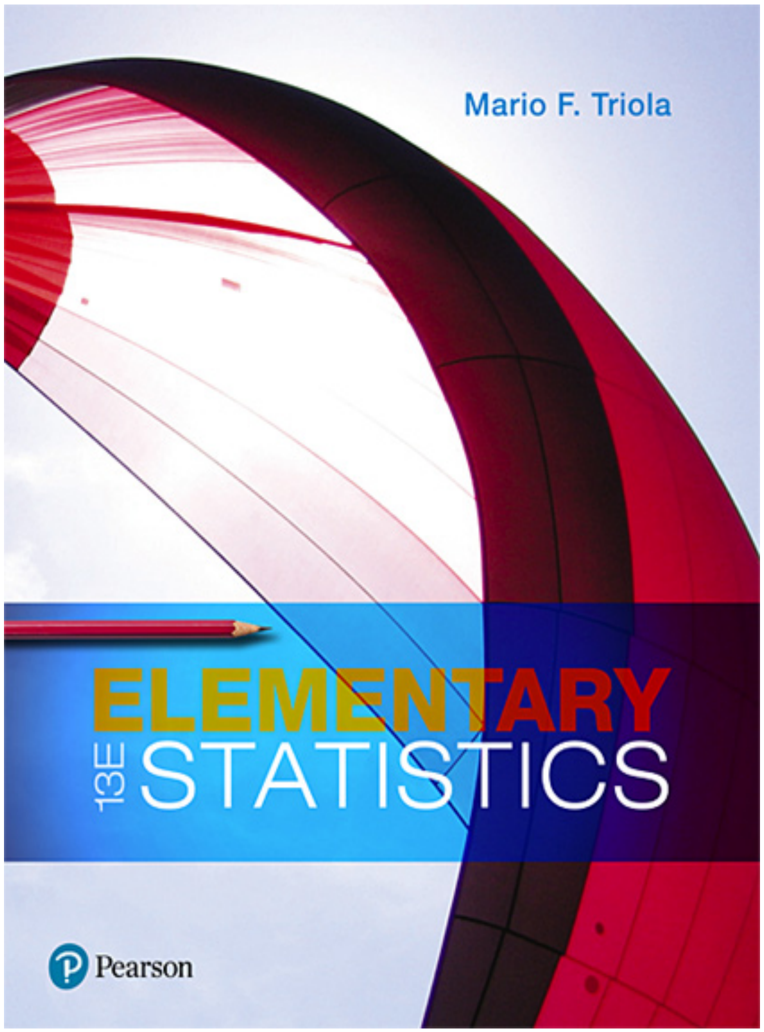
Elementary Statistics

Thirteenth Edition



Chapter 4

Probability

Probability

4-1 Basic Concepts of Probability

4-2 Addition Rule and Multiplication Rule

4-3 Complements and Conditional Probability, and Bayes' Theorem

4-4 Counting

4-5 Probabilities Through Simulations (available at TriloaStats.com)

Key Concept

We extend the use of the multiplication rule to include the probability that among several trials, we get **at least one** of some specified event. We consider **conditional probability:** the probability of an event occurring when we have additional information that some other event has already occurred. We provide a brief introduction to Bayes' theorem.

Complements: The Probability of "At Least One"

When finding the probability of some event occurring "at least once," we should understand the following:

- "At least one" has the same meaning as "one or more."
- The **complement** of getting "at least one" particular event is that you get **no** occurrences of that event.

The Probability of "At Least One" (1 of 2)

Finding the probability of getting **at least one** of some event:

1. Let A = getting **at least one** of some event.
2. Then $\overline{A}$ = getting **none** of the event being considered.

The Probability of "At Least One" (2 of 2)

3. Find $P(\bar{A})$ = probability that event *A* does not occur.

4. Subtract the result from 1. That is, evaluate this expression:

P(**at least one** occurrence of event *A*) = 1 − *P*(**no** occurrences of event *A*)

Example: Accidental iPad Damage (1 of 4)

A study by SquareTrade found that 6% of damaged iPads were damaged by “bags/backpacks.” If 20 damaged iPads are randomly selected, find the probability of getting **at least one** that was damaged in a bag/backpack. Is the probability high enough so that we can be reasonably sure of getting at least one iPad damaged in a bag/backpack?

