

STAT2361

ملاحظات المحاضرات

من إعداد موقع

BZU-HUB



Chapter 2

ملاحظة: المحاضرات من شرح الدكتور محمد مضية

Chapter two

Descriptive Statistics

Tabular and Graphical Methods

2.1 : How to organize (Summarize) Qualitative Data ?

- Table $\equiv$ Frequency Distribution
- Graphs $\equiv$ Bar Graph , Pie Chart

• Frequency Distribution

اسم الفئة ← Class (category)	Frequency # of element in each class



* Example: Drinks a Student purchase in BZU

Water	Coffee	Coffee
Coffee	tea	XL

Raw Data → Organize

Drink	Frequency	Relative Frequency	
		R.F	%
Soft Drinks	7	$\frac{7}{40} = 0.175$	17.5%
Hot Drinks	12	$\frac{12}{40} = 0.3$	30%
Energy Drinks	3	$\frac{3}{40}$	7.5%
Water	10	$\frac{10}{40}$	25%
Juice	8	$\frac{8}{40}$	20%
Total	40	$\frac{40}{40} = 1$	100%



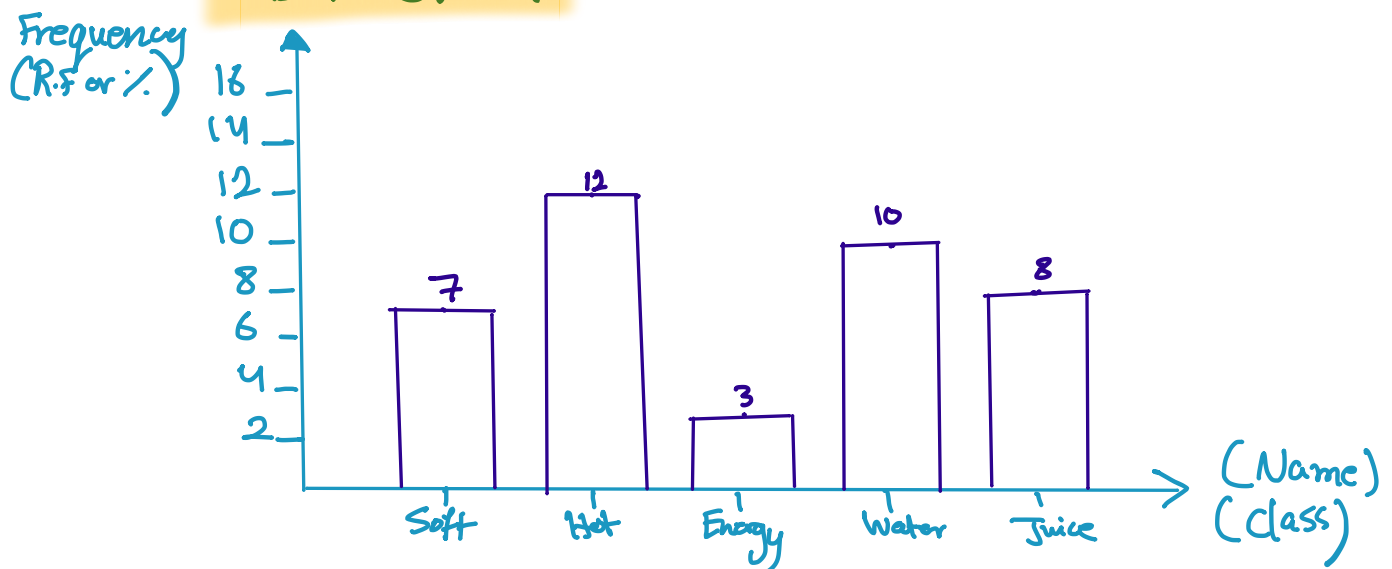
- Variable Drink purchase $\equiv$ Qualitative (5 classes (categories))

- # of elements = 40

- Relative Frequency (R.F) = $\frac{\text{Class Frequency}}{\text{Total Frequency}}$

$$\% = (R.F) \times (100\%)$$

- Bar Chart:



