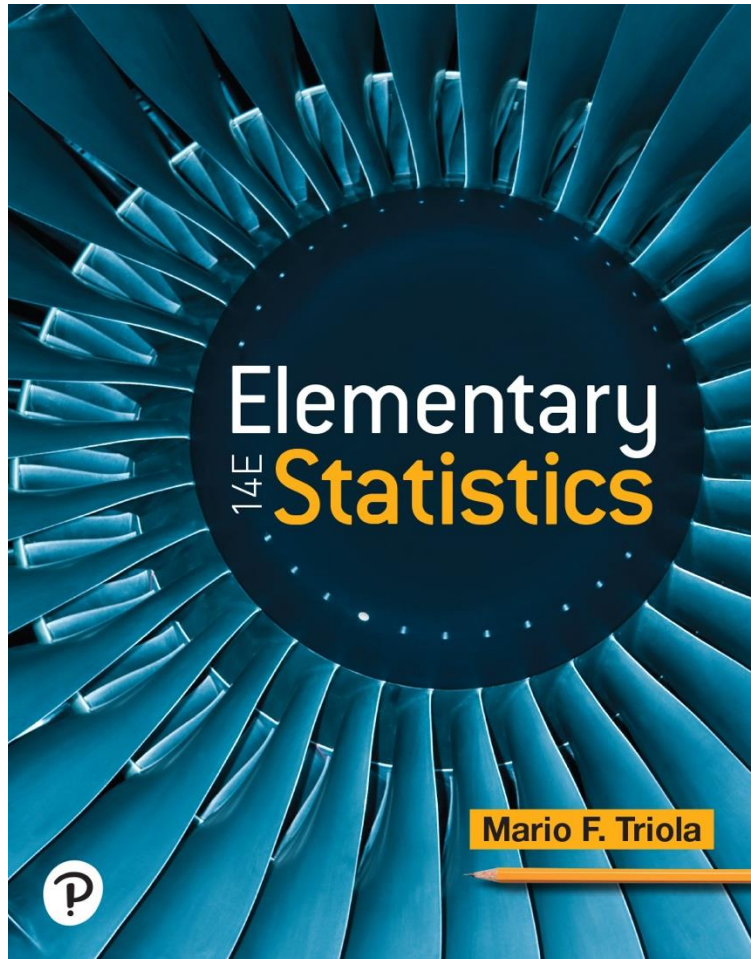


Elementary Statistics

Fourteenth Edition



Chapter 1

Introduction to Statistics

Introduction to Statistics

1-1 Statistical and Critical Thinking

1-2 Types of Data

Key Concept

Know and understand the meanings of the terms *statistic* and *parameter*.

Describe different data types. The type of data is one of the key factors that determine the statistical methods used in the analysis of the data.

Parameter

- Parameter
 - a numerical measurement describing some characteristic of a **population**

Statistic

- Statistic
 - a numerical measurement describing some characteristic of a **sample**

Quantitative Data

- Quantitative (or numerical) data
 - consist of **numbers** representing counts or measurements.

Examples: weight, age

Categorical Data

- Categorical (or qualitative or attribute) data
 - consist of names or labels (not numbers that represent counts or measurements).

Examples: marital status, state of residence

Working with Quantitative Data

Quantitative data can be further described by distinguishing between **discrete** and **continuous** types.

Discrete Data

- Discrete data
 - result when the data values are quantitative and the number of values is finite, or “countable.”

Example: The number of tosses of a coin before getting tails

Continuous Data

- Continuous (numerical) data
 - result from infinitely many possible quantitative values, where the collection of values is not countable.

Example: The lengths of distances from 0 cm to 12 cm

Levels of Measurement

- Another way of classifying data is to use four levels of measurement: nominal, ordinal, interval, and ratio.

Nominal Level

- Nominal level of measurement
 - characterized by data that consist of names, labels, or categories only. The data cannot be arranged in some order (such as low to high).

Example: Survey responses of **yes**, **no**, and **undecided**

Ordinal Level

- Ordinal level of measurement
 - involves data that can be arranged in some order, but differences (obtained by subtraction) between data values either cannot be determined or are meaningless.

Example: Course grades A, B, C, D, or F

Interval Level

- Interval level of measurement
 - involves data that can be arranged in order, and the differences between data values can be found and are meaningful. However, there is no **natural** zero starting point at which none of the quantity is present.

Example: Years 1000, 2000, 1776, and 1492

Ratio Level

- Ratio level of measurement
 - data can be arranged in order, differences can be found and are meaningful, and there is a natural zero starting point (where zero indicates that none of the quantity is present). Differences and ratios are both meaningful.

Example: Class times of 50 minutes and 100 minutes

Summary - Levels of Measurement

- **Nominal** - categories only
- **Ordinal** - categories with some order
- **Interval** - differences but no natural zero point
- **Ratio** - differences and a natural zero point

Big Data

- Big data
 - refers to data sets so large and so complex that their analysis is beyond the capabilities of traditional software tools. Analysis of big data may require software simultaneously running in parallel on many different computers.
- Data science
 - involves applications of statistics, computer science, and software engineering, along with some other relevant fields (such as sociology or finance).

Missing Data

- A data value is **missing completely at random** if the likelihood of its being missing is independent of its value or any of the other values in the data set. That is, any data value is just as likely to be missing as any other data value.
- A data value is **missing not at random** if the missing value is related to the reason that it is missing.