Biostatistics

Lecture 1

Prepared by Dr. Ibrahim AL-Jaafari

What is Statistics ?

- ***Statistics*** is a group of methods used to collect, analyze, present, and interpret data and to make decisions.

Examples of Some Statistics

- The number of educated people in AL-Baha City.
- The number of homeless in the United States of America.
- The number of cars in Saudi Arabia.
- The cost of health care services in Saudi Arabia

What is Bio-Statistics ?

- ***Bio-Statistics*** is a group of methods used to collect, analyze, present, and interpret of biological, medical, and public health data to make decisions.

Examples of Some Biostatistics Data

- The number of diagnostic tests in the hospital
- The number of hepatitis cases in the region
- The number of cancer deaths in Saudi Arabia.
- The number of infectious diseases in Saudi Arabia.

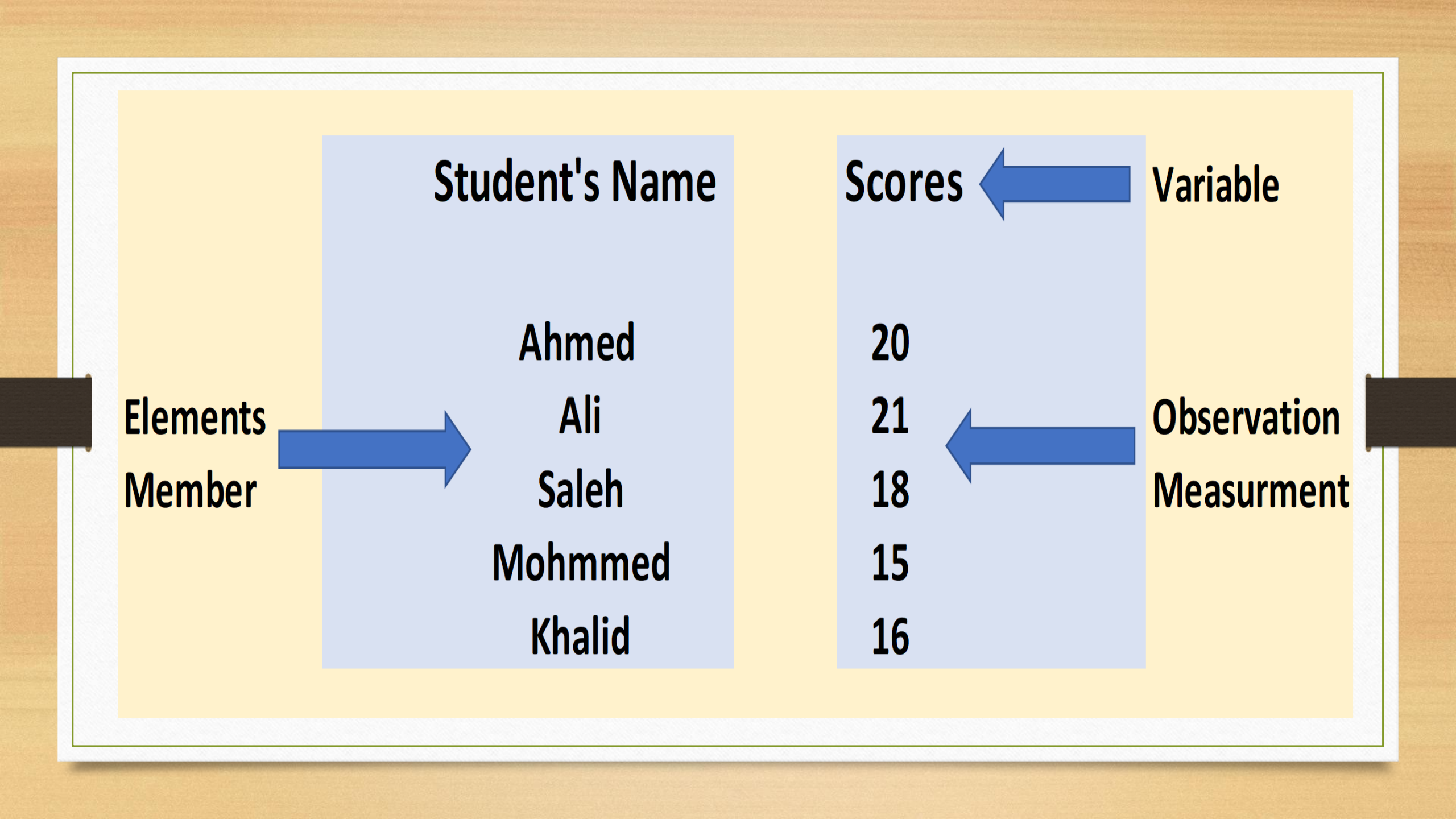
What are The Types of Statistics ?

