{x} = \frac{\Sigma x}{n}$$

$$= \frac{12}{3} = 4$$

$$\bar{y} = \frac{9}{3} = 3$$

$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$
0	0	0
2	1	2
-2	-1	2
		<u>4</u>

+2  
 $S_{xy} > 0 \Rightarrow$  Positive  
 Relation

# # Strength of the Relationship

Ⓐ  $S_{xy} = 2$

Ⓑ  $S_{xy} = 3$

هكذا لا يعني انه (B) أقوى من (A)

منال لو اخذنا العلاقة بين القوة والوزن

High weight  
cm kg

High weight

قوة أخرى

m kg

سوف يكون الجواب ( $S_{xy}$ ) مختلف

مع انه نفس السؤال لذلك هذا

الرقم ( $S_{xy}$ ) لا يحدد القوة وانما فقط

Prof  
Nisar  
Nour.hmp

جانب العندة

# ترتیب الفیہ میں تلاش

# Correlation

Notation:  $r_{xy}$  = Sample correlation coefficient

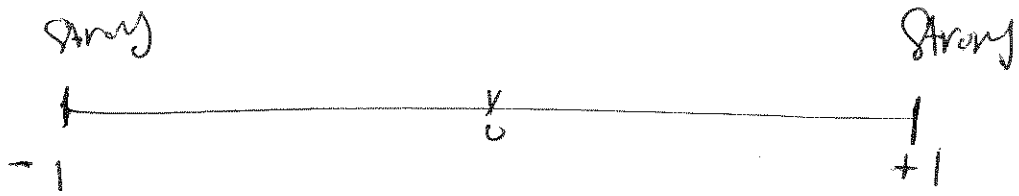
$\rho_{xy}$  = Population " "

\* formula

$$r_{xy} = \frac{\sum xy}{S_x S_y} \quad \left. \vphantom{\frac{\sum xy}{S_x S_y}} \right\} \text{Sample}$$

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad \left. \vphantom{\frac{\sigma_{xy}}{\sigma_x \sigma_y}} \right\} \text{Population}$$

$$\# \quad -1 \leq r_{xy} \leq 1$$



كلما اقتربنا من  $+1$  أو  $-1$  فزاد القوه  
وكلما اقتربنا من  $0$  قل القوه

$r_{xy} = 1 \Rightarrow$  Complete (Perfect) positive relation

$r_{xy} = -1 \Rightarrow$  " " " Negative " "

$r_{xy} = 0 \Rightarrow$  " " " No Relation " "

# # Regression ? نقطة

Method: Least squares method

هو الذي يتم من خلاله إيجاد الخط الذي يربط بين المتغيرات

من

$$\hat{y} = b_0 + b_1 X$$

$b_0 \Rightarrow$  constant value of A تقدير

$b_1 \Rightarrow$  slope of B =

$$\hat{y} = A + B X$$

نقطة التقاطع  
مع المحور Y

$\hat{y}$  = estimated y

$\hat{y}$  = نقطة  
y hat

$$b_1 = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$

or

$$b_1 = \frac{S_{xy}}{(n-1)(s_x)^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Ex:  $S_{xy} = 2$

$s_x = 2$

$\bar{x} = 4$  ,  $\bar{y} = 3$  ,  $n = 3$

find the equation

$$y = b_0 + b_1 x$$

$$b_1 = \frac{S_{xy}}{(n-1)(s_x)^2} = \frac{2}{(3-1)(2)^2} = \frac{2}{(2)(4)} = \frac{2}{8} = \boxed{\frac{1}{4}}$$

~~$y = b_0 + \frac{1}{4}x$~~   
 ~~$b_0 = 3 - \frac{1}{4} \cdot 4$~~   
 ~~$b_0 = 3 - 1$~~   
 ~~$b_0 = 2$~~   
 $y = b_0 +$

$$\begin{aligned} b_0 &= \bar{y} - b_1 \bar{x} \\ &= 3 - \frac{1}{4} \cdot 4 \\ &= 3 - 1 \\ &= 2 \end{aligned}$$

