

# Chapter 1 :- Data and Statistics :-

## → Statistics :-

Science and art of collecting, analyzing, summarizing and interpreting data.

← العلم المختص في تجميع وتحليل وتفسير البيانات.

## ⇒ Data :-

Facts and Figures collected

← الحقائق والأرقام المجمعة للدراسة أو التوثيق

### \* Data Sources

#### 1) Existing Source :-

بيانات موجودة مسبقاً / ready to use من مصادر مستعدة.

مثل (Data bases) ← لا تحتاج لهدف للدراسة

#### 2) Collected Source :-

مربوبة بشكل أساسي في (observational data)

بيانات بجمعها حتى أصل (Statistical study)

← يجمع لهدف للدراسة.

Data  $\xrightarrow{\text{تكون}}$  Raw  $\xrightarrow{\text{ثولها}}$  organized data.

• Example :

kid	Gender	Age	Fav. color
1	M	3	Blue
2	F	6	Pink
3	F	4	Blue
4	M	5	White
5	F	4	Black

$\Rightarrow$  organized data.

(Data Set)  $\Rightarrow$  all data collected in a certain study.

- Element = عنصر (5 element)
- Variable = متغير (3 variables)
- observation = مجموعة كميات (5 observation)

$\Leftarrow$  كل Element مرتب مع مجموعة من observations

## تصنيف بيانات

① Qualitative data (بيانات نوعية)

↳ Nominal scale (البيانات غير مرتبة)

↳ ordinal scale

② Quantitative data (بيانات كمية)

↳ Interval scale (البيانات مرتبة وحسابية)

↳ Ratio scale

## ⇒ Scales of measurement:-

① Nominal scale (Names)

② Ordinal scale (Names and order)

③ Interval scale (Names, order, subtraction<sup>طرح</sup>)

④ Ratio Scale (Names, order, subtraction, ratio<sup>نسبة</sup>)

في المثال السابق ←

Gender → Qualitative → Nominal (M, F)

Age → Quantitative → Ratio (1, 2, 3, 4...)

Fav. color → Qualitative → Nominal (red, black...)

## Example 2:

- Phone n. → Qualitative → Nominal
- Address → Qualitative → Nominal
- Salary → Quantitative → Ratio
- Data of Birth → Quantitative → Interval
- Rating (good, Bad) → Qualitative → ordinal

Data ⇒ (تصنيف الوصاية بالرمز)

- ① Cross-sectional data (بيانات تجمع مرة واحدة (تقسيم لوقت))
- ② Time series data (بيانات تجمع على فترات زمنية)

## Examples -

- ① Temperature in different cities in PS on 13/9/2021  
→ Cross Sectional data.
- ② Gold prices from 1/9/2021 until 13/9/2021  
→ Time series data

\* درجة الحرارة أذا طلبتها في فترة محددة (يوم واحد) ~~Cross Sectional~~ بيكون Cross Sectional  
أما سعر الذهب طلبتو في فترات زمنية (Time series)

## Data Acquisition Error:-

← خطأ يحدث أثناء نقل البيانات

⇒ Example 24

20, 21, 22, 23, (42), ...

• في حالات سهو انوا اكتشف، الخطأ وفي مرات يكون صعب.

1930, 194, (1915), ...

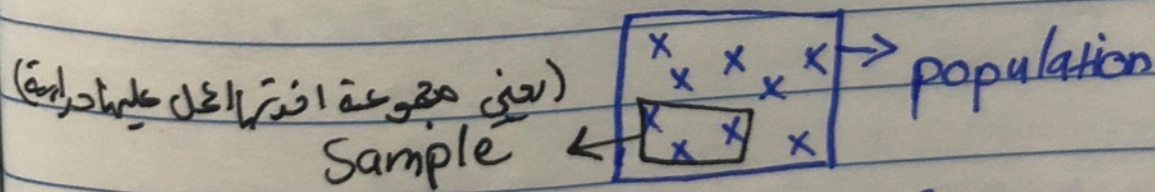
1950

\* Population: Set of all elements to be studied

مجموعة تحتوي جميع العناصر المراد دراستها.

\* Sample: Subset from the population.

مجموعة جزئية من الـ population



\* Census, Population Census: دراسة لكل المجتمع

⇒ Study on the whole population

\* Survey, Sample survey: دراسة للمجموعة الجزئية

Study on the sample

\* علم الإحصاء كثير يتم بالـ Sample

# Statistical Inference: (التعميم)

The process of making conclusions about the population on a sample survey.

إذا اخترت العينة صعب يتبع أعمم بعد م اعل

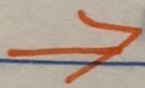
جوت ودراتك على ال Sample

$$\bar{X} = 37 \rightarrow \mu = 37$$

Sample

population

(إذا كان اختيار العينة بشكل قد روى، يربط تعميمها).



