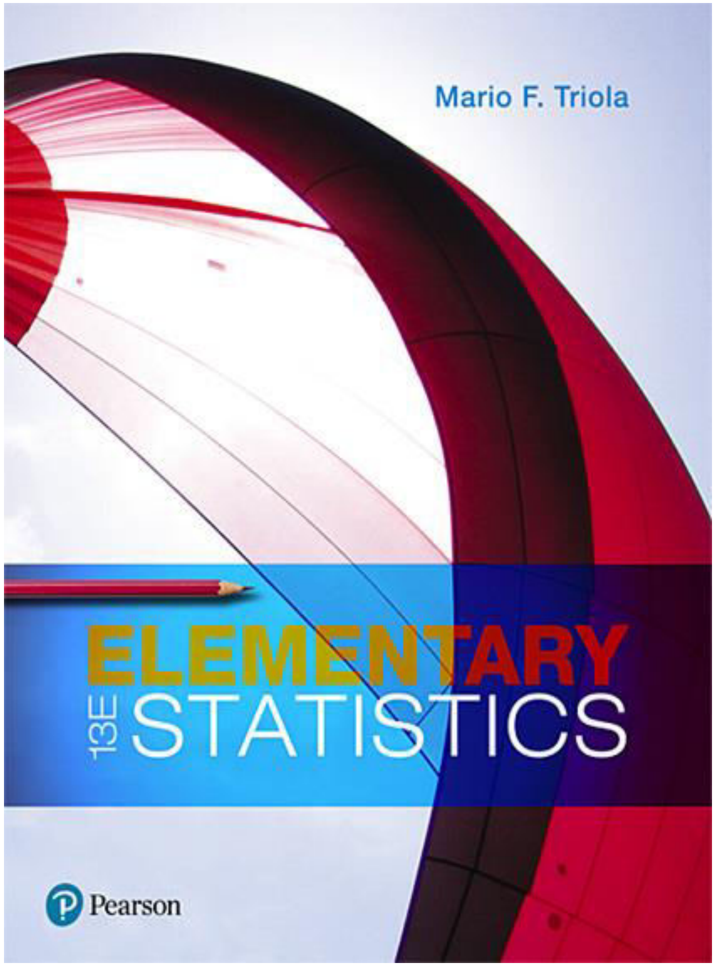
Elementary Statistics

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

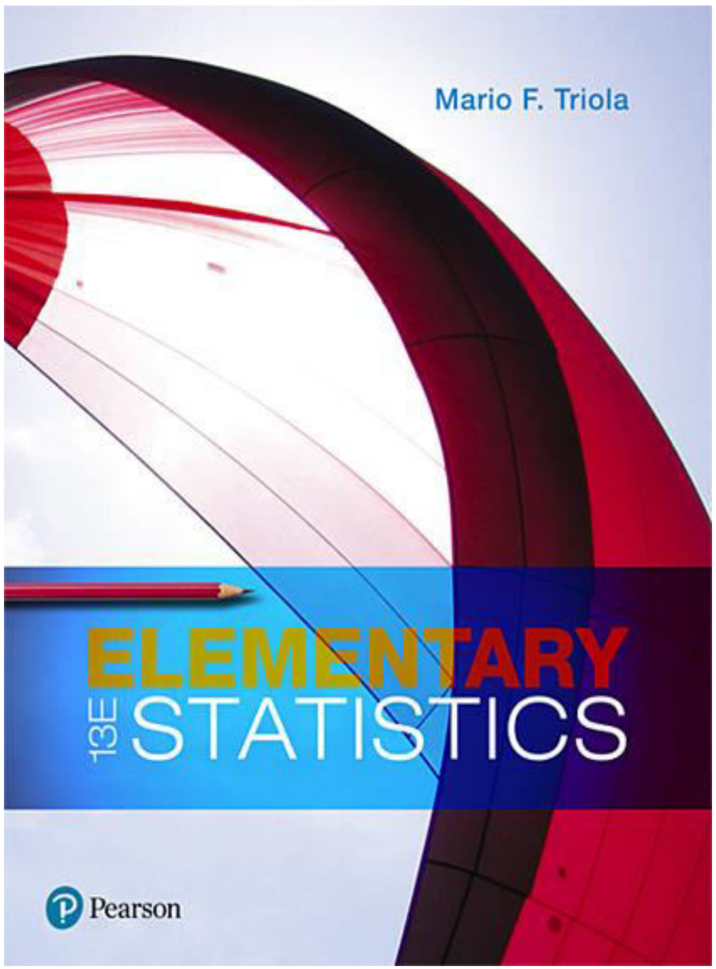
4-5 Probabilities Through Simulations (available at TriolaStats.com)

Probabilities Through Simulations

The website **Triola Stats** includes a downloadable section that discusses the use of simulation methods for finding probabilities. Simulations are also discussed in the Technology Project near the end of this chapter.