• Pie Chart:

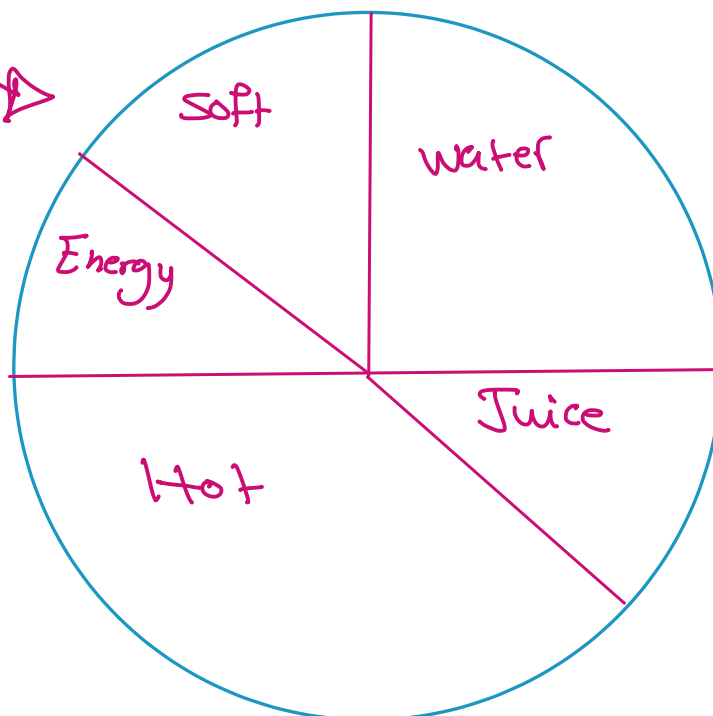
angel ← Setor يكون

$$\begin{aligned}
 &\text{Total Frequency (40)} \quad \swarrow \quad 360^\circ \\
 &\text{Class Frequency} \quad \searrow \quad x \\
 &\rightarrow x^\circ = \frac{\text{Class Frequency}}{\text{Total Frequency}} * 360^\circ \\
 &\quad = (R.F) (360^\circ)
 \end{aligned}$$

- Soft = $\frac{7}{40} * 360^\circ = 63^\circ$
- Hot = $\frac{12}{40} * 360^\circ = 108^\circ$
- Energy = $\frac{3}{40} * 360^\circ = 27^\circ$
- Water = $\frac{10}{40} * 360^\circ = 90^\circ$
- Juice = $\frac{8}{40} * 360^\circ = 72^\circ$



الرسمه تفریجه



● Scales of Measurements

• Scale (Drink) $\equiv$ Nominal

Qualitative

Nominal
↳ Drinks

ordinal
↳ Rank
Evaluation



Quantitative

Interval
↳ Difference
Ex: (100, 50) $\Rightarrow$ 50

Ratio
↳ $\frac{100}{50} = 2$ time

? Ratio لا، Interval لا كيف فيزي في ال

↳ The difference come from their ability to dip below zero

* Interval Scales

↳ Hold no true zero and can represent values below zero

• For example:

You can measure temperatures below zero°

* Ratio Variables

↳ never fall below zero

• For example:

Height and weight measure from zero and above, but never fall down it.

7. Leverock's Waterfront Steakhouse in Maderia Beach, Florida, uses a questionnaire to ask customers how they rate the server, food quality, cocktails, prices, and atmosphere at the restaurant. Each characteristic is rated on a scale of outstanding (O), very good (V), good (G), average (A), and poor (P). Use descriptive statistics to summarize the following data collected on food quality. What is your feeling about the food quality ratings at the restaurant?

