STATISTICS IS THE GRAMMAR OF SCIENCE

PROBABILITY AND STATISTICS

LECTURE – 1

INTRODUCTION TO STATISTICS

PREPARED BY
HAZBER SAMSON
FAST NUCES ISLAMABAD

INTRODUCTION TO STATISTICS

STATISTICS Statistics is a Science used to collect, organize, present, analyze and interpret data to make decisions.

TYPES OF STATISTICS Basically there are two types of statistics.

1. Descriptive Statistics 2. Inferential Statistics

DESCRIPTIVE STATISTICS descriptive statistics consists of methods used to organize, preset and summarize data by using tables, graphs and summary sheets.

INFERENTIAL STATISTICS Inferential statistics consists of methods used to find out something about a population based on a sample.

IMPORTANCE OF STATISTICS

1. Statistics is used in different fields of Sciences like Social Sciences, Plant Sciences, Physical Sciences and Medical sciences.
2. It plays an important role in planning and helps the government in formulating the policies.
3. It plays an important role in Business, industry and Banking sector.
4. statistics present the data in comparable form so it helps us comparing past results with present.
5. Forecasting is the main objective of statistics. Forecasting means estimating future values with the help of past data.

BASIC TERMS IN STATISTICS

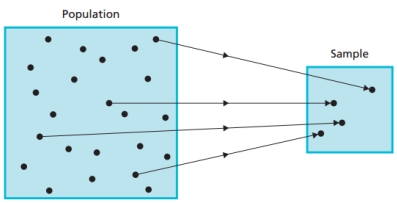
POPULATION A collection of all possible individuals, objects or measurements of interest is called a population. For example set of students in a college.

SAMPLE A portion or part of a population of interest is called a sample. For example set of students of Ist year class from a University.

PARAMETER Any numerical; value computed from the population is called a parameter. Parameters are fixed quantities.

STATISTIC Any numerical value computed from the sample is called a statistic. Statistics are variable quantities because they vary from sample to sample.

FIGURE 1.1 Relationship between population and sample



DATA AND TYPES OF DATA

DATA Numerical facts or observations are collectively called data. A single numerical fact is called datum. There are different types of data.

UNIVARIATE DATA Data that represents observations on one variable that are measured in the same units is called Univariate data.

BIVARIATE DATA Data that represents observations on two variables that are measured in the different units is called Bivariate data.

MUTIVARIATE DATA Data that represents observations on more than two variables that are measured in the different units is called Multivariate data.

PRIMARY DATA Data that have been originally collected and have not undergone any sort of statistical treatment is called a primary data. For example marks of all the students in a class.

SECONDARY DATA Data that have undergone any sort of statistical treatment is called secondary data. For example percentage marks of each student in the class.

SOURCES OF PRIMARY DATA There are different sources of Primary Data.

- Through investigation
- Through Questionnaire
- Through Local Sources
- Through Telephone
- Through Internet

SOURCES OF SECONDARY DATA There are different sources of Secondary Data.

- Government Organizations
- Semi-government Organizations
- Research Organizations
- Newspapers
- Internet

CROSS-SECTION VERSUS TIME-SERIES DATA Based on the time over which they are collected, data can be classified as either cross-section or 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.

VARIABLE AND TYPES OF VARIABLES

VARIABLE A variable is a characteristic under study that assumes different values for different elements. In contrast to a variable, constant assumes only one value ie its value is fixed.

Types of Variables Basically there are two types of variables.

1. Qualitative Variables 2. Quantitative Variables

QUALITATIVE VARIABLE A variable that can not be measured numerically but can be classified into two or more nonnumeric categories is called a qualitative variable or categorical variable or an attribute. The data collected on qualitative variable is called qualitative data. For example gender, religious affiliation and eye color etc.