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 (1 of 2)

In this section we present the **addition rule** as a tool for finding $P(A \text{ or } B)$, which is the probability that either event A occurs or event B occurs (or they both occur) as the single outcome of a procedure. The word "or" in the addition rule is associated with the addition of probabilities.

Key Concept (2 of 2)

This section also presents the basic **multiplication rule** used for finding $P(A \text{ and } B)$, which is the probability that event A occurs and event B occurs. The word "and" in the multiplication rule is associated with the multiplication of probabilities.

Compound Event

- Compound Event
  - A **compound event** is any event combining two or more simple events.

Addition Rule

Notation for Addition Rule

$P(A \text{ or } B) = P$(in a single trial, event A occurs or event B occurs or they both occur)

Intuitive Addition Rule

To find $P(A \text{ or } B)$, add the number of ways event A can occur and the number of ways event B can occur, but **add in such a way that every outcome is counted only once.** $P(A \text{ or } B)$ is equal to that sum, divided by the total number of outcomes in the sample space.

Formal Addition Rule

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

where $P(A \text{ and } B)$ denotes the probability that A and B both occur at the same time as an outcome in a trial of a procedure.

Disjoint Events and the Addition Rule

- Disjoint (or mutually exclusive)
  - Events *A* and *B* are **disjoint** (or **mutually exclusive**) if they cannot occur at the same time. (That is, disjoint events do not overlap.)

Example: Disjoint Events (1 of 2)

Disjoint events:

Event A—Randomly selecting someone for a clinical trial who is a male

Event B—Randomly selecting someone for a clinical trial who is a female (The selected person **cannot** be both.)

Example: Disjoint Events (2 of 2)

Events that are **not** disjoint:

Event A—Randomly selecting someone taking a statistics course

Event B—Randomly selecting someone who is a female (The selected person **can** be both.)

Summary

Here is a summary of the key points of the addition rule:

1. To find $P(A \text{ or } B)$, first associate the word **or** with addition.
2. To find the value of $P(A \text{ or } B)$, add the number of ways A can occur and the number of ways B can occur, but be careful to add without double counting.

Complementary Events and the Addition Rule

We use $\overline{A}$ to indicate that event A does not occur. Common sense dictates this principle: We are certain (with probability 1) that either an event A occurs *or* it does not occur, so it follows that $P(A \text{ or } \overline{A}) = 1$. Because events A and $\overline{A}$ must be disjoint, we can use the addition rule to express this principle as follows:

$$P(A \text{ or } \overline{A}) = P(A) + P(\overline{A}) = 1$$

Rule of Complementary Events

$$P(A) + P(\bar{A}) = 1$$

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

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

Example: Sleepwalking (1 of 2)

Based on a journal article, the probability of randomly selecting someone who has sleepwalked is 0.292, so P(sleepwalked) = 0.292 (based on data from "Prevalence and Comorbidity of Nocturnal Wandering in the U.S. General Population," by Ohayon et al., **Neurology**, Vol. 78, No. 20). If a person is randomly selected, find the probability of getting someone who has **not** sleepwalked.

Example: Sleepwalking (2 of 2)

Solution

Using the rule of complementary events, we get

$$P(\text{has } \textbf{not} \text{ sleepwalked}) = 1 - P(\text{sleepwalked})$$
$$= 1 - 0.292 = 0.708$$

The probability of randomly selecting someone who has not sleepwalked is 0.708.

Multiplication Rule

Notation

P(A **and** B) = P(event A occurs in a first trial and event B occurs in a second trial)

$P(B \mid A)$ represents the probability of event B occurring after it is assumed that event A has already occurred.