G V O V G A O V O G O V A
 V O O P V O G A O O O G O V A
 V A G O V P V O O O G A G O V A
 O G A O V O G V V G A G O V A

Food Quality	Frequency
outstanding (O)	19
Very Good (V)	13
good (G)	10
average (A)	6
Poor (P)	2
Total	50



2.2 How to organize (Summarize) Quantitative Data?

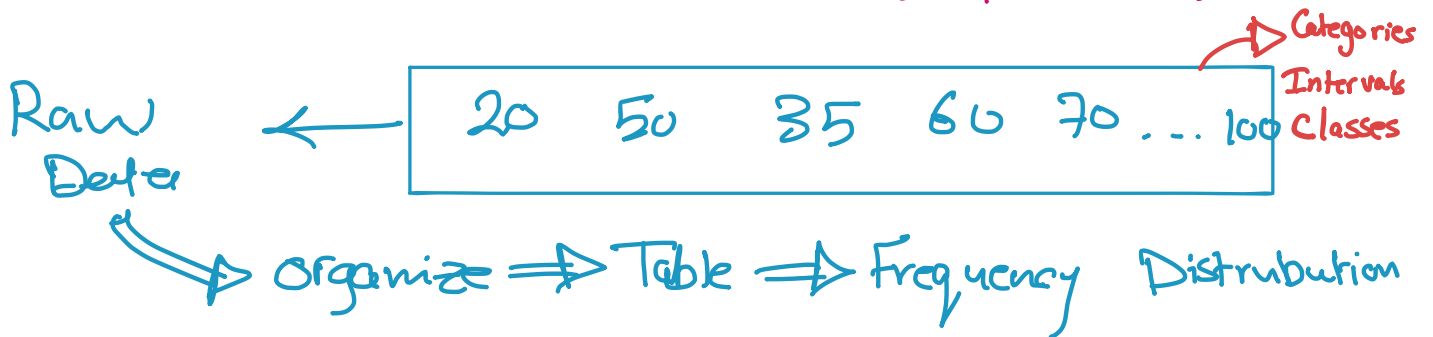
- Table $\equiv$ Frequency Distribution
- Graphs $\equiv$ Histogram, Ogive

* Example:-

Daily Expenditure of a BZU Student

↳ Quantitative, Ratio

Ratio ← Income لأنّه يقدّر شي فكمية انه
أشرف (-) + منظم الأعيان يكونه ال



• Ex: $[20-35]$ فترة ملاحظة ، بين 20 ، 35 ما عليه

* Classes (Intervals)

- Class limits (Lower $\equiv L$)
(Upper $\equiv U$)

- Equal widths

- Class width (w)
- # of classes (n)

- Disjoint

- All exclusive

* علاقة عكسية ، كل ما قلنا ال Width
بزيد عدد ال Classes



*To Construct a Frequency Distribution

- Find the Class width
- Choose a number of classes
- Approximated Class width



$$W = \frac{\text{Range}}{\text{\# of Classes}} = \frac{\text{largest Value} - \text{Smallest Value}}{\text{\# of Classes}}$$

- Round up to the nearest integer

Applications

15. A doctor's office staff studied the waiting times for patients who arrive at the office with a request for emergency service. The following data with waiting times in minutes were collected over a one-month period.

2 5 10 12 4 4 5 17 11 8 9 8 12 21 6 8 7 13 18 3

Use classes of 0-4, 5-9, and so on in the following: 5 classes (صل الدكتور غير)

- Show the frequency distribution.
- Show the relative frequency distribution.
- Show the cumulative frequency distribution.
- Show the cumulative relative frequency distribution.
- What proportion of patients needing emergency service wait 9 minutes or less?

