# Chapter 1

# Introduction to Statistics

## Beginning Definitions & Terminology

- **Statistics** is the "science of data". It includes the methods of collecting, organizing & summarizing, and analyzing data (or information).

## Beginning Definitions & Terminology

- Generally we break the study of statistics into two parts:
  - **Descriptive Statistics**: The organizing & summarizing part; and
  - **Inferential Statistics**: The analyzing (making inferences) part.

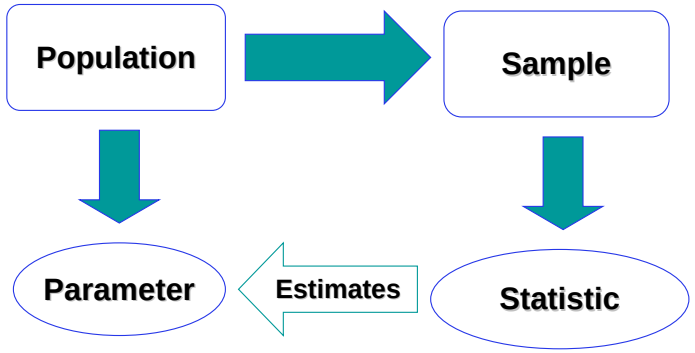
## Beginning Definitions & Terminology

- **Population**: The complete collection of all data to be studied.
- **Sample**: A subset of the population.
- **Census**: A sample of the entire population.

## Beginning Definitions & Terminology

- A **parameter** is a value calculated from the population.
- A **statistic** is a value calculated from a sample.

## The Big Picture




## Beginning Definitions & Terminology

- **Data** are observations (information) that have been collected or observed.
- Data can be classified as:
  - **Quantitative**: Numerical data that represents *counts* or *measurements*. (ex: weights of adult male elephants)
  - **Qualitative**: Categorical data. Non-numerical. (ex: The colors of Skittles candies)

## Beginning Definitions & Terminology

- Data can also be classified as:
  - **Discrete**: Data with finite or countable values. (ex: the number of eggs collected each day in a chicken coop)
  - **Continuous**: Data with values that are not countable and correspond to some continuous scale. (ex: Amount of milk collected per day on a dairy farm)

## Levels of Measurement for Data

- **Nominal (Name-only)**

Data consists of names, labels, or categories.

Example: Names of this year's movies

## Levels of Measurement for Data

- **Ordinal**

Data can be arranged in some order, but differences between values cannot be determined or are meaningless.

Example: Movie ratings of excellent, good, or bad

## Levels of Measurement for Data

- **Interval**

Like ordinal, but differences between data values are meaningful (make sense). However, there is no natural starting value (zero).

Example: Temperatures in degrees Fahrenheit

## Levels of Measurement for Data

- **Ratio**

Like interval, but with an inherent zero.

Examples: weights, lengths, volumes, prices, speed

## Levels of Measurement for Data

Table 1-1 on page 10 provides a good summary of the levels of measurement for data.

## Critically Thinking About Statistics

- Success in the introductory statistics course typically requires more common sense than mathematical expertise.

## Misuses & Abuses of Statistics

"There are three kinds of lies: lies, damned lies, and statistics" – Benjamin Disraeli

"Statistics can be used to support anything – especially statisticians" – Franklin P. Jones

## Misuses & Abuses of Statistics

- How the data is collected has a direct impact on its reliability and usefulness.
- A **voluntary response sample** (or **self-selected sample**) is one in which the respondents themselves decide whether to be included.

## Misuses & Abuses of Statistics

- Areas to look out for when analyzing statistics:
  - Sample size
  - Graphs & Pictographs
  - Percentages
  - Form and order of questions
  - Confounding

## Design of Experiments

- In an **observational study** we observe and measure specific characteristics, but we don't attempt to *modify* the subjects being studied.
- In a **controlled experiment** we apply some *treatment* and then proceed to observe its effects on the subjects.

## Controlling Experiments

- The **placebo effect** occurs when an untreated subject believes that he or she is receiving a real treatment and reports an improvement in symptoms.
- The placebo effect can be minimized by the use of **blinding** where the subject doesn't know whether he or she is receiving a treatment or a placebo.

## Sampling Techniques

- If sample data are not collected in an appropriate way, the data may be completely useless.

## Sampling Techniques

- **Random Sampling**

Each data value in population has equally likely chance of being sampled.

Example: Randomly pick phone numbers to call

## Sampling Techniques

- A **simple random sample** of size *n* is selected in such a way that every sample of size *n* has the same chance of being chosen.

## Sampling Techniques

- **Systematic Sampling**

Select every *k*th member.

Example: Selecting every 10th person coming out of a restaurant

## Sampling Techniques

- **Stratified Sampling**

Break population up into two or more strata and randomly sample from *each*.

Example: Selecting 30 males & 30 females for an opinion poll

## Sampling Techniques

- **Cluster Sampling**

Break population up in several clusters, randomly select some of those clusters & select *all* members in them.

Example: Randomly select 15 classrooms at PSTCC & select all students in them

## Sampling Techniques

- **Convenience Sampling**

Select whatever is convenient.

Example: An instructor surveys the 27 students in her classroom.

## Types of Errors

- **Sampling Error**
  the difference between a sample result and the true population result; such an error results from chance sample fluctuations
- **Nonsampling Error**
  sample data that are incorrectly collected, recorded, or analyzed (such as by selecting a biased sample, using a defective instrument, or copying the data incorrectly)