

Introduction to Statistics

Md. Sabbir Hossain

MNS, BRAC UNIVERSITY

June 9, 2023

Introduction

Statistics

What is Meant by Statistics?

- Statistics is the science of collecting, organizing, presenting, analyzing, and interpreting numerical data for the purpose of assisting in making a more effective decision.
- According to Fisher (1947), the science of statistics is essentially a branch of applied mathematics and may be regarded as mathematics, applied to observational data.
- Croston and Cowden (1948) defined statistics as the subject of collection, presentation and analysis of numerical data.
- As Yule and Kendal (1950) opined, Statistics means quantitative data, which are affected to a marked extent by multiplicity of causes.
- American Heritage Dictionary defines statistics as: “The mathematics of the collection, organization and interpretation of numerical data especially the analysis of population characteristics by inference from sampling.”

Who Uses Statistics?

- Statistical techniques are used extensively by marketing, accounting, quality control, engineers, consumers, professional sports people, hospital administrators, educators, politicians, physicians, etc...

Why Statistics is Studied?

- We face numerical data all the time
 - newspapers, sports magazines, business magazines
- Statistical techniques are used to make decisions that effect our daily lives
- Insurance companies use statistics to set rates for different insurance
- City corporation may want to examine the contamination level of different lake water
- Medical professional assess the performance of a new drug over the existing one
- No matter what is your future line of work, you will make decisions that will involve data

Keywords

- **Population::** Population is the entire set of items from which you draw data for a statistical study.
- **Sample:** A representative part of the population is called sample.



Figure 1: Graphical presentation of the Sample Vs Population

Keywords (Cont....)

- **Parameter:** Parameters are numbers that describe the properties of entire **populations**.
- **Statistic:** Statistic is a measure which describes a **sample**.
- **Example:** For Example, the average income for the United States is a population parameter. Conversely, the average income for a sample drawn from the U.S. is a sample statistic. Both values represent the mean income, but one is a parameter vs a statistic.

Types of Statistics

- **Descriptive Statistics:** Methods of organizing, summarizing, and presenting data in an informative way.
- **Example 1:** A social survey found that 49% of the people knew the name of the first book of the Bible. The statistic 49 describes the number out of every 100 persons who knew the answer.
- **Example 2:** According to Consumer Reports, Whirlpool washing machine owners reported 9 problems per 100 machines during 1995. The statistic 9 describes the number of problems out of every 100 machines.

Types of Statistics (Cont.....)

- **Inferential Statistics:** Taking a sample from a population and making estimates/assumptions about a population, based on a sample.
- **Example 1:** TV networks constantly monitor the popularity of their programs by hiring research organizations to sample the preferences of TV viewers.
- **Example 2:** The accounting department of a large firm will select a sample of the invoices to check for accuracy for all the invoices of the company.

Types of Variables

Variable is a characteristics that can assume any set of prescribed values.

Example: age, height, weight, eye color, total population of a country

- Qualitative Variable (Attribute)
- Quantitative Variable

Types of Variables (Cont....)

- **Qualitative variable** or **Attribute**: the characteristic or variable being studied is non-numeric.
 - **Examples**: gender, religious affiliation, type of automobile owned, eye color.
- **Quantitative variable**: the variable which can be measured and reported numerically.
 - **Examples**: balance in your savings account, time remaining in class, number of children in a family.

Types of Variables (Cont....)

Quantitative variables can be classified as either **discrete** or **continuous**.

- **Discrete variables:** can only assume certain values and there are usually “gaps” between values.
 - **Examples:** the number of bedrooms in a house (1,2,3,..., etc...).
- **Continuous variables:** can assume any value within a specific range.
 - **Examples:** The time it takes to fly from Florida to New York.

