ANSWERS TO TRY IT EXERCISES

openstax™

Introductory Statistics

Chapter 1. Sampling and Data

Section 1. Definitions of Statistics, Probability and Key Terms

TRY IT1.1:
Determine what the key terms refer to in the following study. We want to know the average (mean) amount of money spent on school uniforms each year by families with children at Knoll Academy. We randomly survey 100 families with children in the school. Three of the families spent $65, $75, and $95, respectively.
Solution:
The population is all families with children attending Knoll Academy.
The sample is a random selection of 100 families with children attending Knoll Academy.
The parameter is the average (mean) amount of money spent on school uniforms by families with children at Knoll Academy.
The statistic is the average (mean) amount of money spent on school uniforms by families in the sample.
The variable is the amount of money spent by one family. Let X = the amount of money spent on school uniforms by one family with children attending Knoll Academy.
The data are the dollar amounts spent by the families. Examples of the data are $65, $75, and $95.

Section 2. Data, Sampling, and Variation in Data and Sampling

TRY IT 1.5:
The data are the number of machines in a gym. You sample five gyms. One gym has 12 machines, one gym has 15 machines, one gym has ten machines, one gym has 22 machines, and the other gym has 20 machines. What type of data is this?
Solution:
quantitative discrete data

TRY IT 1.6:
The data are the areas of lawns in square feet. You sample five houses. The areas of the lawns are 144 sq. feet, 160 sq. feet, 190 sq. feet, 180 sq. feet, and 210 sq. feet. What type of data is this?
Solution:
quantitative continuous data

TRY IT 1.8:
The data are the colors of houses. You sample five houses. The colors of the houses are white, yellow, white, red, and white. What type of data is this?
Solution:
qualitative data

TRY IT 1.9:
Determine the correct data type (quantitative or qualitative) for the number of cars in a parking lot. Indicate whether quantitative data are continuous or discrete.
Solution:
qualitative discrete

TRY IT 1.10:

The registrar at State University keeps records of the number of credit hours students complete each semester. The data he collects are summarized in the histogram. The class boundaries are 10 to less than 13, 13 to less than 16, 16 to less than 19, 19 to less than 22, and 22 to less than 25.

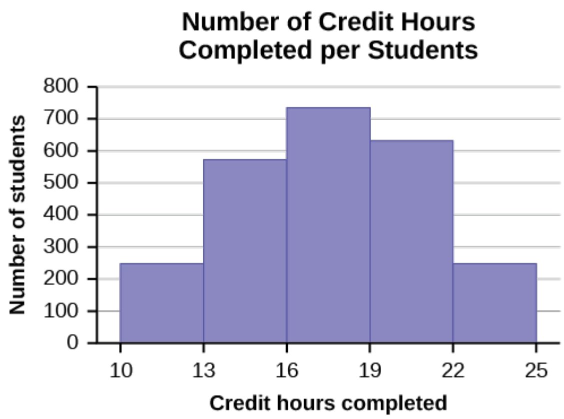


Figure 1.4
What type of data does this graph show?

Solution:

A histogram is used to display quantitative data: the numbers of credit hours completed. Because students can complete only a whole number of hours (no fractions of hours allowed), this data is quantitative discrete.

TRY IT 1.11:

You are going to use the random number generator to generate different types of samples from the data.

This table displays six sets of quiz scores (each quiz counts 10 points) for an elementary statistics class.

#1	#2	#3	#4	#5	#6
5	7	10	9	8	3
10	5	9	8	7	6
9	10	8	6	7	9
9	10	10	9	8	9
7	8	9	5	7	4

#1	#2	#3	#4	#5	#6
9	9	9	10	8	7
7	7	10	9	8	8
8	8	9	10	8	8
9	7	8	7	7	8
8	8	10	9	8	7

Table1.6

Instructions: Use the Random Number Generator to pick samples.

1. Create a stratified sample by column. Pick three quiz scores randomly from each column.
   - Number each row one through ten.
   - On your calculator, press Math and arrow over to PRB.
   - For column 1, Press 5:randInt(and enter 1,10). Press ENTER. Record the number. Press ENTER 2 more times (even the repeats). Record these numbers. Record the three quiz scores in column one that correspond to these three numbers.
   - Repeat for columns two through six.
   - These 18 quiz scores are a stratified sample.
2. Create a cluster sample by picking two of the columns. Use the column numbers: one through six.
   - Press MATH and arrow over to PRB.
   - Press 5:randInt(and enter 1,6). Press ENTER. Record the number. Press ENTER and record that number.
   - The two numbers are for two of the columns.
   - The quiz scores (20 of them) in these 2 columns are the cluster sample.
3. Create a simple random sample of 15 quiz scores.
   - Use the numbering one through 60.
   - Press MATH. Arrow over to PRB. Press 5:randInt(and enter 1, 60).
   - Press ENTER 15 times and record the numbers.
   - Record the quiz scores that correspond to these numbers.
   - These 15 quiz scores are the systematic sample.
4. Create a systematic sample of 12 quiz scores.
   - Use the numbering one through 60.
   - Press MATH. Arrow over to PRB. Press 5:randInt(and enter 1, 60).
   - Press ENTER. Record the number and the first quiz score. From that number, count ten quiz scores and record that quiz score. Keep counting ten quiz scores and recording the quiz score until you have a sample of 12 quiz scores. You may wrap around (go back to the beginning).

Determine the type of sampling used (simple random, stratified, systematic, cluster, or convenience).

a. A soccer coach selects six players from a group of boys aged eight to ten, seven players from a group of boys aged 11 to 12, and three players from a group of boys aged 13 to 14 to form a recreational soccer team.
b. A pollster interviews all human resource personnel in five different high tech companies.
c. A high school educational researcher interviews 50 high school female teachers and 50 high school male teachers.
d. A medical researcher interviews every third cancer patient from a list of cancer patients at a local hospital.
e. A high school counselor uses a computer to generate 50 random numbers and then picks students whose names correspond to the numbers.
f. A student interviews classmates in his algebra class to determine how many pairs of jeans a student owns, on the average.

Solution:
a. stratified; b. cluster; c. stratified; d. systematic; e. simple random; f. convenience

TRY IT 1.12:
Determine the type of sampling used (simple random, stratified, systematic, cluster, or convenience).
A high school principal polls 50 freshmen, 50 sophomores, 50 juniors, and 50 seniors regarding policy changes for after school activities.
Solution:
stratified

TRY IT 1.13:
A local radio station has a fan base of 20,000 listeners. The station wants to know if its audience would prefer more music or more talk shows. Asking all 20,000 listeners is an almost impossible task.
The station uses convenience sampling and surveys the first 200 people they meet at one of the station's music concert events. 24 people said they'd prefer more talk shows, and 176 people said they'd prefer more music.
Do you think that this sample is representative of (or is characteristic of) the entire 20,000 listener population?
Solution:
The sample probably consists more of people who prefer music because it is a concert event. Also, the sample represents only those who showed up to the event earlier than the majority. The sample probably doesn't represent the entire fan base and is probably biased towards people who would prefer music.

Section 3. Frequency, Frequency Tables, and Levels of Measurement

TRY IT 1.14:

Table 1.13 shows the amount, in inches, of annual rainfall in a sample of towns.

Rainfall (Inches)	Frequency	Relative Frequency	Cumulative Relative Frequency
2.95–4.97	6	65/50 = 0.12	0.12
4.97–6.99	7	7/50 = 0.14	0.12 + 0.14 = 0.26
6.99–9.01	15	15/50 = 0.30	0.26 + 0.30 = 0.56
9.01–11.03	8	8/50 = 0.16	0.56 + 0.16 = 0.72
11.03–13.05	9	9/50 = 0.18	0.72 + 0.18 = 0.90
13.05–15.07	5	5/50 = 0.10	0.90 + 0.10 = 1.00
	Total = 50	Total = 1.00	

Table1.13

From Table 1.13, find the percentage of rainfall that is less than 9.01 inches.

Solution:
0.56 or 56%

TRY IT 1.15:
From Table 1.13, find the percentage of rainfall that is between 6.99 and 13.05 inches.
Solution:
0.30 + 0.16 + 0.18 = 0.64 or 64%

TRY IT 1.16:
From Table 1.13, find the number of towns that have rainfall between 2.95 and 9.01 inches.
Solution:
6 + 7 + 15 = 28 towns

TRY IT 1.17:
Table 1.13 represents the amount, in inches, of annual rainfall in a sample of towns. What fraction of towns surveyed get between 11.03 and 13.05 inches of rainfall each year?
Solution:
9/50

TRY IT 1.18:

Table 1.16 contains the total number of fatal motor vehicle traffic crashes in the United States for the period from 1994 to 2011.

Year	Total Number of Crashes	Year	Total Number of Crashes
1994	36,254	2004	38,444
1995	37,241	2005	39,252
1996	37,494	2006	38,648
1997	37,324	2007	37,435
1998	37,107	2008	34,172
1999	37,140	2009	30,862
2000	37,526	2010	30,296
2001	37,862	2011	29,757
2002	38,491	Total	653,782
2003	38,477		

Table1.16

Answer the following questions.

a. What is the frequency of deaths measured from 2000 through 2004?
b. What percentage of deaths occurred after 2006?
c. What is the relative frequency of deaths that occurred in 2000 or before?
d. What is the percentage of deaths that occurred in 2011?
e. What is the cumulative relative frequency for 2006? Explain what this number tells you about the data.

Solution:

a. 190,800 (29.2%)
b. 24.9%

c. 260,086/653,782 or 39.8%
d. 4.6%
e. 75.1% of all fatal traffic crashes for the period from 1994 to 2011 happened from 1994 to 2006.

Section 4. Experimental Design and Ethics

TRY IT 1.21:

You are concerned about the effects of texting on driving performance. Design a study to test the response time of drivers while texting and while driving only. How many seconds does it take for a driver to respond when a leading car hits the brakes?

a. Describe the explanatory and response variables in the study.
b. What are the treatments?
c. What should you consider when selecting participants?
d. Your research partner wants to divide participants randomly into two groups: one to drive without distraction and one to text and drive simultaneously. Is this a good idea? Why or why not?
e. Identify any lurking variables that could interfere with this study.
f. How can blinding be used in this study?

Solution:

a. Explanatory: presence of distraction from texting; response: response time measured in seconds
b. Driving without distraction and driving while texting
c. Answers will vary. Possible responses: Do participants regularly send and receive text messages? How long has the subject been driving? What is the age of the participants? Do participants have similar texting and driving experience?
d. This is not a good plan because it compares drivers with different abilities. It would be better to assign both treatments to each participant in random order.
e. Possible responses include: texting ability, driving experience, type of phone.
f. The researchers observing the trials and recording response time could be blinded to the treatment being applied.

TRY IT 1.22:

Describe the unethical behavior, if any, in each example and describe how it could impact the reliability of the resulting data. Explain how the problem should be corrected.

A study is commissioned to determine the favorite brand of fruit juice among teens in California.

a. The survey is commissioned by the seller of a popular brand of apple juice.
b. There are only two types of juice included in the study: apple juice and cranberry juice.
c. Researchers allow participants to see the brand of juice as samples are poured for a taste test.

d. Twenty-five percent of participants prefer Brand X, 33% prefer Brand Y and 42% have no preference between the two brands. Brand X references the study in a commercial saying "Most teens like Brand X as much as or more than Brand Y."

Solution:

a. This is not necessarily a problem. The study should be monitored carefully, however, to ensure that the company is not pressuring researchers to return biased results.
b. If the researchers truly want to determine the favorite brand of juice, then researchers should ask teens to compare different brands of the same type of juice. Choosing a sweet juice to compare against a sharp-flavored juice will not lead to an accurate comparison of brand quality.
c. Participants could be biased by the knowledge. The results may be different from those obtained in a blind taste test.
d. The commercial tells the truth, but not the whole truth. It leads consumers to believe that Brand X was preferred by more participants than Brand Y while the opposite is true.

Chapter 2. Descriptive Statistics

Section 1. Stem-and-Leaf Graphs (Stemplots), Line Graphs, and Bar Graphs

TRY IT 2.1:

For the Park City basketball team, scores for the last 30 games were as follows (smallest to largest):

32; 32; 33; 34; 38; 40; 42; 42; 43; 44; 46; 47; 47; 48; 48; 48; 49; 50; 50; 51; 52; 52; 52; 53; 54; 56; 57; 57; 60; 61

Construct a stem plot for the data.

Solution:

Stem	Leaf
3	2 2 3 4 8
4	0 2 2 3 4 6 7 7 8 8 8 9
5	0 0 1 2 2 2 3 4 6 7 7
6	0 1

TRY IT 2.2:

The following data show the distances (in miles) from the homes of off-campus statistics students to the college. Create a stem plot using the data and identify any outliers:

0.5; 0.7; 1.1; 1.2; 1.2; 1.3; 1.3; 1.5; 1.5; 1.7; 1.7; 1.8; 1.9; 2.0; 2.2; 2.5; 2.6; 2.8; 2.8; 2.8; 3.5; 3.8; 4.4; 4.8; 4.9; 5.2; 5.5; 5.7; 5.8; 8.0

Solution:

Stem	Leaf
0	5 7
1	1 2 2 3 3 5 5 7 7 8 9
2	0 2 5 6 8 8 8
3	5 8
4	4 8 9
5	2 5 7 8
6	
7	
8	0

The value 8.0 may be an outlier. Values appear to concentrate at one and two miles.

TRY IT 2.3:

The table shows the number of wins and losses the Atlanta Hawks have had in 42 seasons. Create a side-by-side stem-and-leaf plot of these wins and losses.

Losses	Wins	Year	Losses	Wins	Year
34	48	1968–1969	41	41	1989–1990
34	48	1969–1970	39	43	1990–1991

Losses	Wins	Year	Losses	Wins	Year
46	36	1970–1971	44	38	1991–1992
46	36	1971–1972	39	43	1992–1993
36	46	1972–1973	25	57	1993–1994
47	35	1973–1974	40	42	1994–1995
51	31	1974–1975	36	46	1995–1996
53	29	1975–1976	26	56	1996–1997
51	31	1976–1977	32	50	1997–1998
41	41	1977–1978	19	31	1998–1999
36	46	1978–1979	54	28	1999–2000
32	50	1979–1980	57	25	2000–2001
51	31	1980–1981	49	33	2001–2002
40	42	1981–1982	47	35	2002–2003
39	43	1982–1983	54	28	2003–2004
42	40	1983–1984	69	13	2004–2005
48	34	1984–1985	56	26	2005–2006
32	50	1985–1986	52	30	2006–2007

Losses	Wins	Year	Losses	Wins	Year
25	57	1986–1987	45	37	2007–2008
32	50	1987–1988	35	47	2008–2009
30	52	1988–1989	29	53	2009–2010

Table2.6

Solution:

Atlanta Hawks Wins and Losses	
Number of Wins	**Number of Losses**
3	1 9
9 8 8 6 5	2 5 5 9
8 7 6 6 5 5 4 3 1 1 1 1 0	3 0 2 2 2 2 4 4 5 6 6 6 9 9 9
8 8 7 6 6 6 3 3 3 2 2 1 1 0	4 0 0 1 1 2 4 5 6 6 7 7 8 9
7 7 6 3 2 0 0 0 0	5 1 1 1 2 3 4 4 6 7
	6 9

TRY IT 2.4:

In a survey, 40 people were asked how many times per year they had their car in the shop for repairs. The results are shown in Table 2.8. Construct a line graph.

Number of times in shop	Frequency
0	7
1	10

Number of times in shop	Frequency
2	14
3	9

Table2.8

Solution:

Figure2.

TRY IT 2.5:

The population in Park City is made up of children, working-age adults, and retirees. Table 2.10shows the three age groups, the number of people in the town from each age group, and the proportion (%) of people in each age group. Construct a bar graph showing the proportions.

Age groups	Number of people	Proportion of population
Children	67,059	19%
Working-age adults	152,198	43%
Retirees	131,662	38%

Table2.10

Solution:

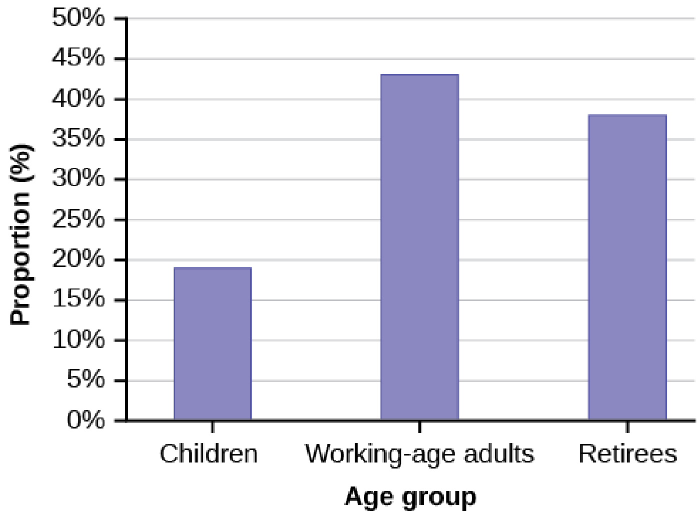


Figure 3.

TRY IT 2.6:

Park city is broken down into six voting districts. The table shows the percent of the total registered voter population that lives in each district as well as the percent total of the entire population that lives in each district. Construct a bar graph that shows the registered voter population by district.

District	Registered voter population	Overall city population
1	15.5%	19.4%
2	12.2%	15.6%
3	9.8%	9.0%
4	17.4%	18.5%
5	22.8%	20.7%
6	22.3%	16.8%

Table2.12

Solution:

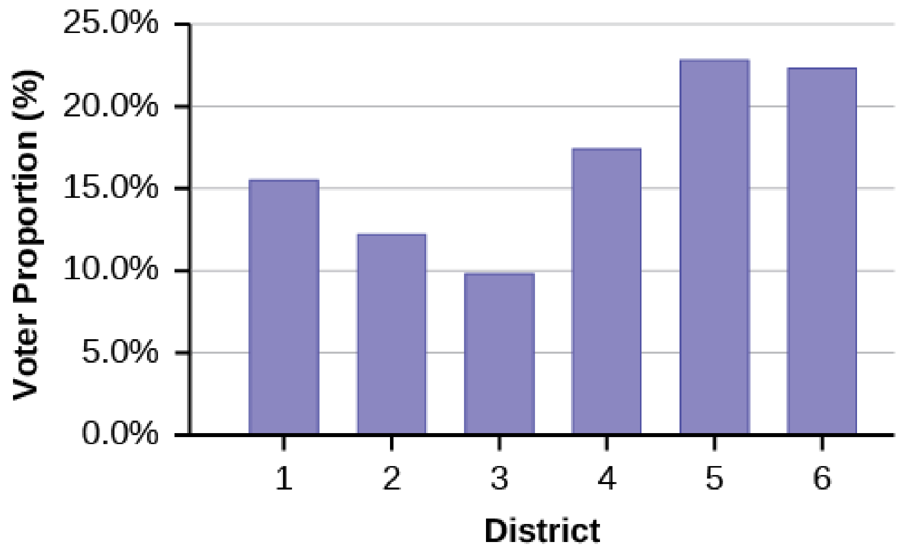


Figure 4.

Section 2. Histograms, Frequency Polygons, and Time Series Graphs

TRY IT 2.7:
The following data are the shoe sizes of 50 male students. The sizes are discrete data since shoe size is measured in whole and half units only. Construct a histogram and calculate the width of each bar or class interval. Suppose you choose six bars.
9; 9; 9.5; 9.5; 10; 10; 10; 10; 10; 10; 10.5; 10.5; 10.5; 10.5; 10.5; 10.5; 10.5; 10.5
11; 11; 11; 11; 11; 11; 11; 11; 11; 11; 11; 11; 11; 11.5; 11.5; 11.5; 11.5; 11.5; 11.5; 11.5
12; 12; 12; 12; 12; 12; 12; 12.5; 12.5; 12.5; 12.5; 14
Solution:

Smallest value: 9

Largest value: 14

Convenient starting value: 9 – 0.05 = 8.95

Convenient ending value: 14 + 0.05 = 14.05

14.05−8.956=0.8514.05−8.956=0.85

The calculations suggests using 0.85 as the width of each bar or class interval. You can also use an interval with a width equal to one

TRY IT 2.8:
The following data are the number of sports played by 50 student athletes. The number of sports is discrete data since sports are counted.

1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1
2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2
3; 3; 3; 3; 3; 3; 3; 3

20 student athletes play one sport. 22 student athletes play two sports. Eight student athletes play three sports.

Fill in the blanks for the following sentence. Since the data consist of the numbers 1, 2, 3, and the starting point is 0.5, a width of one places the 1 in the middle of the interval 0.5 to _____, the 2 in the middle of the interval from _____ to _____, and the 3 in the middle of the interval from _____ to _____.