Done 😊

## Chapter 2 :-

Descriptive Statistics :  
Tabular and Graphical Presentations:

**2.1** ⇒ Summarizing Qualitative data:-

Example :-

1 star	Hated it	very Poor
2 stars	didn't like it	Poor
3 stars	liked it	fair
4 stars	Really liked it	good
5 stars	loved it	Excellent

أنا بقييم الشيء بقدر ما أعجبني، لنجوم يعني مثلا إذا  
فيعتبر نجمة واحدة فأنا (ما أحببت) وما الكه ذلك ...

أما إذا حسب نجوم فأنا حسب المنتج .

يعني مثلا (Hated it → very Poor)

← تخرج بعدد stars مع نسبة حبي للمنتج .

⇒ We want to Rate a product :-

← بقدر اقتار المنتج الكه بدي يا انا من المصدر

(Book, Clothing, Song, Perfume, ...)

Electronic device :-

Take a random Sample

→ Sample :

4 2 3 5 3 2 1 4 5 3 4

← كل ما كانت ال Sample أكبر يكون أحسن (بصير أده)

\* ① What is the population?

All people who tried the electronic device

← ياخذ جميع الناس الي جربوا الجهاز حتى بس اي أخذتها ب Samples

\* ② How many elements do we have in the population / population size?

We don't know / the population size is unknown.

← هون سألني عن العناصر في ال population أنا عندي

العناصر ال Sample عندي هين بيها unknown

\* ③ What is the sample size?

The sample size = 12

← هون سألني عن العناصر في ال Sample وبالطال موجوده 12

\* ④ How many elements do we have?

هون سألني عن عدد العناصر  $N = 12 \Rightarrow$

بشكل عام فانا بتطلع على الي موجود هندي 12



\* 5) <sup>①</sup> Who many variable do we have?

<sup>②</sup> What are they? <sup>③</sup> Are they Qualitative or Quantitative? <sup>④</sup> What is the scale of measurement?

① We have 1 variable ② the variable is rating of the electronic device

③ Qualitative (ما يربط أطرح وأقسم التقسيم)

④ Ordinal (تقدر أرتبها هنا الأحسن للأسوء)

\* 6) Construct frequency distribution

Rating <sup>← التقسيم</sup>	Frequency <sup>← عدد مرات التكرار</sup>
1	1
2	3
3	3
4	3
5	2
Total →	12

Note 8

The frequency distribution is a statistics (tabular presentation)

\* Relative frequency

$$= \frac{\text{Frequency}}{\text{Sample size}}$$

⑦ Construct the relative frequency distribution

Rating	Relative freq.
1	$\frac{1}{12} = 0.08$
2	$\frac{3}{12} = 0.25$
3	$\frac{3}{12} = 0.25$
4	$\frac{3}{12} = 0.25$
5	$\frac{2}{12} = 0.17$
<u>total</u>	<u>1</u>

• با فرضیاتین بحر لافایله.

← دایماً لازم الـ Relative freq. نـ و نـ 1

\* Percent freq. =  $\frac{\text{freq.}}{\text{Sample size}} \times 100$

⑧ Construct a percent freq. distribution?

Rating      Percent freq.

1       $\frac{1}{12} \times 100 = 8.33$

2       $\frac{3}{12} \times 100 = 25$

3       $\frac{3}{12} \times 100 = 25$

4       $\frac{3}{12} \times 100 = 25$

5       $\frac{2}{12} \times 100 = 16.67$

total      100

100 Percent freq. ←

⑨ Construct a paragraph. (use the freq.)

