Statistics Engineering

Course Topics:

- Introduction, Data Summary and Presentation
- Probability: Addition rule, conditional probability, multiplication rule and Bayes Theorem.
- Discrete random variables. Probability mass function. Mean and variance of discrete random variables.
- Probability Distribution functions: Uniform, Binomial, Geometric and Negative Binomial, Hyper-geometric and Poisson Distribution.
- Continuous random variables. Probability Density functions.
- Normal Distribution. Approximation to Binomial and Poisson Distribution. Exponential distribution. Other continuous distributions.
- Joint probability function. Multiple discrete and continuous random variables.
- Covariance and correlation. Bivariate Normal Distribution. Linear combination of random variables. Functions of random variables.
- Parameter estimation. Properties of estimators. Method of Moments.
- Method of Maximum likelihood.
- Interval estimation. Inference on the mean of a population: variance known or unknown. Inference on the variance of a normal population
- Hypothesis testing about the mean and Proportion: Small and Large Sample
- Hypothesis testing: Two Populations

Recommended Textbook(s):

- William Mendenhall and Terry Sincich, **Statistics for Engineering and the Sciences**, Prentice Hall, sixth edition., 2016.
- Douglas C. Montgomery and George C. Runger, Applied **Statistics and Probability for Engineers**, Third Edition, 2003.

- Gujarat Technological University, **Numerical And Statistical Methods for Civil Engineering**, Second Edition, 2017 .

1- Introduction, Data Summary and Presentation

A successful engineer or scientist is one who is proficient at collecting information, evaluating it, and drawing conclusions from it. This requires proper training in statistics. According to *The Random House College Dictionary*, statistics is "**the science that deals with the collection, classification, analysis, and interpretation of information or data**." In short, **statistics** is the **science of data.**

1.1. **Definition:**

1- Statistics is the science of data. This involves collecting, classifying, summarizing, organizing, analyzing, and interpreting data.

The science of statistics is commonly applied to two types of problems:

- Summarizing, describing, and exploring data.
- Using sample data to infer the nature of the data set from which the sample was selected.

- **Descriptive and Inferential Statistics**

Descriptive statistics consists of the collection, organization, summarization, and presentation of data. For example, the average age, income and other characteristics of the population.

Inferential statistics consists of generalizing from samples to populations, performing estimations and hypothesis tests, determining relationships among variables, and making predictions.

A summary of the elements of both descriptive and inferential statistical problems is given as the following.

Four Elements of Descriptive Statistical Problems

1. The population or sample of interest

2. One or more variables (characteristics of the population or sample units) that are to be investigated

3. Tables, graphs, or numerical summary tools

4. Identification of patterns in the data.

Five Elements of Inferential Statistical Problems

1. The population of interest.

2. One or more variables (characteristics of the experimental units) that are to be investigated.

3. The sample of experimental units.

4. The inference about the population based on information contained in the sample.

5. A measure of reliability for the inference.

- A **measure of reliability** is a statement (usually quantified) about the degree of uncertainty associated with a statistical inference.

1.2. Fundamental Elements of Statistics:

In statistical terminology, the data set that we want to describe, the one that characterizes a phenomenon of interest to us, is called a population. Then, we can define am sample as a subset of data selected from a population. Sometimes, the words population and sample are used to represent the objects upon which the measurements are taken (i.e., the experimental units). In a particular study, the meaning attached to these terms will be clear by the context in which they are used.

There for the **statistical population** can be define a data set (usually large, sometimes conceptual) that is our target of interest. But A **sample** is a subset of data selected from the target population.

- **Variables and Types of Data**

  In studying populations and samples, we focus on one or more characteristics or properties of the experimental units in the population. The science of statistics refers to these characteristics as variables.

For example, in the drinking-water quality study, two variables of interest to engineers are the chlorine-residual (measured in parts per million) and the number of fecal coliforms in a 100-milliliter water specimen.