Descriptive Statistics

Descriptive statistics consists of methods for organizing, displaying, and describing data by using tables, graphs, and summary measures

Example of Descriptive Statistics

- **The test scores of students enrolled in a biostatistics class.**
- **In statistical terminology, the whole set of numbers that represents the scores of students is called a data set.**
- **The name of each student is called an element, and the score of each student is called an observation.**


Student's Name
Scores
Variable
Ahmed
Ali
Saleh
Mohmmed
Khalid
20
21
18
15
16
Elements
Member
Observation
Measurment

In statistics, the collection of all elements of interest is called a **population**.

E.g. The number of all diabetic patients in AL-Baha City.

The selection of a few elements from this population is called a **sample**

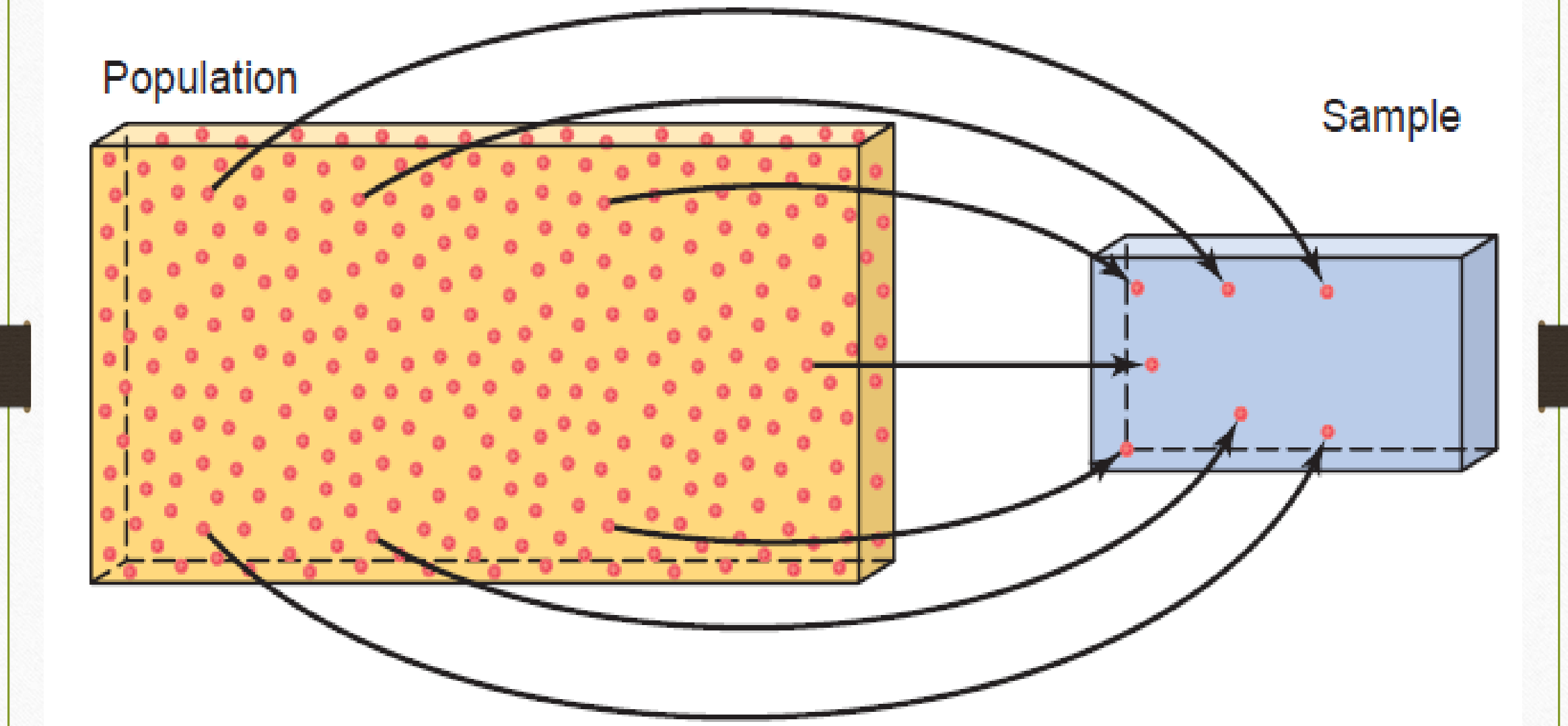
E.g. The number of diabetic patients in a primary health care (A)

Population / consists of all elements, individuals, items or objects whose characteristics are being studied

Sample / A portion of the population selected for study

Census / A survey that includes every member of the population.

Sample Survey / The technique of collecting information from a portion of population


Population
Sample

What are The Types of Statistics ?

Inferential Statistics

Inferential statistics consists of methods that use sample results to help make decisions or predictions about a population.

Example of Inferential Statistics

- The starting salary of a college graduate. We may select a sample of 1000 recent college graduates.
- Look at to the starting salaries to make a decision based on this information.

What is The Meaning of Variable ?

- It is a characteristic under study that assumes different values (scores of students in biostatistics) for different elements (students in college of applied health sciences)

What are The Types of Variables ?

