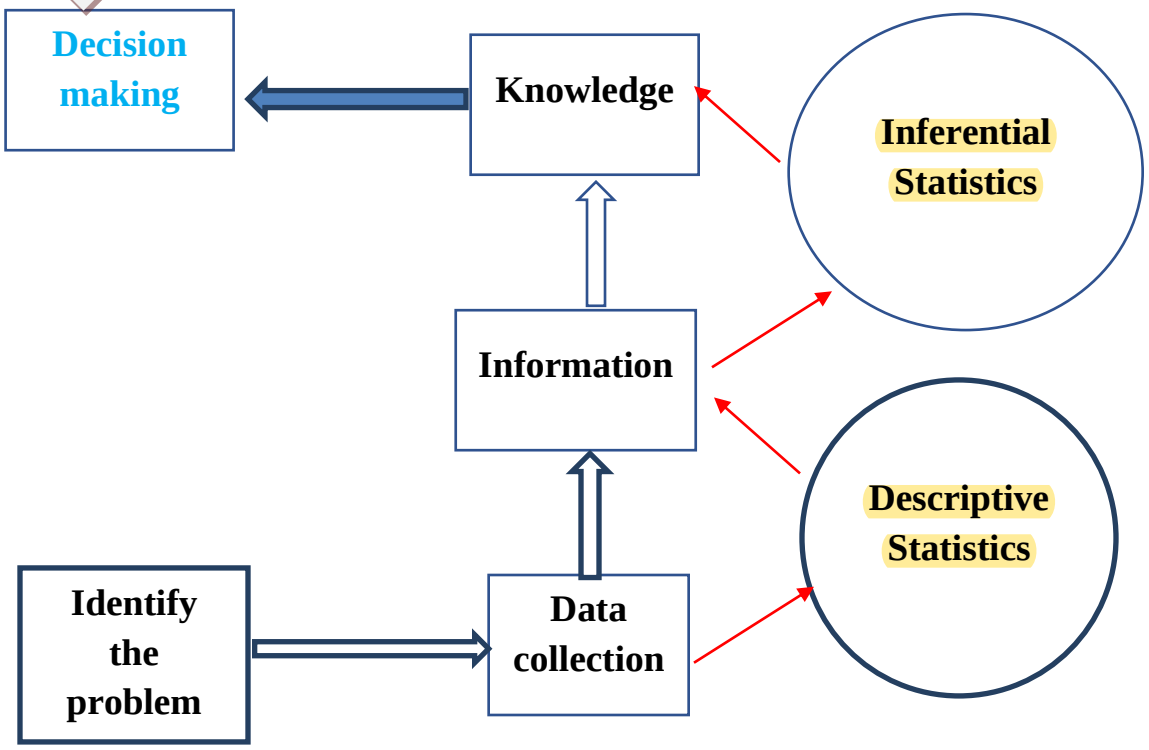
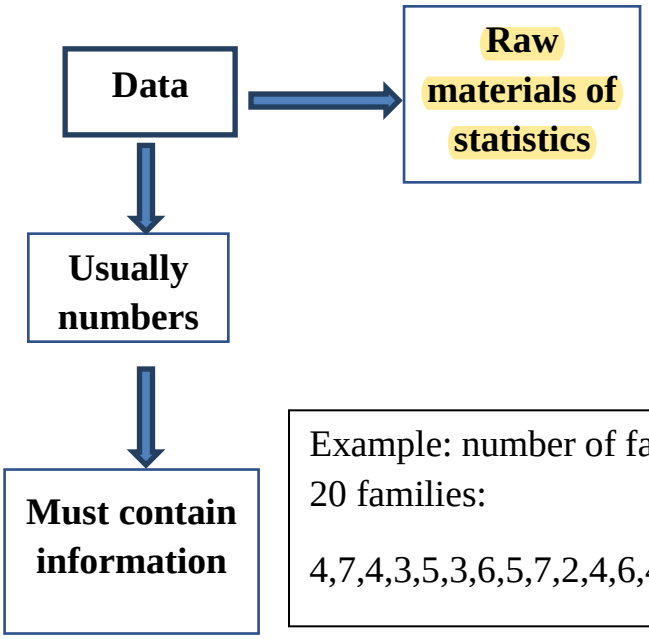
# Lecture 1