Example: Accidental iPad Damage (2 of 4)

Solution

Step 1: Let A = at least 1 of the 20 damaged iPads was damaged in a bag/backpack.

Step 2: Identify the event that is the complement of $\overline{A}$.

$\overline{A}$ = **not** getting at least 1 iPad damaged in a bag/backpack among 20

= all 20 iPads damaged in a way other than bag/backpack

Example: Accidental iPad Damage (3 of 4)

Solution

Step 3: Find the probability of the complement by evaluating $P(\bar{A})$.

$P(\bar{A}) = P(\text{all 20 iPads damaged in a way other than bag/backpack})$

$$= 0.94 \cdot 0.94 \cdot \ldots \cdot 0.94$$

$$= 0.94^{20} = 0.290$$

Step 4: Find $P(A)$ by evaluating $1 - P(\bar{A})$.

$$P(A) = 1 - P(\bar{A}) = 1 - 0.290 = 0.710$$

Example: Accidental iPad Damage (4 of 4)

Interpretation

For a group of 20 damaged iPads, there is a 0.710 probability of getting at least 1 iPad damaged in a bag/backpack.

This probability is not **very** high, so to be **reasonably sure** of getting at least 1 iPad damaged in a bag/backpack, more than 20 damaged iPads should be used.

Conditional Probability (1 of 2)

- Conditional Probability
  - A **conditional probability** of an event is a probability obtained with the additional information that some other event has already occurred.

Conditional Probability (2 of 2)

Notation

$P(B \mid A)$ denotes the conditional probability of event B occurring, given that event A has already occurred.

Intuitive Approach for Finding $P(B \mid A)$

The conditional probability of *B* occurring given that *A* has occurred can be found by **assuming that event A has occurred** and then calculating the probability that event *B* will occur.

Formal Approach for Finding *P*(*B* | *A*)

The probability $P(B \mid A)$ can be found by dividing the probability of events *A* and *B* both occurring by the probability of event *A:*

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$$

Example: Pre-Employment Drug Screening (1 of 11)

Find the following using the table:

a. If 1 of the 555 test subjects is randomly selected, find the probability that the subject had a positive test result, given that the subject actually uses drugs. That is, find *P*(positive test result | subject uses drugs).

	Positive Test Result (Test shows drug use.)	**Negative Test Result** (Test shows no drug use.)
Subject Uses Drugs	45 (True Positive)	5 (False Negative)
Subject Does Not Use Drugs	25 (False Positive)	480 (True Negative)

Example: Pre-Employment Drug Screening (2 of 11)

Find the following using the table:

b. If 1 of the 555 test subjects is randomly selected, find the probability that the subject actually uses drugs, given that he or she had a positive test result. That is, find P(subject uses drugs | positive test result).

	Positive Test Result (Test shows drug use.)	**Negative Test Result** (Test shows no drug use.)
Subject Uses Drugs	45 (True Positive)	5 (False Negative)
Subject Does Not Use Drugs	25 (False Positive)	480 (True Negative)

Example: Pre-Employment Drug Screening (3 of 11)

Solution

a. **Intuitive Approach:** We want P(positive test result | subject uses drugs). If we assume that the selected subject actually uses drugs, we are dealing only with the 50 subjects in the first row of the table. Among those 50 subjects, 45 had positive test results, so we get this result:

$$P(\text{positive test result} \mid \text{subject uses drugs})$$

$$= \frac{45}{50} = 0.900$$

Example: Pre-Employment Drug Screening (4 of 11)

Solution

Formal Approach: The same result can be found by using the formula for $P(B \mid A)$ given with the formal approach. $P(B \mid A) = P(\text{positive test result} \mid \text{subject uses drugs})$ where B = positive test result and A = subject uses drugs.

