CHAPTER ONE

INTRODUCTION TO STATISTICS

Yebelay Berehan

Biostatistician

yebelay.ma@gmail.com

2022-06-07

Introductions

What is statistics?

- Statistics is the science of learning from data, and of measuring, controlling, and communicating **uncertainty**.

- refers to the collection, organizing, presentation, analyzing, and interpretation of numerical data **to make inferences** and

- **reach decisions** in all branches of economics, business,land administration, medicine, and other social and physical sciences.

- The word statistics is defined in **two ways** in the plural and singular sense, depending on its use.

- **In plural sense:** Statistics is defined as the **collection of numerical facts or figures** (or **the raw data themselves**).
  - Example: **Vital statistics** (numerical data on *marriage*, *births*, *deaths*, etc).
  - The annual birth rate or death rate of Ethiopia.
- **statistics are aggregate of facts**.

- **Single and isolated figures are not statistics** as they cannot be compared and are unrelated.

- **In singular sense:** Statistics is the art of science that deals with the method of **data collection, data organization, data presentation, data analysis and interpretation of results**.

- It refers to **a subject matter** that is concerned with **extracting relevant information** from available data **with the aim to make sound decisions.**

Classification of Statistics

- **Descriptive Statistics:** procedures used to **organize and summarize masses of data.**
- **Organizing data** using **classification, tabulation, Frequency Distributions**.
- **Presenting data** using Graphical and Diagrammatic Presentation

- **Summarize data** using **Measures of Central Tendency** and **Measures of Dispersion**.

- *Data processing design to the collection, organization, analyzing, and presentation the important features of the data* **without attempting to infer/conclusion about the data**.

- It describes the **nature or characteristics of the observed data** (usually sample) **without making conclusion or generalization**.

Example:

- The daily temperature **range** of DM was $25^o c$ last week

- The **maximum amount** of coffee export of Ethiopia (as observed from the last 20 years) was in the year 2004.

- **75%** of the instructors in DMU are Male.

- The scores of 50 students in introduction to statistics exam are found to **rage from 20 to 90**.

- **Inferential Statistics: are methods for using sample data to make general conclusions (inferences) about populations**.

- **Performing:**

  - **Hypothesis testing**
  - **Determining relationship between variables**
  - **Making prediction**

- **Example:**

  - The result obtained from the analysis of the income of **1000 randomly selected citizens** in Ethiopia suggests that the average monthly income of a citizen is **estimated to be 600 birr**.

  - From the survey **obtained on 15 randomly selected towns** of Ethiopia is **estimated that 0.1% of the urban Dwellers are Victims of AIDS virus**.

Stages In Statistical Investigation

- There are five steps of in statistical investigation

  - Collection of raw data

  - Organizing of data

  - Presentation of the data

  - Analyzing of data

  - Interpretation of the result

Stages In Statistical Investigation

Collection of data :

- Data collection is the **process of collecting or gathering information or data about the variable of interest**.

- Data are **raw materials or inputs** for **statistical investigation**.

- Data may be obtained from **primary and secondary source**.

Organization of data:

- Organization of data includes **three major steps**.
    - **Editing:** checking and omitting inconsistencies or irrelevancies
    - **Classification:** task of grouping the collecting and edited data
    - **Tabulation:** put the classified data in the form of table using rows and columns

Presentation of data

- The **purpose of presentation** in the statistical analysis is to **display what is contained in the data** in the form of **charts, pictures diagrams and graph** for an easy and better understanding of the data.

Analyzing data

- In a statistical investigation, the process of analyzing data includes finding the **various statistical constants** from the collected mass of data such as **measure of central tendencies(average), measure of dispersion** an so on.

- In its extreme case analysis **requires the knowledge of advanced mathematics**.

Interpretation of the result

- Involve **interpreting constants computed in analyzing data for the information of valid conclusion and inferences**.

- It is the **most difficult and skill requiring stage**.

- At this stage the statistics seems to be very much viable to be **misused**.

- **Correct interpretation** of result will lead to a **valid conclusion of the study and hence can aid in taking correct decisions**.

- **Improper(incorrect) interpretation** may lead to **wrong conclusion** and makes the whole objects of the study useless.

Definition of Some Basic terms

- **Data:** is the raw material of statistics.

- **Data (singular):** The value of the variable associated with one element of a population or sample. This value may be a number, a word, or a symbol.

- **Data (plural):** The set of values collected for the variable from each of the elements belonging to the sample.

There are two types of data:

- **Quantitative data:** data of certain group of individuals which is expressed numerically.