## *What is Statistics? / Why we study Statistics?*

### *Goal of Statistics*




### *What is data?*




Example: number of family members of 20 families:

4,7,4,3,5,3,6,5,7,2,4,6,4,5,3,7,3,5,4,7

**Types of Statistics**

- Statistics has two aspects: **theoretical** and **applied.**
- **Theoretical or mathematical statistics** deals with the development, derivation, and proof of statistical theorems, formulas, rules, and laws.
- **Applied statistics** involves the applications of those theorems, formulas, rules, and laws to **solve real-world problems**.
- **Your course is designed with applied statistics** and not with theoretical statistics.

**Types of Applied Statistics**

Broadly speaking, applied statistics can be divided into two areas:

1. descriptive statistics and
2. inferential statistics.

**Descriptive Statistics**

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

**Inferential Statistics**

- A major portion of statistics deals with making decisions, inferences, predictions, and forecasts about populations based on results obtained from samples.
- Inferential statistics consists of methods that use sample results to help make decisions or predictions about a population.

**Population versus Sample**

- In statistics, the collection of **all elements** of interest is called a **population**. A **representative part of elements** selected from this population is called a **sample**.

**Examples**

| Problem of Interest | Population |
|---|---|
| Per month family income of JU students | All the students of JU |
| Per month production of garments industry in BD | All the garment industries in BD |
| Per tree coconut production in BD | All the coconut trees in BD |

## Representative Sample

- A sample that represents the characteristics of the population as closely as possible is called a representative sample.

## Census and Sample Survey

- **A survey** that includes **every member** of the population is called a census.
- The **technique** of collecting information **from a portion of the population** is called a **Sample survey.**

## Variable

- A *variable* is a characteristic under study that assumes different values for different elements. In contrast to a variable, the value of a *constant* is fixed.

**Example**

Human being related variables: Height, weight, age, number of family members, Gender, Marital status

**Other Examples of variables are:**

- the incomes of households,
- the number of houses built in a city per month during the past year,
- number of workers per garment industry,
- the gross profits of companies,
- the number of insurance policies sold by a salesperson per day during the past month,

## Types of variables

A variable may be classified as **quantitative** or **qualitative.**

## Quantitative Variable

- A variable that can be measured numerically is called a *quantitative variable.*
- If you ask question to the respondent about a quantitative variable, respondent will respond by a number.
- The data collected on a quantitative variable are called *quantitative data.*

***Examples***

Incomes, heights, weight, gross sales, prices of homes, number of cars owned, and number of accidents.

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

## Discrete Variable

A variable whose values are countable is called a *discrete variable*. In other words, a discrete variable can assume only certain values with no intermediate values.

**Examples:**

- number of family members,
- number of cars sold on any day at a car dealership,
- number of accidents in a day,
- number of people visiting a bank on any day,
- number of cattle owned by a farmer, and
- number of students in a class.

## Continuous Variables

Some variables cannot be counted, and they can assume any numerical value between two numbers. Such variables are called **continuous variables**.

**Examples:**

- time taken to complete an examination,
- per month family income,
- height of a person,
- weight of a person,
- amount of soda in a 12-ounce can and
- yield of potatoes (in pounds) per acre.

**Simply,** values of a quantitative variable can be derived by two methods: by **counting or by measuring**.

- Values of variable that can be derived by **counting** is called **discrete** variable
- Values of variable that can be derived by **measuring** is called **continuous** variable

## Qualitative or Categorical Variable

A variable that **cannot assume a numerical value** but can be **classified** into two or more **nonnumeric categories** is called a *qualitative* or *categorical variable*. The data collected on such a variable are called *qualitative data*.