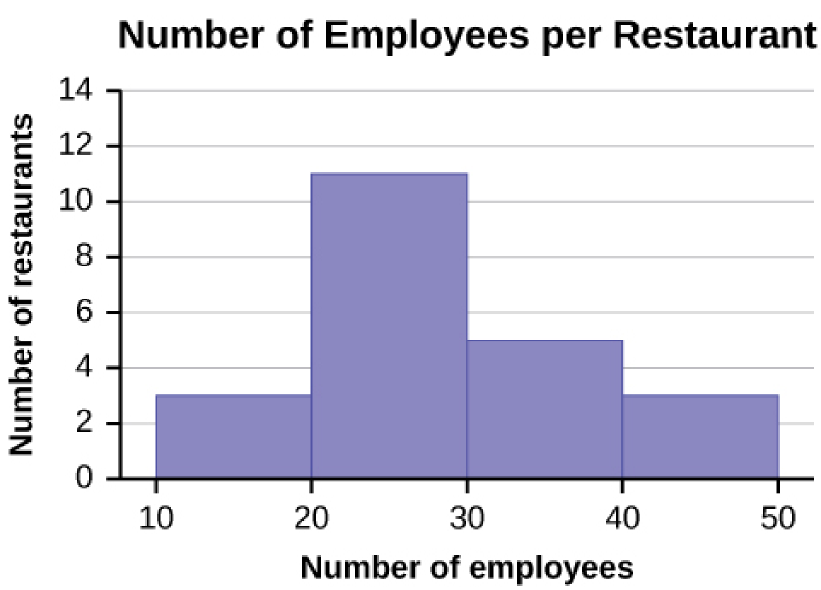
Solution:
1.5
1.5 to 2.5
2.5 to 3.5

TRY IT 2.9:

The following data represent the number of employees at various restaurants in New York City. Using this data, create a histogram.

22; 35; 15; 26; 40; 28; 18; 20; 25; 34; 39; 42; 24; 22; 19; 27; 22; 34; 40; 20; 38; and 28
Use 10–19 as the first interval.

Solution:



TRY IT 2.10:

Construct a frequency polygon of U.S. Presidents' ages at inauguration shown in Table 2.15.

Age at Inauguration	Frequency
41.5–46.5	4
46.5–51.5	11
51.5–56.5	14
56.5–61.5	9
61.5–66.5	4
66.5–71.5	2

Table2.15
Solution:

The first label on the *x*-axis is 39. This represents an interval extending from 36.5 to 41.5. Since there are no ages less than 41.5, this interval is used only to allow the graph to touch the *x*-axis. The point labeled 44 represents the next interval, or the first "real" interval from the table, and contains four scores. This reasoning is followed for each of the remaining intervals with the point 74 representing the interval from 71.5 to 76.5. Again, this interval contains no data and is only used so that the graph will touch the *x*-axis. Looking at the graph, we say that this distribution is skewed because one side of the graph does not mirror the other side.

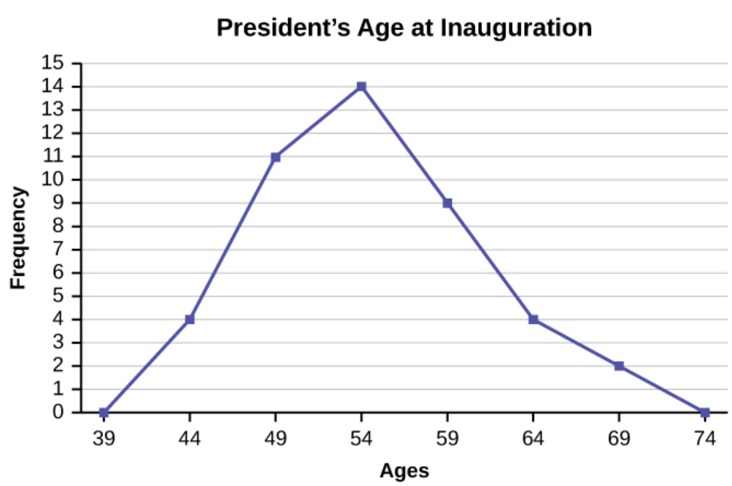


Figure 4.

TRY IT 2.12:

The following table is a portion of a data set from www.worldbank.org. Use the table to construct a time series graph for CO_2 emissions for the United States.

CO2 Emissions

	Ukraine	United Kingdom	United States
2003	352,259	540,640	5,681,664
2004	343,121	540,409	5,790,761
2005	339,029	541,990	5,826,394
2006	327,797	542,045	5,737,615
2007	328,357	528,631	5,828,697
2008	323,657	522,247	5,656,839
2009	272,176	474,579	5,299,563

Table2.20

Solution:

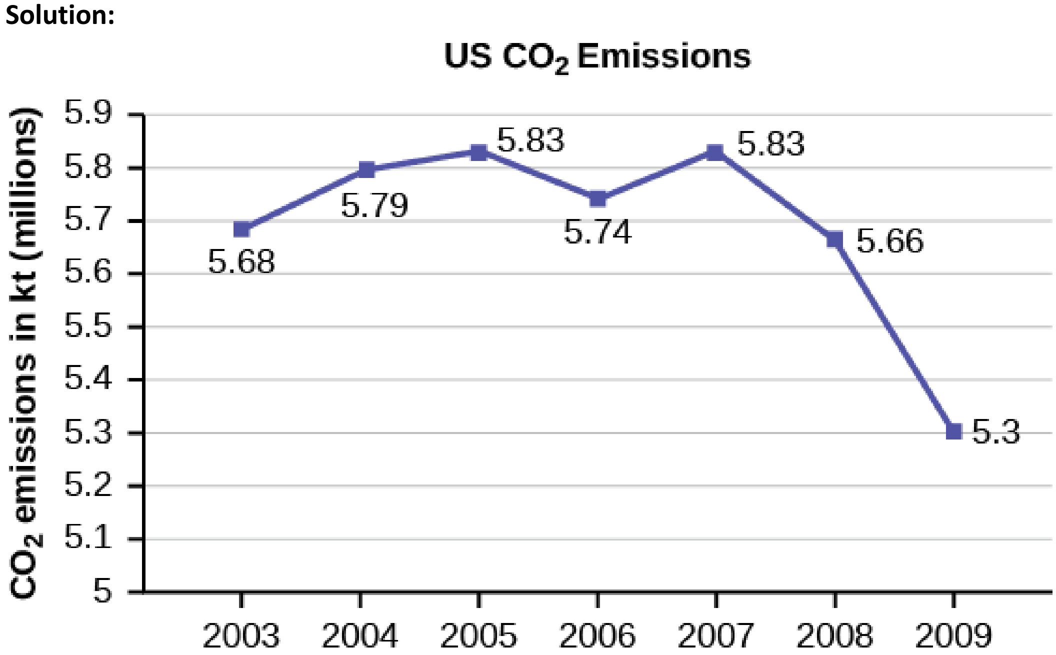


Figure 6

Section 3. Measures of the Location of the Data

TRY IT 2.13

For the following 11 salaries, calculate the *IQR* and determine if any salaries are outliers. The salaries are in dollars.

$33,000; $64,500; $28,000; $54,000; $72,000; $68,500; $69,000; $42,000; $54,000; $120,000; $ 40,500

Solution:

Order the data from smallest to largest.

$28,000 $33,000 $40,500 $42,000 $54,000 $54,000 $64,500 $68,500$69,000 $72,000 $120,000

Median = $54,000

Q_1 = $40,500

Q_3 = $69,000

IQR = $69,000 – $40,500 = $28,500

(1.5)(*IQR*) = (1.5)($28,500) = $42,750

Q_1 – (1.5)(*IQR*) = $40,500 – $42,750 = –$2,250

Q_3 + (1.5)(*IQR*) = $69,000 + $42,750 = $111,750

No salary is less than –$2,250. However, $120,000 is more than $11,750, so $120,000 is a potential outlier.

TRY IT 2.14:

Find the interquartile range for the following two data sets and compare them.

Test Scores for Class *A*
69; 96; 81; 79; 65; 76; 83; 99; 89; 67; 90; 77; 85; 98; 66; 91; 77; 69; 80; 94
Test Scores for Class *B*
90; 72; 80; 92; 90; 97; 92; 75; 79; 68; 70; 80; 99; 95; 78; 73; 71; 68; 95; 100

Solution:

Class *A*

Order the data from smallest to largest.

65 66 67 69 69 76 77 77 79 80 81 83 85 89 90 91 94 96 98 99

Median=$\frac{80+81}{2}$=80.5

Q1= $\frac{69+76}{2}$ =72.5

Q3= $\frac{90+91}{2}$ =90.5

IQR = 90.5 – 72.5 = 18

Class *B*

Order the data from smallest to largest.

68 68 70 71 72 73 75 78 79 80 80 90 90 92 92 95 95 97 99 100

Median= $\frac{80+80}{2}$ =80

Q1= $\frac{72+73}{2}$ =72.5

Q3= $\frac{92+95}{2}$ =93.5

IQR = 93.5 – 72.5 = 21

The data for Class *B* has a larger *IQR*, so the scores between Q_3 and Q_1(middle 50%) for the data for Class *B* are more spread out and not clustered about the median.

TRY IT 2.15:

Forty bus drivers were asked how many hours they spend each day running their routes (rounded to the nearest hour). Find the 65th percentile.

Amount of time spent on route (hours)	Frequency	Relative Frequency	Cumulative Relative Frequency
2	12	0.30	0.30
3	14	0.35	0.65

Amount of time spent on route (hours)	Frequency	Relative Frequency	Cumulative Relative Frequency
4	10	0.25	0.90
5	4	0.10	1.00

Table2.23

Solution:

The 65th percentile is between the last three and the first four.

The 65th percentile is 3.5.

TRY IT 2.16:

Refer to the Table 2.23. Find the third quartile. What is another name for the third quartile?

Solution:

The third quartile is the 75th percentile, which is four. The 65th percentile is between three and four, and the 90th percentile is between four and 5.75. The third quartile is between 65 and 90, so it must be four

TRY IT 2.17:

Listed are 29 ages for Academy Award winning best actors *in order from smallest to largest.*

18; 21; 22; 25; 26; 27; 29; 30; 31; 33; 36; 37; 41; 42; 47; 52; 55; 57; 58; 62; 64; 67; 69; 71; 72; 73; 74; 76; 77
Calculate the 20th percentile and the 55th percentile.

Solution:

k = 20. Index = i = (k/100)*(n+1=(20/100)*(29 + 1) = 6. The age in the sixth position is 27. The 20th percentile is 27 years.

k = 55. Index = i = (k/100)*(n+1)= (55/100)*(29 + 1) = 16.5. Round down to 16 and up to 17. The age in the 16th position is 52 and the age in the 17thposition is 55. The average of 52 and 55 is 53.5. The 55th percentile is 53.5 years.

TRY IT 2.18:

Listed are 30 ages for Academy Award winning best actors in order from smallest to largest.

18; 21; 22; 25; 26; 27; 29; 30; 31, 31; 33; 36; 37; 41; 42; 47; 52; 55; 57; 58; 62; 64; 67; 69; 71; 72; 73; 74; 76; 77
Find the percentiles for 47 and 31.

Solution:

Percentile for 47: Counting from the bottom of the list, there are 15 data values less than 47. There is one value of 47.

x = 15 and *y* = 1. $\frac{x+0.5y}{n}(100) = \frac{15+0.5(1)}{29}(100) = 53.45$. 47 is the 53rd percentile.

Percentile for 31: Counting from the bottom of the list, there are eight data values less than 31. There are two values of 31.

x = 15 and *y* = 2. $\frac{x+0.5y}{n}(100) = \frac{15+0.5(2)}{29}(100) = 31.03$. 31 is the 31st percentile.

TRY IT 2.19:

For the 100-meter dash, the third quartile for times for finishing the race was 11.5 seconds. Interpret the third quartile in the context of the situation.

Solution:

Twenty-five percent of runners finished the race in 11.5 seconds or more. Seventy-five percent of runners finished the race in 11.5 seconds or less. A lower percentile is good because finishing a race more quickly is desirable.

TRY IT 2.20:

On a 60 point written assignment, the 80th percentile for the number of points earned was 49. Interpret the 80th percentile in the context of this situation.

Solution:

Eighty percent of students earned 49 points or fewer. Twenty percent of students earned 49 or more points. A higher percentile is good because getting more points on an assignment is desirable.

TRY IT 2.21:

During a season, the 40th percentile for points scored per player in a game is eight. Interpret the 40th percentile in the context of this situation.

Solution:

Forty percent of players scored eight points or fewer. Sixty percent of players scored eight points or more. A higher percentile is good because getting more points in a basketball game is desirable.

Section 4. Box Plots

TRY IT 2.23

The following data are the number of pages in 40 books on a shelf. Construct a box plot using a graphing calculator, and state the interquartile range.

136; 140; 178; 190; 205; 215; 217; 218; 232; 234; 240; 255; 270; 275; 290; 301; 303; 315; 317; 318; 326; 333; 343; 349; 360; 369; 377; 388; 391; 392; 398; 400; 402; 405; 408; 422; 429; 450; 475; 512

Solution:

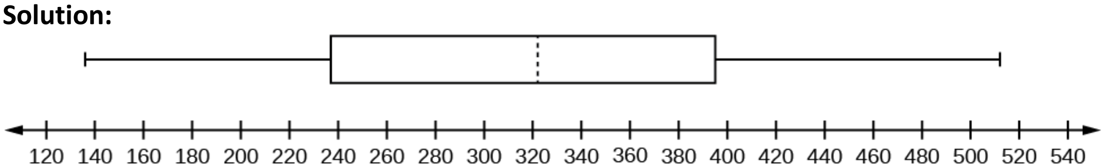


Figure 3.

IQR = 158

TRY IT 2.24

The following data set shows the heights in inches for the boys in a class of 40 students.

66; 66; 67; 67; 68; 68; 68; 68; 68; 69; 69; 69; 70; 71; 72; 72; 72; 73; 73; 74
The following data set shows the heights in inches for the girls in a class of 40 students.
61; 61; 62; 62; 63; 63; 63; 65; 65; 65; 66; 66; 66; 67; 68; 68; 68; 69; 69; 69
Construct a box plot using a graphing calculator for each data set, and state which box plot has the wider spread for the middle 50% of the data.

Solution:

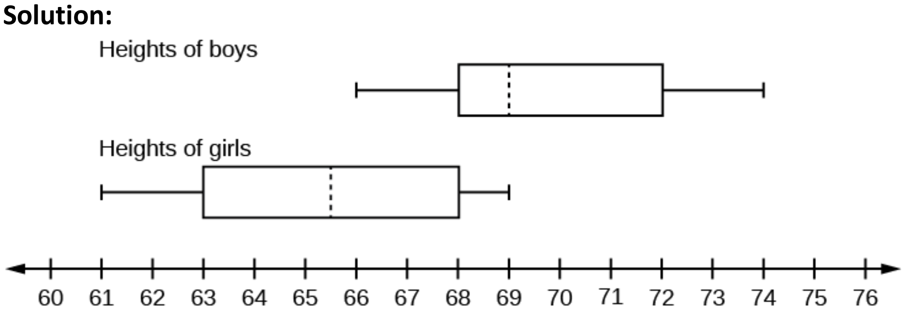


Figure 5.

IQR for the boys = 4

IQR for the girls = 5

The box plot for the heights of the girls has the wider spread for the middle 50% of the data.

TRY IT 2.25

Follow the steps you used to graph a box-and-whisker plot for the data values shown.

0; 5; 5; 15; 30; 30; 45; 50; 50; 60; 75; 110; 140; 240; 330

Solution:

The data are in order from least to greatest. There are 15 values, so the eighth number in order is the median: 50. There are seven data values written to the left of the median and 7 values to the right. The five values that are used to create the boxplot are:

- Min: 0
- Q_1: 15
- Med: 50
- Q_3: 110
- Max: 330

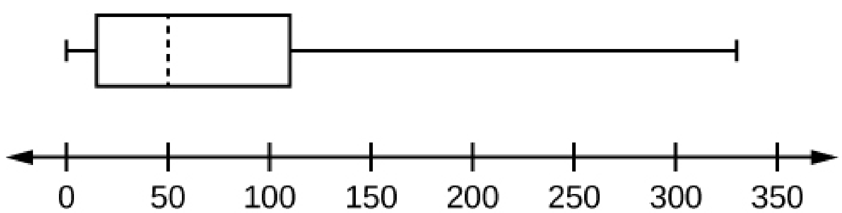


Figure 7.

Section 5. Measures of the Center of the Data

TRY IT 2.26

The following data show the number of months patients typically wait on a transplant list before getting surgery. The data are ordered from smallest to largest. Calculate the mean and median.

3; 4; 5; 7; 7; 7; 7; 8; 8; 9; 9; 10; 10; 10; 10; 10; 11; 12; 12; 13; 14; 14; 15; 15; 17; 17; 18; 19; 19; 19 ; 21; 21; 22; 22; 23; 24; 24; 24; 24

Solution:
Mean: 3 + 4 + 5 + 7 + 7 + 7 + 7 + 8 + 8 + 9 + 9 + 10 + 10 + 10 + 10 + 10 + 11 + 12 + 12 + 13 + 14 + 14 + 15 + 15 + 17 + 17 + 18 + 19 + 19 + 19 + 21 + 21 + 22 + 22 + 23 + 24 + 24 + 24 = 544
$\frac{544}{39}$ = 13.95
Median: Starting at the smallest value, the median is the 20th term, which is 13.

TRY IT 2.27

In a sample of 60 households, one house is worth $2,500,000. Twenty-nine houses are worth $280,000, and all the others are worth $315,000. Which is the better measure of the “center”: the mean or the median?

Solution:

The median is the better measure of the “center” than the mean because 59 of the values are less than $315,000 and one is $2,500,000. The $2,500,000 is an outlier. The median, $315,000, gives us a better sense of the middle of the data than the mean, $334,500.

TRY IT 2.28

The number of books checked out from the library from 25 students are as follows:

0; 0; 0; 1; 2; 3; 3; 4; 4; 5; 5; 7; 7; 7; 7; 8; 8; 8; 9; 10; 10; 11; 11; 12; 12
Find the mode.

Solution:
The most frequent number of books is 7, which occurs four times. Mode = 7.

TRY IT 2.29

Five credit scores are 680, 680, 700, 720, 720. The data set is bimodal because the scores 680 and 720 each occur twice. Consider the annual earnings of workers at a factory. The mode is $25,000 and occurs 150 times out of 301. The median is $50,000 and the mean is $47,500. What would be the best measure of the “center”?

Solution:
Because $25,000 occurs nearly half the time, the mode would be the best measure of the center because the median and mean don’t represent what most people make at the factory.

TRY IT 2.30

Maris conducted a study on the effect that playing video games has on memory recall. As part of her study, she compiled the following data:

Hours Teenagers Spend on Video Games	Number of Teenagers
0–3.5	3
3.5–7.5	7
7.5–11.5	12

Hours Teenagers Spend on Video Games	Number of Teenagers
11.5–15.5	7
15.5–19.5	9

Table2.27

What is the best estimate for the mean number of hours spent playing video games?

Solution:
Find the midpoint of each interval, multiply by the corresponding number of teenagers, add the results and then divide by the total number of teenagers
The midpoints are 1.75, 5.5, 9.5, 13.5,17.5.
Mean = $\frac{(1.75)(3)+(5.5)(7)+(9.5)(12)+(13.5)(7)+(17.5)(9)}{3+7+12+7+9} = \frac{409.75}{38}$ =10.78

Section 6. Skewness and the Mean, Median and Mode

TRY IT 2.31

Discuss the mean, median, and mode for each of the following problems. Is there a pattern between the shape and measure of the center?

a.

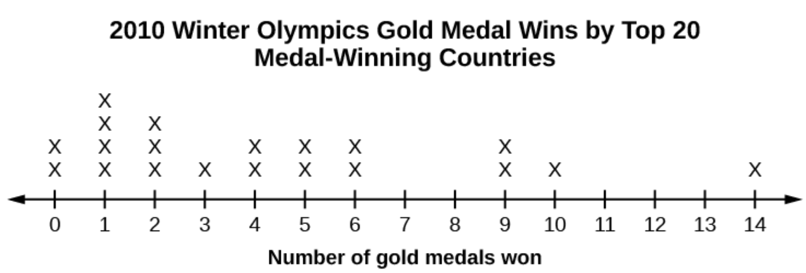


Figure 2.22

b.

The Ages Former U.S Presidents Died	
4	6 9
5	3 6 7 7 7 8

The Ages Former U.S Presidents Died

6	0 0 3 3 4 4 5 6 7 7 7 8
7	0 1 1 2 3 4 7 8 8 9
8	0 1 3 5 8
9	0 0 3 3
Key: 8|0 means 80.	

Table2.28

c.

Figure 2.23

Solution:

a. mean = 4.25, median = 3.5, mode = 1; The mean > median > mode which indicates skewness to the right. (data are 0, 1, 2, 3, 4, 5, 6, 9, 10, 14 and respective frequencies are 2, 4, 3, 1, 2, 2, 2, 2, 1, 1)
b. mean = 70.1 , median = 68, mode = 57, 67 bimodal; the mean and median are close but there is a little skewness to the right which is influenced by the data being bimodal. (data are 46, 49, 53, 56, 57, 57, 57, 58, 60, 60, 63, 63, 64, 64, 65, 66, 67, 67, 67, 68, 70, 71, 71, 72, 73, 74, 77, 78, 78, 79, 80, 81, 83, 85, 88, 90, 90 93, 93).
c. These are estimates: mean =16.095, median = 17.495, mode = 22.495 (there may be no mode); The mean < median < mode which indicates skewness to the left. (data are the midponts of the intervals: 2.495, 7.495, 12.495, 17.495, 22.495 and respective frequencies are 2, 3, 4, 7, 9).