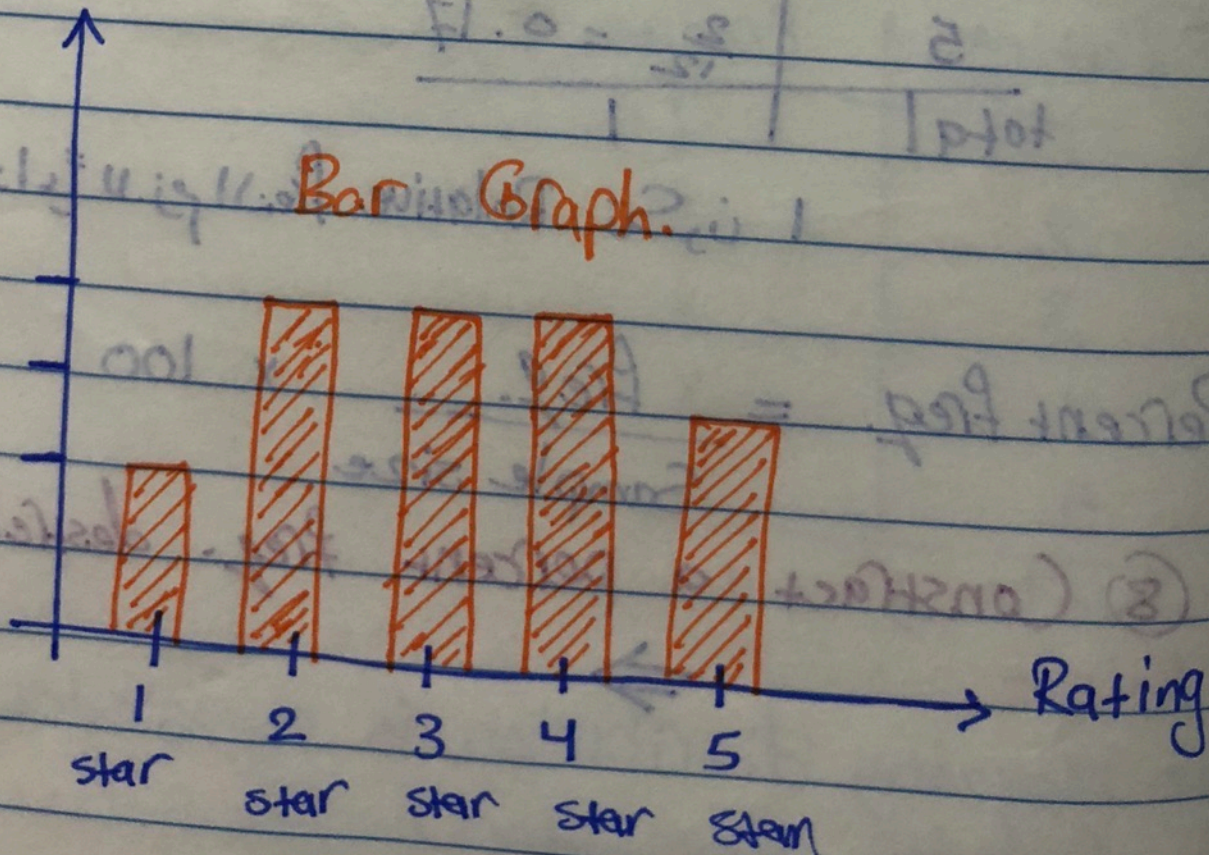
freq.

Bar Graph.

3

2

1



في سؤال 9، ملأنا عمل رسم ديكارتية بالأعمدة بخط في  
محور السينات (x) ← Rating وفي الـ (y) محور الصادات  
إزاها كان محددية ايش هوي طالب ← Freq.

Relative Freq. ←

Percent Freq. ←

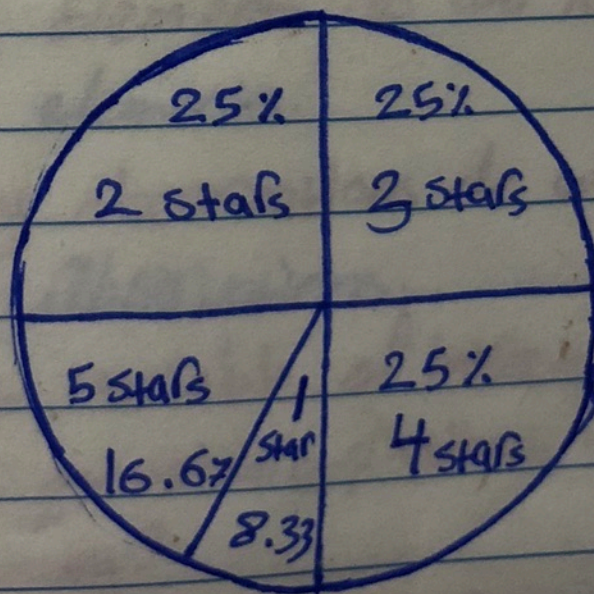
• أما في السؤال فهو بي حددي ائو بي بالبرصة Freq.

### 10) Construct a Pie Chart (تقسيم بالقطاعات الدائرية)

← ملأنا طلب مني تقسيم بالقطاع الدائري دائياً جاذ

الـ Percent Frequency لانو بجمع حسب النسبة % -

16.67 → 5 / 25 → 2, 3, 4 → 8.33 → 1



## 2.2 ⇒ Summarizing for Quantitative Data

### Example 8

You are given waiting time in minutes and a certain clinic

Sample ⇒	<u>2</u>	<u>5</u>	<u>10</u>	<u>12</u>	<u>4</u>	0-4
	<u>4</u>	<u>5</u>	<u>17</u>	<u>11</u>	<u>8</u>	5-9
	<u>9</u>	<u>8</u>	<u>12</u>	<u>21</u>	<u>6</u>	10-14
	<u>8</u>	<u>7</u>	<u>13</u>	<u>18</u>	<u>13</u>	15-19
						20-24

① What is the Sample size?

$$\Rightarrow N = 20$$

② How many Elements do we have?

⇒ 20 elements.

③ How many observations do we have?

⇒ 20 observations.