Intuitive Multiplication Rule

To find the probability that event *A* occurs in one trial and event *B* occurs in another trial, multiply the probability of event *A* by the probability of event *B*, but **be sure that the probability of event *B* is found by assuming that event *A* has already occurred.**

Formal Multiplication Rule

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

Independence and the Multiplication Rule

- Independent
  - Two events *A* and *B* are **independent** if the occurrence of one does not affect the **probability** of the occurrence of the other. (Several events are independent if the occurrence of any does not affect the probabilities of the occurrence of the others.) If *A* and *B* are not independent, they are said to be **dependent.**

Example: Screening Drugs and the Basic Multiplication Rule (1 of 6)

50 test results from the subjects who use drugs are shown below:

Positive Test Results : 45

Negative Test Results : 5

Total : 50

Example: Screening Drugs and the Basic Multiplication Rule (2 of 6)

a. If 2 of these 50 subjects are randomly selected **with replacement,** find the probability the first selected person had a positive test result and the second selected person had a negative test result.

b. Repeat part (a) by assuming that the two subjects are selected **without** replacement.

Example: Screening Drugs and the Basic Multiplication Rule (3 of 6)

Solution

a. With Replacement: First selection (with 45 positive results among 50 total results):

$$P(\text{positive test result}) = \frac{45}{50}$$

Second selection (with 5 negative test results among the same 50 total results):

$$P(\text{negative test result}) = \frac{5}{50}$$

Example: Screening Drugs and the Basic Multiplication Rule (4 of 6)

Solution

We now apply the multiplication rule as follows:

P(1st selection is positive and 2nd is negative)

$$= \frac{45}{50} \cdot \frac{5}{50} = 0.0900$$

Example: Screening Drugs and the Basic Multiplication Rule (5 of 6)

Solution

b. Without Replacement: Without replacement of the first subject, the calculations are the same as in part (a), except that the second probability must be adjusted to reflect the fact that the first selection was positive and is not available for the second selection. After the first positive result is selected, we have 49 test results remaining, and 5 of them are negative.

The second probability is therefore $\frac{5}{49}$.

Example: Screening Drugs and the Basic Multiplication Rule (6 of 6)

Solution

P(1st selection is positive and 2nd is negative)

$$= \frac{45}{50} \cdot \frac{5}{49} = 0.0918$$

Sampling

In the world of statistics, sampling methods are critically important, and the following relationships hold:

- Sampling **with replacement:** Selections are **independent** events.
- Sampling **without replacement:** Selections are **dependent** events.

Treating Dependent Events and Independent

5% Guideline for Cumbersome Calculations

When sampling without replacement and the sample size is no more than 5% of the size of the population, treat the selections as being **independent** (even though they are actually dependent).

Example: Drug Screening and the 5% Guideline for Cumbersome Calculations (1 of 3)

Assume that three adults are randomly selected **without replacement** from the 247,436,830 adults in the United States. Also assume that 10% of adults in the United States use drugs. Find the probability that the three selected adults all use drugs.

Example: Drug Screening and the 5% Guideline for Cumbersome Calculations (2 of 3)

Solution

Because the three adults are randomly selected without replacement, the three events are dependent, but here we can treat them as being independent by applying the 5% guideline for cumbersome calculations. The sample size of 3 is clearly no more than 5% of the population size of 247,436,830.

Example: Drug Screening and the 5% Guideline for Cumbersome Calculations (3 of 3)

Solution

We get P(all 3 adults use drugs)

= P(first uses drugs **and** second uses drugs **and** third uses drugs)

= P(first uses drugs) · P(second uses drugs) · P(third uses drugs)

= (0.10)(0.10)(0.10)

= 0.00100

There is a 0.00100 probability that all three selected adults use drugs.

Redundancy: Important Application of the Multiplication Rule

The principle of **redundancy** is used to increase the reliability of many systems.

Our eyes have passive redundancy in the sense that if one of them fails, we continue to see. An important finding of modern biology is that genes in an organism can often work in place of each other. Engineers often design redundant components so that the whole system will not fail because of the failure of a single component.

Example: Airbus 310; Redundancy for Better Safety (1 of 5)

Modern aircraft are now highly reliable, and one design feature contributing to that reliability is the use of **redundancy,** whereby critical components are duplicated so that if one fails, the other will work. For example, the Airbus 310 twin-engine airliner has three independent hydraulic systems, so if any one system fails, full flight control is maintained with another functioning system.

