* 1. **Statistics**

**Statistics** is the science of conducting studies to collect, organize, summarize, analyze, and draw conclusions from data.

**Why should students study statistics:**

1. To be able to read and understand the various statistical studies performed in their fields.
2. To be able to conduct research in their fields, since statistical procedures are basic to research.
3. To be able to use the knowledge gained from studying

Statistics to become better consumers and citizens.

* 1. **Descriptive and Inferential Statistics**

**Variable:** A variable is a characteristic or attribute that describes the event and can assume different values.

e.g. Average height of 13 year old boys/girls in US ---*Height is the variable*.

**Random variables:** variables whose values are determined by chance.

e.g. No. of students enrolling in Math 103 during a semester ---The admissions office studies its records over the past semesters and determines that, on average 34 students enroll for Math 103. Even though there is no sure way to know, how many students will enroll in any given semester but the office know the general pattern over the years and will assign a room for class meetings accordingly.

**Data:** Data are the values (measurements or observations) that the variable can assume.

**Data set:** A collection of data values. Each value in the data set is called a ***data value*** or a ***datum.***

Statistics is sometimes divided into two main areas, depending on how data are used:

**1.Descriptive statistics:** consists of the collection, organization, summarization, and presentation of data (in the form of charts, graphs, or tables).

e.g. National census conducted by US government every 10 years about average age, income, and other characteristics of US population.

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

Inferential statistics uses probability i.e. chance of an event occurring. e.g. people who play cards, dice, bingo, and lotteries win or lose according to the laws of probability.

**Population:** Consists of all subjects (human or otherwise) that are being studied.

**Sample:** is a group of subjects selected from a population.

e.g. a group of 13 year olds to collect data on height of 13 year olds in US.

* 1. **Variables and Types of Data**

# Qualitative variables are variables that can be placed into distinct categories, according to some characteristic or attribute.

e.g. subjects living in a particular geographic area ---geographic area is qualitative variable.

**Quantitative variables are** numerical and can be ordered or ranked.

e.g. weight of newborn babies in San Diego in 2004---weight is quantitative variable.

Quantitative variables can be further classified into two groups:

**1.Discrete variables:** assume values that can be counted.

e.g. no. of students in class.

**2. Continuous variables:** can assume all values between any two specific values. They are obtained by measuring.

e.g. temperature----it can assume all values between any two given temperatures.

**Data**

**Qualitative Quantitative**

**Discrete Continuous**

**\***Another classification of variables is based on how they are categorized, counted, or measured. This type of classification uses measurement scales, and four common types of scales are used:

1. The **nominal level of measurement** classifies data into mutually exclusive (non-overlapping), exhausting categories in which no order or ranking can be imposed on the data.

e.g. daughters and sons in a family.

1. The **ordinal level of measurement** classifies data into categories that can be ranked; however, precise differences between the ranks do not exist.

e.g. Floats in Rose parade ranked first place, second place etc.

1. The **interval level of measurement** ranks data, and precise differences between units of measures do exist; however there is no meaningful zero.

e.g. Temperature ----there is meaningful difference between 25o F and 26o F. 0o F does not mean no heat at all.

1. The **ratio level of measurement** possesses all the characteristics of interval measurement, and there exists a true zero. (e.g. weight can be zero)

e.g. Ratio between age of a father( 50 years old) and his son(25 years old) is 2 to 1.

* 1. **Data collection and Sampling Techniques**

**Data collection:**

Data can be collected in a variety of ways like conducting surveys, surveying records, direct observation of situations. One of the most common methods is through the use of surveys. Surveys can be done by using following methods:

1. **Telephone surveys** ---

less costly, people may be more open in their opinions since there is no face-to-face contact.

**↓** Some people may not have phones or will not be home when the call was made; so not all people got a chance to get surveyed.

1. **Mailed questionnaire surveys---**

can be used to cover a wider geographic area , less expensive, respondents can be anonymous if they want.

↓ people may not respond, some people may have difficulty understanding the questions.

**3.Personal interview surveys** ---

in depth response to questions

↓expensive because interviewers will have to be trained, interviewer may be biased in his selection of respondents.

**Sampling Techniques:**

Researchers use samples to collect data and information about a particular variable from a large population to save money and time. To obtain unbiased samples- i.e., give each subject in the population an equally likely chance of being selected—statisticians use four basic methods of sampling:

1. **Random Sampling:**  Random samples are selected by using chance methods or random numbers.

e.g. State lottery numbers are selected at random.

Crashers in Math 103—selected at random

1. **Systematic Sampling:** Systematic samples are obtained by numbering each subject of the population and then selecting every *n*th subject.

e.g. Let there are 1000 students in a community college and a sample of 20 students is needed. Since 1000÷20 =50, then n=50 and every 50th student will be selected.

1. **Stratified Sampling:** Stratified samples are obtained by dividing the population into groups (called strata) according to some characteristic that is important to the study, then sampling from each group.

e.g. A School newspaper writer wants to know students’ opinion about the idea of having school uniform. Again she wants to know how opinions of boys differ from those of girls. She will select students from both groups to use in the sample.

1. **Cluster Sampling:** Researchers select cluster samples by using intact groups called clusters.

e.g. Selecting a few hospitals in San Diego at random to conduct research on cancer patients.

**Observational and Experimental Studies:**

Following are two types of statistical studies:

1. **Observational studies:** In an observational study, the researcher merely observes what is happening or what has happened in the past and tries to draw conclusions based on these observations.

e.g. How real estate prices have increased in San Diego over the last 10 years. The researcher merely observed the prices and there was no research intervention.

1. **Experimental Studies:** In an experimental study, the researcher manipulates (controls) one of the variables and tries to determine how the manipulation influences other variables.

e.g. A karate instructor told one group of students to practice as much as they could and gave specific directions to another group to practice 1 hours each day. The performance of second group was better than the first one after one month. The researchers conducting the study manipulated the variable ----type of directions—given to each group.