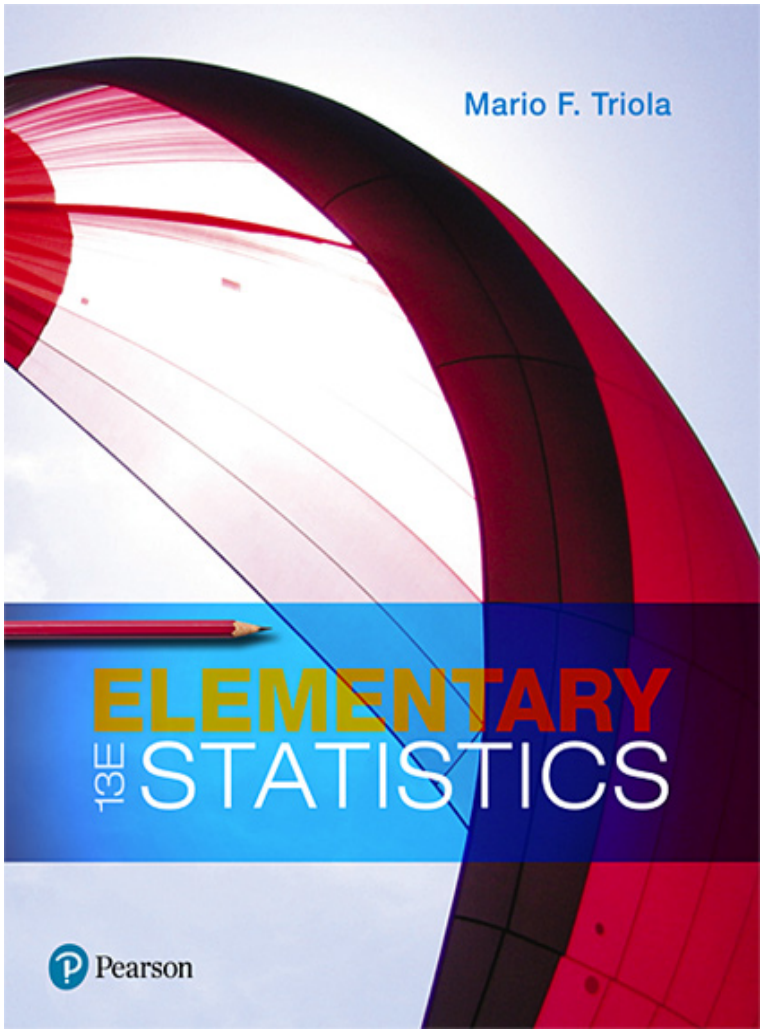
Elementary Statistics

Thirteenth Edition



Chapter 1

Introduction to Statistics

Introduction to Statistics

1-1 Statistical and Critical Thinking

1-2 Types of Data

1-3 Collecting Sample Data

Key Concept

The method used to collect sample data influences the quality of the statistical analysis.

Of particular importance is the **simple random sample**.

If sample data are not collected in an appropriate way, the data may be so utterly useless that no amount of statistical torturing can salvage them.

The Gold Standard

Randomization with placebo/treatment groups is sometimes called the “gold standard” because it is so effective. (A placebo such as a sugar pill has no medicinal effect.)

Basics of Collecting Data

Statistical methods are driven by the data that we collect. We typically obtain data from two distinct sources: **observational studies** and **experiments**.

Experiment

- Experiment
  - apply some **treatment** and then proceed to observe its effects on the individuals. (The individuals in experiments are called experimental units, and they are often called subjects when they are people.)

Observational Study

- Observational study
  - observing and measuring specific characteristics without attempting to **modify** the individuals being studied

Example: Ice Cream and Drownings (1 of 2)

- Observational Study:

  Observe past data to conclude that ice cream causes drownings (based on data showing that increases in ice cream sales are associated with increases in drownings). The mistake is to miss the lurking variable of temperature and the failure to see that as the temperature increases, ice cream sales increase and drownings increase because more people swim.

Example: Ice Cream and Drownings (2 of 2)

- Experiment:
  - Conduct an **experiment** with one group treated with ice cream while another group gets no ice cream. We would see that the rate of drowning victims is about the same in both groups, so ice cream consumption has no effect on drownings.
  - Here, the experiment is clearly better than the observational study.

Design of Experiments (1 of 4)

- Replication
  - Replication is the repetition of an experiment on more than one individual.
  - Good use of replication requires sample sizes that are large enough so that we can see effects of treatments.

Design of Experiments (2 of 4)

- Blinding
  - Blinding is a technique in which the subject doesn't know whether he or she is receiving a treatment or a placebo.
  - Blinding is a way to get around the placebo effect, which occurs when an untreated subject reports an improvement in symptoms.

Design of Experiments (3 of 4)

- Double-Blind
  - Blinding occurs at two levels:
    1. The subject doesn't know whether he or she is receiving the treatment or a placebo.
    2. The experimenter does not know whether he or she is administering the treatment or placebo.