Scales of Measurement

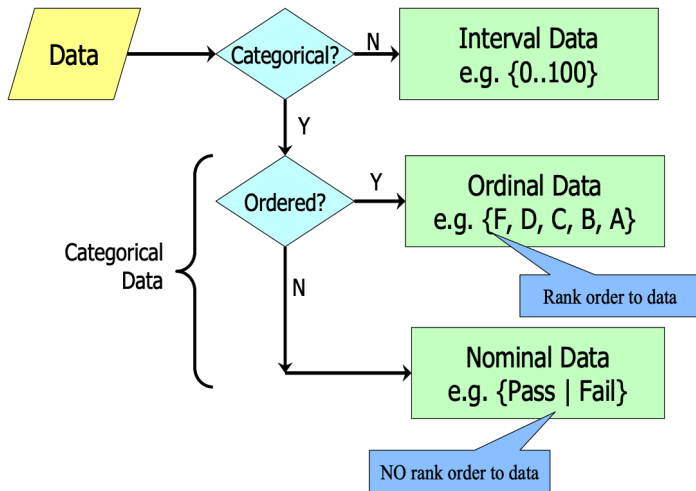
- **Measurement** means assigning numbers or other symbols to characteristics of objects according to certain prescribed rules.
- **Scales of measurement** is how variables are defined and categorised.
- **Arithmetic** and **statistical** operations for summarizing and presenting data depend on the levels of measurement

Scales of Measurement (Cont....)

Four scales of measurement :

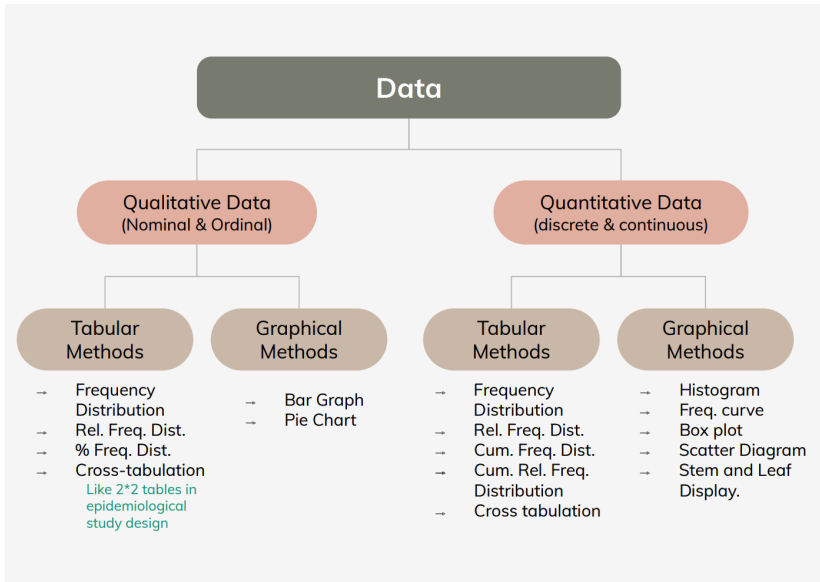
- **Nominal scale:** Data that can only be classified into categories and cannot be arranged in an ordering scheme. Example: eye color, gender, religious affiliation.
- **Ordinal scale:** involves data that may be arranged in some order, but differences between data values cannot be determined or are meaningless. Example: During a taste test of 4 colas, cola C was ranked number 1, cola B was ranked number 2, cola A was ranked number 3, and cola D was ranked number 4.
- **Interval scale:** similar to the ordinal level, with the additional property that meaningful amounts of differences between data values can be determined. There is no natural zero point. Example: Temperature on the Fahrenheit scale.
- **Ratio scale:** the interval level with an inherent zero starting point. Differences and ratios are meaningful for this level of measurement. Example: money, marks etc.

Information Hierarchy



Tabular of and graphical presentation

Tabular and Graphical Procedures



Frequency Distribution

- **Frequency distributions** can show either the actual number of observations falling in each range or the percentage of observations. In the latter instance, the distribution is called a relative frequency distribution.
- Frequency distribution tables can be used for both categorical and numeric variables.
- For continuous data this following step should be followed
 - Determine the range of the data set.
 - Divide the range by the number of the classes that we want our data in and then round up.
 - Create class intervals using class width.
 - Obtain the frequency for each class.