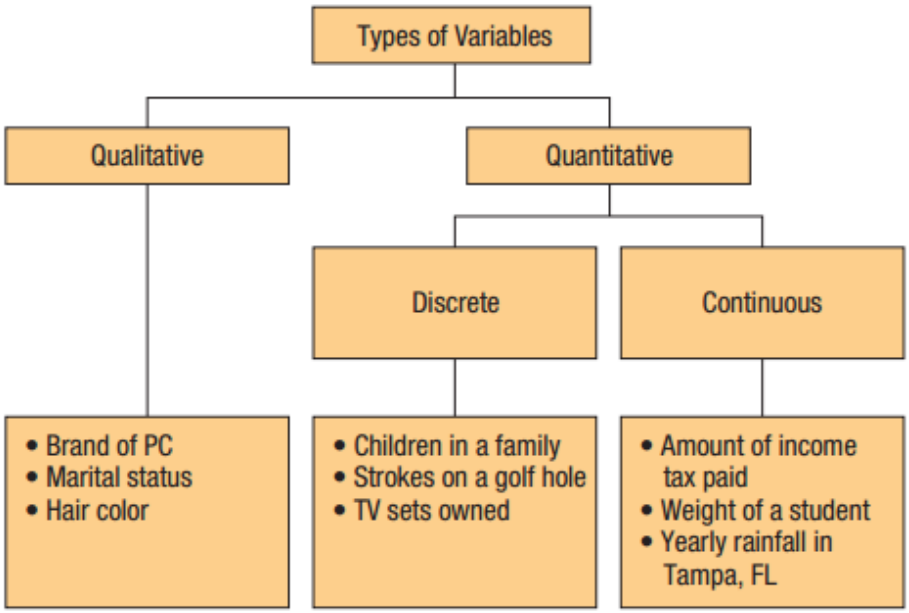
QUANTITATIVE VARIABLE A variable that can be measured numerically is called a quantitative data. The data collected on a quantitative variable is called quantitative data. For example Incomes, heights, gross sales, prices of homes, number of accidents etc.

There are two types of quantitative variables
1. Discrete Variable 2. Continuous Variable

DISCRETE VARIABLE A variable is said to be discrete if it can assume only certain values in a given interval. These values are countable.
For example Number of bedrooms in a house, Number of doctors in a hospital etc

CONTINUOUS VARIABLE A variable is said to be continuous if it can assume any value in a given interval. These values are not countable.
For example temperature at a place, height of a person, speedometer etc

FIGURE 1.2 Summary of the types of variables



MEASUREMENT SCALES OF VARIABLES

Measurement means assigning numbers to observations or objects and scaling is a process of measuring. Basically there are four scales of measurement w.r.t any data set

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

RATIO SCALE Suppose that X is a variable taking two values $x_1 \; and \; x_2$. If it satisfy the following properties it can be referred as ratio scale.

$P-1$: Ratio is meaningful ie x_2/x_1 is meaningful.
$P-2$: Distance is meaningful ie $x_2 - x_1$ is meaningful.
$P-3$: There exists natural ordering. ie $x_2 \geq x_1$

EXAMPLES of ratio scale are Age, Weight, Height, Time, salary, Distance.

It is a special kind of an interval scale where the scale of measurement has a true zero point as its origin.
The Ratio scale data has the following properties.

1. Data categories are mutually exclusive and exhaustive.
2. Data categories are scaled according to the amount of the characteristic they possess.
3. Equal difference in the characteristic are represented by equal difference ij the numbers assigned to the categories.
4. The point 0 reflects the absence of the characteristic.

INTERVAL SCALE Suppose that X is a variable taking two values $x_1 \; and \; x_2$. If it satisfy the 2nd and 3rd properties of ratio scale than it can be referred as Interval scale.

$P-1$: Distance is meaningful ie $x_2 - x_1$ is meaningful.
$P-2$: There exists natural ordering. ie $x_2 \geq x_1$

EXAMPLES of interval scale are Temperature, IQ Score

A measurement scale possessing a constant interval size but not a true zero point, is called an interval scale.

The Interval scale data has the following properties.

1. Data categories are mutually exclusive and exhaustive.
2. Data categories are scaled according to the amount of the characteristic they possess.
3. Equal difference in the characteristic are represented by equal difference in the numbers assigned to the categories.

ORDINAL SCALE Suppose that X is a variable taking two values x_1 *and* x_2. If it satisfy only 3rd property of ratio scale than it can be referred as Ordinal scale.
$P-1$: There exists natural ordering. ie $x_2 \geq x_1$

EXAMPLES of ordinal scale are
(1) Grades of students *(A,B,C,D,E,F)*
(2) Ranking of Universities *(1^{st}, 2^{nd}, 3^{rd})*
(3) Socio Economic Status (*Poor, Middle class, Rich*)

The ordinal scale data has the following properties.