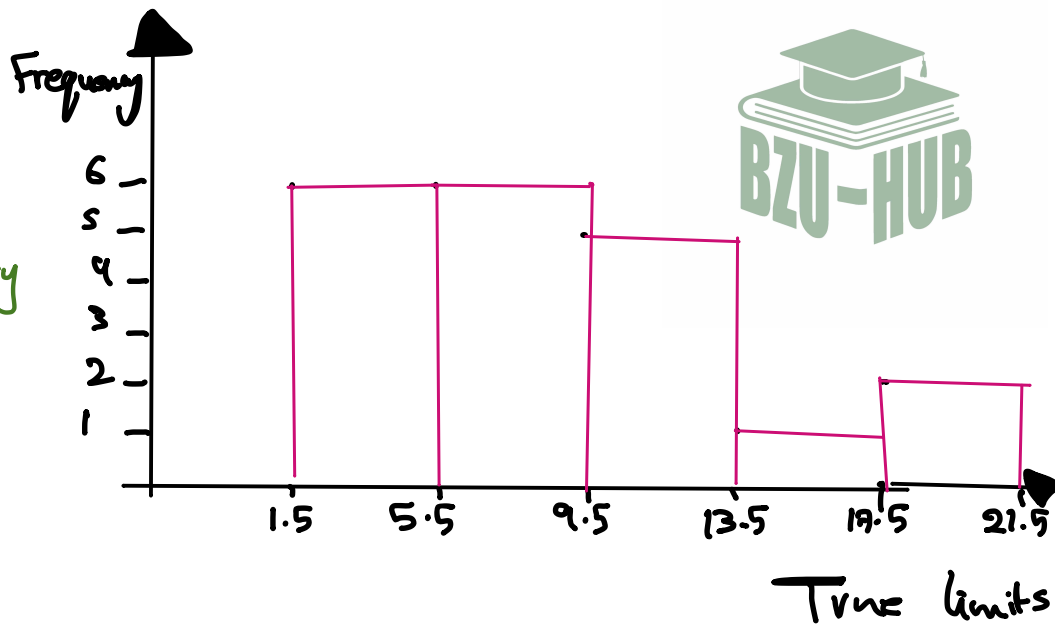
$$W = \frac{21 - 2}{5} = \frac{19}{5} = 3.8 = \boxed{4} + \frac{1}{2}$$

Class	Frequency	Relative Frequency	True Limits	Upper limit
2- <u>5</u>	6	$6 \div 20 = 30\%$	1.5 - 5.5	5
5- <u>9</u>	6	30%	5.5 - 9.5	9
10- <u>13</u>	5	25%	9.5 - 13.5	13
14- <u>17</u>	1	5%	13.5 - 17.5	17
18- <u>21</u>	2	10%	17.5 - 21.5	21
Total	20	100%		

Cumulative Frequency (C.F)	Relative (C.F)
6	6/20
$6 + 6 = 12$	12/20
$12 + 5 = 17$	17/20
$17 + 1 = 18$	18/20
$18 + 2 = 20$	20/20