④ What is the variable of interest? It is  
① qualitative or quantitative? ② what is the  
Scale of measure?

① Waiting time in minutes ← Variable

② Quantitative

③ Ratio

## 5) Construct a frequency distribution

(use the classes: 0-4, 5-9, and so on)

Class	Frequency
0-4	4
5-9	8
10-14	5
15-19	2
20-24	1
Total	20

← This is a descriptive statistics.  
(Frequency Distribution)

\* We have 5 classes.

• Class width = 5 (عرض، رصت)

← کم عدد ستوریال class اولو 5 for all classes

0 - 4

lower limit  $\uparrow$  of the class      upper limit  $\uparrow$  of the class

$$\Rightarrow \text{Class width} = \text{Upper class limit} - \text{lower class limit} + 1$$

In general  $\Rightarrow$

Data  $\rightarrow$  Frequency distribution

① Choose No. of classes from 5 to 20

② Calculate the appropriate class width

$$= \frac{\text{largest value} - \text{Smallest value}}{\text{No. of classes}}$$

⇒ Class width = approx. Class width boundary

\* عيني ما يطلع في رقم يقربوا للقيمة الاكبر

(الكسر يقربوا لاقرن عند صحيح)

3) Choose upper and lower class limits

← انا لختم صون لوما حد في الترتيب ابي بي باه

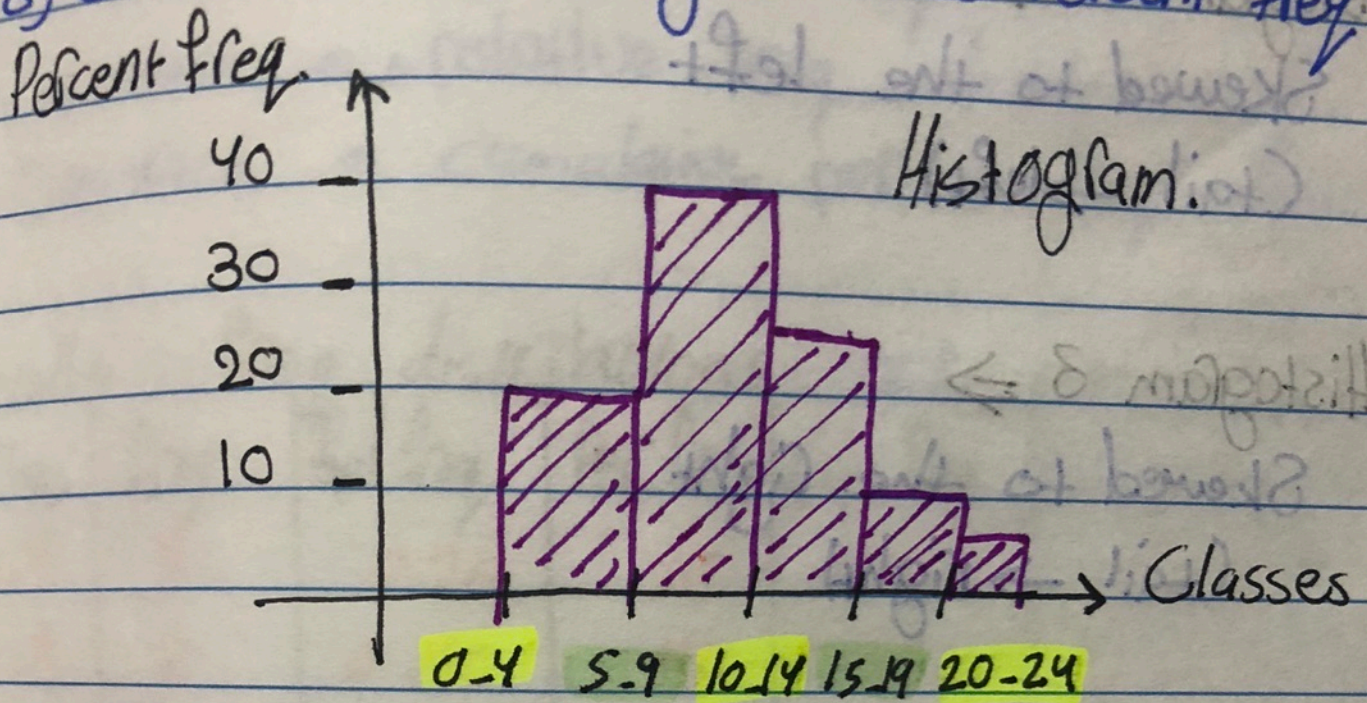
انا بتبع ترتيب من عندي وتتبع الخطوات ابي عندي فويتا

6) Construct a Relative Frequency Dist.

