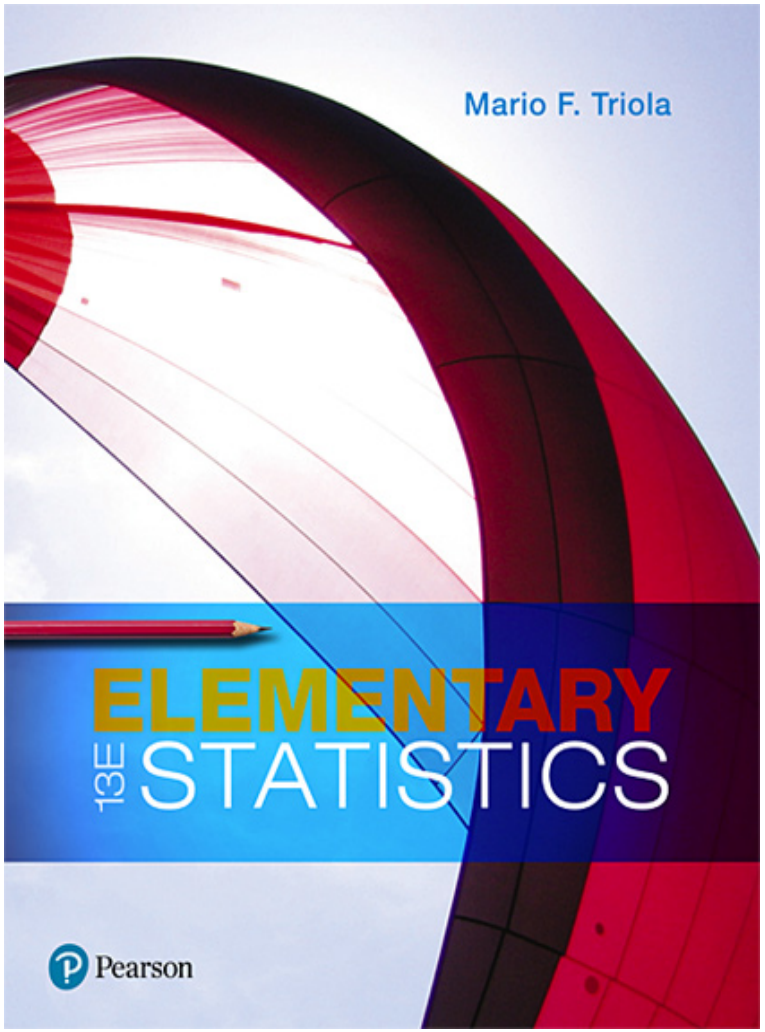
Elementary Statistics

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 process involved in conducting a statistical study consists of “prepare, analyze, and conclude.”

Statistical thinking involves critical thinking and the ability to make sense of results. Statistical thinking demands so much more than the ability to execute complicated calculations.

Data

- Data
  - Collections of observations, such as measurements, genders, or survey responses

Statistics

- **S**tatistics
  - The science of planning studies and experiments, obtaining data, and organizing, summarizing, presenting, analyzing, and interpreting those data and then drawing conclusions based on them.

Population

- Population
  - The complete collection of **all** measurements or data that are being considered. Typically, a population is the complete collection of data that we would like to make inferences about.

Census versus Sample

- Census
  - The collection of data from **every** member of a population
- Sample
  - A **subcollection** of members selected from a population

Example: Residential Carbon Monoxide Detectors (1 of 2)

In the journal article “Residential Carbon Monoxide Detector Failure Rates in the United States”, it was stated that there are 38 million carbon monoxide detectors installed in the United States. When 30 of them were randomly selected and tested, it was found that 12 of them failed to provide an alarm in hazardous carbon monoxide conditions.