Example: Pre-Employment Drug Screening (5 of 11)

Solution

Use P(subject uses drugs and had a positive test result) $= \frac{45}{555}$ and

P(subject uses drugs) $= \frac{50}{555}$ to get the following results:

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$$

Example: Pre-Employment Drug Screening (6 of 11)

Solution

becomes

P(positive test result | subject uses drugs)

$$= \frac{P(\text{subject uses drugs and had a positive test result})}{P(\text{subject uses drugs})}$$

$$= \frac{\frac{45}{555}}{\frac{50}{555}} = 0.900$$

Example: Pre-Employment Drug Screening (7 of 11)

Solution

By comparing the intuitive approach to the formal approach, it should be clear that the intuitive approach is much easier to use, and it is also less likely to result in errors.

Example: Pre-Employment Drug Screening (8 of 11)

Solution

b. Here we want *P*(subject uses drugs | positive test result). This is the probability that the selected subject uses drugs, **given that the subject had a positive test result.** If we assume that the subject had a positive test result, we are dealing with the 70 subjects in the first column of the table.

Example: Pre-Employment Drug Screening (9 of 11)

Solution

Among those 70 subjects, 45 use drugs, so

P(subject uses drugs | positive test result)

$$= \frac{45}{70}$$
$$= 0.643$$

Example: Pre-Employment Drug Screening (10 of 11)

Interpretation

The first result of P(positive test result | subject uses drugs) = 0.900 indicates that a subject who uses drugs has a 0.900 probability of getting a positive test result.

Example: Pre-Employment Drug Screening (11 of 11)

Interpretation

The second result of *P*(subject uses drugs | positive test result) = 0.643 indicates that for a subject who gets a positive test result, there is a 0.643 probability that this subject actually uses drugs. Note that *P*(positive test result | subject uses drugs) ≠ *P*(subject uses drugs | positive test result).

Confusion of the Inverse

In the prior example, P(positive test result | subject uses drugs) $\neq$ P(subject uses drugs | positive test result). This example proves that in general, $P(B \mid A) \neq P(A \mid B)$. (There could be individual cases where $P(A \mid B)$ and $P(B \mid A)$ are equal, but they are generally not equal.) To incorrectly think that $P(B \mid A)$ and $P(A \mid B)$ are equal or to incorrectly use one value in place of the other is called **confusion of the inverse.**

Example: Confusion of the Inverse (1 of 2)

Consider these events:

D: It is dark outdoors.

M: It is midnight.

Example: Confusion of the Inverse (2 of 2)

In the following, we conveniently ignore the Alaskan winter and other such anomalies.

$P(D \mid M) = 1$ (It is certain to be dark given that it is midnight.)

$P(M \mid D) = 0$ (The probability that it is exactly midnight given that it dark is almost zero.)

Here, $P(D \mid M) \neq P(M \mid D)$.

Confusion of the inverse occurs when we incorrectly switch those probability values or think that they are equal.

Bayes' Theorem

We extend the discussion of conditional probability to include applications of **Bayes' theorem** (or **Bayes' rule**), which we use for revising a probability value based on additional information that is later obtained.

Example: Interpreting Medical Test Results (1 of 12)

Assume cancer has a 1% prevalence rate, meaning that 1% of the population has cancer. Denoting the event of having cancer by C, we have $P(C) = 0.01$ for a subject randomly selected from the population.

Example: Interpreting Medical Test Results (2 of 12)

This result is included with the following performance characteristics of the test for cancer.

- $P(C) = 0.01$ (There is a 1% prevalence rate of the cancer.)
- The false positive rate is 10%. That is, *P*(positive test result given that cancer is not present) = 0.10.
- The true positive rate is 80%. That is, *P*(positive test result given that cancer is present) = 0.80.

Example: Interpreting Medical Test Results (3 of 12)

Find $P(C \mid \text{positive test result})$. That is, find the probability that a subject actually has cancer given that he or she has a positive test result.

Example: Interpreting Medical Test Results (4 of 12)

Solution

Using the given information, we can construct a hypothetical population with the above characteristics.