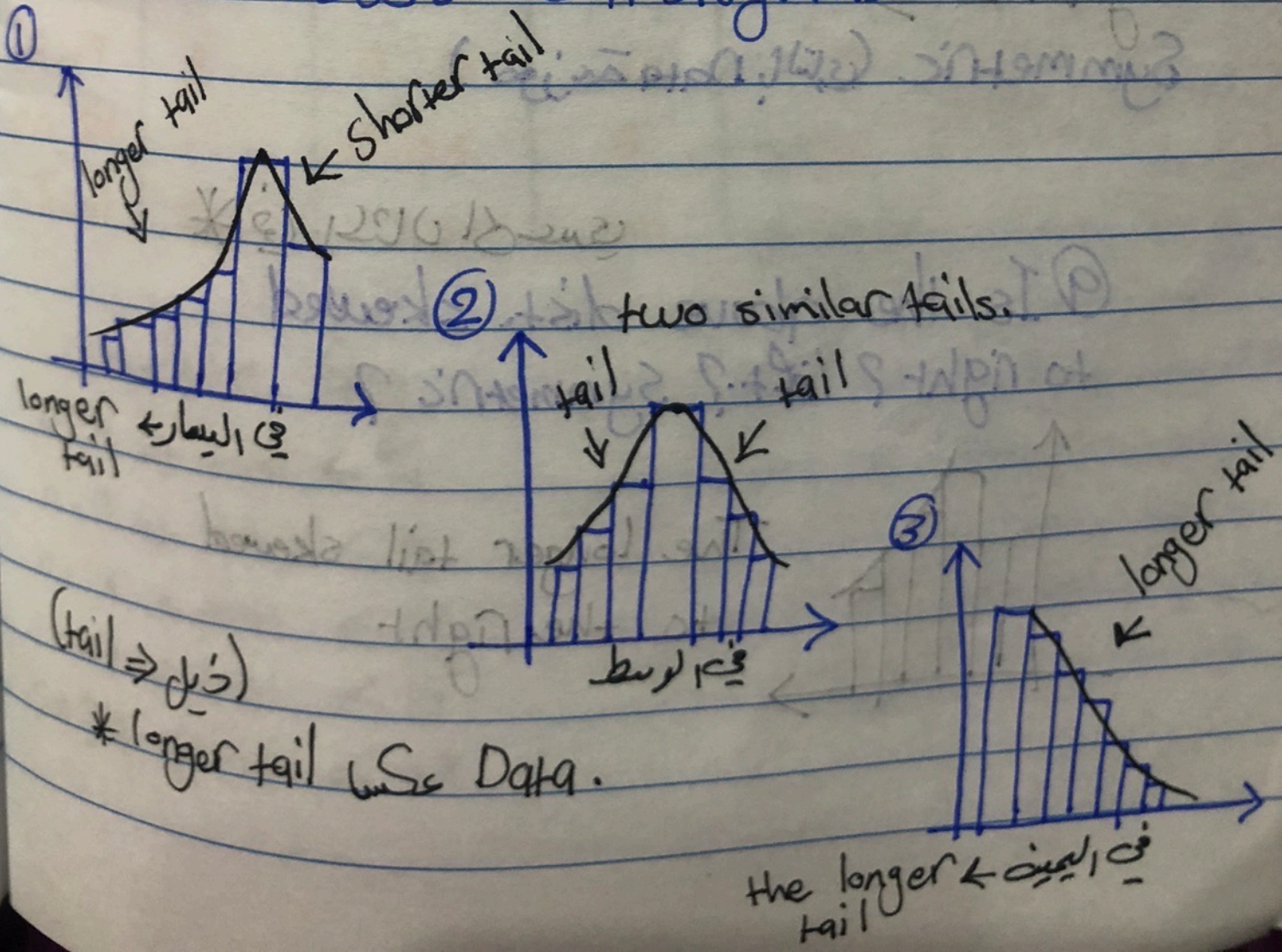
Classes	Freq.	Relative freq.	Percent freq. Dist.
0-4	4	$\frac{4}{20} = 0.20$	$0.20 \times 100 = 20$
5-9	8	$\frac{8}{20} = 0.40$	$0.40 \times 100 = 40$
10-14	5	$\frac{5}{20} = 0.25$	$0.25 \times 100 = 25$
15-19	2	$\frac{2}{20} = 0.10$	$0.10 \times 100 = 10$
20-24	1	$\frac{1}{20} = 0.05$	$0.05 \times 100 = 5$
Total	20	1.00	100

7) Construct a percent Dist. ↑

# 8) Construct a Histogram (use Percent freq.)



## \* SKEWNESS - 3 Histograms



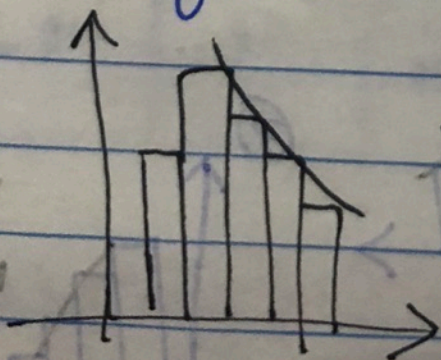


- Histogram 1  $\Rightarrow$   
Skewed to the left  
(tail  $\rightarrow$  left)

- Histogram 3  $\Rightarrow$   
Skewed to the right  
(tail  $\rightarrow$  right)

- Histogram 2  $\Rightarrow$   
Symmetric (بالمتساوية Data as is)

\* في توزيع التكرار  
 ① Is the above dist. skewed to right? / left? / symmetric?