Example: Airbus 310; Redundancy for Better Safety (2 of 5)

For this example, we will assume that for a typical flight, the probability of a hydraulic system failure is 0.002.

a. If the Airbus 310 were to have one hydraulic system, what is the probability that the aircraft’s flight control would work for a flight?

b. Given that the Airbus 310 actually has three independent hydraulic systems, what is the probability that on a typical flight, control can be maintained with a working hydraulic system?

Example: Airbus 310; Redundancy for Better Safety (3 of 5)

Solution

a. The probability of a hydraulic system failure is 0.002, so the probability that it does **not** fail is 0.998. That is, the probability that flight control can be maintained is as follows:

$$
\begin{aligned}
P(1 \text{ hydraulic system } \textbf{does not fail}) &= 1 - P(\text{failure}) \\
&= 1 - 0.002 \\
&= 0.998
\end{aligned}
$$

Example: Airbus 310; Redundancy for Better Safety (4 of 5)

Solution

b. With three independent hydraulic systems, flight control will be maintained if the three systems do not all fail. The probability of all three hydraulic systems failing is $0.002 \cdot 0.002 \cdot 0.002 = 0.000000008$.

It follows that the probability of maintaining flight control is as follows:

P(it does **not** happen that all three hydraulic systems fail) $= 1 - 0.000000008 = 0.999999992$

Example: Airbus 310; Redundancy for Better Safety

Interpretation

With only one hydraulic system we have a 0.002 probability of failure, but with three independent hydraulic systems, there is only a 0.000000008 probability that flight control cannot be maintained because all three systems failed.

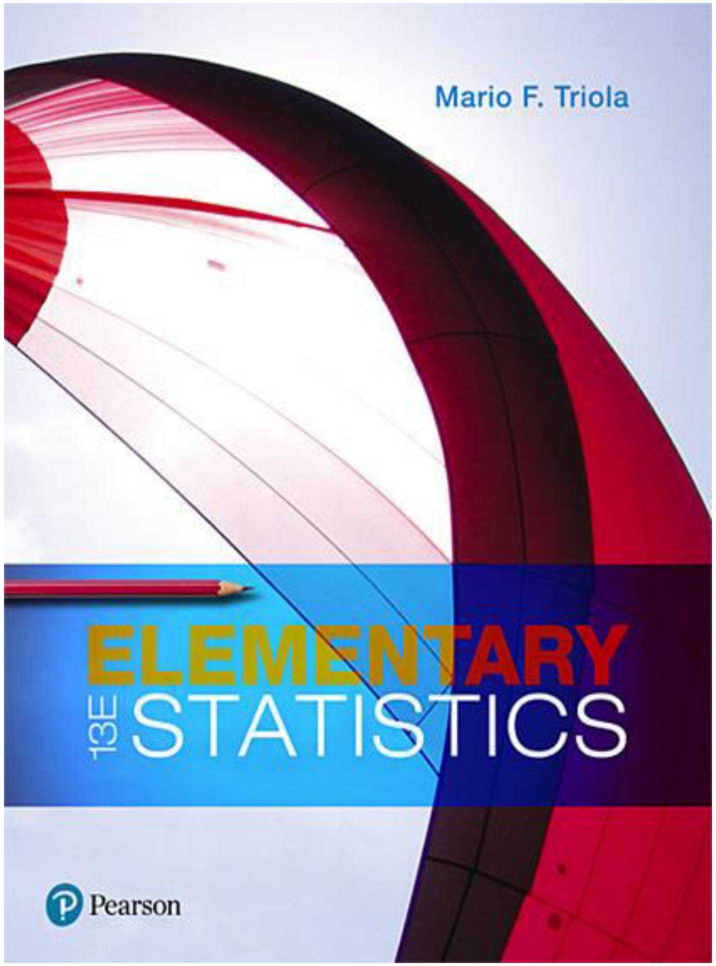
By using three hydraulic systems instead of only one, risk of failure is decreased not by a factor of $\frac{1}{3}$, but by a factor of $\frac{1}{250,000}$. By using three independent hydraulic systems, risk is dramatically decreased and safety is dramatically increased.

Summary of Addition Rule and Multiplication Rule

- **Addition Rule for P(*A* or *B*):** The word **or** suggests addition, and when adding $P(A)$ and $P(B)$, we must add in such a way that every outcome is counted only once.
- **Multiplication Rule for P(*A* and *B*):** The word **and** for two trials suggests multiplication, and when multiplying $P(A)$ and $P(B)$, we must be sure that the probability of event B takes into account the previous occurrence of event A.

