Basic Terms of Statistics

Rupika Abeynayake (B.Sc, M.Sc, Ph.D,)

Professor in Applied Statistics

OBJECTIVES

- Discuss the nature, purpose, and types of statistics

- Discuss variables, levels of measurement, and their relationships to statistical analysis.

Introduction to Data Analysis

- Statistics:

  A branch of applied mathematics that deals with collecting, organizing, and interpreting data using well-defined procedure

The Uses of Data Analysis

- Analyses for Description Vs. Inference
- Analyses concerning the number of variables
- Analyses for different purposes

1. Analyses for Description Vs. Inference

- **<u>Descriptive Statistics</u>** used to describe or characterize data by summarizing them into more understandable terms without losing or distorting much of the information.

- **<u>Inferential Statistics</u>** consists of a set of statistical techniques that provide predictions about population based on information in a sample from that population.

2. Analyses concerning the number of variables.

- Univariate Statistics involve one variable at a time.
- Bivariate Statistics involve two variables examined simultaneously.
- Multivariate Statistics involve three or more variables in the same analysis.

3. Analyses for different purposes

- Sample Description.
- Data Cleaning.
- Evaluation of Measuring Tools.
- Evaluation of the Need for Transformations.
- Addressing Research Questions.

Population & Sample

- **Population** is the set of observations or scores about which the researcher wishes to draw conclusions.

- Population characteristics are called
  Parameters (e.g., μ, σ, σ^2, ρ)

- **Sample** is a part of the population.

- Sample characteristics are called **statistics**
  (e.g., $\overline{X}$, S, S^2, r)

Research Variables

- A **Variable** is a characteristic being measured that varies among the persons, places, or objects being studied

  Examples: Gender, eye color, intelligence, age, height, weight, blood pressure, and heart rate.

Independent and Dependent Variables

- The independent variable is the cause of, or influence on, the dependent variable

- Example:

  Does a low-cholesterol diet reduce the risk of heart disease?

- Dependent Variable: Heart disease
- Indep. Variable: The amount of cholesterol

Discrete and Continuous Variables

- **Discrete variable** has a finite number of values between any two points

  The variable for the number of times hospitalized is discrete, because a number such as 1.5 is not a meaningful value

- **Continuous variable** can assume an infinite number of values between any 2 points. Weight is an example of a continuous variable.

Measurement of a Variable

Measurement is the process of assigning numbers to the characteristics you want to measure according to acceptable rules

There are some well-known rules for assigning numbers to variables. A particular set of rules defined a scale of measurement, and different sets of rules define different scales of measurement

Measurement Scales

Four kinds of scale of measurement are important for quantifying variables in the behavioral sciences:

1. **Nominal Scale**
2. **Ordinal Scale**
3. **Interval Scale**
4. **Ratio Scale**

1. Nominal Scale

- This type of scale allows a researcher to classify characteristics of the persons, places or objects into categories
- Sometimes variables measured on nominal scales are called categorical or qualitative

Examples:

Group membership (1 = Experimental, 2=Placebo)

A person’s gender (0 = Female, 1 = Male)

Blood type, marital status, and religion.

2. Ordinal Scale

- In this case, the characteristics can be put into categories and the categories also can be ordered in some meaningful way.

- The distance between the categories, however, is unknown.

Ordinal Scale, Continued…

- For example, in a swimming race the results are reported in terms of which swimmer was first, who was second, and who was third
  - However, it is irrelevant whether the winning swimmer won by one length or by several lengths.

Ordinal Scale, Continued…

- **Examples:**

Socioeconomic Status

1 = Low

2 = Middle

3 = High

Health Status

1 = Poor

2 = Fair

3 = Good

4 = Excellent

3. Interval Scale

In this case, the distance between these ordered category values are equal because there is some accepted physical unit of measurement.

In the Fahrenheit thermometer, mercury rises in equal intervals called degrees.

3. Interval Scale, Continued…

- However, the zero point is arbitrary, chosen because Daniel Fahrenheit, the inventor, decided that zero point on this scale would be 32 degree below the freezing point of water.

3. Interval Scale, Continued

- Because the units are in equal intervals, it is possible to add and subtract across an interval scale.

- You can say that 100^0 F is warmer than 50^0, but you cannot say that 100^0 F is twice as hot as 50^0 F.

4. Ratio Scale

- The most precise level of measurement consists of meaningfully ordered characteristics with equal intervals between them and the presence of a zero point that is not arbitrary but determined by nature.

4. Ratio Scale, Continued…

- On the Kelvin temperature scale, zero represents the absence of molecular motion. Because the zero point is not arbitrary, it is possible to multiply and divide across a ratio scale.

4. Ratio Scale, Continued…

- It is possible to say that 100^0 K is twice as hot as 50^0 K.

  - Examples: Weight, Length, blood pressure

  - It is possible to say that 40 inches is twice as long as 20 inches.