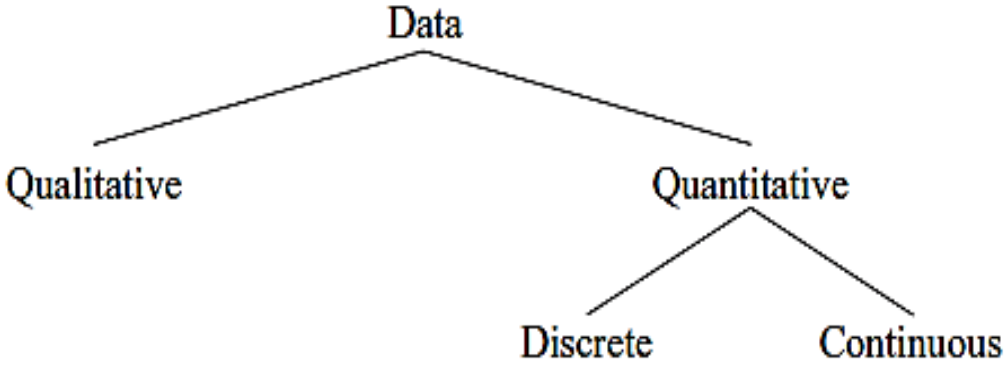
A variable is a characteristic or property of an individual experimental unit. Variables can be classified as

- **Qualitative variables** are variables that can be placed into distinct categories, according to some characteristic or attribute. For example, gender (male or female).
- **Quantitative variables** are numerical and can be ordered or ranked. For example, the variable *age* heights, weights, and body temperatures. Quantitative variables can be further classified into two groups:

✓ **Discrete variables** can be assigned values such as 0, 1, 2, 3 (integer values) and are said to be countable. For examples: the number of children in a family, the number of students in a classroom.

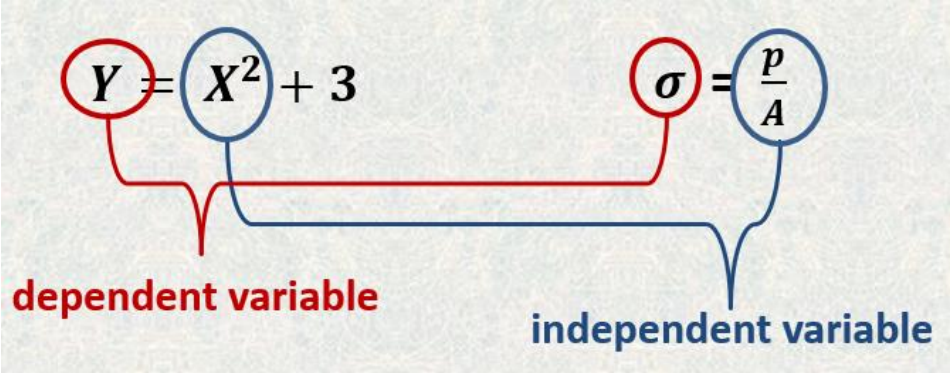
✓ **Continuous variables** can assume an infinite number of values between any two specific values. They are obtained by measuring. They often include fractions and decimals.

The classification of variables can be summarized as follows:



Statistically, variables can be divided into two types: independent and one dependent variables.

- The **independent variable** in an experimental study is the one that is being manipulated by the researcher.
- the **dependent** variable is the resultant variable or the outcome variable.



1.3. Collecting Data: Sampling

Once you decide on the type of data—quantitative or qualitative—appropriate for the problem at hand, you'll need to collect the data. Generally, you can obtain the data in three different ways:

1. Data from a *published source*
2. Data from a *designed experiment*
3. Data from an *observational study* (e.g., a *survey*)

Sometimes, the data set of interest has already been collected for you and is available in a published source, such as a book, journal, newspaper, or Web site. For example, a transportation engineer may want to examine and summarize the automobile accident death rates in the 50 accident in Iraq. second, more common, method of collecting data in engineering and the sciences involves conducting a designed experiment, in which the researcher exerts strict control over the units (people, objects, or events) in the study. For example, an often-cited medical study investigated the potential of aspirin in preventing heart attacks.

Finally, observational studies can be employed to collect data. In an observational study, the researcher observes the experimental units in their natural setting and records the variable(s) of interest. For example, an industrial engineer might observe and record the level of productivity of a sample of assembly line workers.