- **Quantitative variables** / It can be measured numerically

Types of quantitative variables

1- Continuous variable / Any numerical values over a certain interval or intervals (Should has a decimal)

E.g. (Length, age, height, weight, time, and blood cholesterol)

What are The Types of Variables ?

- 2- Discrete variable / Only a certain numerical values with no intermediate values (Never has a decimal).

- E.g. (Number of cars, houses, and accidents)

What are The Types of Variables ?

- **Qualitative variables** / It can not be measured numerically

Types of qualitative variables

1- Nominal variable / It is another name for a categorical variable

E.g. (Occupation, type of disease, type of cars)

What are The Types of Variables ?

2- Ordinal variables / They have ordered categories

E.g. (Severity of disease, socioeconomic class, level of education)

Low – Middle - High

Mild – Moderate - Severe

What are The Types of Variables ?

3- Binary variables / They have only two categories

E.g. (Gender : Male and Female)

Question answer (True, False)

Question (Yes, No)

ORGANIZING AND GRAPHING QUALITATIVE DATA

- Frequency Distributions
- Relative Frequency and Percentage Distributions
- Graphical Presentation of Qualitative Data
  - Bar Graphs
  - Pie Charts

Frequency Distributions

Definition

A ***frequency distribution*** for qualitative data lists all categories and the number of elements that belong to each of the categories.

Frequency Table

Variable → Type of College	Number of Students ← Frequency Column
Public Health	200
Category → Nursing	300 ← Frequency
Pharmacy	400

Sum = 900

Example for Frequency Table

- A sample of 30 employees from Al-Baha city was selected, and these employees were asked for their monthly salary. The responses of these employees are recorded next where ***high income***, ***middle income***, and ***low income***.

Construct a frequency distribution table for these data ?

High	Low	High	Low	Low	Middle
Low	Middle	High	Middle	Middle	High
Middle	High	Middle	Middle	Middle	Middle
High	Low	High	Middle	Low	Middle
High	Low	Low	High	Middle	High

Solution for Frequency Table

Income	Tally	Frequency (f)
High	𝍸 𝍸	10
Middle	𝍸 𝍸 ||	12
Low	𝍸 |||	8
		Sum = 30

Relative Frequency and Percentage Distributions

Calculating Relative Frequency of a Category

$$\text{Relative frequency of a category} = \frac{\text{Frequency of that category}}{\text{Sum of all frequencies}}$$

Relative Frequency and Percentage Distributions

Calculating Percentage

Percentage = (Relative frequency) * 100

Solution

Income	Relative Frequency	Percentage
High	10/30 = 0.333	0.333(100) = 33.3
Middle	12/30 = 0.40	0.40(100) = 40.0
Low	8/30 = 0.267	0.267(100) = 26.7
	Sum = 1.00	**Sum = 100**

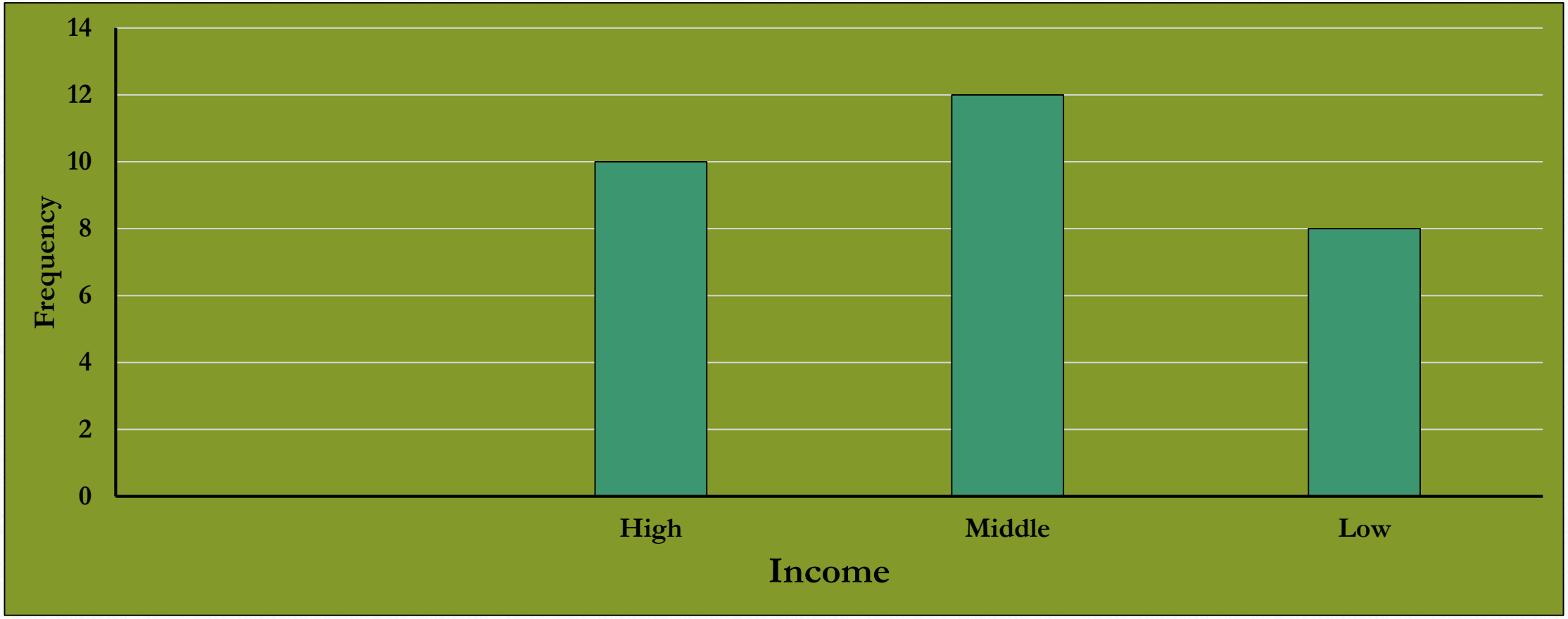
Graphical Presentation of Qualitative Data