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 $P(B \mid A)$

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

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) $= \dfrac{45}{555}$ and

P(subject uses drugs) $= \dfrac{50}{555}$ 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) ≠ 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.

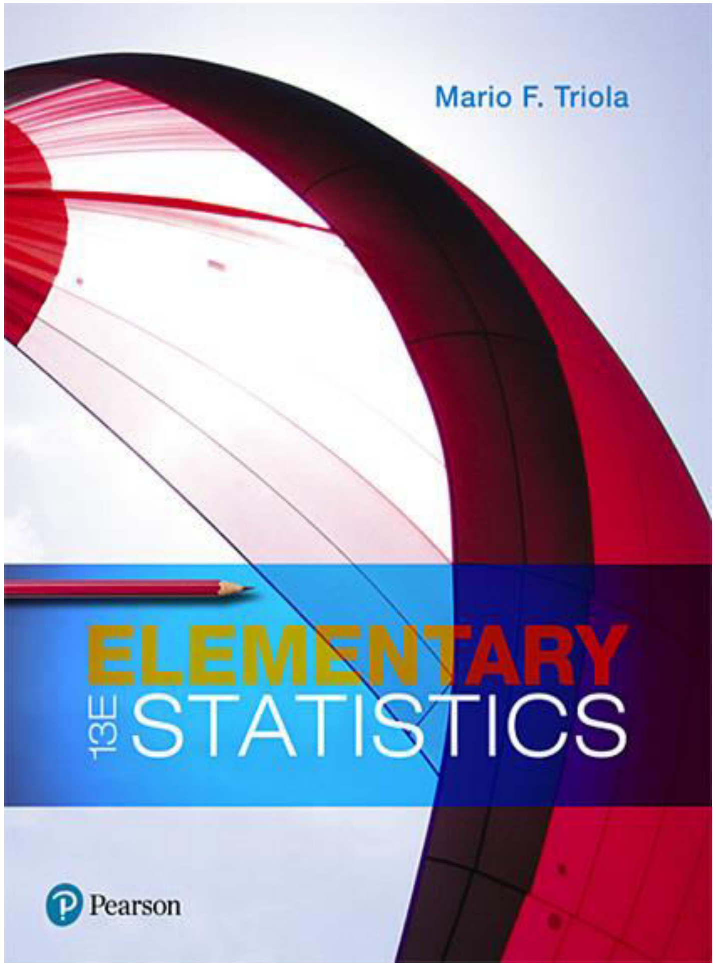
Prior and Posterior Probability (2 of 2)

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.

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

The single most important objective of this section is to learn how to **interpret** probability values, which are expressed as values between 0 and 1. A small probability, such as 0.001, corresponds to an event that rarely occurs.

Basics of Probability

- An **event** is any collection of results or outcomes of a procedure.
- A **simple event** is an outcome or an event that cannot be further broken down into simpler components.
- The **sample space** for a procedure consists of all possible **simple** events. That is, the sample space consists of all outcomes that cannot be broken down any further.

Example: Simple Events and Sample Spaces (1 of 5)

In the following display, we use "b" to denote a baby boy and "g" to denote a baby girl.

Procedure	Example of Event	Sample Space: Complete List of Simple Events
Single birth	1 girl (simple event)	{b, g}
3 births	2 boys and 1 girl (bbg, bgb, and gbb are all simple events resulting in 2 boys and 1 girl)	{bbb, bbg, bgb, bgg, gbb, gbg, ggb, ggg}

Example: Simple Events and Sample Spaces (2 of 5)

Solution

Simple Events:

- With one birth, the result of 1 girl is a **simple event** and the result of 1 boy is another simple event. They are individual simple events because they cannot be broken down any further.

Example: Simple Events and Sample Spaces (3 of 5)

Solution

Simple Events:

- With three births, the result of 2 girls followed by a boy (ggb) is a simple event.
- When rolling a single die, the outcome of 5 is a simple event, but the outcome of an even number is not a simple event.

Example: Simple Events and Sample Spaces (4 of 5)

Solution

Not a Simple Event: With three births, the event of "2 girls and 1 boy" is **not a simple event** because it can occur with these different simple events: ggb, gbg, bgg.