Section 7. Measure of the Spread of the Data

TRY IT 2.32

On a baseball team, the ages of each of the players are as follows:

21; 21; 22; 23; 24; 24; 25; 25; 28; 29; 29; 31; 32; 33; 33; 34; 35; 36; 36; 36; 36; 38; 38; 38; 40

Use your calculator or computer to find the mean and standard deviation. Then find the value that is two standard deviations above the mean.

Solution:

$\mu = 30.68$

$s = 6.09$

$(\bar{X} + 2s) = 30.68 + (2)(6.09) = 42.86.$

TRY IT 2.33

The following data show the different types of pet food stores in the area carry.
6; 6; 6; 6; 7; 7; 7; 7; 7; 8; 9; 9; 9; 9; 10; 10; 10; 10; 10; 11; 11; 11; 11; 12; 12; 12; 12; 12; 12;
Calculate the sample mean and the sample standard deviation to one decimal place using a TI-83+ or TI-84 calculator.

Solution:

$\mu = 9.3$

$s = 2.2$

TRY IT 2.34

Find the standard deviation for the data from the previous example

Class	Frequency, *f*
0–2	1
3–5	6
6–8	10

Class	Frequency, *f*
9–11	7
12–14	0
15–17	2

Table2.32

Solution:

First, press the **STAT** key and select **1:Edit**

Figure 2.26

Input the midpoint values into **L1** and the frequencies into **L2**

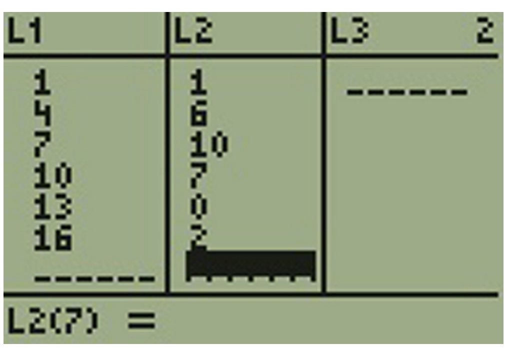


Figure 2.27

Select **STAT**, **CALC**, and **1: 1-Var Stats**

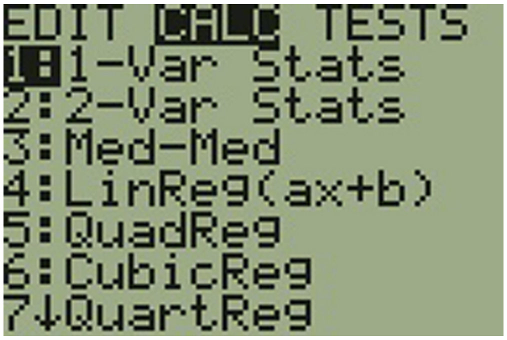


Figure 2.28

Select **2nd** then **1** then , **2nd** then **2 Enter**

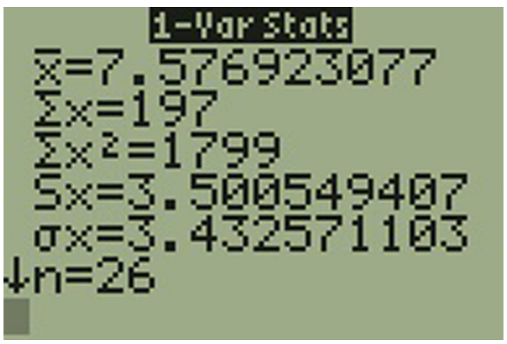


Figure 2.29

You will see displayed both a population standard deviation, σ_x, and the sample standard deviation, s_x.

TRY IT 2.35

Two swimmers, Angie and Beth, from different teams, wanted to find out who had the fastest time for the 50 meter freestyle when compared to her team. Which swimmer had the fastest time when compared to her team?

Swimmer	Time (seconds)	Team Mean Time	Team Standard Deviation
Angie	26.2	27.2	0.8
Beth	27.3	30.1	1.4

Table2.35

Solution:

For Angie: $z = \frac{26.2 - 27.2}{0.8} = -1.25$

For Beth: $z = \frac{27.3 - 30.1}{1.4} = -2$

Chapter 3. Probability Topics

Section 1. Terminology

TRY IT 3.1

The sample space S is all the ordered pairs of two whole numbers, the first from one to three and the second from one to four (Example: (1, 4)).

a. S = ________________________________

Let event A = the sum is even and event B = the first number is prime.

b. A = _______________________, B = _______________________
c. $P(A)$ = ______________, $P(B)$ = _________________
d. A AND B = ______________________, A OR B = _________________
e. $P(A \text{ AND } B)$ = __________, $P(A \text{ OR } B)$ = ______________
f. B' = ______________, $P(B')$ = ______________
g. $P(A) + P(A')$ = _____________
h. $P(A|B)$ = ____________, $P(B|A)$ = ______________; are the probabilities equal?

Solution:

a. S = {(1,1), (1,2), (1,3), (1,4), (2,1), (2,2), (2,3), (2,4), (3,1), (3,2), (3,3), (3,4)}
b. A = {(1,1), (1,3), (2,2), (2,4), (3,1), (3,3)}

B = {(2,1), (2,2), (2,3), (2,4), (3,1), (3,2), (3,3), (3,4)}
c. $P(A) = \frac{1}{2}$, $P(B) = \frac{2}{3}$
d. A AND B = {(2,2), (2,4), (3,1), (3,3)}

A OR B = {(1,1), (1,3), (2,1), (2,2), (2,3), (2,4), (3,1), (3,2), (3,3), (3,4)}
e. $P(A \text{ AND } B) = \frac{1}{3}$, $P(A \text{ OR } B) = \frac{5}{6}$
f. B' = {(1,1), (1,2), (1,3), (1,4)}, $P(B') = \frac{1}{3}$
g. $P(B) + P(B') = 1$
h. $P(A|B) = \frac{\mathrm{P(AANDB)}}{P(B)} = \frac{1}{2}$, $P(B|A) = \frac{\mathrm{P(AANDB)}}{P(B)}$ =23P(A AND B)P(B) = $\frac{2}{3}$, No.

Section 2. Independent and Mutually Exclusive Events

TRY IT 3.4

You have a fair, well-shuffled deck of 52 cards. It consists of four suits. The suits are clubs, diamonds, hearts and spades. There are 13 cards in each suit consisting of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, J (jack), Q (queen), K (king) of that suit. Three cards are picked at random.

a. Suppose you know that the picked cards are *Q* of spades, *K* of hearts and *Q* of spades. Can you decide if the sampling was with or without replacement?
b. Suppose you know that the picked cards are *Q* of spades, *K* of hearts, and *J* of spades. Can you decide if the sampling was with or without replacement?

Solution：

a. With replacement
b. No

TRY IT 3.5

You have a fair, well-shuffled deck of 52 cards. It consists of four suits. The suits are clubs, diamonds, hearts, and spades. There are 13 cards in each suit consisting of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, *J* (jack), *Q* (queen), and *K* (king) of that suit. *S* = spades, *H* = Hearts, *D* = Diamonds, *C* = Clubs. Suppose that you sample four cards without replacement. Which of the following outcomes are possible? Answer the same question for sampling with replacement.

a. *QS*, 1*D*, 1*C*, *QD*
b. *KH*, 7*D*, 6*D*, *KH*
c. *QS*, 7*D*, 6*D*, *KS*

Solution:

without replacement: 1. Possible; 2. Impossible, 3. Possible

with replacement: 1. Possible; 2. Possible, 3. Possible

TRY IT 3.6

Draw two cards from a standard 52-card deck with replacement. Find the probability of getting at least one black card.

Solution:

The sample space of drawing two cards with replacement from a standard 52-card deck with respect to color is {*BB*, *BR*, *RB*, *RR*}.

Event *A* = Getting at least one black card = {*BB*, *BR*, *RB*}

$P(A) = \frac{3}{4} = 0.75$

TRY IT 3.7

A box has two balls, one white and one red. We select one ball, put it back in the box, and select a second ball (sampling with replacement). Find the probability of the following events:

a. Let *F* = the event of getting the white ball twice.
b. Let *G* = the event of getting two balls of different colors.
c. Let *H* = the event of getting white on the first pick.
d. Are *F* and *G* mutually exclusive?
e. Are *G* and *H* mutually exclusive?

Solution:

a. $P(F) = \frac{1}{4}$
b. $P(G) = \frac{1}{2}$
c. $P(H) = \frac{1}{2}$
d. Yes
e. No

TRY IT 3.8

Let event *A* = learning Spanish. Let event *B* = learning German. Then *A* AND *B* = learning Spanish and German. Suppose *P*(*A*) = 0.4 and *P*(*B*) = 0.2. *P*(*A* AND *B*) = 0.08. Are events *A* and *B* independent? Hint: You must show ONE of the following:

- $P(A|B) = P(A)$
- $P(B|A) = P(B)$
- $P(A \text{ AND } B) = P(A)P(B)$

Solution:

$P(A|B) = \frac{P(AANDB)}{P(B)} = \frac{0.08}{0.2} = 0.4 = \mathrm{P(A)}$

The events are independent because $P(A|B) = P(A)$.

TRY IT 3.9

In a bag, there are six red marbles and four green marbles. The red marbles are marked with the numbers 1, 2, 3, 4, 5, and 6. The green marbles are marked with the numbers 1, 2, 3, and 4.

- *R* = a red marble
- *G* = a green marble
- *O* = an odd-numbered marble
- The sample space is *S* = {*R*1, *R*2, *R*3, *R*4, *R*5, *R*6, *G*1, *G*2, *G*3, *G*4}.

S has ten outcomes. What is *P*(*G* AND *O*)?

Solution:

Event *G* and *O* = {*G*1, *G*3}

P(G and O) = $\frac{2}{10}$ = 0.2

TRY IT 3.10

A student goes to the library. Let events B = the student checks out a book and D = the student checks out a DVD. Suppose that $P(B)$ = 0.40, $P(D)$ = 0.30 and P(B AND D) = 0.20.

a. Find $P(B|D)$.
b. Find $P(D|B)$.
c. Are B and D independent?
d. Are B and D mutually exclusive?

Solution:

a. $P(B|D)$ = 0.6667
b. $P(D|B)$ = 0.5
c. No
d. No

TRY IT 3.11

In a basketball arena,

- 70% of the fans are rooting for the home team.
- 25% of the fans are wearing blue.
- 20% of the fans are wearing blue and are rooting for the away team.
- Of the fans rooting for the away team, 67% are wearing blue.

Let A be the event that a fan is rooting for the away team.
Let B be the event that a fan is wearing blue.
Are the events of rooting for the away team and wearing blue independent? Are they mutually exclusive?

Solution:

$P(B|A)$ = 0.67

$P(B)$ = 0.25

So $P(B)$ does not equal $P(B|A)$ which means that B and A are not independent (wearing blue and rooting for the away team are not independent). They are also not mutually exclusive, because P(B AND A) = 0.20, not 0.

TRY IT 3.12

Mark is deciding which route to take to work. His choices are *I* = the Interstate and *F* = Fifth Street.

- *P*(*I*) = 0.44 and *P*(*F*) = 0.56
- *P*(*I* AND *F*) = 0 because Mark will take only one route to work.

What is the probability of *P*(*I* OR *F*)?

Solution:

Because *P*(*I* AND *F*) = 0,

P(*I* OR *F*) = *P*(*I*) + *P*(*F*) - *P*(*I* AND *F*) = 0.44 + 0.56 - 0 = 1

TRY IT 3.13

A box has two balls, one white and one red. We select one ball, put it back in the box, and select a second ball (sampling with replacement). Let *T* be the event of getting the white ball twice, *F* the event of picking the white ball first, *S* the event of picking the white ball in the second drawing.

a. Compute *P*(*T*).
b. Compute *P*(*T*|*F*).
c. Are *T* and *F* independent?.
d. Are *F* and *S* mutually exclusive?
e. Are *F* and *S* independent?

Solution:

a. $P(T) = \frac{1}{4}$
b. $P(T|F) = \frac{1}{2}$
c. No
d. No
e. Yes

Section 3. Two Basic Rules of Probability

TRY IT 3.15

Helen plays basketball. For free throws, she makes the shot 75% of the time. Helen must now attempt two free throws. *C* = the event that Helen makes the first shot. *P*(*C*) = 0.75. *D* = the event Helen makes the second shot. *P*(*D*) = 0.75. The probability that Helen makes the second free throw given that she made the first is 0.85. What is the probability that Helen makes both free throws?

Solution:

$P(D|C) = 0.85$

$P(C \text{ AND } D) = P(D \text{ AND } C)$
$P(D \text{ AND } C) = P(D|C)P(C) = (0.85)(0.75) = 0.6375$
Helen makes the first and second free throws with probability 0.6375.

TRY IT 3.16

A school has 200 seniors of whom 140 will be going to college next year. Forty will be going directly to work. The remainder are taking a gap year. Fifty of the seniors going to college play sports. Thirty of the seniors going directly to work play sports. Five of the seniors taking a gap year play sports. What is the probability that a senior is taking a gap year?

Solution:

$P = \frac{200 - 140 - 40}{200} = \frac{20}{200} = 0.1$

TRY IT 3.17

A student goes to the library. Let events *B* = the student checks out a book and *D* = the student check out a DVD. Suppose that $P(B) = 0.40$, $P(D) = 0.30$ and $P(D|B) = 0.5$.

a. Find $P(B \text{ AND } D)$.
b. Find $P(B \text{ OR } D)$.

Solution:

a. $P(B \text{ AND } D) = P(D|B)P(B) = (0.5)(0.4) = 0.20$.
b. $P(B \text{ OR } D) = P(B) + P(D) - P(B \text{ AND } D) = 0.40 + 0.30 - 0.20 = 0.50$

TRY IT 3.18

A school has 200 seniors of whom 140 will be going to college next year. Forty will be going directly to work. The remainder are taking a gap year. Fifty of the seniors going to college play sports. Thirty of the seniors going directly to work play sports. Five of the seniors taking a gap year play sports. What is the probability that a senior is going to college and plays sports?

Solution:

Let *A* = student is a senior going to college.

Let *B* = student plays sports.

$P(B) = 140/200$

$P(B|A) = 50/140$

$P(A \text{ AND } B) = P(B|A)P(A)$

$P(A \text{ AND } B) = (\frac{140}{200})(\frac{50}{140}) = 1/4$

TRY IT 3.19

A student goes to the library. Let events B = the student checks out a book and D = the student checks out a DVD. Suppose that $P(B) = 0.40$, $P(D) = 0.30$ and $P(D|B) = 0.5$.

a. Find $P(B')$.
b. Find $P(D \text{ AND } B)$.
c. Find $P(B|D)$.
d. Find $P(D \text{ AND } B')$.
e. Find $P(D|B')$.

Solution:

a. $P(B') = 0.60$
b. $P(D \text{ AND } B) = P(D|B)P(B) = 0.20$
c. $P(B|D) = \frac{P(BANDD)}{P(D)} = 0.2/0.3 = 0.66$
d. $P(D \text{ AND } B') = P(D) - P(D \text{ AND } B) = 0.30 - 0.20 = 0.10$
e. $P(D|B') = P(D \text{ AND } B')P(B') = (P(D) - P(D \text{ AND } B))(0.60) = (0.10)(0.60) = 0.06$

Section 4. Contingency Tables

TRY IT 3.20

Table 3.3 shows the number of athletes who stretch before exercising and how many had injuries within the past year.

	Injury in last year	**No injury in last year**	**Total**
Stretches	55	295	350
Does not stretch	231	219	450
Total	286	514	800

Table3.3

a. What is P(athlete stretches before exercising)?
b. What is P(athlete stretches before exercising|no injury in the last year)?

Solution:

a. P(athlete stretches before exercising) = $\frac{350}{800}$ = 0.4375
b. P(athlete stretches before exercising|no injury in the last year) = $\frac{295}{514}$ = 0.5739

TRY IT 3.21

Table 3.6 shows a random sample of 200 cyclists and the routes they prefer. Let M = males and H = hilly path.

Gender	Lake Path	Hilly Path	Wooded Path	Total
Female	45	38	27	110
Male	26	52	12	90
Total	71	90	39	200

Table3.6

a. Out of the males, what is the probability that the cyclist prefers a hilly path?
b. Are the events "being male" and "preferring the hilly path" independent events?

Solution:

a. $P(H|M)$ = 52/90 = 0.5778
b. For M and H to be independent, show $P(H|M) = P(H)$

   $P(H|M)$ = 0.5778, $P(H)$ = 90/200 = 0.45

   $P(H|M)$ does not equal $P(H)$ so M and H are NOT independent.

TRY IT 3.23

Table 3.10 relates the weights and heights of a group of individuals participating in an observational study.

Weight/Height	Tall	Medium	Short	Totals
Obese	18	28	14	

Weight/Height	Tall	Medium	Short	Totals
Normal	20	51	28	
Underweight	12	25	9	
Totals				

Table3.10

a. Find the total for each row and column
b. Find the probability that a randomly chosen individual from this group is Tall.
c. Find the probability that a randomly chosen individual from this group is Obese and Tall.
d. Find the probability that a randomly chosen individual from this group is Tall given that the individual is Obese.
e. Find the probability that a randomly chosen individual from this group is Obese given that the individual is Tall.
f. Find the probability a randomly chosen individual from this group is Tall and Underweight.
g. Are the events Obese and Tall independent?

Solution:

Weight/Height	Tall	Medium	Short	Totals
Obese	18	28	14	60
Normal	20	51	28	99
Underweight	12	25	9	46
Totals	50	104	51	205

a. Row Totals: 60, 99, 46. Column totals: 50, 104, 51.
b. *P*(Tall) = 50/205 = 0.244
c. *P*(Obese AND Tall) = 18/205 = 0.088
d. *P*(Tall|Obese) = 18/60 = 0.3
e. *P*(Obese|Tall) = 18/50 = 0.36
f. *P*(Tall AND Underweight = 12/205 = 0.0585
g. No. *P*(Tall) does not equal *P*(Tall|Obese).

Section 5. Tree and Venn Diagrams

TRY IT 3.24

In a standard deck, there are 52 cards. 12 cards are face cards (event *F*) and 40 cards are not face cards (event *N*). Draw two cards, one at a time, with replacement. All possible outcomes are shown in the tree diagram as frequencies. Using the tree diagram, calculate *P*(*FF*).

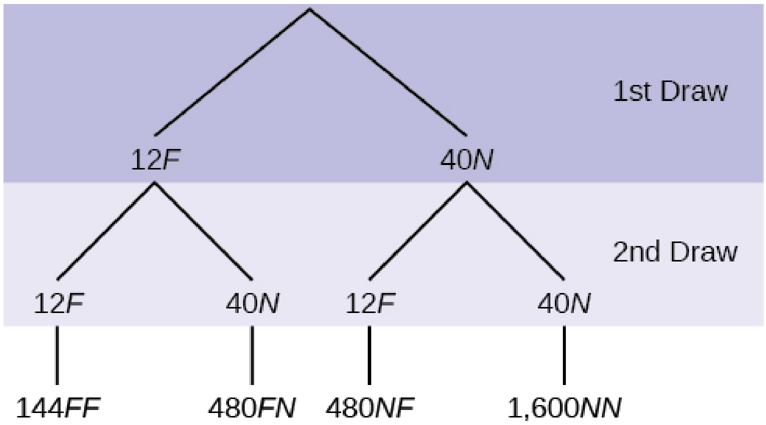


Figure 3.3

Solution:

Total number of outcomes is 144 + 480 + 480 + 1600 = 2,704.

$P(FF) = \frac{144}{144+480+480+1600} = \frac{144}{2704} = \frac{9}{169}$

TRY IT 3.25

In a standard deck, there are 52 cards. Twelve cards are face cards (*F*) and 40 cards are not face cards (*N*). Draw two cards, one at a time, without replacement. The tree diagram is labeled with all possible probabilities.

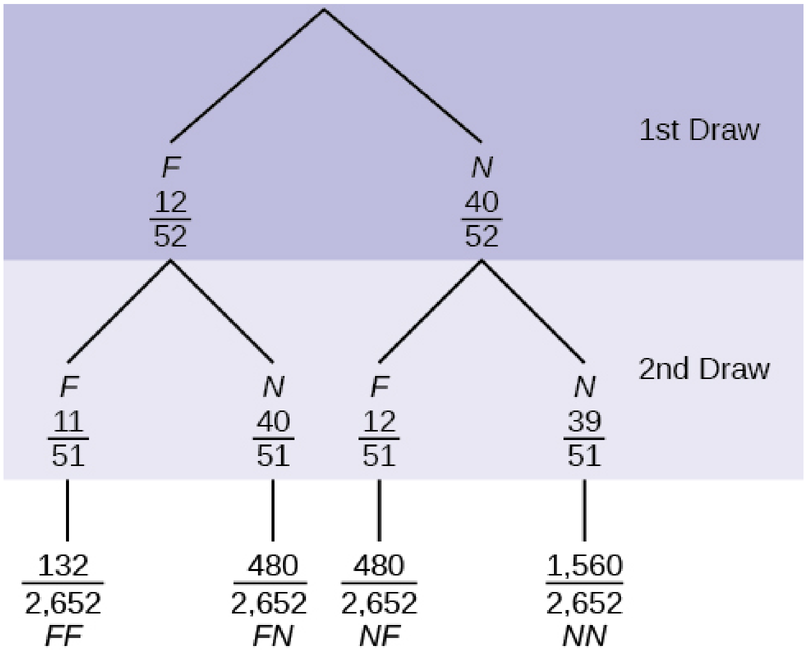


Figure 3.5

a. Find *P*(*FN* OR *NF*).
b. Find *P*(*N*|*F*).
c. Find *P*(at most one face card).
   Hint: "At most one face card" means zero or one face card.
d. Find *P*(at least on face card).
   Hint: "At least one face card" means one or two face cards.