Frequency Distribution (Continuous)

Patient no.	Hb (g/dL)	Patient no.	Hb (g/dL)	Patient no.	Hb (g/dL)
1	12.0	11	11.2	21	14.9
2	11.9	12	13.6	22	12.2
3	11.5	13	10.8	23	12.2
4	14.2	14	12.3	24	11.4
5	12.3	15	12.3	25	10.7
6	13.0	16	15.7	26	12.5
7	10.5	17	12.6	27	11.8
8	12.8	18	9.1	28	15.1
9	13.2	19	12.9	29	13.4
10	11.2	20	14.6	30	13.1

Frequency Distribution (Continuous)

Dummy Table			Tall Marks Table		
Hb (g/dL)	Tall marks	No. Of patients	Hb (g/dL)	Tall marks	No. Of patients
<div> <div>Width</div> <div>9.0 – 9.9</div> </div>			9.0 – 9.9	I	1
10.0 – 10.9			10.0 – 10.9	III	3
11.0 – 11.9			11.0 – 11.9	IIII 1	6
12.0 – 12.9			12.0 – 12.9	IIII III	10
13.0 – 13.9			13.0 – 13.9	III II	5
14.0 – 14.9			14.0 – 14.9	III	3
15.0 – 15.9			15.0 – 15.9	II	2
Total			Total	-	30

Frequency Distribution (Categorical)

- The blood types of 25 blood donors are given below. **Summarize** the data using a frequency distribution.

AB	B	A	O	B
O	B	O	A	O
B	O	B	B	B
A	O	AB	AB	O
A	B	AB	O	A



Class (Blood type)	Frequent, F
A	5
B	8
O	8
AB	4
Total	N= 25

Note: The classes for the distribution are the blood types.

Contingency table

A sample of 500 shoppers was selected in a large metropolitan area to determine various information concerning consumer behavior. Among the questions asked was “do you enjoy shopping for clothing?” the results are summarized in the following cross classified table:

Table 1: Frequency distribution of preference of shopping for clothing of the consumer

Enjoy shopping for clothing	Sex		Total
	Male	Female	
Yes	136	224	360
No	104	36	140
Total	240	260	500

Histogram

Table 2: Frequency distribution of male and female by age group

Age group	Male	Female
0 – 5	3243	1621
5 – 10	2842	1413
10 – 15	2398	1192
15 – 20	2125	1056
20 – 25	1776	880
25 – 30	1450	716
30 – 35	1173	580
35 – 40	936	461
40 – 45	773	378
45 – 50	633	306
50 – 55	503	240
55 – 60	391	184
60 – 65	749	130
65+	749	208

Histogram (Cont...)

Figure: Histogram of number of female by age group

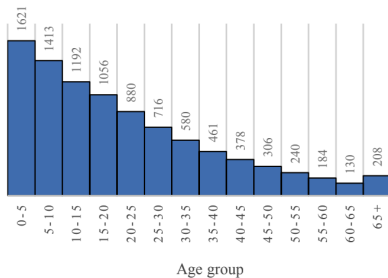
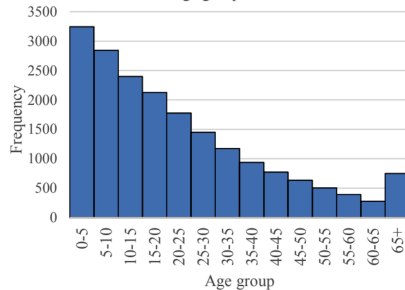