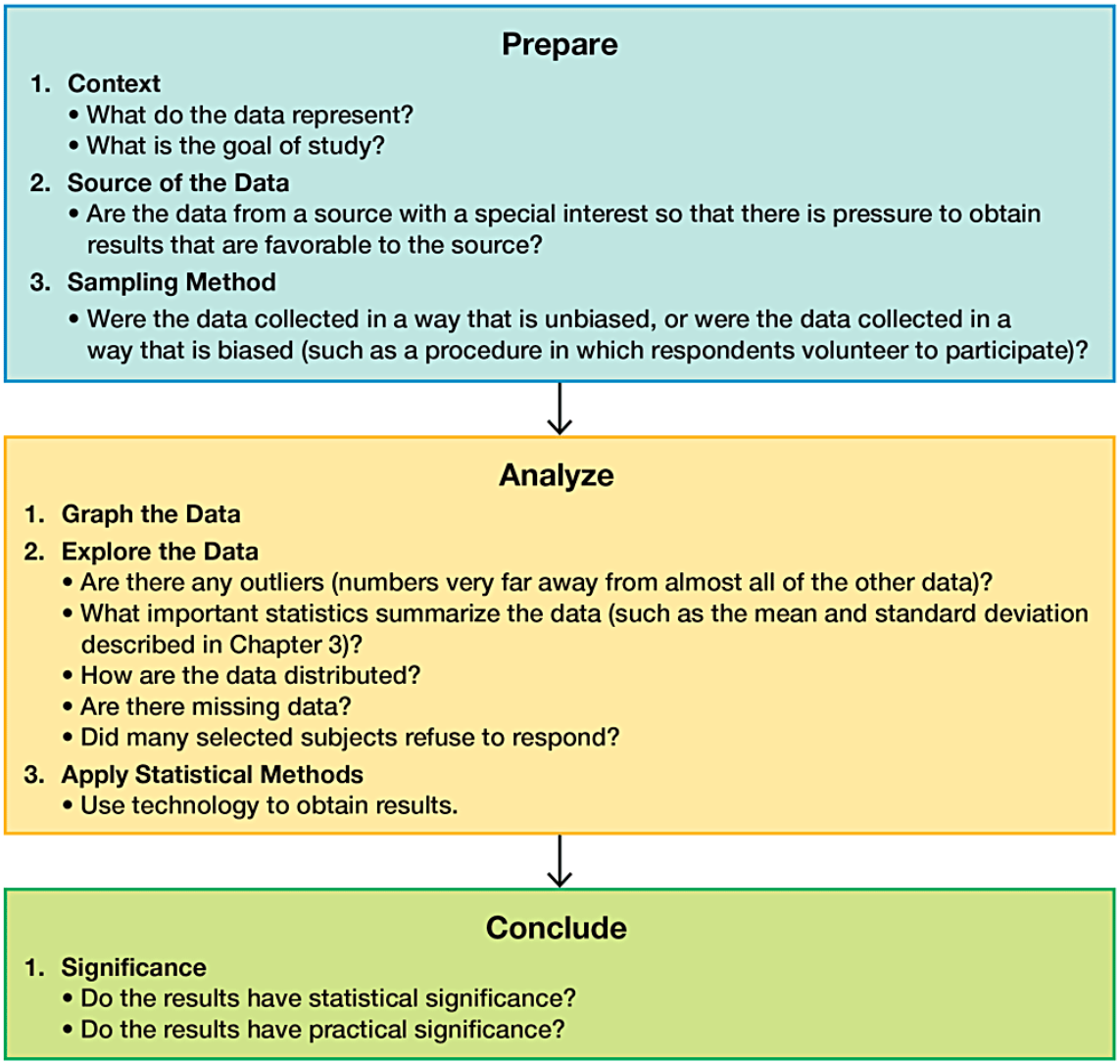
Example: Residential Carbon Monoxide Detectors (2 of 2)

In this case, the population and sample are as follows:

- **Population:** All 38 million carbon monoxide detectors in the United States
- **Sample:** The 30 carbon monoxide detectors that were selected and tested

The objective is to use the sample data as a basis for drawing a conclusion about the population of all carbon monoxide detectors, and methods of statistics are helpful in drawing such conclusions.

Statistical and Critical Thinking



Prepare (1 of 2)

Pleasure Boats and Manatee Fatalities from Boat Encounters

Pleasure Boats (ten of thousands)	99	99	97	95	90	90	87	90	90
Manatee Fatalities	92	73	90	97	83	88	81	73	68

- **Context**
  - The table includes the number of registered pleasure boats in Florida (tens of thousands) and the number of manatee fatalities from encounters with boats in Florida for each of several recent years.
  - The format of the table suggests the following goal: Determine whether there is a **relationship** between numbers of boats and numbers of manatee deaths from boats.