Solution:

a. $P(FN \text{ OR } NF) = \frac{480}{2652} + \frac{480}{2652} = \frac{960}{2652} = \frac{80}{221}$
b. $P(N|F) = \frac{40}{51}$
c. $P(\text{at most one face card}) = \frac{480 + 480 + 1560}{2652} = \frac{2520}{2652}$
d. $P(\text{at least one face card}) = \frac{480 + 480 + 132}{2652} = \frac{1092}{2652}$

TRY IT 3.26

Suppose there are four red balls and three yellow balls in a box. Two balls are drawn from the box without replacement. What is the probability that one ball of each coloring is selected?

Solution:

$\left(\frac{4}{7}\right)\left(\frac{3}{6}\right) + \left(\frac{3}{7}\right)\left(\frac{4}{6}\right)$

TRY IT 3.27

Suppose an experiment has outcomes black, white, red, orange, yellow, green, blue, and purple, where each outcome has an equal chance of occurring. Let event *C* = {green, blue, purple} and event *P* = {red, yellow, blue}. Then *C* AND *P* = {blue} and *C*OR *P* = {green, blue, purple, red, yellow}. Draw a Venn diagram representing this situation.

Solution:

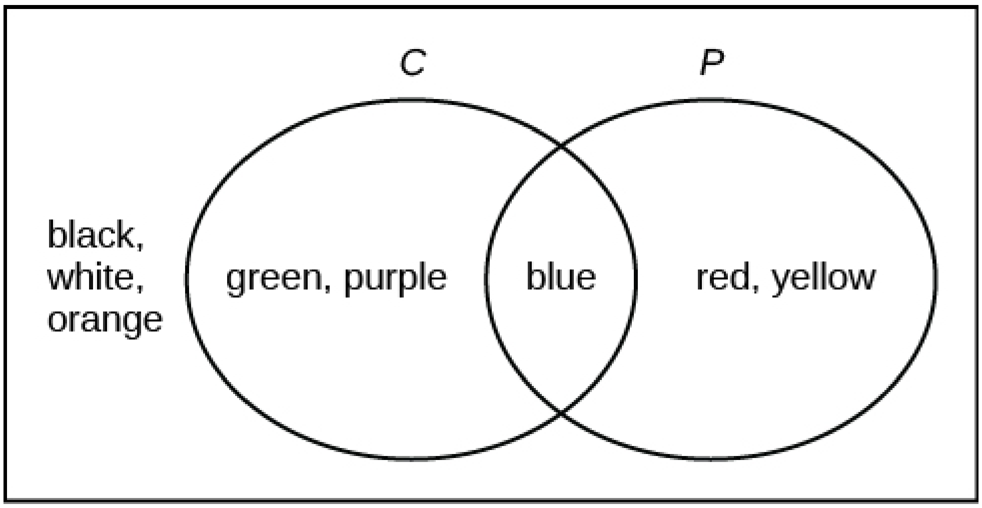


Figure 6.

TRY IT 3.28

Roll a fair, six-sided die. Let *A* = a prime number of dots is rolled. Let *B* = an odd number of dots is rolled. Then *A* = {2, 3, 5} and *B* = {1, 3, 5}. Therefore, *A* AND *B* = {3, 5}. *A* OR *B* = {1, 2, 3, 5}. The sample space for rolling a fair die is *S* = {1, 2, 3, 4, 5, 6}. Draw a Venn diagram representing this situation.

Solution:

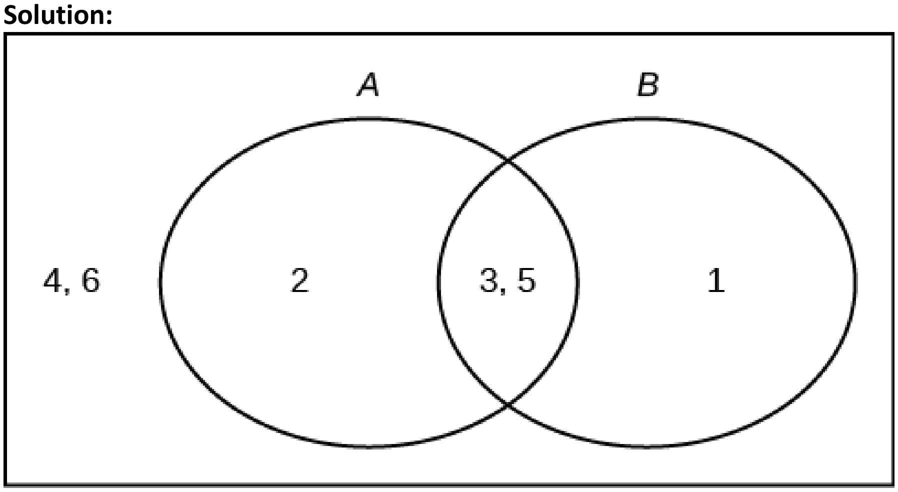


Figure 8.

TRY IT 3.29

Fifty percent of the workers at a factory work a second job, 25% have a spouse who also works, 5% work a second job and have a spouse who also works. Draw a Venn diagram showing the relationships. Let *W* = works a second job and *S* = spouse also works.

Solution:

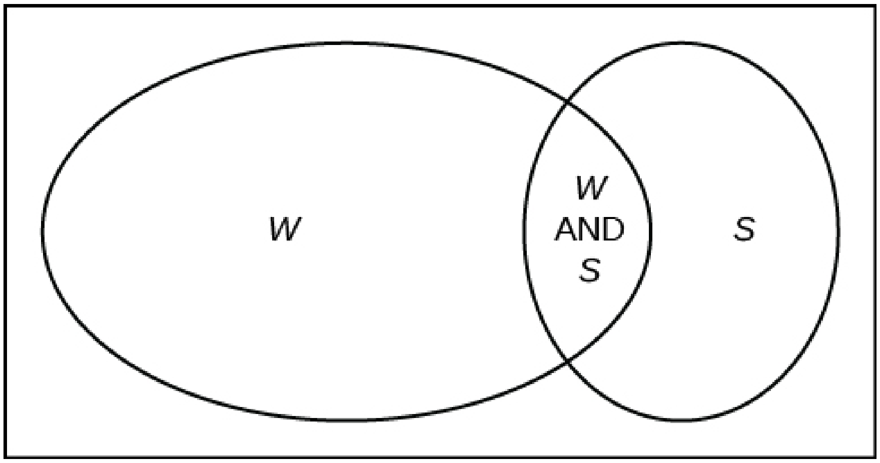


Figure 10.

TRY IT 3.30

In a bookstore, the probability that the customer buys a novel is 0.6, and the probability that the customer buys a non-fiction book is 0.4. Suppose that the probability that the customer buys both is 0.2.

a. Draw a Venn diagram representing the situation.
b. Find the probability that the customer buys either a novel or anon-fiction book.
c. In the Venn diagram, describe the overlapping area using a complete sentence.
d. Suppose that some customers buy only compact disks. Draw an oval in your Venn diagram representing this event.

Solution:

a. and d. In the following Venn diagram below, the blue oval represent customers buying a novel, the red oval represents customer buying non-fiction, and the yellow oval customer who buy compact disks.

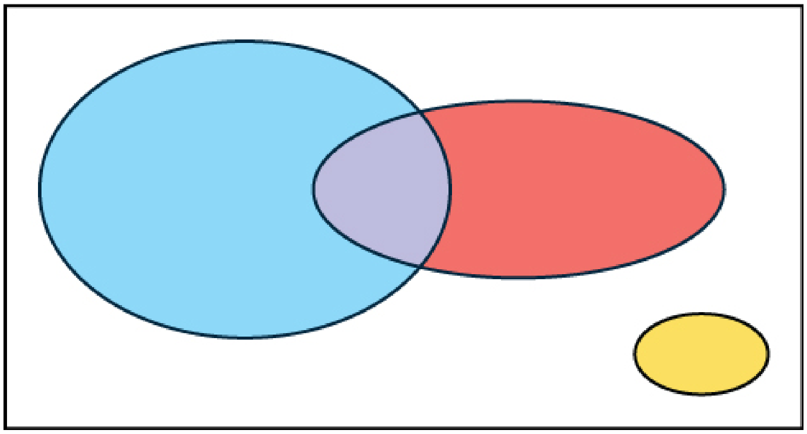


Figure 12.

b. P(novel or non-fiction) = P(Blue OR Red) = P(Blue) + P(Red) - P(Blue AND Red) = 0.6 + 0.4 - 0.2 = 0.8.
c. The overlapping area of the blue oval and red oval represents the customers buying both a novel and a nonfiction book.

Chapter 4. Discrete Random Variables

Section 1. Probability Distribution Function (PDF) for a Discrete Random Variable

TRY IT 4.1

A hospital researcher is interested in the number of times the average post-op patient will ring the nurse during a 12-hour shift. For a random sample of 50 patients, the following information was obtained. Let X = the number of times a patient rings the nurse during a 12-hour shift. For this exercise, x = 0, 1, 2, 3, 4, 5. $P(x)$ = the probability that X takes on value x. Why is this a discrete probability distribution function (two reasons)?

X	***P*(*x*)**
0	$P(x = 0) = \frac{4}{50}$
1	$P(x = 1) = \frac{8}{50}$
2	$P(x = 2) = \frac{16}{50}$
3	$P(x = 3) = \frac{14}{50}$

X	*P*(*x*)
4	$P(x = 4) = \frac{6}{50}$
5	$P(x = 5) = \frac{2}{50}$

Table4.3

Solution:

Each *P*(*x*) is between 0 and 1, inclusive, and the sum of the probabilities is 1, that is: $\frac{4}{50} + \frac{8}{50} + \frac{16}{50} + \frac{14}{50} + \frac{6}{50} + \frac{2}{50} = 1$

TRY IT 4.2

Jeremiah has basketball practice two days a week. Ninety percent of the time, he attends both practices. Eight percent of the time, he attends one practice. Two percent of the time, he does not attend either practice. What is *X* and what values does it take on?

Solution:

X is the number of days Jeremiah attends basketball practice per week. *X* takes on the values 0, 1, and 2.

TRY IT 4.4

A hospital researcher is interested in the number of times the average post-op patient will ring the nurse during a 12-hour shift. For a random sample of 50 patients, the following information was obtained. What is the expected value?

X	*P*(*x*)
0	$P(x = 0) = \frac{4}{50}$
1	$P(x = 1) = \frac{8}{50}$
2	$P(x = 2) = \frac{16}{50}$
3	$P(x = 3) = \frac{14}{50}$

X	*P*(*x*)
4	$P(x = 4) = \frac{6}{50}$
5	$P(x = 5) = \frac{2}{50}$

Table4.7

Solution:

$(0)\frac{4}{50} + (1)\frac{8}{50} + (2)\frac{16}{50} + (3)\frac{14}{50} + (4)\frac{6}{50} + (5)\frac{2}{50} = 0 + \frac{8}{50} + \frac{32}{50} + \frac{42}{50} + \frac{24}{50} + \frac{10}{50} = \frac{116}{50} = 2.32$

TRY IT 4.5

You are playing a game of chance in which four cards are drawn from a standard deck of 52 cards. You guess the suit of each card before it is drawn. The cards are replaced in the deck on each draw. You pay $1 to play. If you guess the right suit every time, you get your money back and $256. What is your expected profit of playing the game over the long term?

Solution:

Let *X* = the amount of money you profit. The *x*-values are –$1 and $256.

The probability of guessing the right suit each time is $(\frac{1}{4})(\frac{1}{4})(\frac{1}{4})(\frac{1}{4}) = (\frac{1}{256}) = 0.0039$

The probability of losing is $1 - (\frac{1}{256}) = (\frac{255}{256}) = 0.9961$

(0.0039)256 + (0.9961)(–1) = 0.9984 + (–0.9961) = 0.0023 or 0.23 cents.

TRY IT 4.6

Suppose you play a game with a spinner. You play each game by spinning the spinner once. $P(\text{red}) = \frac{2}{5}$, $P(\text{blue}) = \frac{2}{5}$, and $P(\text{green}) = \frac{1}{5}$. If you land on red, you pay $10. If you land on blue, you don't pay or win anything. If you land on green, you win $10. Complete the following expected value table.

	x	*P*(*x*)	*x***P*(*x*)
Red			$-\frac{20}{5}$

	x	$P(x)$	$x*P(x)$
Blue		$\frac{2}{5}$	
Green	10		

Table4.11

Solution:

	x	$P(x)$	$x*P(x)$
Red	–10	$\frac{2}{5}$	$-\frac{20}{5}$
Blue	0	$\frac{2}{5}$	$\frac{0}{5}$
Green	10	$\frac{1}{5}$	$\frac{10}{5}$

TRY IT 4.8

	x	$P(x)$	$x \cdot (Px)$	$(x - \mu)^2 \cdot P(x)$
win	100	0.0108	1.08	$[100 - (-8.812)]^2 \cdot 0.0108 = 127.8726$
loss	–10	0.9892	–9.892	$[-10 - (-8.812)]^2 \cdot 0.9892 = 1.3961$

On May 11, 2013 at 9:30 PM, the probability that moderate seismic activity (one moderate earthquake) would occur in the next 48 hours in Japan was about 1.08%. As in Example 4.8, you bet that a moderate earthquake will occur in Japan during this period. If you win the bet, you win $100. If you lose the bet, you pay $10. Let X = the amount of profit from a bet. Find the mean and standard deviation of X.

Solution:

Mean = Expected Value = μ = 1.08 + (–9.892) = –8.812

If you make this bet many times under the same conditions, your long term outcome will be an average loss of $8.81 per bet.

Standard Deviation = $\sqrt{127.7826 + 1.3961} \approx 11.3696.$

TRY IT 4.9

The state health board is concerned about the amount of fruit available in school lunches. Forty-eight percent of schools in the state offer fruit in their lunches every day. This implies that 52% do not. What would a "success" be in this case?

Solution:

a school that offers fruit in their lunch every day

TRY IT 4.10

A trainer is teaching a dolphin to do tricks. The probability that the dolphin successfully performs the trick is 35%, and the probability that the dolphin does not successfully perform the trick is 65%. Out of 20 attempts, you want to find the probability that the dolphin succeeds 12 times. State the probability question mathematically.

Solution:
$P(x = 12)$

TRY IT 4.11

A fair, six-sided die is rolled ten times. Each roll is independent. You want to find the probability of rolling a one more than three times. State the probability question mathematically.

Solution:

$P(x > 3)$

TRY IT 4.12

Sixty-five percent of people pass the state driver's exam on the first try. A group of 50 individuals who have taken the driver's exam is randomly selected. Give two reasons why this is a binomial problem.

Solution:
This is a binomial problem because there is only a success or a failure, and there are a definite number of trials. The probability of a success stays the same for each trial.

TRY IT 4.13

About 32% of students participate in a community volunteer program outside of school. If 30 students are selected at random, find the probability that at most 14 of them participate in a community volunteer program outside of school. Use the **TI-83+** or **TI-84** calculator to find the answer.

Solution:
$P(x \leq 14) = 0.9695$

TRY IT 4.14

According to a Gallup poll, 60% of American adults prefer saving over spending. Let *X* = the number of American adults out of a random sample of 50 who prefer saving to spending.

a. What is the probability distribution for *X*?
b. Use your calculator to find the following probabilities:
   i. the probability that 25 adults in the sample prefer saving over spending
   ii. the probability that at most 20 adults prefer saving
   iii. the probability that more than 30 adults prefer saving
c. Using the formulas, calculate the (i) mean and (ii) standard deviation of *X*.

Solution:

a. $X \sim B(50, 0.6)$
b. Using the **TI-83, 83+, 84** calculator with instructions as provided in Example:
   i. $P(x = 25)$ = **binompdf**(50, 0.6, 25) = 0.0405
   ii. $P(x \leq 20)$ = **binomcdf**(50, 0.6, 20) = 0.0034
   iii. $P(x > 30)$ = 1 - **binomcdf**(50, 0.6, 30) = 1 – 0.5535 = 0.4465
c.
   i. Mean = np = 50(0.6) = 30
   ii. Standard Deviation = $\sqrt{npq} = \sqrt{50(0.6)(0.4)} \approx 3.4641$

TRY IT 4.15

During the 2013 regular NBA season, DeAndre Jordan of the Los Angeles Clippers had the highest field goal completion rate in the league. DeAndre scored with 61.3% of his shots. Suppose you choose a random sample of 80 shots made by DeAndre during the 2013 season. Let *X* = the number of shots that scored points.

a. What is the probability distribution for *X*?
b. Using the formulas, calculate the (i) mean and (ii) standard deviation of *X*.
c. Use your calculator to find the probability that DeAndre scored with 60 of these shots.
d. Find the probability that DeAndre scored with more than 50 of these shots.

Solution:

a. $X \sim B(80, 0.613)$
b.
   i. Mean = np = 80(0.613) = 49.04
   ii. Standard Deviation = $\sqrt{npq} = \sqrt{80(0.613)(0.387)}$ 4.3564
c. Using the **TI-83, 83+, 84** calculator with instructions as provided in Example:
   $P(x = 60)$ = **binompdf**(80, 0.613, 60) = 0.0036
d. $P(x > 50) = 1 - P(x \leq 50)$ = 1 – **binomcdf**(80, 0.613, 50) = 1 – 0.6282 = 0.3718

TRY IT 4.16

A lacrosse team is selecting a captain. The names of all the seniors are put into a hat, and the first three that are drawn will be the captains. The names are not replaced once they are drawn (one person cannot be two captains). You want to see if the captains all play the same position. State whether this is binomial or not and state why.

Solution:

This is not binomial because the names are not replaced, which means the probability changes for each time a name is drawn. This violates the condition of independence.

Section 4. Geometric Distribution

TRY IT 4.17

You throw darts at a board until you hit the center area. Your probability of hitting the center area is $p = 0.17$. You want to find the probability that it takes eight throws until you hit the center. What values does X take on?

Solution:

1, 2, 3, 4, … *n*. It can go on indefinitely.

TRY IT 4.18

An instructor feels that 15% of students get below a C on their final exam. She decides to look at final exams (selected randomly and replaced in the pile after reading) until she finds one that shows a grade below a C. We want to know the probability that the instructor will have to examine at least ten exams until she finds one with a grade below a C. What is the probability question stated mathematically?

Solution:
$P(x \geq 10)$

TRY IT 4.19

You need to find a store that carries a special printer ink. You know that of the stores that carry printer ink, 10% of them carry the special ink. You randomly call each store until one has the ink you need. What are p and q?

Solution:
$p = 0.1$
$q = 0.9$

TRY IT 4.20

The probability of a defective steel rod is 0.01. Steel rods are selected at random. Find the probability that the first defect occurs on the ninth steel rod. Use the **TI-83+** or **TI-84** calculator to find the answer.

Solution:

P(x = 9) = 0.0092

TRY IT 4.21

The literacy rate for a nation measures the proportion of people age 15 and over who can read and write. The literacy rate for women in Afghanistan is 12%. Let *X* = the number of Afghani women you ask until one says that she is literate.

a. What is the probability distribution of *X*?
b. What is the probability that you ask five women before one says she is literate?
c. What is the probability that you must ask ten women?
d. Find the (i) mean and (ii) standard deviation of *X*.

Solution:

a. $X \sim G(0.12)$
b. $P(x = 5)$ = **geometpdf**(0.12, 5) = 0.0720
c. $P(x = 10)$ = **geometpdf**(0.12, 10) = 0.0380
d.
   I. Mean = $\mu = \frac{1}{p} = \frac{1}{0.12} \approx 8.333$
   II. Standard Deviation = $\sigma = \sqrt{\frac{1-p}{p^2}} = \sqrt{\frac{1-0.12}{0.12^2}} \approx 7.8174$

TRY IT 4.22

A bag contains letter tiles. Forty-four of the tiles are vowels, and 56 are consonants. Seven tiles are picked at random. You want to know the probability that four of the seven tiles are vowels. What is the group of interest, the size of the group of interest, and the size of the sample?

Solution:
The group of interest is the vowel letter tiles. The size of the group of interest is 44. The size of the sample is seven.

TRY IT 4.23

A gross of eggs contains 144 eggs. A particular gross is known to have 12 cracked eggs. An inspector randomly chooses 15 for inspection. She wants to know the probability that, among the 15, at most three are cracked. What is *X*, and what values does it take on?