$\boxed{64}$

$$y^A = b_0 + b_1 x$$

$$\boxed{y^A = 2 + \frac{1}{4}x}$$



هذه المعادلة تستخدم لتقدير الناتج من

البيانات التجريبية من طرف البنك الوطني

$$X \Rightarrow \text{incom}$$

$$Y = \text{CAR}$$

السعر ان احب هو 1000 وحدة

Data: هذه المعادلة الناتجة من

$$\hat{y} = 2 + \frac{1}{4} X$$

$$\hat{y}^1 = 2 + \frac{1}{4} (1000)$$

$$\hat{y}^2 = 2 + 250$$

$$= 252$$

طريقة استخدام الآلة الحاسبة في حال وجود

2 Variables.

[1] Clear the Data

Shift + Mode  $\Rightarrow$  4-3  $\Rightarrow$  [AC]

[2] Mode

Comp	SD	REG
1	2	3

لا بد من 2 Variable 3 نحدد رقم 3 ولونز به اعداد

في Variable واحد نحدد رقم 2

الخطوة 3 وبعد ذلك نظهر لنا ✓

Lin	Log	EXP
1	2	3

اصفوا (5) الرقم 1 (Line)

3) Enter the Data

x	y
10	12
2	4
6	3

بجای آنکه در آن Data را وارد

بدخل اول آنرا رقم  $\frac{10}{x}$  و بجای و اول آن در جنب  $M7$   
 بعداً بدخل  $\frac{18}{y}$  و بعد از آن آنجا  $M7$

و هم با این داتا حساب مندا امکان است که آن به نظر

بعد از آن  $n=3$

اکل

Shift + 1  $\Rightarrow$  و هم کالری

$\Sigma x^2$   
1

$\Sigma x$   
2

n  
3

چند اکتوب

Shift + 1 + Replay (اولیٰ زبانی موجودہ)   
 (c Rank 0)

$$\sum y^2$$

$$\sum y$$

$$\sum xy$$

گنتا، اکی بے ای

Shift + 2

$$\bar{x}$$

$$\sigma_x$$

$$s_x$$

Shift + 2 + Replay

$$\bar{y}$$

$$\sigma_y$$

$$s_y$$

Shift + 2 + Rep(y) + Rep(y)

A	B	r
1	2	3

---

A  $\Rightarrow$  b0

B  $\Rightarrow$  b1

r  $\Rightarrow$  Correlation

Ex

x	y
10	18
15	28
12	20
20	35
25	40
13	25
14	30
11	20
7	15
4	10

find

$n, \bar{X}, \bar{y}, S_x, S_y, S_{xy}, r_{xy}, b_0, b_1,$   
 $CV(x), CV(y), IQR_x, \text{IQR}_y$

هنا السؤال بجزءه الأول أكابر  
صعب ما هو شرط في الصفات  
التي

$$\textcircled{1} n = 10$$

$$\textcircled{2} \bar{x} = 13.1$$

$$\textcircled{3} \bar{y} = 24.1$$

$$\textcircled{4} S_x = 6.05$$

$$\textcircled{5} S_y = 9.28$$

$$\textcircled{6} S_{xy} = ??$$

$$\textcircled{7} r_{xy} = 0.98$$

$$\textcircled{8} b_0 = 4.47 = A$$

$$\textcircled{9} b_1 = B = 1.5$$

$$\textcircled{10} CV(x) = \frac{S_x}{\bar{x}} \times 100\% = \frac{6.05}{13.1} \times 100\% = 46.2\%$$

$$\textcircled{11} CV(y) = \frac{S_y}{\bar{y}} \times 100\% = \frac{9.28}{24.1} \times 100\% = 38.5\%$$

$IQR_x = Q_{3x} - Q_{1x} =$   $\textcircled{1}$  برین ال Data  
کند  
انتخاب اول X

$$r_{xy} = \frac{S_{xy}}{S_x S_y}$$
$$0.98 = \frac{S_{xy}}{(9.28)(6.05)}$$
$$S_{xy} = 55.02$$

$\textcircled{7}$