Prepare (2 of 2)

Pleasure Boats and Manatee Fatalities from Boat Encounters

Pleasure Boats (ten of thousands)	99	99	97	95	90	90	87	90	90
Manatee Fatalities	92	73	90	97	83	88	81	73	68

- **Source of the Data**
  - The data in the table are from the Florida Department of Highway Safety and Motor Vehicles and the Florida Marine Research Institute. The sources certainly appear to be reputable.
- **Sampling Method**
  - The data were obtained from official government records known to be reliable. The sampling method appears to be sound.

Voluntary Response Sample (1 of 2)

- Voluntary Response Sample
  - **Voluntary Response Sample** or **Self-Selected Sample** is one in which the respondents themselves decide whether to be included.

Voluntary Response Sample (2 of 2)

The following types of polls are common examples of voluntary response samples. By their very nature, all are seriously flawed because we should not make conclusions about a population on the basis of samples with a strong possibility of bias:

- Internet polls, in which people online can decide whether to respond
- Mail-in polls, in which people can decide whether to reply
- Telephone call-in polls, in which newspaper, radio, or television announcements ask that you voluntarily call a special number to register your opinion

Example: Voluntary Response Sample (1 of 2)

Nightline asked viewers to call with their opinion about whether the UN headquarters should remain in the United States. Viewers then decided themselves whether to call with their opinions, and 67% of 186,000 respondents said that the UN should be moved out of the United States.

In a separate, independent survey, 500 respondents were randomly selected and surveyed, and 38% of this group wanted the UN to move out of the United States.

Example: Voluntary Response Sample (2 of 2)

The two polls produced dramatically different results. Even though the **Nightline** poll involved 186,000 volunteer respondents, the much smaller poll of 500 randomly selected respondents is more likely to provide better results because of the far superior sampling method.

Analyze

After completing our preparation by considering the context, source, and sampling method, we begin to **analyze** the data.

- **Graph and Explore**
  - An analysis should begin with appropriate graphs and explorations of the data.
- **Apply Statistical Methods**
  - A good statistical analysis **does not** require strong computational skills. A good statistical analysis **does** require using common sense and paying careful attention to sound statistical methods.

Conclude (1 of 2)

The final step in our statistical process involves conclusions, and we should develop an ability to distinguish between statistical significance and practical significance.

- **Statistical Significance**
  - **Statistical significance** is achieved in a study if the likelihood of an event occurring by chance is 5% or less.
    - Getting 98 girls in 100 random births *is* statistically significant because such an extreme outcome is not likely to result from random chance.
    - Getting 52 girls in 100 births *is not* statistically significant because that event could easily occur with random chance.

Conclude (2 of 2)

- **Practical Significance**
  - It is possible that some treatment or finding is effective, but common sense might suggest that the treatment or finding does not make enough of a difference to justify its use or to be practical.

Example Statistical Significance Versus Practical Significance (1 of 2)

ProCare Industries once supplied a product named Gender Choice that supposedly increased the chance of a couple having a baby with the gender that they desired. In the absence of any evidence of its effectiveness, the product was banned by the Food and Drug Administration (FDA) as a "gross deception of the consumer."

- Suppose that the product was tested with 10,000 couples who wanted to have baby girls, and the results consist of 5200 baby girls born in the 10,000 births. This result is statistically significant because the likelihood of it happening due to chance is only 0.003%, so chance doesn't seem like a feasible explanation.

Example Statistical Significance Versus Practical Significance (2 of 2)

- That 52% rate of girls is statistically significant, but it lacks practical significance because 52% is only slightly above 50%. Couples would not want to spend the time and money to increase the likelihood of a girl from 50% to 52%. (**Note**: In reality, the likelihood of a baby being a girl is about 48.8%, not 50%.)

Analyzing Data: Potential Pitfalls (1 of 2)

- **Misleading Conclusions**
  - When forming a conclusion based on a statistical analysis, we should make statements that are clear even to those who have no understanding of statistics and its terminology.
- **Sample Data Reported Instead of Measured**
  - When collecting data from people, it is better to take measurements yourself instead of asking subjects to **report** results.
- **Loaded Questions**
  - If survey results are not worded carefully, the results of a study can be misleading.

Analyzing Data: Potential Pitfalls (2 of 2)

- **Order of Questions**
  - Sometimes survey questions are unintentionally loaded by the order of the items being considered.
- **Nonresponse**
  - A nonresponse occurs when someone either refuses to respond or is unavailable.
- **Percentages**
  - Some studies cite misleading percentages. Note that 100% of some quantity is **all** of it, but if there are references made to percentages that exceed 100%, such references are often not justified.