Solution:
Let X = the number of cracked eggs in the sample of 15. X takes on the values 0, 1, 2, …, 12.

TRY IT 4.24

A palette has 200 milk cartons. Of the 200 cartons, it is known that ten of them have leaked and cannot be sold. A stock clerk randomly chooses 18 for inspection. He wants to know the probability that among the 18, no more than two are leaking. Give five reasons why this is a hypergeometric problem.

Solution:

1. There are two groups.
2. You are concerned with a group of interest.
3. You sample without replacement.
4. Each pick is not independent.
5. You are not dealing with Bernoulli Trials.

TRY IT 4.25

An intramural basketball team is to be chosen randomly from 15 boys and 12 girls. The team has ten slots. You want to know the probability that eight of the players will be boys. What is the group of interest and the sample?

Solution:

The group of interest is the 15 boys. The sample consists of the ten slots on the intramural basketball team.

TRY IT 4.26

The average number of fish caught in an hour is eight. Of interest is the number of fish caught in 15 minutes. The time interval of interest is 15 minutes. What is the average number of fish caught in 15 minutes?

Solution:

(15/60)(8) = 2 fish

TRY IT 4.27

An electronics store expects to have ten returns per day on average. The manager wants to know the probability of the store getting fewer than eight returns on any given day. State the probability question mathematically.

Solution:

$P(x < 8)$

TRY IT 4.28

An emergency room at a particular hospital gets an average of five patients per hour. A doctor wants to know the probability that the ER gets more than five patients per hour. Give the reason why this would be a Poisson distribution.

Solution:

This problem wants to find the probability of events occurring in a fixed interval of time with a known average rate. The events are independent.

TRY IT 4.29

A customer service center receives about ten emails every half-hour. What is the probability that the customer service center receives more than four emails in the next six minutes? Use the **TI-83+** or **TI-84** calculator to find the answer.

Solution:

$P(x > 4) = 0.0527$

TRY IT 4.30

According to a recent poll by the Pew Internet Project, girls between the ages of 14 and 17 send an average of 187 text messages each day. Let X = the number of texts that a girl aged 14 to 17 sends per day. The discrete random variable X takes on the values x= 0, 1, 2 …. The random variable X has a Poisson distribution: $X \sim P(187)$. The mean is 187 text messages.

a. What is the probability that a teen girl sends exactly 175 texts per day?
b. What is the probability that a teen girl sends at most 150 texts per day?
c. What is the standard deviation?

Solution:

a. $P(x = 175)$ = **poissonpdf**(187, 175) ≈ 0.0203
b. $P(x \leq 150)$ = **poissoncdf**(187, 150) ≈ 0.0030
c. Standard Deviation = $\sigma=\sqrt{\mu} = \sqrt{187} \approx 13.6748$

TRY IT 4.31

Atlanta's Hartsfield-Jackson International Airport is the busiest airport in the world. On average there are 2,500 arrivals and departures each day.

a. How many airplanes arrive and depart the airport per hour?
b. What is the probability that there are exactly 100 arrivals and departures in one hour?
c. What is the probability that there are at most 100 arrivals and departures in one hour?

Solution:

a. Let X = the number of airplanes arriving and departing from Hartsfield-Jackson in one hour. The average number of arrivals and departures per hour is 2,500/24 ≈ 104.1667.
b. $X \sim P(104.1667)$, so $P(x = 100)$ = **poissonpdf**(104.1667, 100) ≈ 0.0366.
c. $P(x \leq 100)$ = **poissoncdf**(104.1667, 100) ≈ 0.3651.

TRY IT 4.32

On May 13, 2013, starting at 4:30 PM, the probability of moderate seismic activity for the next 48 hours in the Kuril Islands off the coast of Japan was reported at about 1.43%. Use this information for the next 100 days to find the probability that there will be low seismic activity in five of the next 100 days. Use both the binomial and Poisson distributions to calculate the probabilities. Are they close?

Solution:

Let X = the number of days with moderate seismic activity.

Using the binomial distribution: $P(x = 5)$ = **binompdf**(100, 0.0143, 5) ≈ 0.0115

Using the Poisson distribution:

- Calculate $\mu = np = 100(0.0143) = 1.43$
- $P(x = 5)$ = **poissonpdf**(1.43, 5) = 0.0119

We expect the approximation to be good because n is large (greater than 20) and p is small (less than 0.05). The results are close—the difference between the values is 0.0004.

Chapter 5. Continuous Random Variables

Section 1. Continuous Probability Functions

TRY IT 5.1

Consider the function $f(x) = \frac{1}{8}$ for $0 \le x \le 8$. Draw the graph of $f(x)$ and find $P(2.5 < x < 7.5)$.

Solution:

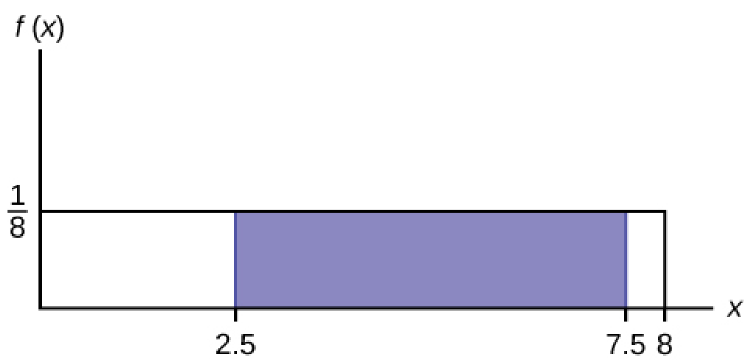


Figure 7.

$P(2.5 < x < 7.5) = 0.625$

Section 2. The Uniform Distribution

TRY IT 5.2

The data that follow are the number of passengers on 35 different charter fishing boats. The sample mean = 7.9 and the sample standard deviation = 4.33. The data follow a uniform distribution where all values between and including zero and 14 are equally likely. State the values of *a* and *b*. Write the distribution in proper notation, and calculate the theoretical mean and standard deviation.

1	12	4	10	4	14	11
7	11	4	13	2	4	6
3	10	0	12	6	9	10
5	13	4	10	14	12	11
6	10	11	0	11	13	2

Table 5.2

Solution:

a is zero; b is 14; $X \sim U(0, 14)$; $\mu = 7$ passengers; $\sigma = 4.04$ passengers

TRY IT 5.3

A distribution is given as $X \sim U(0, 20)$. What is $P(2 < x < 18)$? Find the 90th percentile.

Solution:

$P(2 < x < 18) = 0.8$; 90th percentile = 18

TRY IT 5.4

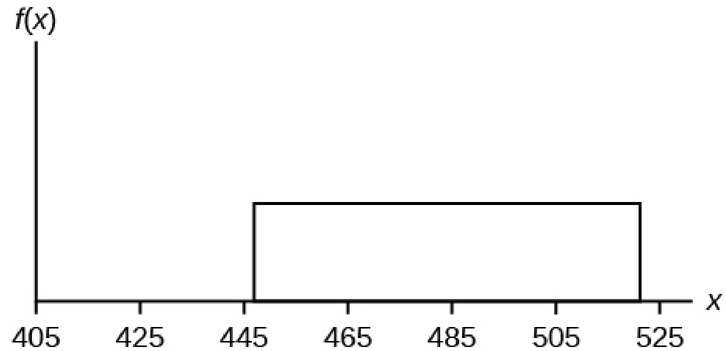
The total duration of baseball games in the major league in the 2011 season is uniformly distributed between 447 hours and 521 hours inclusive.

a. Find *a* and *b* and describe what they represent.
b. Write the distribution.
c. Find the mean and the standard deviation.
d. What is the probability that the duration of games for a team for the 2011 season is between 480 and 500 hours?
e. What is the 65th percentile for the duration of games for a team for the 2011 season?

Solution:

a. *a* is 447, and *b* is 521. *a* is the minimum duration of games for a team for the 2011 season, and *b* is the maximum duration of games for a team for the 2011 season.

b. $X \sim U(447, 521)$.



c. $\mu = 484$, and $\sigma = 21.36$
d. $P(480 < x < 500) = 0.2703$
e. 65th percentile is 495.1 hours.

TRY IT 5.5

Suppose the time it takes a student to finish a quiz is uniformly distributed between six and 15 minutes, inclusive. Let X = the time, in minutes, it takes a student to finish a quiz. Then $X \sim U(6, 15)$.

Find the probability that a randomly selected student needs at least eight minutes to complete the quiz. Then find the probability that a different student needs at least eight minutes to finish the quiz given that she has already taken more than seven minutes.

Solution:

$P(x > 8) = 0.7778$

$P(x > 8 \mid x > 7) = 0.875$

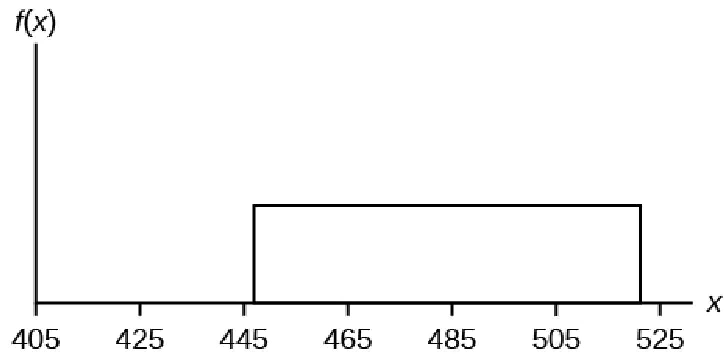
TRY IT 5.6

The amount of time a service technician needs to change the oil in a car is uniformly distributed between 11 and 21 minutes. Let X = the time needed to change the oil on a car.

a. Write the random variable X in words. X = ____________________.
b. Write the distribution.
c. Graph the distribution.
d. Find $P(x > 19)$.
e. Find the 50th percentile.

Solution:

a. Let X = the time needed to change the oil in a car.
b. $X \sim U(11, 21)$.



c.
d. $P(x > 19) = 0.2$
e. the 50th percentile is 16 minutes.

Section 3. The Exponential Distribution

TRY IT 5.7

The amount of time spouses shop for anniversary cards can be modeled by an exponential distribution with the average amount of time equal to eight minutes. Write the distribution, state the probability density function, and graph the distribution.

Solution:

$X \sim Exp(0.125)$; $f(x) = 0.125e^{-0.125x}$;

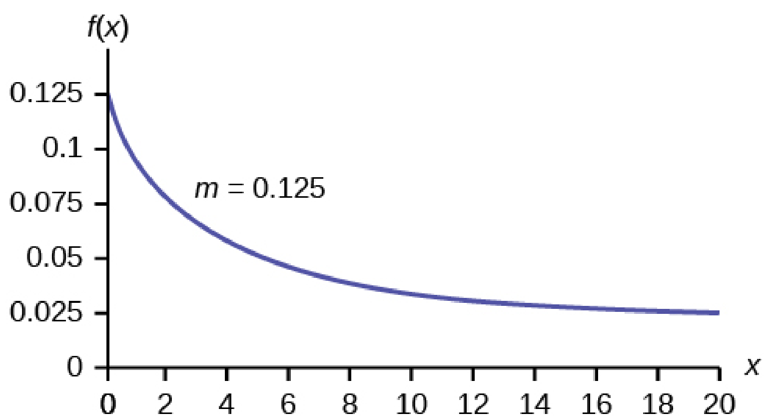


Figure 2.

TRY IT 5.8

The number of days ahead travelers purchase their airline tickets can be modeled by an exponential distribution with the average amount of time equal to 15 days. Find the probability that a traveler will purchase a ticket fewer than ten days in advance. How many days do half of all travelers wait?

Solution:

$P(x < 10) = 0.4866$

50^{th} percentile = 10.40

TRY IT 5.9

On average, a pair of running shoes can last 18 months if used every day. The length of time running shoes last is exponentially distributed. What is the probability that a pair of running shoes last more than 15 months? On average, how long would six pairs of running shoes last if they are used one after the other? Eighty percent of running shoes last at most how long if used every day?

Solution:

$P(x > 15) = 0.4346$

Six pairs of running shoes would last 108 months on average.

80^{th} percentile = 28.97 months

TRY IT 5.10

Suppose that the distance, in miles, that people are willing to commute to work is an exponential random variable with a decay parameter $\frac{1}{20}$. Let X = the distance people are willing to commute in miles. What is m, μ, and σ? What is the probability that a person is willing to commute more than 25 miles?

Solution:

$m = \frac{1}{20}$; $\mu = 20$; $\sigma = 20$; $P(x > 25) = 0.2865$

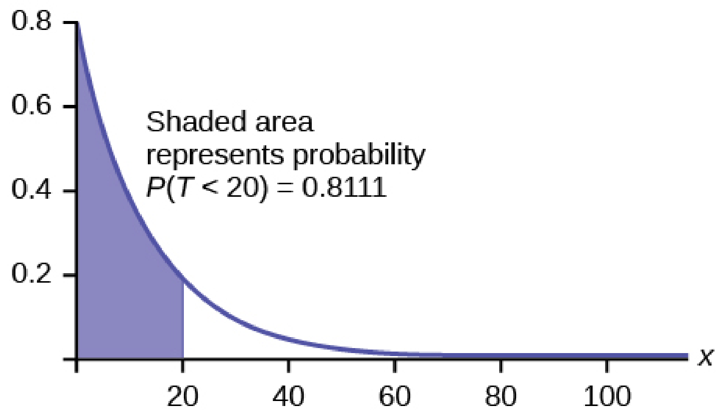
TRY IT 5.11

Suppose that on a certain stretch of highway, cars pass at an average rate of five cars per minute. Assume that the duration of time between successive cars follows the exponential distribution.

a. On average, how many seconds elapse between two successive cars?
b. After a car passes by, how long on average will it take for another seven cars to pass by?
c. Find the probability that after a car passes by, the next car will pass within the next 20 seconds.
d. Find the probability that after a car passes by, the next car will not pass for at least another 15 seconds.

Solution:

a. At a rate of five cars per minute, we expect $\frac{60}{5}$ = 12 seconds to pass between successive cars on average.
b. Using the answer from part a, we see that it takes (12)(7) = 84 seconds for the next seven cars to pass by.
c. Let T = the time (in seconds) between successive cars.
The mean of T is 12 seconds, so the decay parameter is $\frac{1}{12}$ and $T \sim Exp\frac{1}{12}$. The cumulative distribution function of T is $P(T < t) = 1 - e^{-t/12}$. Then $P(T < 20) = 1 - e^{-20/12} \approx$



0.8111.
$P(T > 15) = 1 - P(T < 15) = 1 - (1 - e^{-15/12}) = e^{-15/12} \approx 0.2865$.

TRY IT 5.12

Suppose that the longevity of a light bulb is exponential with a mean lifetime of eight years. If a bulb has already lasted 12 years, find the probability that it will last a total of over 19 years.

Solution:

Let T = the lifetime of the light bulb. Then $T \sim Exp(1/8)$.

The cumulative distribution function is $P(T < t) = 1 - e^{-t/s}$

We need to find $P(T > 19 | T = 12)$. By the **memoryless property**,

$P(T>19 | T = 12) = P(T > 7) = 1 - P(T < 7) = 1 - (1 - e^{-7/8}) = e^{-7/8} \approx 0.4169$.

TRY IT 5.13

In a small city, the number of automobile accidents occur with a Poisson distribution at an average of three per week.

a. Calculate the probability that there are at most 2 accidents occur in any given week.
b. What is the probability that there is at least two weeks between any 2 accidents?

Solution:

a. Let X = the number of accidents per week, so that $X \sim Poisson(3)$. We need to find $P(X \leq 2) \approx 0.4232$

   poissoncdf(3, 2)

b. Let T = the time (in weeks) between successive accidents.
   Since the number of accidents occurs with a Poisson distribution, the time between accidents follows the exponential distribution.
   If there are an average of three per week, then on average there is $\mu = 1/3$ of a week between accidents, and the decay parameter is $m = \frac{1}{\frac{1}{3}} = 3$.
   To find the probability that there are at least two weeks between two accidents, $P(T > 2) = 1 – P(T < 2) = 1 – (1 – e(–3)(2)) = e^{–6} \approx 0.0025$.

   e^(-3*2).

Chapter 6. The Normal Distribution

Section 1. The Standard Normal Distribution

TRY IT 6.1

What is the *z*-score of *x*, when $x = 1$ and $X \sim N(12,3)$?

Solution:

$z = \frac{1-12}{3} \approx –3.67$

TRY IT 6.2

Fill in the blanks.

Jerome averages 16 points a game with a standard deviation of four points. $X \sim N(16,4)$. Suppose Jerome scores ten points in a game. The *z*–score when $x = 10$ is –1.5. This score tells you that $x = 10$ is _____ standard deviations to the ______(right or left) of the mean______(What is the mean?).

Solution:

1.5, left, 16

TRY IT 6.3

Use the information in Example 6.3 to answer the following questions.

a. Suppose a 15 to 18-year-old male from Chile was 176 cm tall from 2009 to 2010. The *z*-score when x = 176 cm is z = _______. This *z*-score tells you that x = 176 cm is ________ standard deviations to the ________ (right or left) of the mean _____ (What is the mean?).
b. Suppose that the height of a 15 to 18-year-old male from Chile from 2009 to 2010 has a *z*-score of $z = -2$. What is the male's height? The *z*-score ($z = -2$) tells you that the male's height is ________ standard deviations to the __________ (right or left) of the mean.

Solution:

Solve the equation $z = (x-\mu)/\sigma$ for x. $x = \mu + (z)(\sigma)$

a. $z = (176-170)/6.28 \approx 0.96$, This *z*-score tells you that x = 176 cm is 0.96 standard deviations to the right of the mean 170 cm.
b. X = 157.44 cm, The *z*-score($z = -2$) tells you that the male's height is two standard deviations to the left of the mean.

TRY IT 6.4

In 2012, 1,664,479 students took the SAT exam. The distribution of scores in the verbal section of the SAT had a mean $\mu = 496$ and a standard deviation $\sigma = 114$. Let X = a SAT exam verbal section score in 2012. Then $X \sim N(496, 114)$.

Find the *z*-scores for $x_1 = 325$ and $x_2 = 366.21$. Interpret each *z*-score. What can you say about $x_1 = 325$ and $x_2 = 366.21$ as they compare to their respective means and standard deviations?

Solution:

The *z*-score for $x_1 = 325$ is $z_1 = -1.5$.

The *z*-score for $x_2 = 366.21$ is $z_2 = -1.14$.

Student 2 scored closer to the mean than Student 1 and, since they both had negative *z*-scores, Student 2 had the better score.

TRY IT 6.5

Suppose X has a normal distribution with mean 25 and standard deviation five. Between what values of x do 68% of the values lie?

Solution:

between 20 and 30.

TRY IT 6.6

The scores on a college entrance exam have an approximate normal distribution with mean, μ = 52 points and a standard deviation, σ = 11 points.

a. About 68% of the *y* values lie between what two values? These values are ________________. The *z*-scores are ________________, respectively.
b. About 95% of the *y* values lie between what two values? These values are ________________. The *z*-scores are ________________, respectively.
c. About 99.7% of the *y* values lie between what two values? These values are ________________. The *z*-scores are ________________, respectively.

Solution:

a) About 68% of the values lie between the values 41 and 63. The *z*-scores are –1 and 1, respectively.
b) About 95% of the values lie between the values 30 and 74. The *z*-scores are –2 and 2, respectively.
c) About 99.7% of the values lie between the values 19 and 85. The *z*-scores are –3 and 3, respectively.

Section 2. Using the Normal Distribution

TRY IT 6.7

If the area to the left of *x* is 0.012, then what is the area to the right?

Solution:

1 − 0.012 = 0.988

TRY IT 6.8

The golf scores for a school team were normally distributed with a mean of 68 and a standard deviation of three.

Find the probability that a randomly selected golfer scored less than 65.

Solution:

normalcdf(0,65,68,3) = 0.1587

TRY IT 6.9

The golf scores for a school team were normally distributed with a mean of 68 and a standard deviation of three. Find the probability that a golfer scored between 66 and 70.

Solution:

normalcdf(66,70,68,3) = 0.4950

TRY IT 6.10

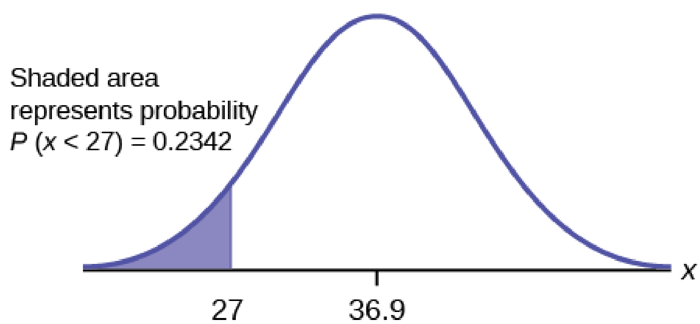
Use the information in Example 6.10 to answer the following questions.

a. Find the 30th percentile, and interpret it in a complete sentence.
b. What is the probability that the age of a randomly selected smartphone user in the range 13 to 55+ is less than 27 years old.