Bar Graph – Bar Chart

Definition

A graph made of bars whose heights represent the frequencies of respective categories is called a ***bar graph***.

Bar Graph – Bar Chart



Graphical Presentation of Qualitative Data

Pie Graph - Pie Chart

Definition

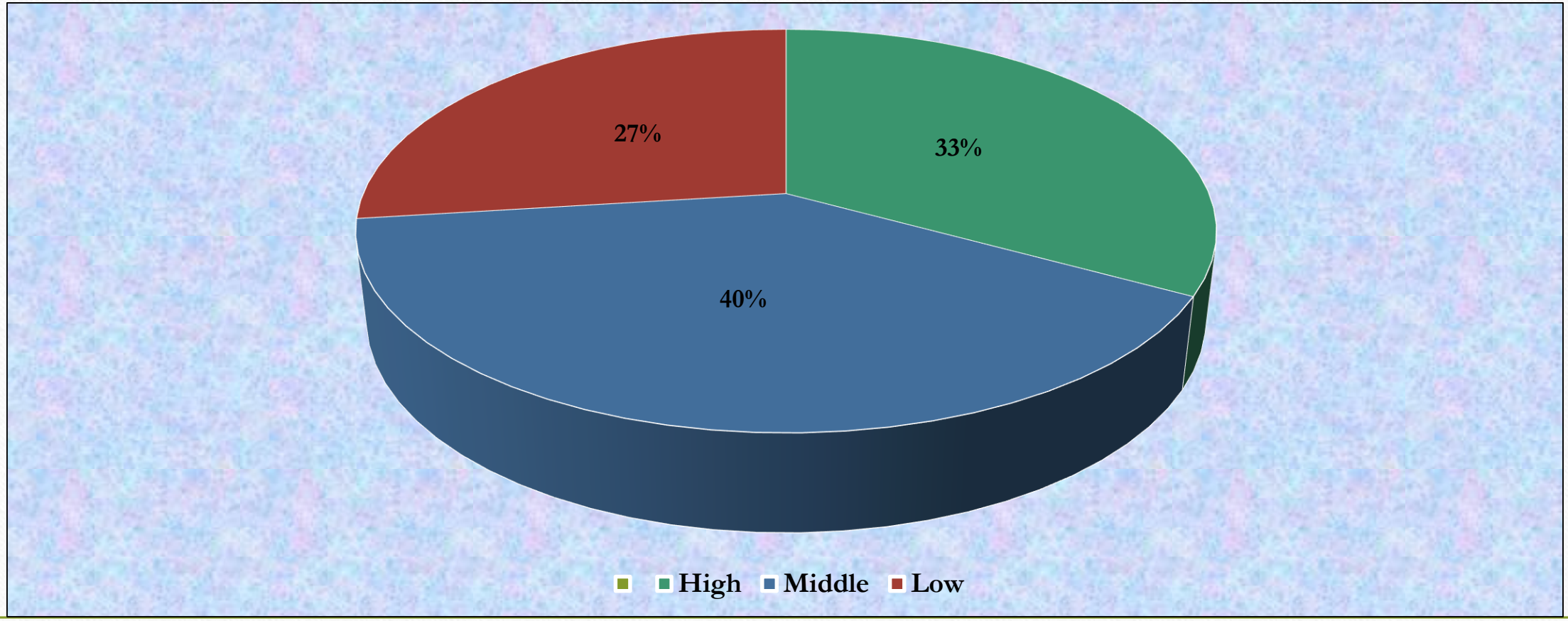
A circle divided into portions that represent the relative frequencies or percentages of a population or a sample belonging to different categories is called a ***pie chart***.