1 Histogram

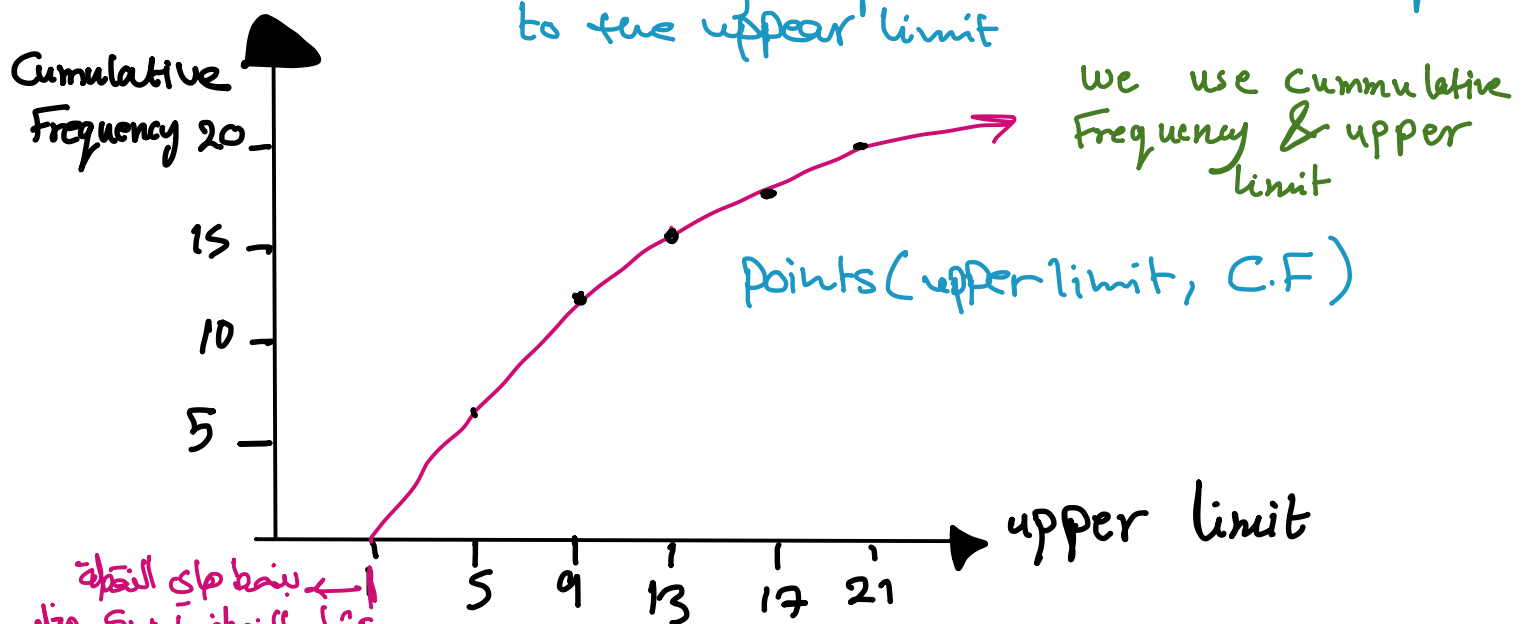
We use Frequency
&
True limits



2 Ogive

Cumulative Frequency distribution

→ how many frequencies less than or equal to the upper limit



Example: Consider the following frequency dist.

Class	Frequency	mid point $\left(\frac{\text{upper} + \text{lower}}{2} \right)$
6-10	10	$(6+10)/2 = 8$
11-15	20	13
16-20	18	18
21-25	12	23
26-30	5	28
Total	65	