Solution:

Let X = a smart phone user whose age is 13 to 55+. $X \sim N(36.9, 13.9)$

a. To find the 30th percentile, find k such that $P(x < k) = 0.30$.
   invNorm(0.30, 36.9, 13.9) = 29.6 years
   Thirty percent of smartphone users 13 to 55+ are at most 29.6 years and 70% are at least 29.6 years.



b. Find $P(x < 27)$
   normalcdf(0,27,36.9,13.9) = 0.2342
   (Note that **normalcdf**(-10^{99},27,36.9,13.9) = 0.2382. The two answers differ only by 0.0040.)

TRY IT 6.11

Two thousand students took an exam. The scores on the exam have an approximate normal distribution with a mean $\mu = 81$ points and standard deviation $\sigma = 15$ points.

a. Calculate the first- and third-quartile scores for this exam.
b. The middle 50% of the exam scores are between what two values?

Solution:

a. Q_1 = 25th percentile = **invNorm**(0.25,81,15) = 70.9
   Q_3 = 75th percentile = **invNorm**(0.75,81,15) = 91.1
b. The middle 50% of the scores are between 70.9 and 91.1.

TRY IT 6.12

Using the information from Example 6.12, answer the following:

a. The middle 40% of mandarin oranges from this farm are between _______ and _______.
b. Find the 16th percentile and interpret it in a complete sentence.

Solution:

a. The middle area = 0.40, so each tail has an area of 0.30.

   1 – 0.40 = 0.60

   The tails of the graph of the normal distribution each have an area of 0.30.

   Find *k1*, the 30th percentile and *k2*, the 70th percentile (0.40 + 0.30 = 0.70).

   k1 = **invNorm**(0.30,5.85,0.24) = 5.72 cm

   k2 = **invNorm**(0.70,5.85,0.24) = 5.98 cm

b. **invNorm**(0.16, 5.85, 0.24) = 5.61; 16% of mandarin oranges from this farm have diameter 5.61 cm or less.

Chapter 7. The Central Limit Theorem

Section 1. The Central Limit Theorem for Sample Means (Averages)

TRY IT 7.1

An unknown distribution has a mean of 45 and a standard deviation of eight. Samples of size *n* = 30 are drawn randomly from the population. Find the probability that the sample mean is between 42 and 50.

Solution:
$P(42 < \bar{x} < 50) = (42,50,45,\frac{8}{\sqrt{30}}) = 0.9797$

TRY IT 7.2

The length of time taken on the SAT for a group of students is normally distributed with a mean of 2.5 hours and a standard deviation of 0.25 hours. A sample size of *n* = 60 is drawn randomly from the population. Find the probability that the sample mean is between two hours and three hours.

Solution:
$P(2 < \bar{x} < 3) = (2,3,2.5,\frac{0.25}{\sqrt{60}}) = 1$

TRY IT 7.3

In an article on Flurry Blog, a gaming marketing gap for men between the ages of 30 and 40 is identified. You are researching a startup game targeted at the 35-year-old demographic. Your idea is to develop a strategy game that can be played by men from their late 20s through their late 30s. Based on the article's data, industry research shows that the average strategy player is 28 years old with a standard deviation of 4.8 years. You take a sample of 100 randomly selected gamers. If your target market is 29- to 35-year-olds, should you continue with your development strategy?

Solution:

You need to determine the probability for men whose mean age is between 29 and 35 years of age wanting to play a strategy game.

$P(29 < \bar{x} < 35)$ = **normalcdf** $(29,35,28,\frac{4.8}{\sqrt{100}})$ = 0.0186

You can conclude there is approximately a 1.9% chance that your game will be played by men whose mean age is between 29 and 35.

TRY IT 7.4

Cans of a cola beverage claim to contain 16 ounces. The amounts in a sample are measured and the statistics are n = 34, $\bar{x}$ = 16.01 ounces. If the cans are filled so that μ = 16.00 ounces (as labeled) and σ = 0.143 ounces, find the probability that a sample of 34 cans will have an average amount greater than 16.01 ounces. Do the results suggest that cans are filled with an amount greater than 16 ounces?

Solution:

We have $P(\bar{x} > 16.01)$ = normalcdf$(16.01,E99,16,\frac{0.143}{\sqrt{34}})$ = 0.3417. Since there is a 34.17% probability that the average sample weight is greater than 16.01 ounces, we should be skeptical of the company's claimed volume. If I am a consumer, I should be glad that I am probably receiving free cola. If I am the manufacturer, I need to determine if my bottling processes are outside of acceptable limits.

Section 2. The Central Limit Theorem for Sums

TRY IT 7.5

An unknown distribution has a mean of 45 and a standard deviation of eight. A sample size of 50 is drawn randomly from the population. Find the probability that the sum of the 50 values is more than 2,400.

Solution:

0.0040

TRY IT 7.6

In a recent study reported Oct.29, 2012 on the Flurry Blog, the mean age of tablet users is 35 years. Suppose the standard deviation is ten years. The sample size is 39.

a. What are the mean and standard deviation for the sum of the ages of tablet users? What is the distribution?
b. Find the probability that the sum of the ages is between 1,400 and 1,500 years.
c. Find the 90th percentile for the sum of the 39 ages.

Solution:

a. $\mu_{\Sigma x} = n\mu_x$ = 1,365 and $\sigma_{\Sigma x} = \sqrt{n}\sigma x$ = 62.4
The distribution is normal for sums by the central limit theorem.
b. $P(1400 < \Sigma_x < 1500)$ = **normalcdf** (1400,1500,(39)(35),($\sqrt{39}$)(10)) = 0.2723
c. Let k = the 90th percentile.
k = **invNorm**(0.90,(39)(35),($\sqrt{39}$) (10)) = 1445.0

TRY IT 7.7

The mean number of minutes for app engagement by a table use is 8.2 minutes. Suppose the standard deviation is one minute. Take a sample size of 70.

a. What is the probability that the sum of the sample is between seven hours and ten hours? What does this mean in context of the problem?
b. Find the 84th and 16th percentiles for the sum of the sample. Interpret these values in context.

Solution

a. $\mu_{\Sigma x} = n\mu_x$ = 70(8.2) = 574 minutes and $\sigma_{\Sigma x}$ = ($\sqrt{n}$)(σx) = ($\sqrt{70}$)(1) = 8.37 minutes
b. Let k = the 95th percentile.
k = **invNorm** (0.95,(70)(8.2),($\sqrt{70}$)(1)) = 587.76 minutes
Ninety five percent of the sums of app engagement times are at most 587.76 minutes.
c. ten hours = 600 minutes
$P(\Sigma x \geq 600)$ = **normalcdf**(600,E99,(70)(8.2),($\sqrt{70}$)(70)(1)) = 0.0009

Section 3. Using the Central Limit Theorem

TRY IT 7.8

Use the information in Example 7.8, but use a sample size of 55 to answer the following questions.

a. Find $P(\bar{x} < 7)$.
b. Find $P(\Sigma x > 170)$.
c. Find the 80th percentile for the mean of 55 scores.
d. Find the 85th percentile for the sum of 55 scores.

Solution:

a. 0.0265
b. 0.2789
c. 3.13
d. 173.84

TRY IT 7.9

Use the information in Example 7.9, but change the sample size to 144.

a. Find $P(20 < \bar{x} < 30)$.
b. Find $P(\Sigma x$ is at least 3,000).
c. Find the 75th percentile for the sample mean excess time of 144 customers.
d. Find the 85th percentile for the sum of 144 excess times used by customers.

Solution:

a. 0.8623
b. 0.7377
c. 23.2
d. 3,441.6

TRY IT 7.10

Based on data from the National Health Survey, women between the ages of 18 and 24 have an average systolic blood pressures (in mm Hg) of 114.8 with a standard deviation of 13.1. Systolic blood pressure for women between the ages of 18 to 24 follow a normal distribution.

a. If one woman from this population is randomly selected, find the probability that her systolic blood pressure is greater than 120.
b. If 40 women from this population are randomly selected, find the probability that their mean systolic blood pressure is greater than 120.
c. If the sample were four women between the ages of 18 to 24 and we did not know the original distribution, could the central limit theorem be used?

Solution:

a. $P(x > 120)$ = normalcdf(120,1E99,114.8,13.1) = 0.3457. There is about a 35% chance, that the randomly selected woman will have systolics blood pressure greater than 120.
b. $P(\bar{x} > 120)$ = normalcdf(120, 1E99,114.8,$\frac{13.1}{\sqrt{40}}$)(120, 1E99,114.8,13.140) = 0.006. There is only a 0.6% chance that the average systolic blood pressure for the randomly selected group is greater than 120.
c. The central limit theorem could not be used if the sample size were four and we did not know the original distribution was normal. The sample size would be too small.

TRY IT 7.11

According to Boeing data, the 757 airliner carries 200 passengers and has doors with a height of 72 inches. Assume for a certain population of men we have a mean height of 69.0 inches and a standard deviation of 2.8 inches.

a. What doorway height would allow 95% of men to enter the aircraft without bending?
b. Assume that half of the 200 passengers are men. What mean doorway height satisfies the condition that there is a 0.95 probability that this height is greater than the mean height of 100 men?
c. For engineers designing the 757, which result is more relevant: the height from part a or part b? Why?

Solution:

a. We know that $\mu_x = \mu = 69$ and we have $\sigma_x = 2.8$. The height of the doorway is found to be **invNorm**(0.95,69,2.8) = 73.61
b. We know that $\mu_x = \mu = 69$ and we have $\sigma_x = 0.28$. So, **invNorm**(0.95,69,0.28) = 69.49
c. When designing the doorway heights, we need to incorporate as much variability as possible in order to accommodate as many passengers as possible. Therefore, we need to use the result based on part a.

TRY IT 7.12

In a city, 46 percent of the population favor the incumbent, Dawn Morgan, for mayor. A simple random sample of 500 is taken. Using the continuity correction factor, find the probability that at least 250 favor Dawn Morgan for mayor.

Solutions:

0.0401

Chapter 8. Confidence Intervals

Section 1. A Single Population Mean using the Normal Distribution

TRY IT 8.1

Suppose we have data from a sample. The sample mean is 15, and the error bound for the mean is 3.2.

What is the confidence interval estimate for the population mean?

Solution:

(11.8, 18.2)

TRY IT 8.2

Suppose average pizza delivery times are normally distributed with an unknown population mean and a population standard deviation of six minutes. A random sample of 28 pizza delivery restaurants is taken and has a sample mean delivery time of 36 minutes.

Find a 90% confidence interval estimate for the population mean delivery time.

Solution:

(34.1347, 37.8653)

TRY IT 8.3

Table 8.2 shows a different random sampling of 20 cell phone models. Use this data to calculate a 93% confidence interval for the true mean SAR for cell phones certified for use in the United States. As previously, assume that the population standard deviation is $\sigma = 0.337$.

Phone Model	SAR	Phone Model	SAR
Blackberry Pearl 8120	1.48	Nokia E71x	1.53
HTC Evo Design 4G	0.8	Nokia N75	0.68
HTC Freestyle	1.15	Nokia N79	1.4
LG Ally	1.36	Sagem Puma	1.24
LG Fathom	0.77	Samsung Fascinate	0.57
LG Optimus Vu	0.462	Samsung Infuse 4G	0.2
Motorola Cliq XT	1.36	Samsung Nexus S	0.51
Motorola Droid Pro	1.39	Samsung Replenish	0.3
Motorola Droid Razr M	1.3	Sony W518a Walkman	0.73

Phone Model	SAR	Phone Model	SAR
Nokia 7705 Twist	0.7	ZTE C79	0.869

Table8.2

Solution:

x̄ =0.940 x̄ =0.940

$\frac{\alpha}{2} = \frac{1-CL}{2} = \frac{1-0.93}{2} = 0.035$

$Z_{0.035}$ = 1.812

$EBM = (Z_{0.035})\left(\frac{\sigma}{\sqrt{n}}\right) = (1.812)\left(\frac{0.337}{\sqrt{20}}\right) = 0.1365$

x̄ – *EBM* = 0.940 – 0.1365 = 0.8035

x̄ + *EBM* = 0.940 + 0.1365 = 1.0765

We estimate with 93% confidence that the true SAR mean for the population of cell phones in the United States is between 0.8035 and 1.0765 watts per kilogram.

TRY IT 8.4

Refer back to the pizza-delivery Try It exercise. The population standard deviation is six minutes and the sample mean deliver time is 36 minutes. Use a sample size of 20. Find a 95% confidence interval estimate for the true mean pizza delivery time.

Solution:
(33.37, 38.63)

TRY IT 8.5

Refer back to the pizza-delivery Try It exercise. The mean delivery time is 36 minutes and the population standard deviation is six minutes. Assume the sample size is changed to 50 restaurants with the same sample mean. Find a 90% confidence interval estimate for the population mean delivery time.

Solution:

(34.6041, 37.3958)

TRY IT 8.6

Suppose we know that a confidence interval is (42.12, 47.88). Find the error bound and the sample mean.

Solution:

Sample mean is 45, error bound is 2.88

TRY IT 8.7

The population standard deviation for the height of high school basketball players is three inches. If we want to be 95% confident that the sample mean height is within one inch of the true population mean height, how many randomly selected students must be surveyed?

Solution:

35 Students

Section 2. Single Population Mean using the Student t Distribution

TRY IT 8.8

You do a study of hypnotherapy to determine how effective it is in increasing the number of hours of sleep subjects get each night. You measure hours of sleep for 12 subjects with the following results. Construct a 95% confidence interval for the mean number of hours slept for the population (assumed normal) from which you took the data.

8.2; 9.1; 7.7; 8.6; 6.9; 11.2; 10.1; 9.9; 8.9; 9.2; 7.5; 10.5

Solution:

(8.1634, 9.8032)

TRY IT 8.9

A random sample of statistics students were asked to estimate the total number of hours they spend watching television in an average week. The responses are recorded in Table 8.4. Use this sample data to construct a 98% confidence interval for the mean number of hours statistics students will spend watching television in one week.

0	3	1	20	9
5	10	1	10	4

14	2	4	4	5

Table8.4

Solution:

Solution A

x̄ = 6.133, *s* = 5.514, *n* = 15, and *df* = 15 – 1 = 14

CL = 0.98, so *α* = 1 - *CL* = 1 - 0.98 = 0.02

α/2=0.01tα/2=t0.01=2.624

EBM=tα/2($\frac{s}{\sqrt{n}}$)=2.624($\frac{5.514}{\sqrt{15}}$)~3.736

x̄ – *EBM* = 6.133 – 3.736 = 2.397

x̄ + *EBM* = 6.133 + 3.736 = 9.869

We estimate with 98% confidence that the mean number of all hours that statistics students spend watching television in one week is between 2.397 and 9.869.

Solution B

Enter the data as a list.
Press STAT and arrow over to TESTS.
Arrow down to 8:TInterval.
Press ENTER.
Arrow to Data and press ENTER.
Arrow down and enter the name of the list where the data is stored.
Enter Freq: 1
Enter C-Level: 0.98
Arrow down to Calculate and press Enter.
The 98% confidence interval is (2.3965, 9.8702).

Section 3. A Population Proportion

TRY IT 8.10

Suppose 250 randomly selected people are surveyed to determine if they own a tablet. Of the 250 surveyed, 98 reported owning a tablet. Using a 95% confidence level, compute a confidence interval estimate for the true proportion of people who own tablets.

Solution:

(0.3315,0.4525)

TRY IT 8.11

A student polls his school to see if students in the school district are for or against the new legislation regarding school uniforms. She surveys 600 students and finds that 480 are against the new legislation.

a. Compute a 90% confidence interval for the true percent of students who are against the new legislation, and interpret the confidence interval.

b. In a sample of 300 students, 68% said they own an iPod and a smart phone. Compute a 97% confidence interval for the true percent of students who own an iPod and a smartphone.

Solution:

a. (0.7731, 0.8269); We estimate with 90% confidence that the true percent of all students in the district who are against the new legislation is between 77.31% and 82.69%.

b.

Solution A;

Sixty-eight percent (68%) of students own an iPod and a smart phone.

p′=0.68

q′=1–p′=1–0.68=0.32

Since *CL* = 0.97, we know *α* = 1 – 0.97 = 0.03 and α/2 = 0.015.

The area to the left of $z_{0.015}$ is 0.015, and the area to the right of $z_{0.015}$ is 1 – 0.015 = 0.985.

Using the **TI 83**, **83+**, or **84+** calculator function **InvNorm**(.985,0,1),

$z_{0.015}$=2.17

$$\text{EPB} = (z_{\alpha/2})\sqrt{\frac{p`q`}{n}} = 2.17\sqrt{\frac{0.68(0.32)}{300}} \approx 0.0584$$

p′ – *EPB* = 0.68 – 0.0584 = 0.6216

p′ + *EPB* = 0.68 + 0.0584 = 0.7384

We are 97% confident that the true proportion of all students who own an iPod and a smart phone is between 0.6216 and 0.7384.

Solution B:

Press **STAT** and arrow over to **TESTS**.
Arrow down to **A:1-PropZint**. Press **ENTER**.
Arrow down to x and enter 300*0.68.
Arrow down to n and enter 300.
Arrow down to **C-Level** and enter 0.97.
Arrow down to **Calculate** and press **ENTER**.
The confidence interval is (0.6216, 0.7384).

TRY IT 8.12

Out of a random sample of 65 freshmen at State University, 31 students have declared a major. Use the "plus-four" method to find a 96% confidence interval for the true proportion of freshmen at State University who have declared a major.

Solution:

Solution A

Using "plus four," we have x = 31 + 2 = 33 and n = 65 + 4 = 69.

$p'=33/69\approx0.478$

$q'=1–p'=1–0.478=0.522$

Since CL = 0.96, we know α = 1 – 0.96 = 0.04 and $\alpha/2$ = 0.02.

$z_{0.02}$=2.054z0.02=2.054

$$EPB = (z_{\alpha/2})\sqrt{\frac{p`q`}{n}} = 2.054\sqrt{\frac{(0.478)(0.522)}{69}} \approx 0.124$$

$p' - EPB$ = 0.478 – 0.124 = 0.354

$p' + EPB$ = 0.478 + 0.124 = 0.602

We are 96% confident that between 35.4% and 60.2% of all freshmen at State U have declared a major.

Solution B

Press **STAT** and arrow over to **TESTS**.
Arrow down to **A:1-PropZint**. Press **ENTER**.

Arrow down to *x* and enter 33.
Arrow down to *n* and enter 69.
Arrow down to **C-Level** and enter 0.96.
Arrow down to **Calculate** and press **ENTER**.
The confidence interval is (0.355, 0.602).

TRY IT 8.13

The Berkman Center Study referenced in Example 8.13 talked to teens in smaller focus groups, but also interviewed additional teens over the phone. When the study was complete, 588 teens had answered the question about their Facebook friends with 159 saying that they have more than 500 friends. Use the "plus-four" method to find a 90% confidence interval for the true proportion of teens that would report having more than 500 Facebook friends based on this larger sample. Compare the results to those in Example 8.13.

Solution:

Solution A

Using "plus four," we have *x* = 159 + 2 = 161 and *n* = 588 + 4 = 592.

p′=161/592 ≈ 0.272

q′=1–p′=1–0.272=0.728

Since *CL* = 0.90, we know *α* = 1 – 0.90 = 0.10 and α/2 = 0.05.

$$\text{EPB} = (z_{\alpha/2})\sqrt{\frac{p`q`}{n}} = 1.645\sqrt{\frac{(0.272)(0.728)}{592}} \approx 0.030$$

p′ – *EPB* = 0.272 – 0.030 = 0.242

p′ + *EPB* = 0.272 + 0.034 = 0.302

We are 96% confident that between 24.2% and 30.2% of all teens would report having more than 500 friends on Facebook.

TRY IT 8.14

Suppose an internet marketing company wants to determine the current percentage of customers who click on ads on their smartphones. How many customers should the company survey in order to be 90% confident that the estimated proportion is within five percentage points of the true population proportion of customers who click on ads on their smartphones?

Solution:

271 customers should be surveyed.

Chapter 9. Hypothesis Testing with One Sample

Section 1. Null and Alternative Hypotheses

TRY IT 9.1

A medical trial is conducted to test whether or not a new medicine reduces cholesterol by 25%. State the null and alternative hypotheses.

Solution:

H_0 : The drug reduces cholesterol by 25%. $p = 0.25$

H_a : The drug does not reduce cholesterol by 25%. $p \neq 0.25$

TRY IT 9.2

We want to test whether the mean height of eighth graders is 66 inches. State the null and alternative hypotheses. Fill in the correct symbol (=, ≠, ≥, <, ≤, >) for the null and alternative hypotheses.

a. H_0: μ __ 66
b. H_a: μ __ 66

Solution:

a. $H_0 : \mu = 66$
b. $H_a : \mu \neq 66$

TRY IT 9.3

We want to test if it takes fewer than 45 minutes to teach a lesson plan. State the null and alternative hypotheses. Fill in the correct symbol (=, ≠, ≥, <, ≤, >) for the null and alternative hypotheses.

a. H_0: μ __ 45
b. H_a: μ __ 45

Solution:

a. $H_0: \mu \geq 45$
b. $H_a: \mu < 45$

TRY IT 9.4

On a state driver's test, about 40% pass the test on the first try. We want to test if more than 40% pass on the first try. Fill in the correct symbol (=, ≠, ≥, <, ≤, >) for the null and alternative hypotheses.

a. H_0: p __ 0.40
b. H_a: p __ 0.40

Solution:

a. H_0: $p = 0.40$
b. H_a: $p > 0.40$

TRY IT 9.5

Suppose the null hypothesis, H_0, is: the blood cultures contain no traces of pathogen *X*. State the Type I and Type II errors.