Calculating Angle Sizes for the Pie Chart

Pie Graph - Pie Chart

Income	Relative Frequency	Angle Size
High	10/30 = 0.333	360 (0.333) = 119.9
Middle	12/30 = 0.40	360 (0.40) = 144.0
Low	8/30 = 0.267	360 (0.267) = 96.1
	Sum = 1.00	**Sum = 360**

Pie Graph - Pie Chart



Organizing and Displaying Data
Quantitative Data

- Frequency Distributions
- Constructing Frequency Distribution Tables
- Relative and Percentage Distributions
- Graphing Grouped Data
  - Histograms
  - Polygons

Frequency Distributions

Definition

A ***frequency distribution*** for quantitative data lists all the classes and the number of values that belong to each class. Data presented in the form of a frequency distribution are called ***grouped data***.

Class Boundary

Definition

The ***class boundary*** is given by the midpoint of the upper limit of one class and the lower limit of the next class.

Class Boundary

Age	Frequency	Class Boundaries
0 - 9	20	-0.5 to 9.5
10 - 19	15	9.5 to 19.5
20 - 29	25	19.5 to 29.5
30 - 39	20	29.5 to 39.5
40 - 49	20	39.5 to 49.5
50 - 59	30	49.5 to 59.5
60 - 69	20	59.5 to 69.5

- Class boundary for first class (0-9)

= 9 + 10 / 2 = 9.5

- Class boundary for first class (10-19)

= 19 + 20 / 2 = 19.5

- Class boundary for first class (20-29)

= 29 + 30 / 2 = 29.5

Class Width – Class Size

- The difference between the two boundaries of the class

Class width = Upper Boundary – Lower Boundary

Class Width – Class Size

Age	Frequency	Class Boundaries	Class Width
0 - 9	20	-0.5 to 9.5	10
10 - 19	15	9.5 to 19.5	10
20 - 29	25	19.5 to 29.5	10
30 - 39	20	29.5 to 39.5	10
40 - 49	20	39.5 to 49.5	10
50 - 59	30	49.5 to 59.5	10
60 - 69	20	59.5 to 69.5	10

Class width for the first class (0-9)

$= 9.5 - 0.5 = 10$

$=19.5 - 9.5 = 10$

$= 29.5 - 19.5 = 10$

$= 39.5 - 29.5 = 10$

$= 49.5 - 39.5 = 10$

Class Midpoint – Mark

Calculating Class Midpoint or Mark

$$\text{Class Midpoint or Mark} = \frac{\textit{Lower limit} + \textit{Upper Limit}}{2}$$

Age	Frequency	Class Boundaries	Class Width	Class Midpoint
0 - 9	20	-0.5 to 9.5	10	0 + 9 ÷ 2 = 4.5
10 - 19	15	9.5 to 19.5	10	10 + 19 ÷ 2 = 14.5
20 - 29	25	19.5 to 29.5	10	49 ÷ 2 = 24.5
30 - 39	20	29.5 to 39.5	10	69 ÷ 2 = 34.5
40 - 49	20	39.5 to 49.5	10	89 ÷ 2 = 44.5
50 - 59	30	49.5 to 59.5	10	109 ÷ 2 = 54.5
60 - 69	20	59.5 to 69.5	10	129 ÷ 2 = 64.5

Class Midpoint
Class Mark

- Class Midpoint
- 0 + 9 / 2 = 4.5
- 10 + 19 / 2 = 14.5
- 49 / 2 = 24.5
- 69 / 2 = 34.5

Relative Frequency and Percentage Distributions

What is cumulative frequency distribution ??

The total number of values that fall below the upper boundary of each class

What is cumulative relative frequency ??

Cumulative frequency ÷ Total observations in the data set

What is cumulative percentage ??

Cumulative relative frequency × 100

Relative Frequency and Percentage Distributions

Age	Number of Patients	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percentage
0 - 9	20	20	20÷130 = 0.153	15.3 %
10 - 19	15	20+15 =35	35÷130 = 0.269	26.9 %
20 - 29	25	20+15+25=60	60÷130 = 0.461	46.1 %
30 - 39	20	20+15+25+20=80	80÷130 = 0.615	61.5 %
40 - 49	20	20+15+25+20+20=100	100÷130 = 0.769	76.9 %
50 - 59	30	20+15+25+20+20+30= 130	130÷130 = 1	100

