**Notes – Ch 1 Data and Statistics**

Applications of Statistics in Business and Economics

1. Accounting
2. Finance
3. Marketing
4. Production
5. Economics

**Data** – Data are the facts and figures collected, analyzed, and summarized for presentation and interpretation.

**Data set -** All the data collected in a particular study are referred to as the data set for the study.

**Elements** - Elements are the entities on which data are collected.

**Variable** - A variable is a characteristic of interest for the elements.

**Observation** - Measurements collected on each variable for every element in a study provide the data. The set of measurements obtained for a particular element is called an observation.

**Scales of Measurement** - The scale of measurement determines the amount of information contained in the data and indicates the most appropriate data summarization and statistical analyses.

1. **Nominal Scale** - When the data for a variable consist of labels or names used to identify an attribute of the element, the scale of measurement is considered a nominal scale.
2. **Ordinal Scale** - The scale of measurement for a variable is called an ordinal scale if the data exhibit the properties of nominal data and the order or rank of the data is meaningful.
3. **Interval Scale** - The scale of measurement for a variable is an interval scale if the data have all the properties of ordinal data and the interval between values is expressed in terms of a fixed unit of measure.
4. **Ratio Scale** - The scale of measurement for a variable is a ratio scale if the data have all the properties of interval data and the ratio of two values is meaningful.

Data can be classified as either qualitative or quantitative.

1. **Qualitative data** - Data that include labels or names used to identify an attribute of each element are referred to as qualitaitve data. Qualitative data use either the nominal or ordinal scale of measurement. A qualitative variable is a variable with qualitative data.
2. **Quantitative Data** - Data that use numeric values to indicate how much or how many are referred to as quantitative data. Quantitative data are obtained using either the interval or ratio scale of measurement. A quantitative variable is a variable with quantitative data.

Data can be classified as either Cross-sectional data or time period data

1. **Cross-sectional data** are data collected at the same or approximately the same point in time.
2. **Time series data** are data collected over several time periods.

**Data Source:** Data can be obtained from existing sources or from surveys and experimental studies designed to collect new data.

1. **Existing Data** - In some cases, data needed for a particular application already exist. Companies maintain a variety of databases about their employees, customers, and business operations. Data on employee salaries, ages, and years of experience can usually be obtained from internal personnel records. Other internal records contain data on sales, advertising expenditures, distribution costs, inventory levels, and production quantities. The Internet continues to grow as an important source of data and statistical information. Almost all companies maintain websites that provide general information about the company as well as data on sales, number of employees, number of products, product prices, and product specifications. In addition, a number of companies now specialize in making information available over the Internet. As a result, one can obtain access to stock quotes, meal prices at restaurants, salary data, and an almost infinite variety of information. Government agencies are another important source of existing data.
2. **Statistical Study** - Sometimes the data needed for a particular application are not available through existing sources. In such cases, the data can often be obtained by conducting a statistical study. Statistical studies can be classified as either experimental or observational.
   1. In an **experimental study**, a variable of interest is first identified. Then one or more other variables are identified and controlled so that data can be obtained about how they influence the variable of interest.
   2. **Nonexperimental, or observational,** statistical studies make no attempt to control the variables of interest. A survey is perhaps the most common type of observational study. For instance, in a personal interview survey, research questions are first identified. Then a questionnaire is designed and administered to a sample of individuals. Anyone wanting to use data and statistical analysis as aids to decision making must be aware of the time and cost required to obtain the data.

The use of existing data sources is desirable when data must be obtained in a relatively short period of time. If important data are not readily available from an existing source, the additional time and cost involved in obtaining the data must be taken into account. In all cases, the decision maker should consider the contribution of the statistical analysis to the decision-making process. The cost of data acquisition and the subsequent statistical analysis should not exceed the savings generated by using the information to make a better decision.

**Data Acquisition Errors-** Managers should always be aware of the possibility of data errors in statistical studies. Usingerroneous data can be worse than not using any data at all. An error in data acquisition occurswhenever the data value obtained is not equal to the true or actual value that would be obtainedwith a correct procedure. Such errors can occur in a number of ways. For example, aninterviewer might make a recording error, such as a transposition in writing the age of a24-year-old person as 42, or the person answering an interview question might misinterpret the question and provide an incorrect response. Experienced data analysts take great care in collecting and recording data to ensure that errors are not made. Special procedures can be used to check for internal consistency of the data. For instance, such procedures would indicate that the analyst should review the accuracy of data for a respondent shown to be 22 years of age but reporting 20 years of work experience. Data analysts also review data with unusually large and small values, called outliers, which are candidates for possible data errors.Errors often occur during data acquisition. Blindly using any data that happen to beavailable or using data that were acquired with little care can result in misleading informationand bad decisions. Thus, taking steps to acquire accurate data can help ensure reliableand valuable decision-making information.

**Types of Statistics**-

1. **Descriptive statistics** are the tabular, graphical, and numerical methods used to summarize data.
2. The process of **statistical inference** uses data obtained from a sample to make estimates or test hypotheses about the characteristics of a population.

**Population -** A population is the set of all elements of interest in a particular study.

**Sample -** A sample is a subset of the population.

**Census -** The process of conducting a survey to collect data for the entire population is called a census.

**Sample Survey -** The process of conducting a survey to collect data for a sample is called a sample survey.