	Positive Test Result (Test shows cancer.)	**Negative Test Result (test shows no cancer)**	**Total**
Cancer	8 (True Positive)	2 (False Negative)	10
No Cancer	99 (False Positive)	891 (True Negative)	990

Example: Interpreting Medical Test Results (5 of 12)

Solution

- Assume that we have 1000 subjects. With a 1% prevalence rate, 10 of the subjects are expected to have cancer. The sum of the entries in the first row of values is therefore 10.

	Positive Test Result (Test shows cancer.)	Negative Test Result (test shows no cancer)	Total
Cancer	8 (True Positive)	2 (False Negative)	10
No Cancer	99 (False Positive)	891 (True Negative)	990

Example: Interpreting Medical Test Results (6 of 12)

Solution

- The other 990 subjects do not have cancer. The sum of the entries in the second row of values is therefore 990.

	Positive Test Result (Test shows cancer.)	Negative Test Result (test shows no cancer)	Total
Cancer	8 (True Positive)	2 (False Negative)	10
No Cancer	99 (False Positive)	891 (True Negative)	990

Example: Interpreting Medical Test Results (7 of 12)

Solution

- Among the 990 subjects without cancer, 10% get positive test results, so 10% of the 990 cancer-free subjects in the second row get positive test results. See the entry of 99 in the second row.

	Positive Test Result (Test shows cancer.)	**Negative Test Result (test shows no cancer)**	**Total**
Cancer	8 (True Positive)	2 (False Negative)	10
No Cancer	99 (False Positive)	891 (True Negative)	990

Example: Interpreting Medical Test Results (8 of 12)

Solution

- For the 990 subjects in the second row, 99 test positive, so the other 891 must test negative. See the entry of 891 in the second row.

	Positive Test Result (Test shows cancer.)	Negative Test Result (test shows no cancer)	Total
Cancer	8 (True Positive)	2 (False Negative)	10
No Cancer	99 (False Positive)	891 (True Negative)	990

Example: Interpreting Medical Test Results (9 of 12)

Solution

- Among the 10 subjects with cancer in the first row, 80% of the test results are positive, so 80% of the 10 subjects in the first row test positive. See the entry of 8 in the first row.

	Positive Test Result (Test shows cancer.)	Negative Test Result (test shows no cancer)	Total
Cancer	8 (True Positive)	2 (False Negative)	10
No Cancer	99 (False Positive)	891 (True Negative)	990

Example: Interpreting Medical Test Results (10 of 12)

Solution

- The other 2 subjects in the first row test negative. See the entry of 2 in the first row.

	Positive Test Result (Test shows cancer.)	Negative Test Result (test shows no cancer)	Total
Cancer	8 (True Positive)	2 (False Negative)	10
No Cancer	99 (False Positive)	891 (True Negative)	990

Example: Interpreting Medical Test Results (11 of 12)

Solution

To find $P(C \mid \text{positive test result})$, see that the first column of values includes the positive test results. In that first column, the probability of randomly selecting a subject with cancer is $\frac{8}{107}$ or 0.0748, so $P(C \mid \text{positive test result}) = 0.0748$

	Positive Test Result (Test shows cancer.)	Negative Test Result (test shows no cancer)	Total
Cancer	8 (True Positive)	2 (False Negative)	10
No Cancer	99 (False Positive)	891 (True Negative)	990

Example: Interpreting Medical Test Results (12 of 12)

Interpretation

For the data given in this example, a randomly selected subject has a 1% chance of cancer, but for a randomly selected subject given a test with a positive result, the chance of cancer increases to 7.48%. Based on the data given in this example, a positive test result should not be devastating news, because there is still a good chance that the test is wrong.

Prior and Posterior Probability (1 of 2)

- Prior Probability
  - A **prior probability** is an initial probability value originally obtained before any additional information is obtained.
- Posterior Probability
  - A **posterior probability** is a probability value that has been revised by using additional information that is later obtained.

Prior and Posterior Probability (2 of 2)

Relative to the last example, $P(C) = 0.01$, which is the probability that a randomly selected subject has cancer. $P(C)$ is an example of a **prior probability.**

Using the additional information that the subject has received a positive test result, we found that $P(C \mid \text{positive test result}) = 0.0748$, and this is a **posterior probability** because it uses that additional information of the positive test result.