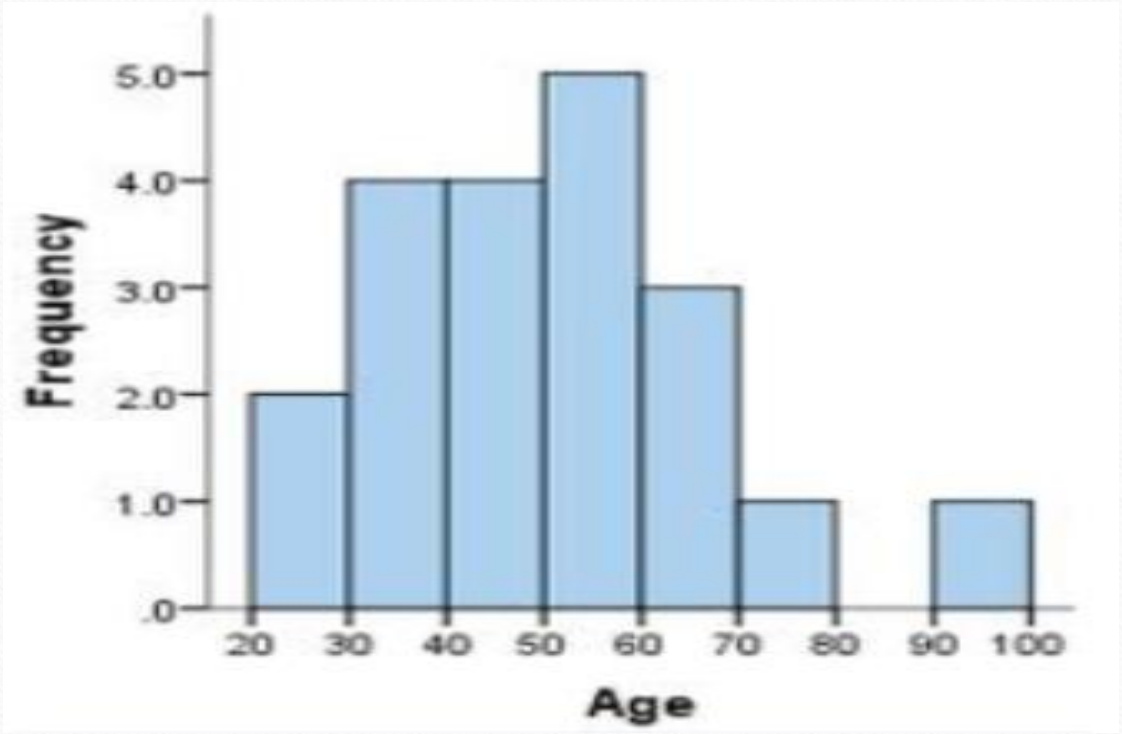
GRAPHING QUANTITATIVE DATA

How to present quantitative data ??

1- Histograms / a graph in which classes are marked on horizontal axis and the frequencies are marked on the vertical axis which represent the height of bars.

In a histogram, the bars are attached to each other.

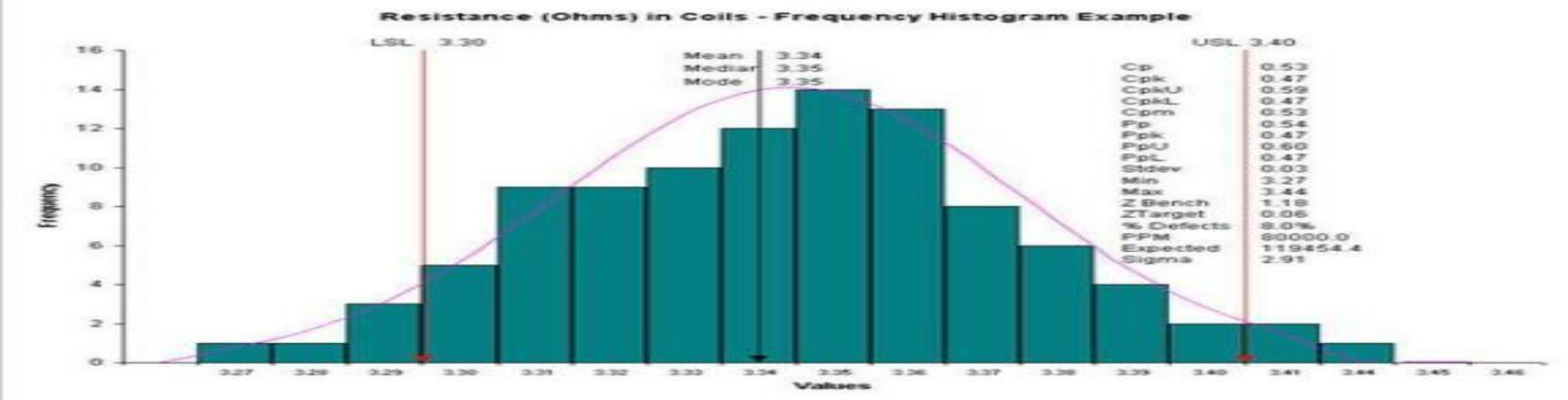
Histograms



Histograms

What are the shapes of histograms ??