The longer tail skewed to the right

تراسی

- ⑩ Construct a cumulative freq. dist.
- ⑪ Construct a relative freq. dist.
- ⑫ Construct a cumulative percent freq. dist.

Recall: Freq. distributions  $\rightleftharpoons$  cumulative freq.

Class	Freq.	Rel. Freq.	Per. Freq.	Cum Freq.	Cum Rel Freq.	Cum Per
0-4	4	0.20	20	4	0.20	20
5-9	8	0.40	40	(4+8) 12	(0.20+0.40) 0.60	(20+40) 60
10-14	5	0.25	25	(12+5) 17	(0.60+0.25) 0.85	(60+25) 85
15-19	2	0.10	10	(17+2) 19	(0.85+0.10) 0.95	(85+10) 95
20-24	1	0.05	5	(19+1) 20	(0.95+0.05) 1.0	(95+5) 100
Total	20	1.00	100	مافی دای		

⑩ ⑪ ⑫

Cumulative freq.

- 0-4  $\Rightarrow$  time  $\leq 4$
- 0-9  $\Rightarrow$  time  $\leq 9$
- 0-14  $\Rightarrow$  time  $\leq 14$
- 15-19  $\Rightarrow$  time  $\leq 19$
- 20-24  $\Rightarrow$  time  $\leq 24$

\* الترتیبی Freq. و cumulative  
 هو ترتیبی از Freq. پس بحسب  
 بالترتیب و اینها موجود می باشد.  
 اینها cumulative فاینا بیج  
 این قبل هو اینها اینها.

### 13) Constant a Ogive (use Accu. Relative Freq.)

acumulative Dist.  $\left( \sum F_i \right)$

Cumulative Rel. Freq.

1.00

0.95

0.85

0.60

0.20

Ogive

waiting  
Time

4 9 14 19 24

5 5 5 5 5

عنوان نکتہ سے لے کر  
بسط لے کر اورا Class  
واحد

waiting time | Cum. Dist.

time $\leq 4$	0.20
time $\leq 9$	0.60
time $\leq 14$	0.85
time $\leq 19$	0.95
time $\leq 24$	1.0

Note  $\Rightarrow$  4.5 ?? lbs Class  $\left( \sum F_i \right)$  (Actual class limit)  
 $\Rightarrow$  5 سے 4.6 اور 4 سے 4.5

## 2.4 ⇒ Cross tabulation and Scatter diagram

\* Cross tabulation :-

**Examples**

We want to study the relationship between age and health insurance:

Raw Data

1	27	Yes
2	34	No
⋮	⋮	⋮
1200	45	Yes

NS - 1200 Elements  
 + 1200 observation  
 Sample size 1200  
 variable → 2  
 age health I.

Age → Quantitative → Ratio

H.I. → Qualitative → nominal

• freq. dist. for H.I.

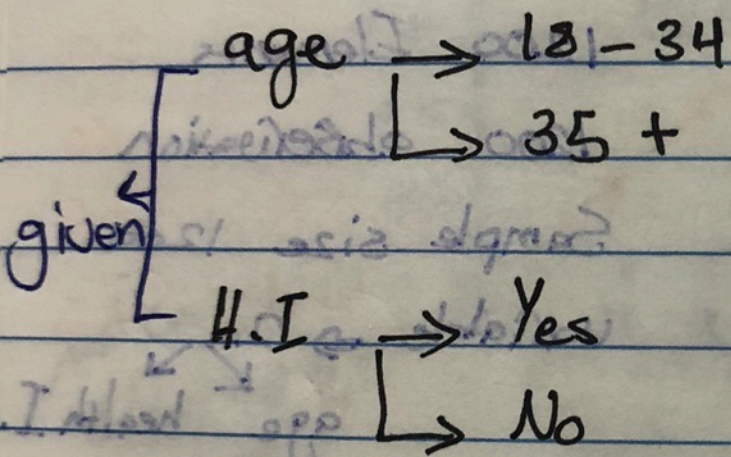
H.I.	freq.
Yes	1020
No	180
Total	1200

Raw Data

• given Raw Data (Joint frequency)

# freq. dist. for age

Age	freq.	
18-34	552	← Young adults
35+	648	← old adults
Total	1200	(Raw data)