**Examples:**

- the gender of a person,
- marital status of a person,
- the brand of a computer,
- the opinions of people, and
- the make of a car.

## Cross-section versus Time-Series Data

### Cross-Section Data

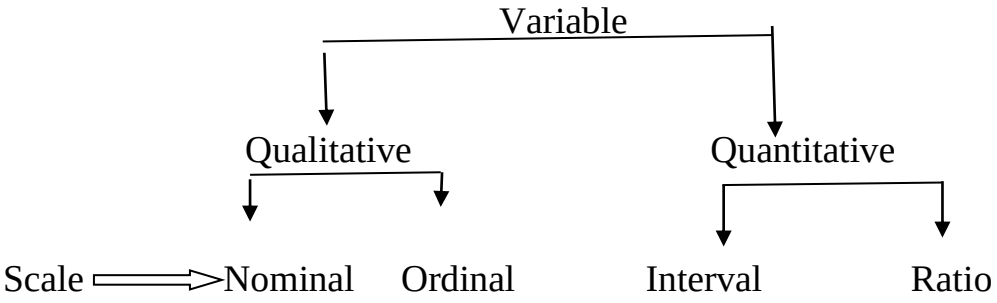
Data collected **on different elements** **at the same point in time** or for the same period of time are called *cross-section data*.

### Time-Series Data

Data collected **on the same element** for the same variable **at different points in time** or for different periods of time are called *time-series data*.

### Scale of Measurement

In statistics, there are four data measurement scales: nominal, ordinal, interval and ratio.




### Nominal Scale

- This scale is the easiest one to understand.
- Nominal scales are used for labeling variables, without any quantitative value.
- "Nominal" scales could simply be called "labels."
- Notice that all of these scales are mutually exclusive (no overlap) and none of them have any numerical significance.
- A good way to remember all of this is that "nominal" sounds a lot like "name" and nominal scales are kind of like "names" or labels.

**Example:** **Gender:** Male and Female
**Hair color:** Brown, Black, Gray, Others
**Marital Status:** Unmarried, married, divorced, separated, widow.

## Ordinal Scale

- With ordinal scales, the ***order* of the values** is what's important and significant, but the **differences between each one is not really known.**
- Take a look at the examples below. In each case, we know that a #4 is better than a #3 or #2, **but we don't know–and cannot quantify–how *much* better it is.**

**Examples:**

How do feel today?

1. Very unhappy,
2. Unhappy,
3. Ok,
4. Happy,
5. Very happy

How satisfied are you with my class?

1. Very unsatisfied,
2. Somewhat unsatisfied,
3. Neutral,
4. Somewhat satisfied,
5. Very satisfied

## Interval scale

- Interval scales are numeric scales in which we know **both the order and the exact differences between the values**.
- Interval **scales are nice** because the **area of statistical analysis** on these data sets **opens up**.
- "Interval" itself means **"space in between,"** which is the important thing to remember–interval scales not only tell us about order, but also about the value between each item.
- The classic example of an interval scale is **Celsius temperature** because the difference between each value is the same. For example, the difference between 60 and 50 degrees is a measurable 10 degrees, as is the difference between 80 and 70 degrees.
- Here's the **problem** with interval scales: they don't have a **"true zero."** For example, there is no such thing as "**no temperature**," at least not with

celsius. In the case of interval scales, **zero doesn't mean the absence of value**, but is actually another number used on the scale, like 0 degrees celsius. Negative numbers also have meaning. Without a true zero, it is impossible to compute ratios. With interval data, **we can subtract, but cannot add, multiply or divide**.

**Ratio Scale**

- Ratio scales are **the ultimate pleasure** when it comes to data measurement scales because they tell us **about the order**, they tell us the exact value between units, AND they also have an **absolute zero**–which allows for a wide range of both descriptive and inferential statistics to be applied.
- Good examples of ratio variables include height, weight, and duration.