Example: Simple Events and Sample Spaces (5 of 5)

Solution

Sample Space: With three births, the **sample space** consists of the eight different simple events listed in the above table.

Procedure	Example of Event	Sample Space: Complete List of Simple Events
Single birth	1 girl (simple event)	{b, g}
3 births	2 boys and 1 girl (bbg, bgb, and gbb are all simple events resulting in 2 boys and 1 girl)	{bbb, bbg, bgb, bgg, gbb, gbg, ggb, ggg}

Three Common Approaches to Finding the Probability of an Event (1 of 6)

Notation for Probabilities

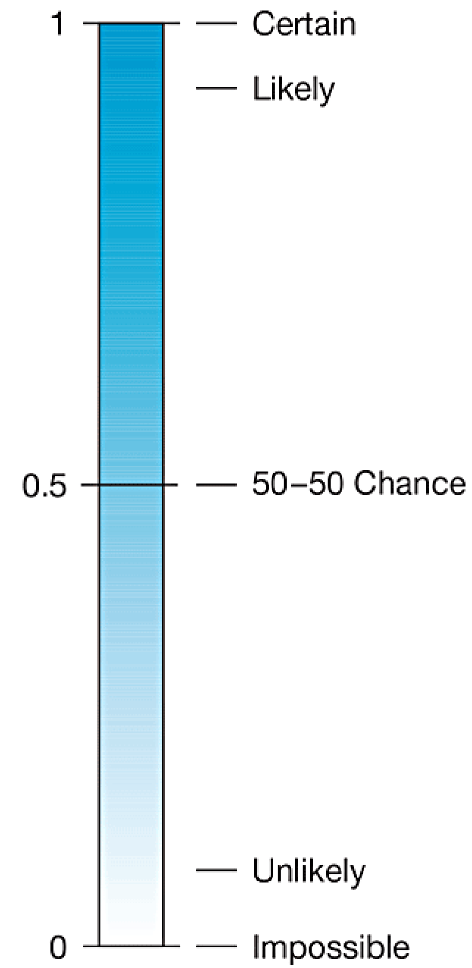
P denotes a probability.

A, *B*, and *C* denote specific events.

P(*A*) denotes the “probability of event *A* occurring.”

Three Common Approaches to Finding the Probability of an Event (2 of 6)

Possible values of probabilities and the more familiar and common expressions of likelihood



Three Common Approaches to Finding the Probability of an Event (3 of 6)

The following three approaches for finding probabilities result in values between 0 and 1: $0 \le P(A) \le 1$.

Three Common Approaches to Finding the Probability of an Event (4 of 6)

1. **Relative Frequency Approximation of Probability** Conduct (or observe) a procedure and count the number of times that event A occurs. $P(A)$ is then **approximated** as follows:

$$P(A) = \frac{\text{number of times } A \text{ occurred}}{\text{number of times the procedure repeated}}$$

Three Common Approaches to Finding the Probability of an Event (5 of 6)

2. **Classical Approach to Probability (Requires Equally Likely Outcomes)** If a procedure has n different simple events that are **equally likely**, and if event A can occur in s different ways, then

$$P(A) = \frac{\text{number of ways } A \text{ occurs}}{\text{number of different simple events}} = \frac{s}{n}$$

Caution When using the classical approach, always confirm that the outcomes are **equally likely**.

Three Common Approaches to Finding the Probability of an Event (6 of 6)

3. **Subjective Probabilities** $P(A)$, the probability of event A, is **estimated** by using knowledge of the relevant circumstances.

Simulations

- **Simulations**
  - Sometimes none of the preceding three approaches can be used. A **simulation** of a procedure is a process that behaves in the same ways as the procedure itself so that similar results are produced. Probabilities can sometimes be found by using a simulation.

Rounding Probabilities

When expressing the value of a probability, either give the **exact** fraction or decimal or round off final decimal results to **three** significant digits.

(**Suggestion:** When a probability is not a simple fraction such as $\frac{2}{3}$ or $\frac{5}{9}$, express it as a decimal so that the number can be better understood.)

Law of Large Numbers (1 of 2)

- **Law of Large Numbers**
  - As a procedure is repeated again and again, the relative frequency probability of an event tends to approach the actual probability.