1. Data categories are mutually exclusive and exhaustive.
2. Data categories have a logical order.

NOMINAL SCALE Nominal scale does not exhibit any property of ratio scale. It is a categorical variable,

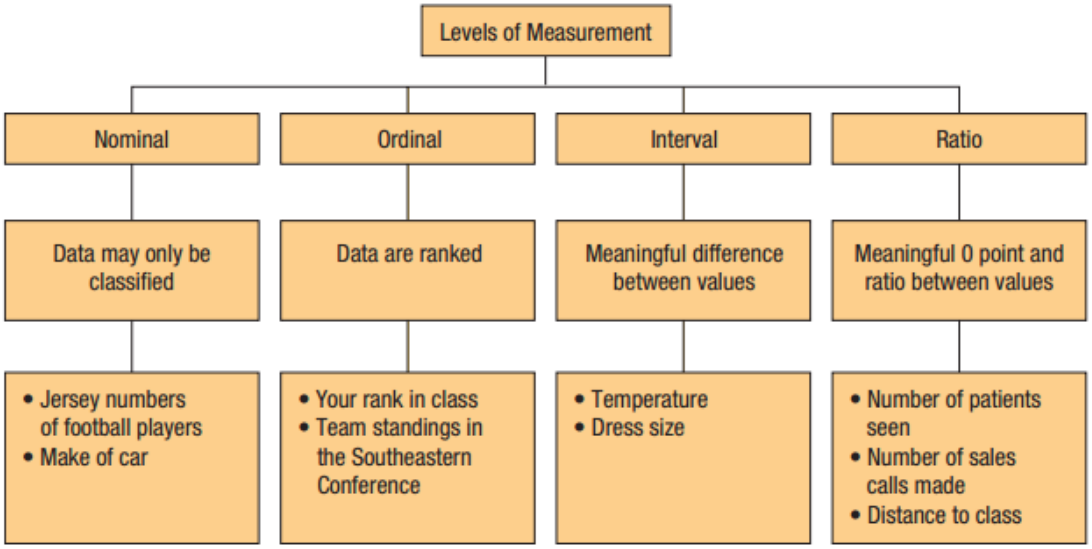
EXAMPLES of Nominal scale are
(1) Gender
(2) Religion
(3) Nationality
(4) Eye color
(5) Marital Status

In this scale observations are classified into mutually exclusive qualitative categories.

The nominal scale data has the following properties.

1. Data categories are mutually exclusive.
2. Data categories have no logical order.

FIGURE 1.3 Summary and Examples of the Characteristics for Levels of Measurement



IMPORTANT NOTATIONS

SUM $\left(\sum\right)$ It is a Greek letter (read as sigma) and is used as a short-hand notation for sum.

In general $$y_1 + y_2 + ... + y_n = \sum_{i=1}^{i=n} y_i$$

For example if $y_1 = 2$, $y_2 = 3$ and $y_3 = 5$

then $\sum_{i=1}^{i=3} y_i = y_1 + y_2 + y_3 = 2 + 3 + 5 = 10$

PRODUCT $\left(\prod\right)$ It is a Greek letter (read as pi) and is used as a short hand notation for product.

In general $$y_1 \times y_2 \times ... \times y_n = \prod_{i=1}^{i=n} y_i$$

For example if $y_1 = 2$, $y_2 = 3$ and $y_3 = 5$

then $\prod_{i=1}^{i=n} y_i = y_1 \times y_2 \times y_3 = 2 \times 3 \times 5 = 30$

SIGNIFICANT FIGURES Those figures in a number which give its information other than its magnitude are called significant figures.

EXAMPLES

1. 26.2 contains 3 significant figures.
2. 1225 contains 4 significant figures.
3. 0.95 contains 2 significant figures.
4. 0.0037 contains 2 significant figures.
5. 0.001592 contains 4 significant figures.

ROUNDING OF FIGURES In order to express figures to smaller number of significant figures we usually round off the numbers. In rounding off the figures we examine the last significant figure. So

1. If it is less than 5 last significant figure will remain the same.
2. If it is greater than 5 than last significant figure is increased by 1.
3. If it is exactly 5 then
   (a) Increase the last significant figure by one if it is an odd number.
   (b) The last significant figure remains same if it is an even number.

EXAMPLES

1. 523 becomes 520
2. 526 becomes 530
3. 1975 becomes 1980
4. 1962 becomes 1962
5. 4.345 becomes 4.32