- Example height, weight, age, income, number of students, Students Grades, Educational level, etc certain group of individuals

- **Qualitative data:** data of certain group of individuals that is not expressed numerically.

- Example: colors, language, nationalities, Religion, health, poverty, occupations, etc certain group of individuals.

- **Variable:** is a factor or characteristic that **can take on different possible values or outcomes**.

- A variable can be qualitative (non-numeric) or quantitative (numeric)

- Example: Income, height, weight, sex, age, diastolic blood pressure etc of a certain group of individual are example of variable.

- **Population:** A complete list of observation(data) of the **entire group of individuals** under consideration.

- A population can be finite or infinite.

- Example: the number of students in this class, the number of population in Debre Markos.

- **Parameter:** is summary or calculated value obtained from a populati...

- **Sample:** a **portion of population** selected for further analysis

- **Sample size:** the **number of items** under investigation in a sample.

- **Sampling:** The **process or method of sample selection** from the population.

- **Statistic:** is summary or calculated value obtained from a sample.

- **Survey:** it a process of obtaining data.

Two types of survey

i). Census survey: a way of obtaining data referring the entire population including a total of population.

ii). Sample survey: a way of obtaining data referring a portion of the entire population consisting only a partial coverage of the population

- **Sampling Frame:** A list of the elements belonging to the population from which the sample will be drawn.

Importance of Statistics

- It simplifies mass of data (condensation)

- It Helps to get **concrete information about any problem**.

- It Helps **to reliable and objective decision making**.

- It presents **facts in a precise and definite form**.

- It used **to facilitate comparison** using measure of central tendency and measure of variation.

- It used to facilitates **prediction or Forecasting**.

- It helps **to formation of suitable policy**.

Types of Variables and Measurement Scales

Types of Variables:

- **Quantitative Variables:** A quantitative variable is a characteristic that can be measured.

- The values of a quantitative variable are numbers indicating how much or how many of something.

- Examples: Family Size, Weight, height and the weights of students.

Types of Quantitative Variables:

- **Discrete Variables:-** are quantitative variables that will assume only values in a discrete set like the set of integers, such as the number of chairs in the classroom.

- There is a gap between any two values.

- Examples:- Family size (x = 1, 2, 3, ...)

- **Continuous Variables:-** are quantitative variables that will assume an infinite number of possible values in any given interval such as room temperature.

- A continuous variable can assume any value along a line interval, including every possible value between any two values.

- Examples:- Height ($140 < x < 190$)

- **Qualitative Variables:** It deal with quality, and cannot be measured.

- These variables arrange individuals into categories according to shared qualities or characteristics.

Examples: Students Grades, Educational level

Exercise

- Identify each of the following examples as **qualitative or quantitative** variables.

- **If the variable is quantitative** then classify as **discrete and continuous**

1. The residence hall for each student in a statistics class.
2. The number of colors used in a statistics textbook.
3. The number of files on a computer's hard disk.
4. The color of the baseball cap worn by each of 20 students.
5. The length of time to complete a mathematics homework assignment.
6. The brand of refrigerator in a home
7. The pH level of the water in a swimming pool.
8. The overall satisfaction rating of a new car.
9. The length of time until a pain reliever begins to work.

Types of Measurement Scales

- **Nominal Variable:** is an **unordered set of categories** identified **only by name**.

- Nominal measurements only permit you to determine whether two individuals are the same or different.

- A qualitative variable that categorizes (or describes, or names) an element of a population.

- **No arithmetic and relational operation can be applied**.

- example: eye colour, gender etc.

- **Ordinal Variable:** is an **ordered set of categories**.

- Ordinal measurements tell you the **direction of difference** between two individuals.

- A qualitative variable that incorporates an **ordered position**, or **ranking**.

- **Arithmetic operations are not applicable** but **relational operations are applicable**.

- *Ordering is the sole property of ordinal scale*.

- Example: finishing position in a race.

- You can say that the **person who finished first was faster than the person who finished second**, **but you _don't_ know how much faster**.

- **Interval scale:** is an **ordered series of equal-sized categories**.

- Interval measurements identify the direction and magnitude of a difference.

- The zero point is located arbitrarily on an interval scale.

- All arithmetic operations **except division** are applicable.

- Relational operations are also possible

- In short, addition and subtraction are meaningful for interval scale variables.

- Notice that the 0° does not mean "no temperature at all".

- **Ratio scale:** is an interval scale where a value of zero indicates none of the variable.

- Ratio measurements identify the direction and magnitude of differences and allow ratio comparisons of measurements.

- All arithmetic and relation operation are applied.