Solution:

Type I error: The researcher thinks the blood cultures do contain traces of pathogen *X*, when in fact, they do not.

Type II error: The researcher thinks the blood cultures do not contain traces of pathogen *X*, when in fact, they do.

TRY IT 9.6

Suppose the null hypothesis, H_0, is: a patient is not sick. Which type of error has the greater consequence, Type I or Type II?

Solution:

The error with the greater consequence is the Type II error: the patient will be thought well when, in fact, he is sick, so he will not get treatment.

TRY IT 9.7

"Red tide" is a bloom of poison-producing algae–a few different species of a class of plankton called dinoflagellates. When the weather and water conditions cause these blooms, shellfish such as clams living in the area develop dangerous levels of a paralysis-inducing toxin. In Massachusetts, the Division of Marine Fisheries (DMF) monitors levels of the toxin in shellfish by regular sampling of shellfish along the coastline. If the mean level of toxin in clams exceeds 800 μg (micrograms) of toxin per kg of clam meat in any area, clam harvesting is banned there until the bloom is over and levels of toxin in clams subside. Describe both a Type I and a Type II error in this context, and state which error has the greater consequence.

Solution:

In this scenario, an appropriate null hypothesis would beH_0: the mean level of toxins is at most 800 μg, H_0 : $\mu_0 \leq 800$ μg.

Type I error: The DMF believes that toxin levels are still too high when, in fact, toxin levels are at most 800 μg. The DMF continues the harvesting ban.

Type II error: The DMF believes that toxin levels are within acceptable levels (are at least 800 μg) when, in fact, toxin levels are still too high (more than 800 μg). The DMF lifts the harvesting ban. This error could be the most serious. If the ban is lifted and clams are still toxic, consumers could possibly eat tainted food.

In summary, the more dangerous error would be to commit a Type II error, because this error involves the availability of tainted clams for consumption.

TRY IT 9.8

Determine both Type I and Type II errors for the following scenario:

Assume a null hypothesis, H_0, that states the percentage of adults with jobs is at least 88%.

Identify the Type I and Type II errors from these four statements.

a. Not to reject the null hypothesis that the percentage of adults who have jobs is at least 88% when that percentage is actually less than 88%
b. Not to reject the null hypothesis that the percentage of adults who have jobs is at least 88% when the percentage is actually at least 88%.
c. Reject the null hypothesis that the percentage of adults who have jobs is at least 88% when the percentage is actually at least 88%.
d. Reject the null hypothesis that the percentage of adults who have jobs is at least 88% when that percentage is actually less than 88%.

Solution:

Type I error: c

Type I error: b

Section 4. Rare Events, the Sample, Decision and Conclusion

TRY IT 9.9

A normal distribution has a standard deviation of 1. We want to verify a claim that the mean is greater than 12. A sample of 36 is taken with a sample mean of 12.5.

H_0: $\mu \leq 12$
H_a: $\mu > 12$

The p-value is 0.0013
Draw a graph that shows the p-value.

Solution:

p-value = 0.0013

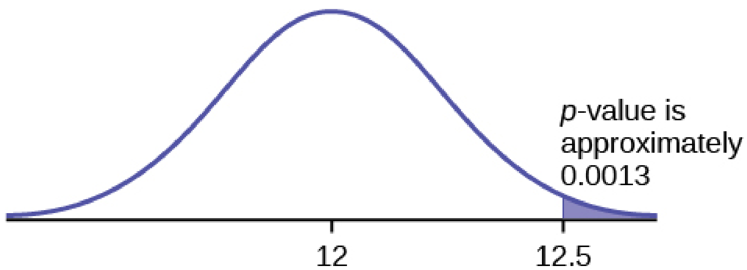


Figure 2.

TRY IT 9.10

It's a Boy Genetics Labs claim their procedures improve the chances of a boy being born. The results for a test of a single population proportion are as follows:

H_0: p = 0.50, H_a: p > 0.50

α = 0.01

p-value = 0.025

Interpret the results and state a conclusion in simple, non-technical terms.

Solution:

Since the p-value is greater than the established alpha value (the p-value is high), we do not reject the null hypothesis. There is not enough evidence to support It's a Boy Genetics Labs' stated claim that their procedures improve the chances of a boy being born.

Section 5. Additional Information and Full Hypothesis Test Examples

TRY IT 9.11

H_0: μ = 10, H_a: μ < 10

Assume the p-value is 0.0935. What type of test is this? Draw the picture of the p-value.

Solution:

left-tailed test

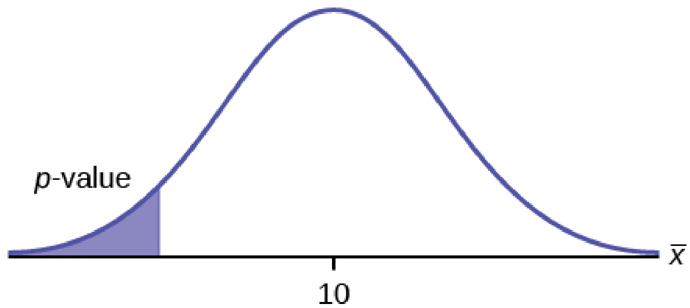


Figure 2.

TRY IT 9.12

H_0: $\mu \leq 1$, H_a: $\mu > 1$

Assume the *p*-value is 0.1243. What type of test is this? Draw the picture of the *p*-value.

Solution:

right-tailed test

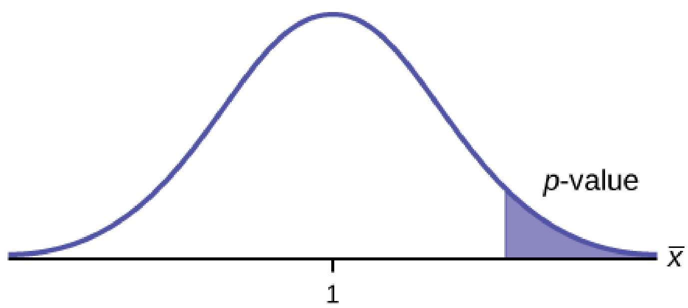


Figure 4.

TRY IT 9.13

H_0: $p = 0.5$, H_a: $p \neq 0.5$

Assume the *p*-value is 0.2564. What type of test is this? Draw the picture of the *p*-value.

Solution:

two-tailed test

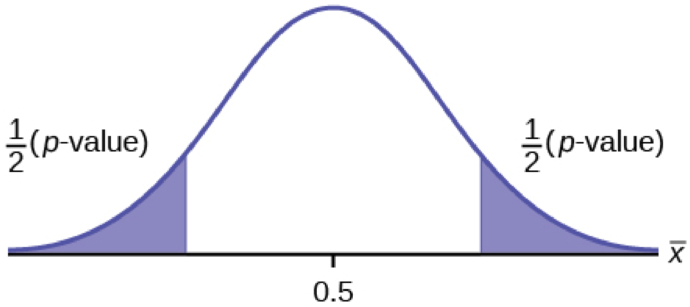


Figure 6.

TRY IT 9.14

The mean throwing distance of a football for Marco, a high school freshman quarterback, is 40 yards, with a standard deviation of two yards. The team coach tells Marco to adjust his grip to get more distance. The coach records the distances for 20 throws. For the 20 throws, Marco's mean distance was 45 yards. The coach thought the different grip helped Marco throw farther than 40 yards. Conduct a hypothesis test using a preset $\alpha = 0.05$. Assume the throw distances for footballs are normal.

First, determine what type of test this is, set up the hypothesis test, find the *p*-value, sketch the graph, and state your conclusion.

Solution:

Since the problem is about a mean, this is a test of a single population mean.

$H_0 : \mu = 40$

$H_a : \mu > 40$

$p = 2.6115 \times 10^{-29}$

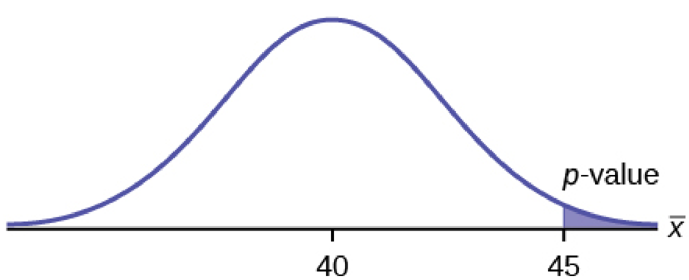


Figure 6.

Because $p < \alpha$, we reject the null hypothesis. There is sufficient evidence to suggest that the change in grip improved Marco's throwing distance.

TRY IT 9.16

It is believed that a stock price for a particular company will grow at a rate of \$5 per week with a standard deviation of \$1. An investor believes the stock won't grow as quickly. The changes in stock price is recorded for ten weeks and are as follows: \$4, \$3, \$2, \$3, \$1, \$7, \$2, \$1, \$1, \$2. Perform a hypothesis test using a 5% level of significance. State the null and alternative hypotheses, find the *p*-value, state your conclusion, and identify the Type I and Type II errors.

Solution:

H_0: $\mu = 5$

H_a: $\mu < 5$

$p = 0.0082$

Because $p < \alpha$, we reject the null hypothesis. There is sufficient evidence to suggest that the stock price of the company grows at a rate less than \$5 a week.

Type I Error: To conclude that the stock price is growing slower than \$5 a week when, in fact, the stock price is growing at \$5 a week (reject the null hypothesis when the null hypothesis is true).

Type II Error: To conclude that the stock price is growing at a rate of \$5 a week when, in fact, the stock price is growing slower than \$5 a week (do not reject the null hypothesis when the null hypothesis is false).

TRY IT 9.17

A teacher believes that 85% of students in the class will want to go on a field trip to the local zoo. She performs a hypothesis test to determine if the percentage is the same or different from 85%. The teacher samples 50 students and 39 reply that they would want to go to the zoo. For the hypothesis test, use a 1% level of significance.

First, determine what type of test this is, set up the hypothesis test, find the *p*-value, sketch the graph, and state your conclusion.

Solution:

Since the problem is about percentages, this is a test of single population proportions.

H_0 : $p = 0.85$

H_a: $p \neq 0.85$

$p = 0.7554$

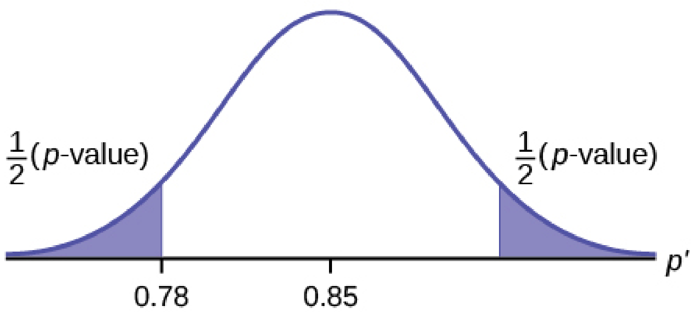


Figure 8.

Because $p > \alpha$, we fail to reject the null hypothesis. There is not sufficient evidence to suggest that the proportion of students that want to go to the zoo is not 85%.

TRY IT 9.18

Marketers believe that 92% of adults in the United States own a cell phone. A cell phone manufacturer believes that number is actually lower. 200 American adults are surveyed, of which, 174 report having cell phones. Use a 5% level of significance. State the null and alternative hypothesis, find the *p*-value, state your conclusion, and identify the Type I and Type II errors.

Solution:

H_0: $p = 0.92$

H_a: $p < 0.92$

p-value = 0.0046

Because $p < 0.05$, we reject the null hypothesis. There is sufficient evidence to conclude that fewer than 92% of American adults own cell phones.

Type I Error: To conclude that fewer than 92% of American adults own cell phones when, in fact, 92% of American adults do own cell phones (reject the null hypothesis when the null hypothesis is true).

Type II Error: To conclude that 92% of American adults own cell phones when, in fact, fewer than 92% of American adults own cell phones (do not reject the null hypothesis when the null hypothesis is false).

Chapter 10. Hypothesis Testing with Two Samples

Section 1. Two Population Means with Unknown Standard Deviations

TRY IT 10.1

Two samples are shown in Table 10.2. Both have normal distributions. The means for the two populations are thought to be the same. Is there a difference in the means? Test at the 5% level of significance.

	Sample Size	Sample Mean	Sample Standard Deviation
Population A	25	5	1
Population B	16	4.7	1.2

Table10.2

Solution:

The *p*-value is 0.4125, which is much higher than 0.05, so we decline to reject the null hypothesis. There is not sufficient evidence to conclude that the means of the two populations are not the same.

TRY IT 10.2

A study is done to determine if Company A retains its workers longer than Company B. Company A samples 15 workers, and their average time with the company is five years with a standard deviation of 1.2. Company B samples 20 workers, and their average time with the company is 4.5 years with a standard deviation of 0.8. The populations are normally distributed.

a. Are the population standard deviations known?
b. Conduct an appropriate hypothesis test. At the 5% significance level, what is your conclusion?

Solution:

a. They are unknown.
b. The *p*-value = 0.0878. At the 5% level of significance, there is insufficient evidence to conclude that the workers of Company A stay longer with the company.

TRY IT 10.5

Weighted alpha is a measure of risk-adjusted performance of stocks over a period of a year. A high positive weighted alpha signifies a stock whose price has risen while a small positive weighted alpha indicates an unchanged stock price during the time period. Weighted alpha is used to identify companies with strong upward or downward trends. The weighted alpha for the top 30 stocks of banks in the northeast and in the west as identified by Nasdaq on May 24, 2013 are listed in Table 10.6 and Table 10.7, respectively.

94.2	75.2	69.6	52.0	48.0	41.9	36.4	33.4	31.5	27.6
77.3	71.9	67.5	50.6	46.2	38.4	35.2	33.0	28.7	26.5
76.3	71.7	56.3	48.7	43.2	37.6	33.7	31.8	28.5	26.0

Table10.6 Northeast

126.0	70.6	65.2	51.4	45.5	37.0	33.0	29.6	23.7	22.6
116.1	70.6	58.2	51.2	43.2	36.0	31.4	28.7	23.5	21.6
78.2	68.2	55.6	50.3	39.0	34.1	31.0	25.3	23.4	21.5

Table10.7 West

Is there a difference in the weighted alpha of the top 30 stocks of banks in the northeast and in the west? Test at a 5% significance level. Answer the following questions:

a. Is this a test of two means or two proportions?
b. Are the population standard deviations known or unknown?
c. Which distribution do you use to perform the test?
d. What is the random variable?
e. What are the null and alternative hypotheses? Write the null and alternative hypotheses in words and in symbols.
f. Is this test right, left, or two tailed?
g. What is the *p*-value?
h. Do you reject or not reject the null hypothesis?
i. At the ___ level of significance, from the sample data, there ______ (is/is not) sufficient evidence to conclude that ______.
j. Calculate Cohen's *d* and interpret it.

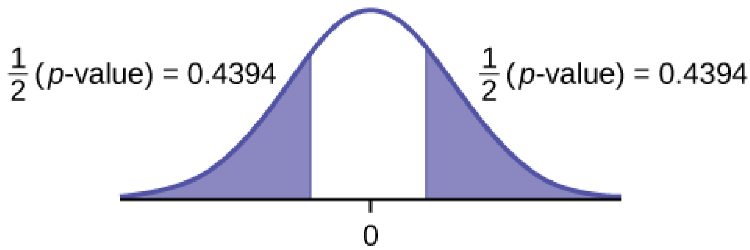
Solution:

a. two means
b. unknown
c. Student's-t
d. $\bar{X}1 - \bar{X}2$
e.
    1. $H_0 : \mu_1 = \mu_2$ Null hypothesis: the means of the weighted alphas are equal.
    2. $H_a : \mu_1 \neq \mu_2$ Alternative hypothesis : the means of the weighted alphas are not equal.
f. two-tailed

g. *p*-value = 0.8787

h. Do not reject the null hypothesis

i. This indicates that the trends in stocks are about the same in the top 30 banks in each



region. 5% level of significance, from the sample data, there is not sufficient evidence to conclude that the mean weighted alphas for the banks in the northeast and the west are different

j. *d* = 0.040, Very small, because 0.040 is less than Cohen's value of 0.2 for small effect size. The size of the difference of the means of the weighted alphas for the two regions of banks is small indicating that there is not a significant difference between their trends in stocks.

Section 2. Two Population Means with Known Standard Deviations

TRY IT 10.6

The means of the number of revolutions per minute of two competing engines are to be compared. Thirty engines are randomly assigned to be tested. Both populations have normal distributions. Table 10.9 shows the result. Do the data indicate that Engine 2 has higher RPM than Engine 1? Test at a 5% level of significance.

Engine	Sample Mean Number of RPM	Population Standard Deviation
1	1,500	50
2	1,600	60

Table10.9

Solution:

The *p*-value is almost zero, so we reject the null hypothesis. There is sufficient evidence to conclude that Engine 2 runs at a higher RPM than Engine 1.

Section 3. Comparing Two Independent Population Proportions

TRY IT 10.8

Two types of valves are being tested to determine if there is a difference in pressure tolerances. Fifteen out of a random sample of 100 of Valve *A* cracked under 4,500 psi. Six out of a random sample of 100 of Valve *B* cracked under 4,500 psi. Test at a 5% level of significance.

Solution:
The *p*-value is 0.0379, so we can reject the null hypothesis. At the 5% significance level, the data support that there is a difference in the pressure tolerances between the two valves.

TRY IT 10.10

A concerned group of citizens wanted to know if the proportion of forcible rapes in Texas was different in 2011 than in 2010. Their research showed that of the 113,231 violent crimes in Texas in 2010, 7,622 of them were forcible rapes. In 2011, 7,439 of the 104,873 violent crimes were in the forcible rape category. Test at a 5% significance level. Answer the following questions:

a. Is this a test of two means or two proportions?

b. Which distribution do you use to perform the test?

c. What is the random variable?

d. What are the null and alternative hypothesis? Write the null and alternative hypothesis in symbols.

e. Is this test right-, left-, or two-tailed?

f. What is the *p*-value?

g. Do you reject or not reject the null hypothesis?

h. At the ___ level of significance, from the sample data, there ______ (is/is not) sufficient evidence to conclude that ____________.

Solution:

a. two proportions
b. normal for two proportions
c. Subscripts: 1 = 2010, 2 = 2011
P'_2 - P'_2
d. Subscripts: 1 = 2010, 2 = 2011
H_0: $p_1 = p_2$ H_0: $p_1 - p_2 = 0$
H_a: $p_1 \neq p_2$ H_a: $p_1 - p_2 \neq 0$
e. two-tailed
f. *p*-value = 0.00086

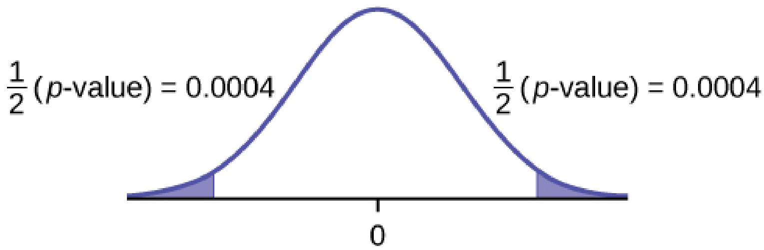


Figure 1.

g. Reject the H_0.

h. At the 5% significance level, from the sample data, there is sufficient evidence to conclude that there is a difference between the proportion of forcible rapes in 2011 and 2010.

Section 4. Matched or Paired Samples

TRY IT 10.11

A study was conducted to investigate how effective a new diet was in lowering cholesterol. Results for the randomly selected subjects are shown in the table. The differences have a normal distribution. Are the subjects' cholesterol levels lower on average after the diet? Test at the 5% level.

Subject	A	B	C	D	E	F	G	H	I
Before	209	210	205	198	216	217	238	240	222
After	199	207	189	209	217	202	211	223	201

Table10.13

Solution:

The p-value is 0.0130, so we can reject the null hypothesis. There is enough evidence to suggest that the diet lowers cholesterol.

TRY IT 10.12

A new prep class was designed to improve SAT test scores. Five students were selected at random. Their scores on two practice exams were recorded, one before the class and one after. The data recorded in Table 10.15. Are the scores, on average, higher after the class? Test at a 5% level.

SAT Scores	Student 1	Student 2	Student 3	Student 4
Score before class	1840	1960	1920	2150
Score after class	1920	2160	2200	2100

Table10.15

Solution:

The p-value is 0.0874, so we decline to reject the null hypothesis. The data do not support that the class improves SAT scores significantly.

TRY IT 10.13

Five ball players think they can throw the same distance with their dominant hand (throwing) and off-hand (catching hand). The data were collected and recorded in Table 10.17. Conduct a hypothesis test to determine whether the mean difference in distances between the dominant and off-hand is significant. Test at the 5% level.