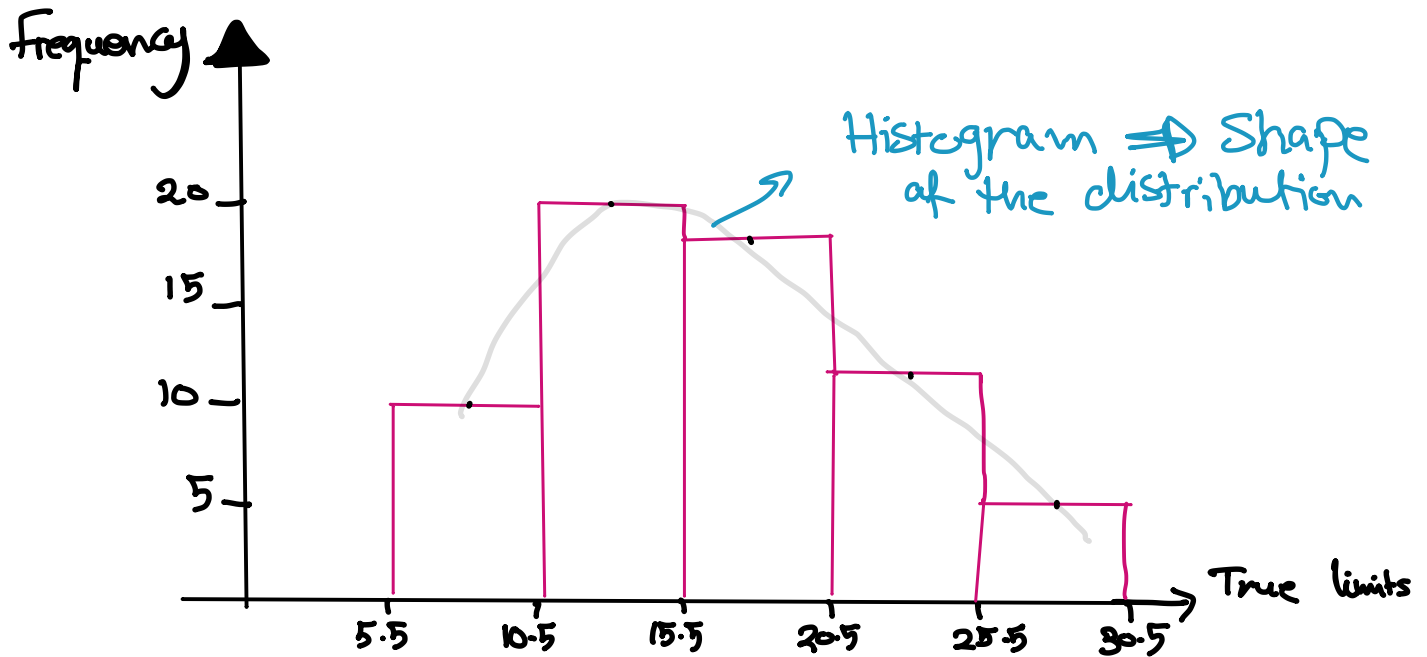


3 Percentage of Data greater than 15

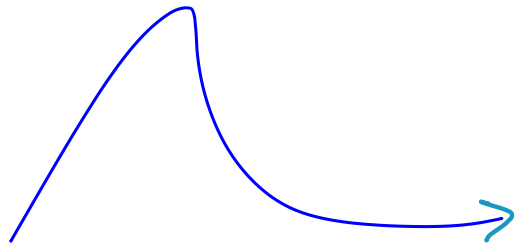
$$\frac{18+12+5}{65} = 53.8$$

لو مسا أو equal لا يجوز
كم مرة ال 15 تكررت
(Cannot be determine)

4 Construct = Histogram



Shape of the distribution

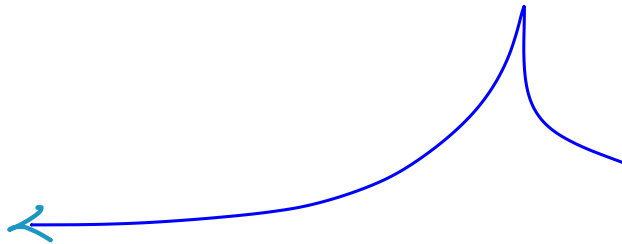


• Right Skewness

• Skewness (cibzio)

Skewness = +

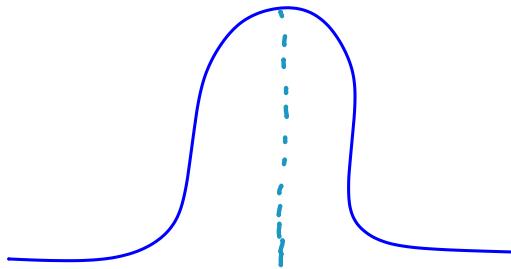
mean > median



• Left Skewness

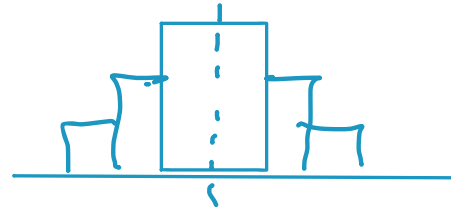
Skewness = -

mean < median



• Symmetric

mean = median = mode



Skewness = Zero



2.4 Crosstabulation and Scatter Diagrams

Qualitative and Quantitative

Quantitative (numeric)

• جدول به علامت بنقد و نکتتست إذا في علاقة بين 2 Variables

* **Crosstabulation**: is a tabular summary of data for 2 variables.

* Example:



TABLE 2.9 QUALITY RATING AND MEAL PRICE FOR 300 LOS ANGELES RESTAURANTS

Restaurant	Quality Rating	Meal Price (\$)
1	Good	18
2	Very Good	22
3	Good	28
4	Excellent	38
5	Very Good	33
6	Good	28
7	Very Good	19
8	Very Good	11
9	Very Good	23
10	Good	13
.	.	.
.	.	.