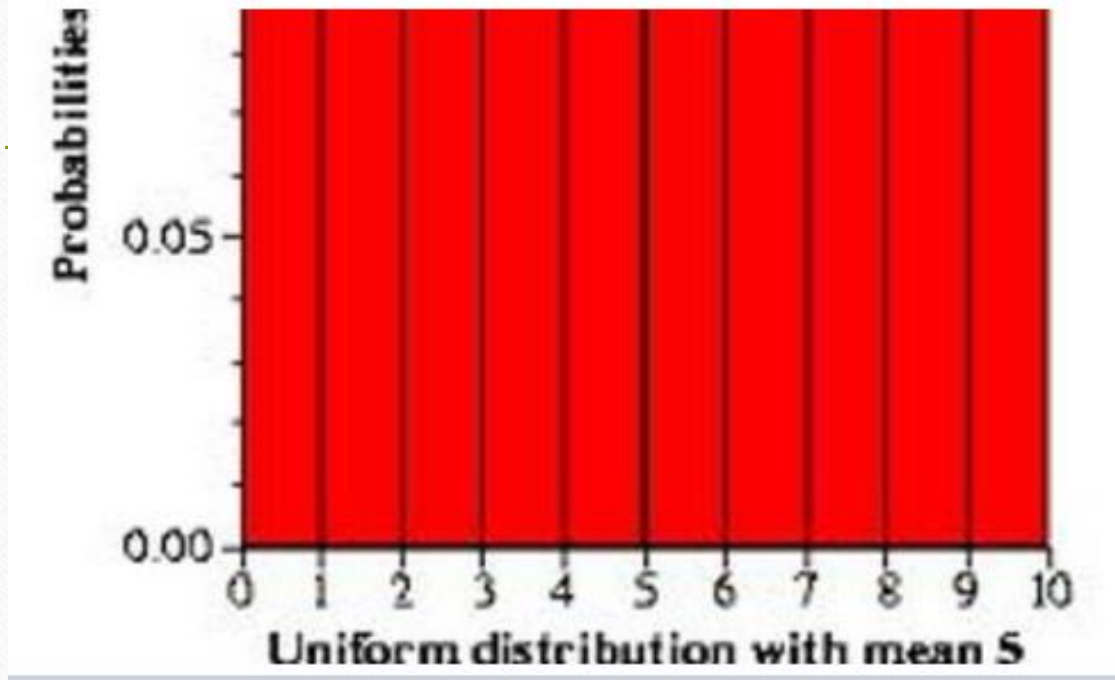
A-) Symmetric histogram / is identical on both sides of its central points.



Histograms

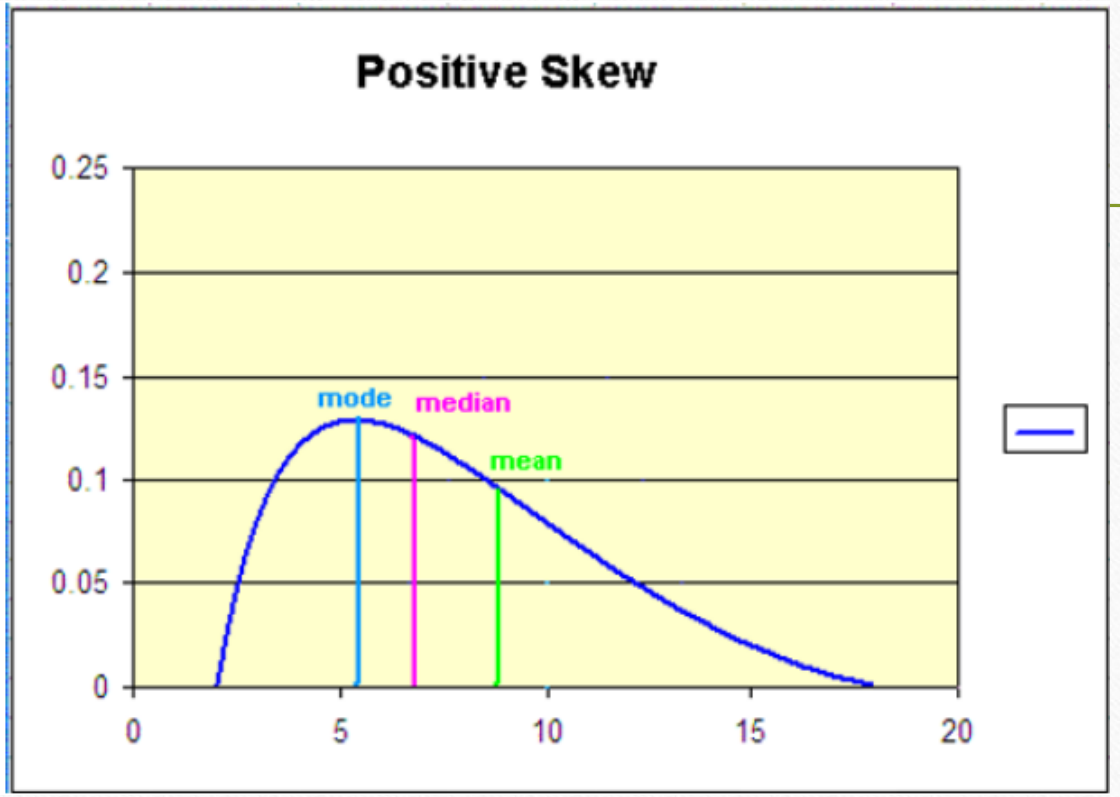
B-) Bimodal symmetric histogram / is identical on both sides of its central points with two modes