Figure: Histogram of the distribution of Male by age group

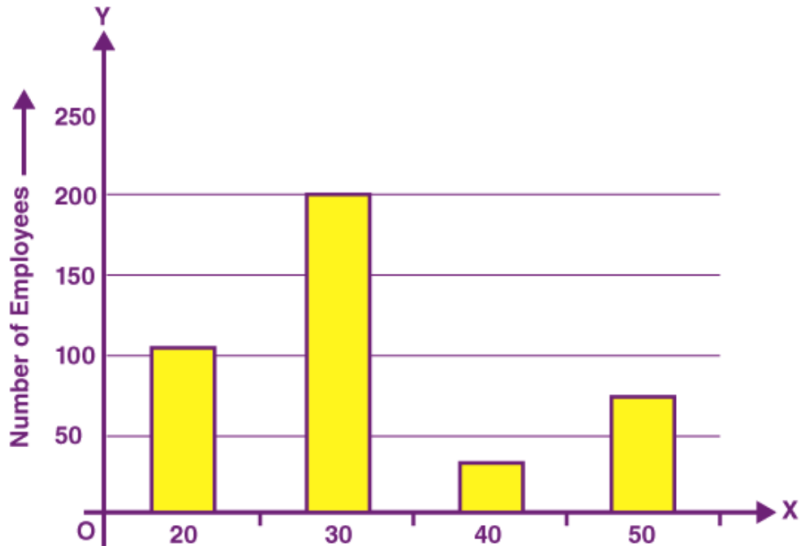


Bar Plot

In a firm of 406 employees, the percentage of monthly salary saved by each employee is given in the following table. Represent it through a bar graph.

Savings (in percentage)	Number of Employees(Frequency)
20	105
30	199
40	29
50	73
Total	406

Bar Plot (Cont..)



Clustered Bar Plot

The following contingency table presents the summarization of two categorical variables.

Table 3

Religion	Sex		Total
	Male	Female	
Muslim	25	15	40
Hindu	20	10	30
Christian	10	5	15
Buddha	18	3	21
Others	6	2	8
Total	79	35	114

Clustered Bar Plot (Cont...)

Fig: Distribution employee by religion and sex

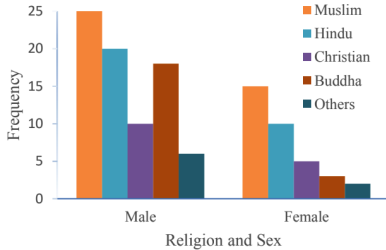


Fig: Distribution employee by religion and sex

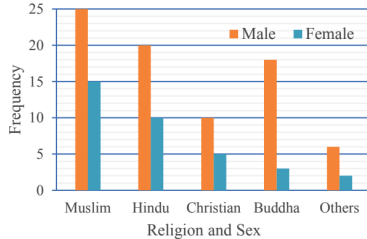
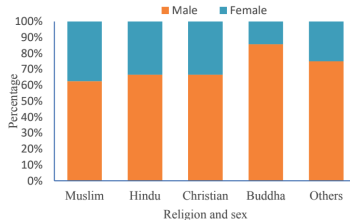


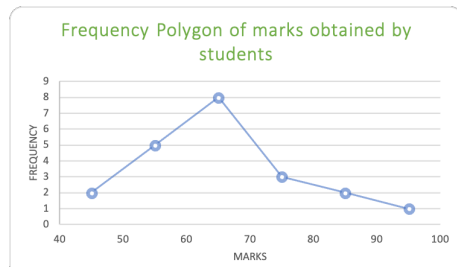
Figure: Percentage distribution of Employee by Religion and Sex



Frequency Polygon

Table 4: Frequency distribution of students by age group

Marks	Mid value	Frequency
40 – 50	45	2
50 – 60	55	6
60 – 70	65	8
70 – 80	75	3
80 – 90	85	2
90 – 100	95	1



Ogive or Cumulative frequency curve

Table 5: Frequency distribution of students by age group

Class interval of the amount of sell	upper limit	frequency	Cumulative frequency
50 – 60	60	2	2
60 – 70	70	8	10
70 – 80	80	18	28
80 – 90	90	22	50
90 – 100	100	12	62
100 – 110	110	8	70
110 – 120	120	4	74

Figure: Cumulative frequency curve