Design of Experiments (4 of 4)

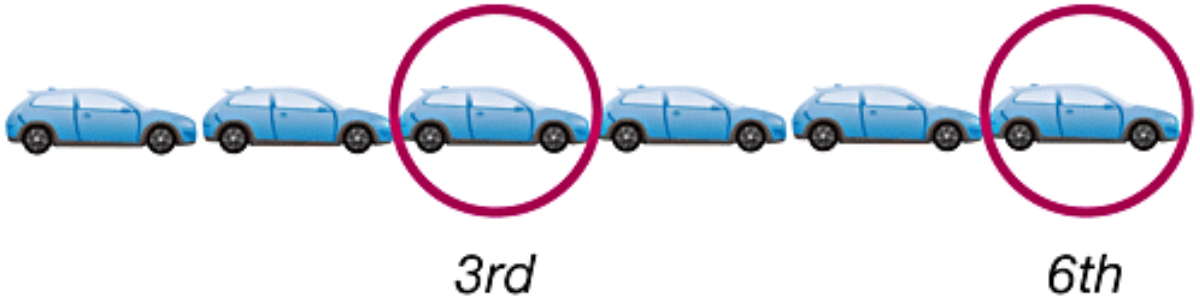
- Randomization
  - Randomization is used when subjects are assigned to different groups through a process of random selection. The logic is to use chance as a way to create two groups that are similar.

Simple Random Sample

- Simple Random Sample
  - A sample of *n* subjects is selected in such a way that every possible **sample of the same size** *n* has the same chance of being chosen.
  - A simple random sample is often called a random sample, but strictly speaking, a **random sample** has the weaker requirement that all members of the population have the same chance of being selected.

Systematic Sampling

- Select some starting point and then select every *k*th element in the population.

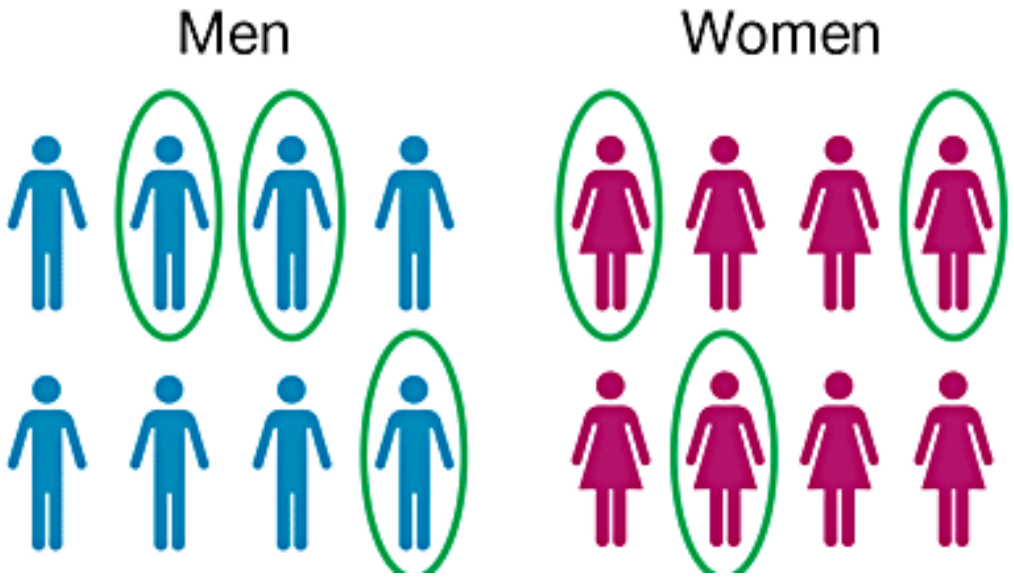


Convenience Sampling

- Use data that are very easy to get.

Stratified Sampling

Subdivide the population into at least two different subgroups (or strata) so that the subjects within the same subgroup share the same characteristics. Then draw a sample from each subgroup (or stratum).



Cluster Sampling

Divide the population area into sections (or clusters), then randomly select some of those clusters, and choose **all** the members from those selected clusters.

Multistage Sampling

Collect data by using some combination of the basic sampling methods.

In a multistage sample design, pollsters select a sample in different stages, and each stage might use different methods of sampling.

Observational Studies

Observe and measure, but do not modify.

Types of Observational Studies

- Cross-sectional study
  - Data are observed, measured, and collected at one point in time, not over a period of time.
- Retrospective (or case control) study
  - Data are collected from a past time period by going back in time (through examination of records, interviews, and so on).
- Prospective (or longitudinal or cohort) study
  - Data are collected in the future from groups sharing common factors (called **cohorts**).

Confounding

- Confounding
  - occurs in an experiment when the experimenter is not able to distinguish between the effects of different factors.
  - Try to plan the experiment so that confounding does not occur.

Controlling Effects of Variables (1 of 2)

- Completely Randomized Experimental Design
  - Assign subjects to different treatment groups through a process of **random selection**.
- Randomized Block Design
  - A block is a group of subjects that are similar, but blocks differ in ways that might affect the outcome of the experiment.

Controlling Effects of Variables (2 of 2)

- Matched Pairs Design
  - Compare two treatment groups by using subjects matched in pairs that are somehow related or have similar characteristics.
- Rigorously Controlled Design
  - Carefully assign subjects to different treatment groups, so that those given each treatment are similar in ways that are important to the experiment.

Sampling Errors (1 of 2)

- No matter how well you plan and execute the sample collection process, there is likely to be some error in the results.
- Sampling error (or random sampling error)
  - occurs when the sample has been selected with a random method, but there is a discrepancy between a sample result and the true population result; such an error results from chance sample fluctuations.

Sampling Errors (2 of 2)

- Nonsampling error
  - Nonsampling is the result of human error, including such factors as wrong data entries, computing errors, questions with biased wording, false data provided by respondents, forming biased conclusions, or applying statistical methods that are not appropriate for the circumstances.
- Nonrandom sampling error
  - Nonrandom sampling error is the result of using a sampling method that is not random, such as using a convenience sample or a voluntary response sample.