## • Cross tabulation

↳ (Car as 2 variables) freq. (row, col)

age / H.I	Yes	No	Total
18-34	450.	102	552
35+	570.	78	648
Total	1020	180	1200

• marginal frequencies  
 • Sample size

- given from <sup>m.f</sup> Raw Data.
- From Raw Data (Joint frequencies)

Def. Row Percentages =  $\frac{\text{Row}}{\text{Total Row}} \times 100$  (نسب أسطر)

age	Yes	No	Total
18-34	$\frac{450}{550} \times 100$ = 81.52	$\frac{102}{552} \times 100$ = 18.48	100
35+	$\frac{570}{648} \times 100$ = 87.96	$\frac{78}{648} \times 100$ = 12.04	100

The higher their age, the higher the Percentage with H.I.

(الناس الكبار هم التي ما نسبت أكثر)

(Row Percentages)

← المجموع فيها يكون يساوي أسطر مابعد الأعمدة

Def. Column Percentages =  $\frac{\text{Column}}{\text{Column Total}} \times 100$  (نسب أعمدة)

age	Yes	No	Total
18-34	$\frac{450}{1020} \times 100$ = 44.12	$\frac{102}{180} \times 100$ = 56.67	100
35+	$\frac{570}{1020} \times 100$ = 55.88	$\frac{78}{180} \times 100$ = 43.33	100
Total	100	100	

the higher the percentage with H.I., The larger the age.

(كل ما زاد العمر زاد عدد المؤسسين)

## Conclusions:-

The higher the age, the higher the percentage of people with H.I, and

Vice versa.

← يعني كلما تزايد العمر بزيادة عدد المرضى وكل ما يقل العمر يقل عدد المرضى

اذا يكون كما يلي على cross tabulation

Raw Data  
↓

Cross tabulation

Row Percentage

Column Percentage

conclusion

conclusion

2 variables → Tabular Presentations

# \* Scatter Diagrams and Trendlines :-

• Example:

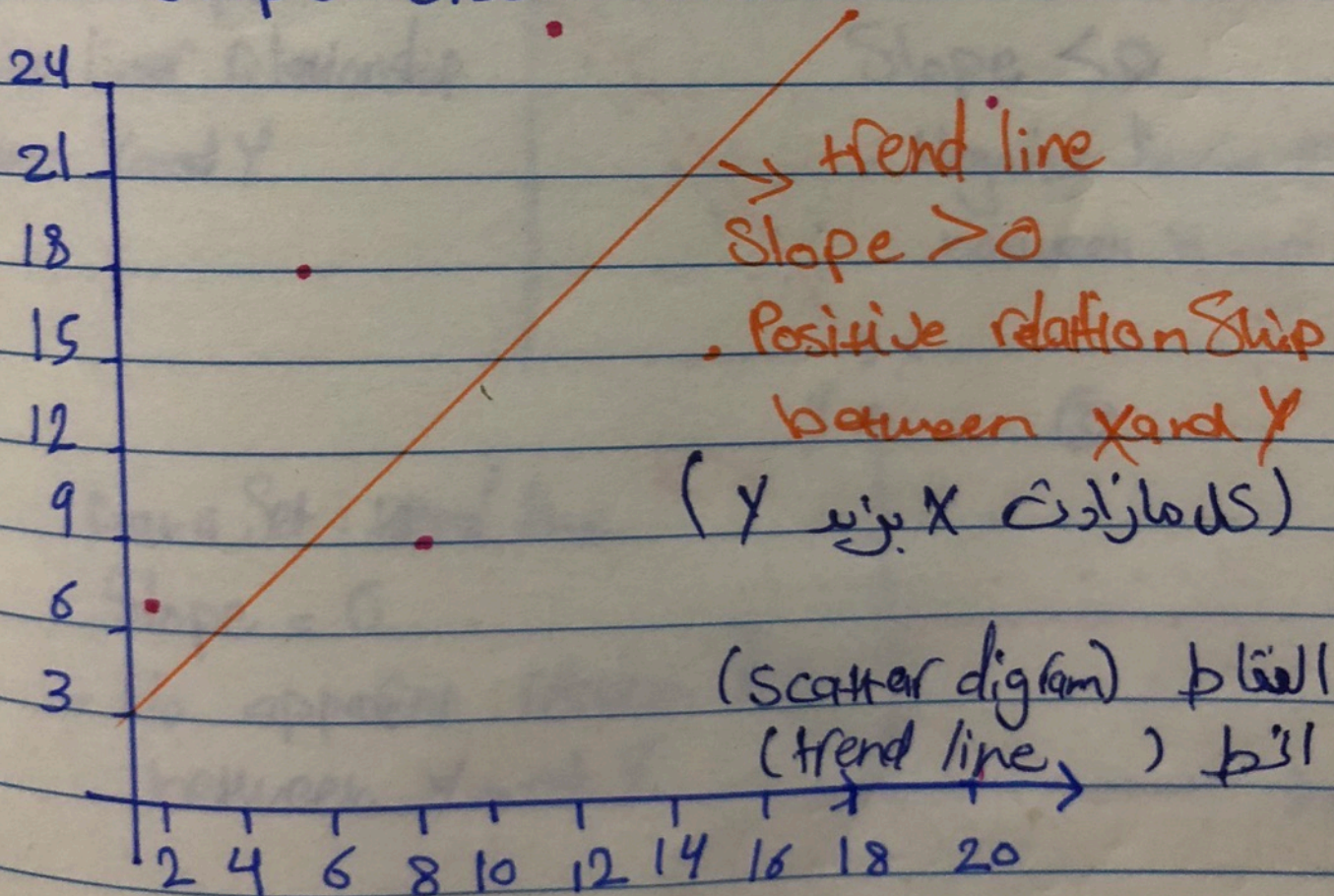
X	2	6	9	13	20
Y	7	18	9	26	23

⇒ 2 variables (X, Y)

5 observations

5 elements

Sample size  $n = 5$

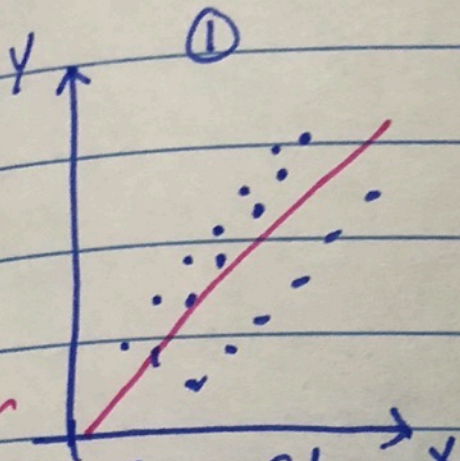


Scatter diagram

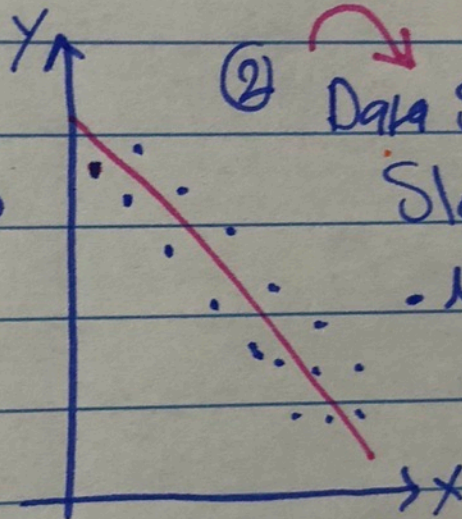
رسم



# Trend lines

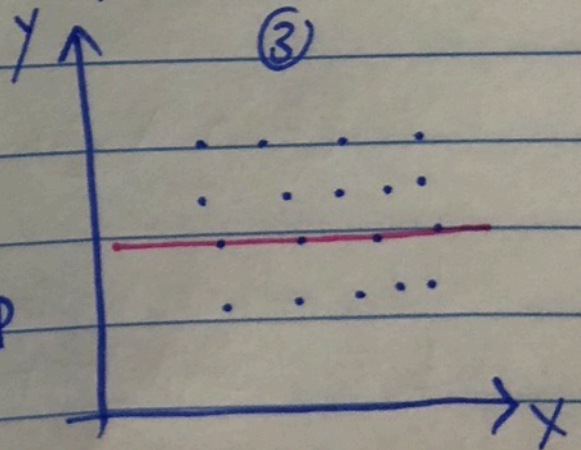


→ Data Set: Slope  $> 0$   
Trend line Slope  $> 0$   
• Positive linear relationship between X and Y



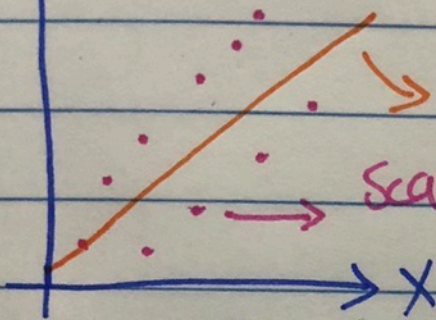
→ Data Set ②: Trend line Slope  $< 0$   
• Negative linear relationship between X and Y

← Data Set: trend line Slope = 0  
• No apparent relationship between X and Y



Ex.

Y ↑



→ trend line

→ scatter diagram

Trend line: Slope  $> 0$

$\Rightarrow \{ X \uparrow \Rightarrow Y \downarrow \}$

↓  $\downarrow$   $\downarrow$   $\downarrow$

Data

↓  
Scatter diagram

↓  
Trend line

positive relationship

negative relationship

no apparent relationship.

Scatter diagram: Presentation, description Statistics 2 variables.

Done 😊😊