□ Construct a Crosstabulation for the data

- Quality Rating: is a qualitative variable with categories: good, very good, Excellent
- Meal price: is a quantitative variable that ranges from \$10 to \$49

Quality Rating	Meal Price				Total
	\$10-19	\$20-29	\$30-39	\$40-49	
Good	42	40	2	0	84
Very Good	34	64	46	6	150
Excellent	2	14	28	22	66
Total	78	118	76	28	300

row $\Rightarrow$ Column $\Downarrow$ Crosstabulation
هم نوع کد رقم بالجدول شو يعني وعلى شو بيل

b) Develop a relative and percent frequency distribution for quality rating

Quality Rating	Relative Frequency	Percent Frequency
Good	$\frac{84}{300}$.28	28
Very Good	$\frac{150}{300}$.50	50
Excellent	$\frac{66}{300}$.22	22
Total	1.00	100

• 28% of the restaurant rating good

c) Develop a relative and percent frequency distribution for meal price

Meal Price	Relative Frequency	Percent Frequency
\$10-19	$\frac{78}{300}$.26	26
\$20-29	.39	39
\$30-39	.25	25
\$40-49	.09	9
Total	1.00	100

• 26% of the meal price are in the lowest class \$10-19

*To know the relationship between 2 variables within Crosstabulation, we convert the entries into row percentages or column percentages

d) Construct a row percentages for each quality rating category

TABLE 2.11 ROW PERCENTAGES FOR EACH QUALITY RATING CATEGORY

Quality Rating	Meal Price				Total
	\$10-19	\$20-29	\$30-39	\$40-49	
Good	$100 * \frac{42}{84}$ 50.0	47.6	2.4	0.0	100
Very Good	22.7	42.7	30.6	4.0	100
Excellent	3.0	21.2	42.4	33.4	100

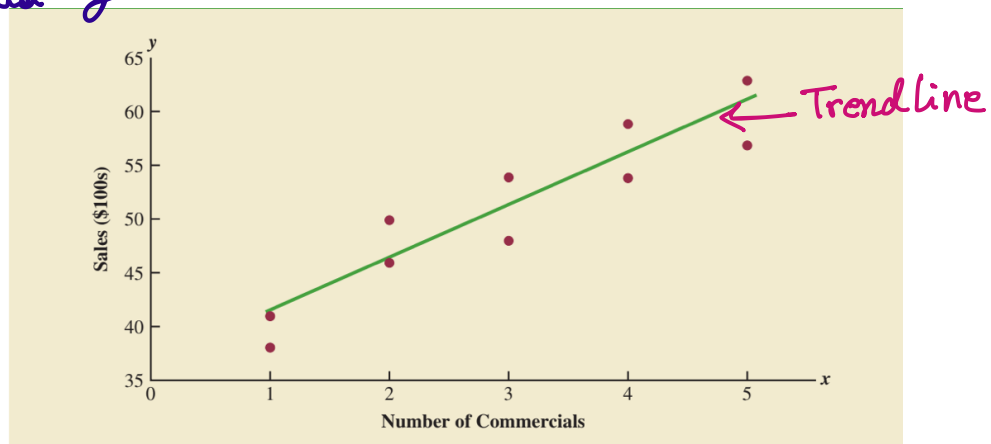
Scatter Diagram and Trendline

- A scatter diagram: is a graphical presentation of the relationship between two quantitative variables.
- Trendline: is a line that provides an approximation of the relationship.

TABLE 2.12 SAMPLE DATA FOR THE STEREO AND SOUND EQUIPMENT STORE

Week	Number of Commercials x	Sales (\$100s) y
1	2	50
2	5	57
3	1	41
4	3	54
5	4	54
6	1	38
7	5	63
8	3	48
9	4	59
10	2	46

- a) Develop a scatter diagram for the relationship between x and y



- b) What is the relationship between x and y ?
The scatter diagram indicates a positive relationship.

- c) Is the relation perfect?

No, because all the points are not on the trend line