Law of Large Numbers (2 of 2)

- **Law of Large Numbers**

CAUTIONS

1. The law of large numbers applies to behavior over a large number of trials, and it does not apply to any one individual outcome.
2. If we know nothing about the likelihood of different possible outcomes, we should not assume that they are equally likely. The actual probability depends on factors such as the amount of preparation and the difficulty of the test.

Example: Relative Frequency: Skydiving (1 of 3)

Find the probability of dying when making a skydiving jump.

Example: Relative Frequency: Skydiving (2 of 3)

Solution

In a recent year, there were about 3,000,000 skydiving jumps and 21 of them resulted in deaths. We use the relative frequency approach as follows:

$$P(\text{skydiving death}) = \frac{\text{number of skydiving deaths}}{\text{number of skydiving jumps}}$$

$$= \frac{21}{3{,}000{,}0000} = 0.000007$$

Example: Relative Frequency: Skydiving (3 of 3)

Solution

Here the classical approach cannot be used because the two outcomes (dying, surviving) are not equally likely.

A subjective probability can be estimated in the absence of historical data.

Example: Texting and Driving (1 of 4)

In a study of U.S. high school drivers, it was found that 3785 texted while driving during the previous 30 days, and 4720 did not text while driving during that same time period (based on data from "Texting While Driving . . . ," by Olsen, Shults, Eaton, **Pediatrics**, Vol. 131, No. 6). Based on these results, if a high school driver is randomly selected, find the probability that he or she texted while driving during the previous 30 days.

Example: Texting and Driving (2 of 4)

Solution

Instead of trying to determine an answer directly from the given statement, first summarize the information in a format that allows clear understanding, such as this format:

3785	texted while driving
4720	did not text while driving
8505	total number of drivers in the sample

Example: Texting and Driving (3 of 4)

Solution

We can now use the relative frequency approach as follows:

$$P(\text{texting while driving}) =$$

$$\frac{\text{drivers who texted while driving}}{\text{number of drivers in the sample}} = \frac{3785}{8505} = 0.445$$

Example: Texting and Driving (4 of 4)

Interpretation

There is a 0.445 probability that if a high school driver is randomly selected, he or she texted while driving during the previous 30 days.

Complementary Events

- **Complement**
  - The **complement** of event A, denoted by $\bar{A}$, consists of all outcomes in which event A does **not** occur.

Example: Complement of Death from Skydiving (1 of 2)

In a recent year, there were 3,000,000 skydiving jumps and 21 of them resulted in death. Find the probability of **not** dying when making a skydiving jump.

Example: Complement of Death from Skydiving (2 of 2)

Solution

Among 3,000,000 jumps there were 21 deaths, so it follows that the other 2,999,979 jumps were survived. We get

$$P(\text{not dying when skydiving}) = \frac{2{,}999{,}979}{3{,}000{,}000}$$
$$= 0.999993$$

The probability of **not** dying when making a skydiving jump is 0.999993.

Identifying Significant Results with Probabilities

The Rare Event Rule for Inferential Statistics

If, under a given assumption, the probability of a particular observed event is very small and the observed event occurs **significantly less than or significantly greater than** what we typically expect with that assumption, we conclude that the assumption is probably not correct.

Identifying Significant Results with Probabilities (2 of 3)

Using Probabilities to Determine When Results Are Significantly High or Significantly Low

- **Significantly high number of successes:** *x* successes among *n* trials is a **significantly high** number of successes if the probability of *x* or more successes is unlikely with a probability of 0.05 or less. That is, *x* is a significantly high number of successes if $P(x \text{ or more}) \leq 0.05$*.

*The value 0.05 is not absolutely rigid.

Identifying Significant Results with Probabilities (3 of 3)

Using Probabilities to Determine When Results Are Significantly High or Significantly Low

- **Significantly low number of successes:** x successes among n trials is a **significantly low** number of successes if the probability of x or fewer successes is unlikely with a probability of 0.05 or less. That is, x is a significantly low number of successes if $P(x \text{ or fewer}) \leq 0.05$*.

*The value 0.05 is not absolutely rigid.

Probability Review

- The probability of an event is a fraction or decimal number between 0 and 1 inclusive.
- The probability of an impossible event is 0.
- The probability of an event that is certain to occur is 1.
- Notation: $P(A)$ = the probability of event A.
- Notation: $P(\bar{A})$ = the probability that event A does **not** occur.