Histograms



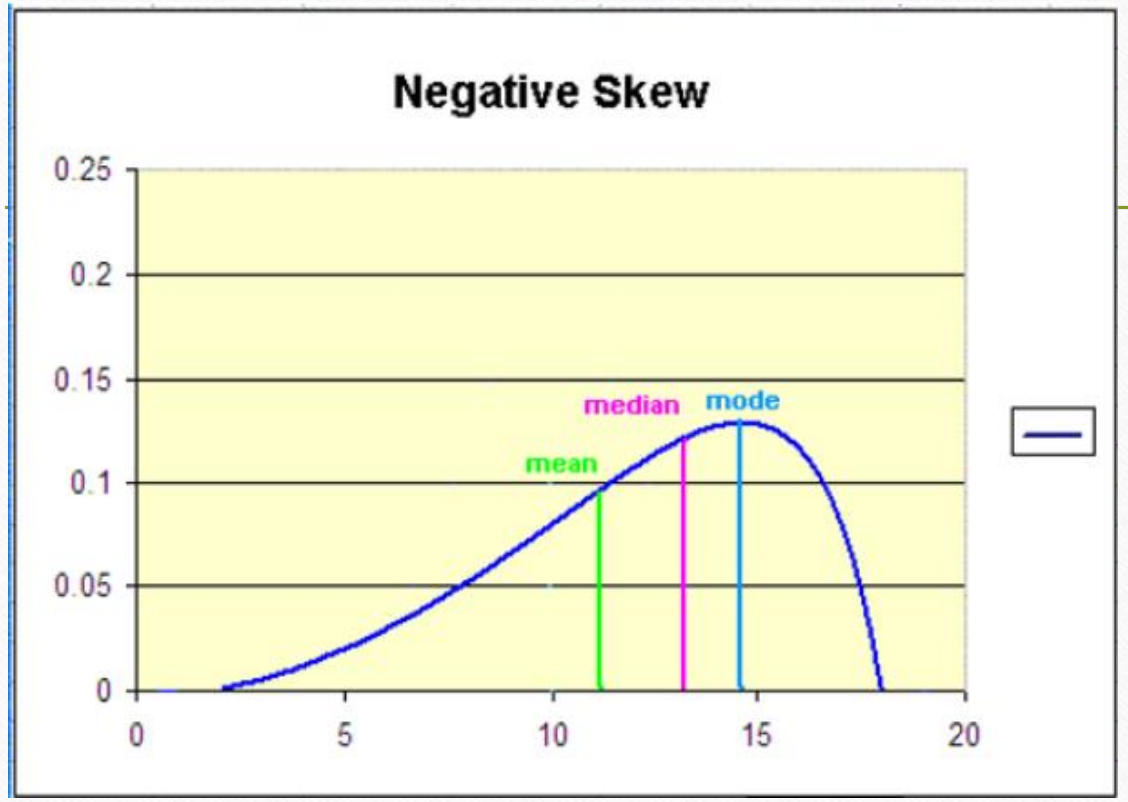
C-) Uniform or rectangular histogram (symmetric) / the frequencies of each class are the same or equal to each other.

Histograms



D-) Skewed to the right histogram (positive skewed) / Most of data is shown in the left side of histogram and the tail on the other right side.

Histograms



E-) Skewed to the left histogram (negative skewed) / Most of data is shown in the right side of histogram and the tail on the other left side.

Polygons

2- Polygons / A graph formed by joining the midpoints of the tops of bars in a histogram with straight lines.

Polygons


No. of Week
20
18
16
14
12
10
8
6
4
2
0
140-150
150-160
160-170
170-180
180-190
190-200
Cost of living index

Stem and Leaf display

3- Stem and Leaf display / each value is divided into two portions -- a stem and a leaf. Then the leaves for each stem are shown separately in a display.

Stem and Leaf display

Construct a stem-and-leaf display for these data

Example / 22 , 26 , 27 , 31 , 33 , 35 , 42 , 44 , 46 , 57 , 58 , 59 , 61 , 63, 64 , 65 ,67

Solution

Stem	Leaf
2	2 6 7
3	1 3 5
4	2 4 6
5	7 8 9
6	1 3 4 5 7

Reference

- Prem S. Mann 1998, Introductory Statistics, 7th edn, New York, USA.

Good Luck for All Students

- Please do not hesitate to contact me if you have any questions.
- Dr. Ibrahim AL-Jaafari
- www.Alghamdi-Biostatistics.com
- Email. Bio-stat@Hotmail.com
- Mobile Number : 0553777925