	Player 1	Player 2	Player 3	Player 4	Player 5
Dominant Hand	120	111	135	140	125
Off-hand	105	109	98	111	99

Table10.17

Solution:

The p-level is 0.0230, so we can reject the null hypothesis. The data show that the players do not throw the same distance with their off-hands as they do with their dominant hands.

Chapter 11. The Chi-Square Distribution

Section 2. Goodness-of-Fit Test

TRY IT 11.1

A factory manager needs to understand how many products are defective versus how many are produced. The number of expected defects is listed in Table 11.5.

Number produced	**Number defective**
0–100	5
101–200	6
201–300	7
301–400	8
401–500	10

Table11.5

A random sample was taken to determine the actual number of defects. Table 11.6 shows the results of the survey.

Number produced	Number defective
0–100	5
101–200	7
201–300	8
301–400	9
401–500	11

Table11.6

State the null and alternative hypotheses needed to conduct a goodness-of-fit test, and state the degrees of freedom.

Solution:

H_0:The number of defaults fits expectations.

H_a:The number of defaults does not fit expectations.

df = 4

TRY IT 11.2

Teachers want to know which night each week their students are doing most of their homework. Most teachers think that students do homework equally throughout the week. Suppose a random sample of 56 students were asked on which night of the week they did the most homework. The results were distributed as in Table 11.8.

	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Number of Students	11	8	10	7	10	5	5

Table11.8

From the population of students, do the nights for the highest number of students doing the majority of their homework occur with equal frequencies during a week? What type of hypothesis test should you use?

Solution:

df = 6

p-value = 0.6093

We decline to reject the null hypothesis. There is not enough evidence to support that students do not do the majority of their homework equally throughout the week.

TRY IT 11.3

The expected percentage of the number of pets students have in their homes is distributed (this is the given distribution for the student population of the United States) as in Table 11.12.

Number of Pets	Percent
0	18
1	25
2	30
3	18
4+	9

Table11.12

A random sample of 1,000 students from the Eastern United States resulted in the data in Table 11.13.

Number of Pets	Frequency
0	210
1	240
2	320
3	140
4+	90

Table11.13

At the 1% significance level, does it appear that the distribution “number of pets” of students in the Eastern United States is different from the distribution for the United States student population as a whole? What is the *p*-value?

Solution:

p-value = 0.0036

We reject the null hypothesis that the distributions are the same. There is sufficient evidence to conclude that the distribution “number of pets” of students in the Eastern United States is different from the distribution for the United States student population as a whole.

TRY IT 11.4

Students in a social studies class hypothesize that the literacy rates across the world for every region are 82%. Table 11.14 shows the actual literacy rates across the world broken down by region. What are the test statistic and the degrees of freedom?

MDG Region	Adult Literacy Rate (%)
Developed Regions	99.0
Commonwealth of Independent States	99.5
Northern Africa	67.3
Sub-Saharan Africa	62.5
Latin America and the Caribbean	91.0
Eastern Asia	93.8
Southern Asia	61.9
South-Eastern Asia	91.9
Western Asia	84.5
Oceania	66.4

Table11.14

Solution:

degrees of freedom = 9

chi^2 test statistic = 26.38

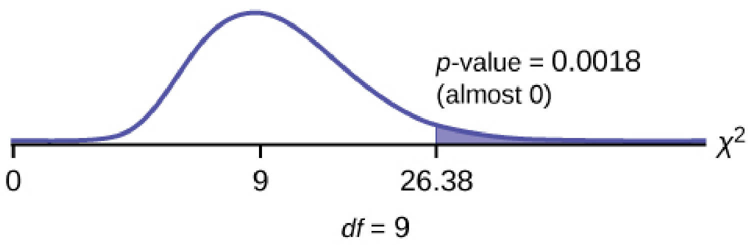


Figure 1.

Press STAT and ENTER. Make sure you clear lists L1, L2, and L3 if they have data in them. Into L1, put the observed frequencies 99, 99.5, 67.3, 62.5, 91, 93.8, 61.9, 91.9, 84.5, 66.4. Into L2, put the expected frequencies 82, 82, 82, 82, 82, 82, 82, 82, 82, 82. Arrow over to list L3 and up to the name area "L3". Enter (L1-L2)^2/L2and ENTER. Press 2nd QUIT. Press 2nd LIST and arrow over to MATH. Press 5. You should see "sum". Enter L3. Rounded to two decimal places, you should see 26.38. Press 2nd DISTR. Arrow down to 7:χ2cdf (or press 7). Press ENTER. Enter 26.38,1E99,9). Rounded to four places, you should see .0018, which is the *p*-value.

The newer TI-84 calculators have in STAT TESTS the test Chi2 GOF. To run the test, put the observed values (the data) into a first list and the expected values (the values you expect if the null hypothesis is true) into a second list. Press STAT TESTS and Chi2 GOF. Enter the list names for the Observed list and the Expected list. Enter the degrees of freedom and press calculate or draw. Make sure you clear any lists before you start.

Section 3. Test of Independence

TRY IT 11.5
A sample of 300 students is taken. Of the students surveyed, 50 were music students, while 250 were not. Ninety-seven were on the honor roll, while 203 were not. If we assume being a music student and being on the honor roll are independent events, what is the expected number of music students who are also on the honor roll?
Solution:
About 16 students are expected to be music students and on the honor roll.

TRY IT 11.6
The Bureau of Labor Statistics gathers data about employment in the United States. A sample is taken to calculate the number of U.S. citizens working in one of several industry sectors over time. Table 11.17 shows the results:

Industry Sector	2000	2010	2020	Total
Nonagriculture wage and salary	13,243	13,044	15,018	41,305
Goods-producing, excluding agriculture	2,457	1,771	1,950	6,178

Industry Sector	2000	2010	2020	Total
Services-providing	10,786	11,273	13,068	35,127
Agriculture, forestry, fishing, and hunting	240	214	201	655
Nonagriculture self-employed and unpaid family worker	931	894	972	2,797
Secondary wage and salary jobs in agriculture and private household industries	14	11	11	36
Secondary jobs as a self-employed or unpaid family worker	196	144	152	492
Total	27,867	27,351	31,372	86,590

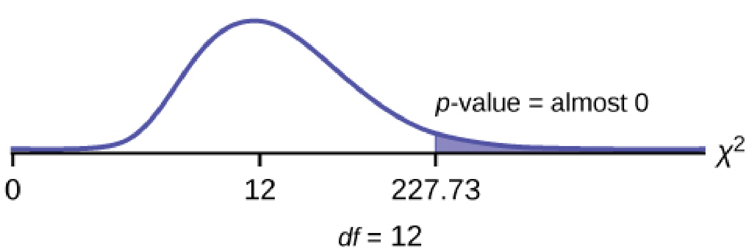
Table11.17

We want to know if the change in the number of jobs is independent of the change in years. State the null and alternative hypotheses and the degrees of freedom.

Solution:

H_0 : The number of jobs is independent of the year.

H_a : The number of jobs is dependent on the year.
df = 12



Press the MATRX key and arrow over to EDIT. Press 1:[A]. Press 3 ENTER 3 ENTER. Enter the table values by row. Press ENTER after each. Press 2nd QUIT. Press STATand arrow over to TESTS. Arrow down to C:χ2-TEST. Press ENTER. You should see Observed:[A] and

Expected:[B]. Arrow down to Calculate. Press ENTER. The test statistic is 227.73 and the *p*–value = 5.90E - 42 = 0. Do the procedure a second time but arrow down to Draw instead of calculate.

TRY IT 11.7

Refer back to the information in Try It. How many service providing jobs are there expected to be in 2020? How many nonagriculture wage and salary jobs are there expected to be in 2020?

Solution:
12,727, 14,965

TRY IT 11.8

Do families and singles have the same distribution of cars? Use a level of significance of 0.05. Suppose that 100 randomly selected families and 200 randomly selected singles were asked what type of car they drove: sport, sedan, hatchback, truck, van/SUV. The results are shown in Table 11.20. Do families and singles have the same distribution of cars? Test at a level of significance of 0.05.

	Sport	Sedan	Hatchback	Truck	Van/SUV
Family	5	15	35	17	28
Single	45	65	37	46	7

Table11.20

Solution:
With a *p*-value of almost zero, we reject the null hypothesis. The data show that the distribution of cars is not the same for families and singles.

TRY IT 11.9

Ivy League schools receive many applications, but only some can be accepted. At the schools listed in Table 11.22, two types of applications are accepted: regular and early decision.

Application Type Accepted	Brown	Columbia	Cornell	Dartmouth	Penn	Yale
Regular	2,115	1,792	5,306	1,734	2,685	1,245
Early Decision	577	627	1,228	444	1,195	761

Table11.22

We want to know if the number of regular applications accepted follows the same distribution as the number of early applications accepted. State the null and alternative hypotheses, the degrees of freedom and the test statistic, sketch the graph of the *p*-value, and draw a conclusion about the test of homogeneity.

Solution:

H_0 : The distribution of regular applications accepted is the same as the distribution of early applications accepted.

H_a : The distribution of regular applications accepted is not the same as the distribution of early applications accepted.
$df = 5$
χ^2 test statistic = 430.06

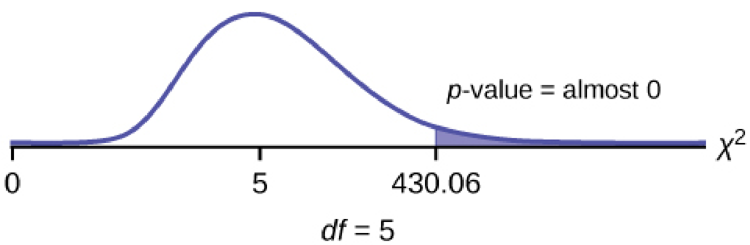


Figure 1.

Press the MATRX key and arrow over to EDIT. Press 1:[A]. Press 3 ENTER 3 ENTER. Enter the table values by row. Press ENTER after each. Press 2nd QUIT. Press STAT and arrow over to TESTS. Arrow down toC:χ2-TEST. Press ENTER. You should see Observed:[A] and Expected:[B]. Arrow down to Calculate. Press ENTER. The test statistic is 430.06 and the *p*-value = 9.80E-91. Do the procedure a second time but arrow down to Draw instead of calculate.

Section 6. Test of a Single Variance

TRY IT 11.10

A SCUBA instructor wants to record the collective depths each of his students dives during their checkout. He is interested in how the depths vary, even though everyone should have been at the same depth. He believes the standard deviation is three feet. His assistant thinks the standard deviation is less than three feet. If the instructor were to conduct a test, what would the null and alternative hypotheses be?

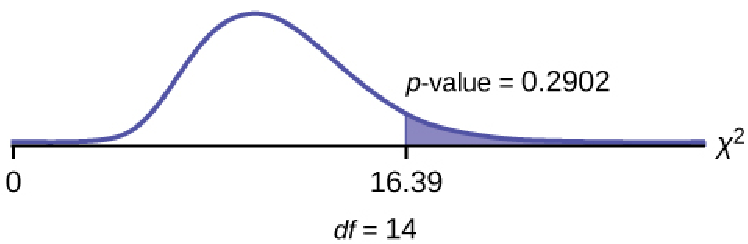
Solution:
H_0: $\sigma^2 = 3^2$
H_a: $\sigma^2 < 3^2$

TRY IT 11.11

The FCC conducts broadband speed tests to measure how much data per second passes between a consumer's computer and the internet. As of August of 2012, the standard deviation of Internet speeds across Internet Service Providers (ISPs) was 12.2 percent. Suppose a sample of 15 ISPs is taken, and the standard deviation is 13.2. An analyst claims that the standard deviation of speeds is more than what was reported. State the null and alternative hypotheses, compute the degrees of freedom, the test statistic, sketch the graph of the *p*-value, and draw a conclusion. Test at the 1% significance level.

Solution:

H_0: $\sigma^2 = 12.2^2$
H_a: $\sigma^2 > 12.2^2$
$df = 14$
chi^2 test statistic = 16.39



The *p*-value is 0.2902, so we decline to reject the null hypothesis. There is not enough evidence to suggest that the variance is greater than 12.2^2.
In 2nd DISTR, use7:χ2cdf. The syntax is (lower, upper, df) for the parameter list. χ2cdf(16.39,10^99,14). The *p*-value = 0.2902.

Chapter 12. Linear Regression and Correlation

Section 1. Linear Equations

TRY IT 12.1

Is the following an example of a linear equation?

$y = -0.125 - 3.5x$

Solution:
Yes

TRY IT 12.2

Is the following an example of a linear equation? Why or why not?

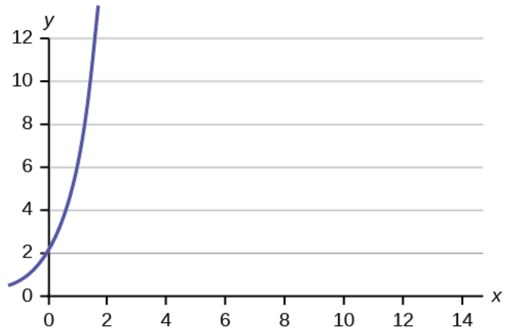


Figure 12.3

Solution:

No, the graph is not a straight line; therefore, it is not a linear equation.

TRY IT 12.3

Emma's Extreme Sports hires hang-gliding instructors and pays them a fee of $50 per class as well as $20 per student in the class. The total cost Emma pays depends on the number of students in a class. Find the equation that expresses the total cost in terms of the number of students in a class.

Solution:

y = 50 + 20x

TRY IT 12.4

Ethan repairs household appliances like dishwashers and refrigerators. For each visit, he charges $25 plus $20 per hour of work. A linear equation that expresses the total amount of money Ethan earns per visit is $y = 25 + 20x$.

What are the independent and dependent variables? What is the *y*-intercept and what is the slope? Interpret them using complete sentences.

Solution:

The independent variable (x) is the number of hours Ethan works each visit. The dependent variable (y) is the amount, in dollars, Ethan earns for each visit.
The y-intercept is 25 (a = 25). At the start of a visit, Ethan charges a one-time fee of $25 (this is when x = 0). The slope is 20 (b = 20). For each visit, Ethan earns $20 for each hour he works.

Section 2. Scatter Plots

TRY IT 12.5

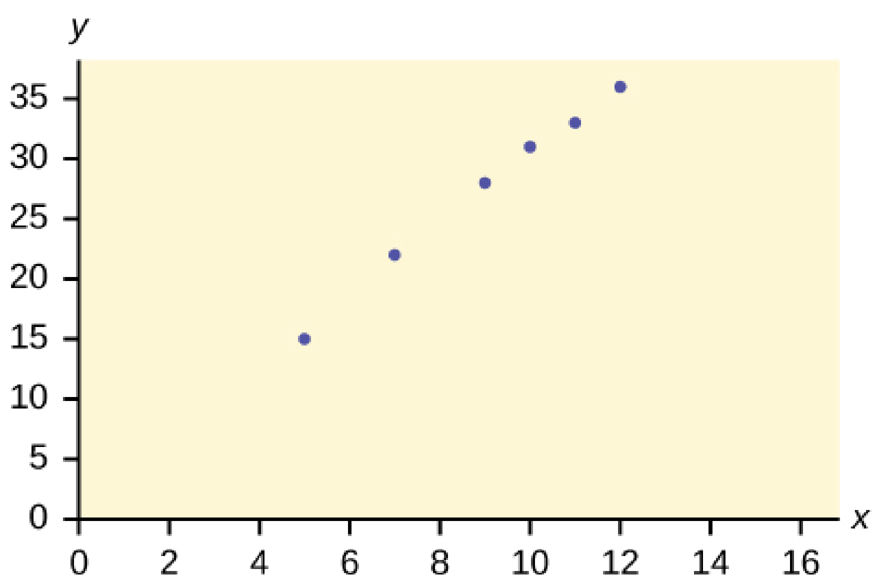
Amelia plays basketball for her high school. She wants to improve to play at the college level. She notices that the number of points she scores in a game goes up in response to the number of hours she practices her jump shot each week. She records the following data:

X (hours practicing jump shot)	*Y* (points scored in a game)
5	15
7	22
9	28
10	31
11	33
12	36

Table12.2

Construct a scatter plot and state if what Amelia thinks appears to be true.

Solution:



Yes, Amelia's assumption appears to be correct. The number of points Amelia scores per game goes up when she practices her jump shot more.

Section 3. The Regression Equation

TRY IT 12.6

SCUBA divers have maximum dive times they cannot exceed when going to different depths. The data in Table 12.4show different depths with the maximum dive times in minutes. Use your calculator to find the least squares regression line and predict the maximum dive time for 110 feet.

X **(depth in feet)**	***Y*** **(maximum dive time)**
50	80
60	55
70	45
80	35
90	25
100	22

Table12.4

Solution:
$\hat{y} = 127.24 – 1.11x$
At 110 feet, a diver could dive for only five minutes.

Section 4. Testing the Significance of the Correlation Coefficent

TRY IT 12.7

For a given line of best fit, you computed that $r = 0.6501$ using $n = 12$ data points and the critical value is 0.576. Can the line be used for prediction? Why or why not?

Solution:

If the scatter plot looks linear then, yes, the line can be used for prediction, because $r >$ the positive critical value.

TRY IT 12.8

For a given line of best fit, you compute that $r = 0.5204$ using $n = 9$ data points, and the critical value is 0.666. Can the line be used for prediction? Why or why not?

Solution:

No, the line cannot be used for prediction, because r < the positive critical value.

TRY IT 12.9

For a given line of best fit, you compute that $r = –0.7204$ using $n = 8$ data points, and the critical value is = 0.707. Can the line be used for prediction? Why or why not?

Solution:

Yes, the line can be used for prediction, because r < the negative critical value.

TRY IT 12.10

For a given line of best fit, you compute that $r = 0$ using $n = 100$ data points. Can the line be used for prediction? Why or why not?

Solution:

No, the line cannot be used for prediction no matter what the sample size is.

Section 5. Prediction

TRY IT 12.11

Data are collected on the relationship between the number of hours per week practicing a musical instrument and scores on a math test. The line of best fit is as follows:

$\hat{y} = 72.5 + 2.8x$
What would you predict the score on a math test would be for a student who practices a musical instrument for five hours a week?

Solution:
86.5

Section 6. Outliers

TRY IT 12.12

Identify the potential outlier in the scatter plot. The standard deviation of the residuals or errors is approximately 8.6.

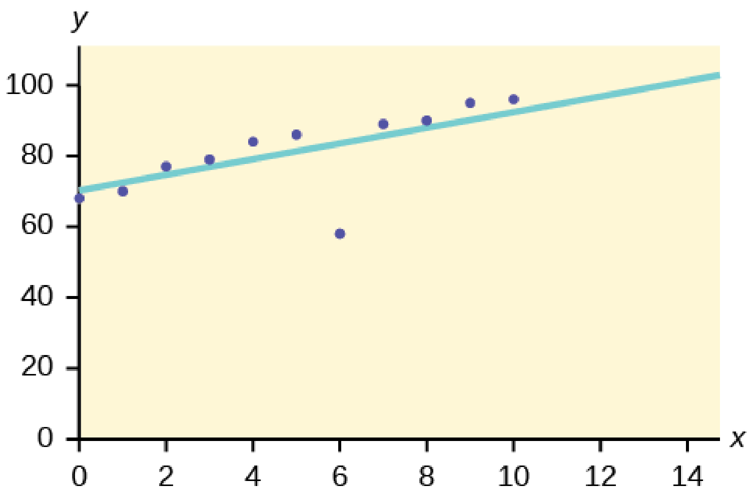


Figure 12.19

Solution:

The outlier appears to be at (6, 58). The expected *y* value on the line for the point (6, 58) is approximately 82. Fifty-eight is 24 units from 82. Twenty-four is more than two standard deviations ($2s = (2)(8.6) = 17.2$). So 82 is more than two standard deviations from 58, which makes (6, 58) a potential outlier.

TRY IT 12.13

The data points for the graph from the third exam/final exam example are as follows: (1, 5), (2, 7), (2, 6), (3, 9), (4, 12), (4, 13), (5, 18), (6, 19), (7, 12), and (7, 21). Remove the outlier and recalculate the line of best fit. Find the value of $\hat{y}$ when $x = 10$.

Solution:

$\hat{y} = 1.04 + 2.96x$; 30.64

TRY IT 12.14

The following table shows economic development measured in per capita income PCINC.

Year	PCINC	Year	PCINC
1870	340	1920	1050
1880	499	1930	1170
1890	592	1940	1364
1900	757	1950	1836

Year	PCINC	Year	PCINC
1910	927	1960	2132

Table12.7

a. What are the independent and dependent variables?
b. Draw a scatter plot.
c. Use regression to find the line of best fit and the correlation coefficient.
d. Interpret the significance of the correlation coefficient.
e. Is there a linear relationship between the variables?
f. Find the coefficient of determination and interpret it.
g. What is the slope of the regression equation? What does it mean?
h. Use the line of best fit to estimate PCINC for 1900, for 2000.
i. Determine if there are any outliers.

Solution:

a. The independent variable (*x*) is the year and the dependent variable (*y*) is the per capita income.

b.

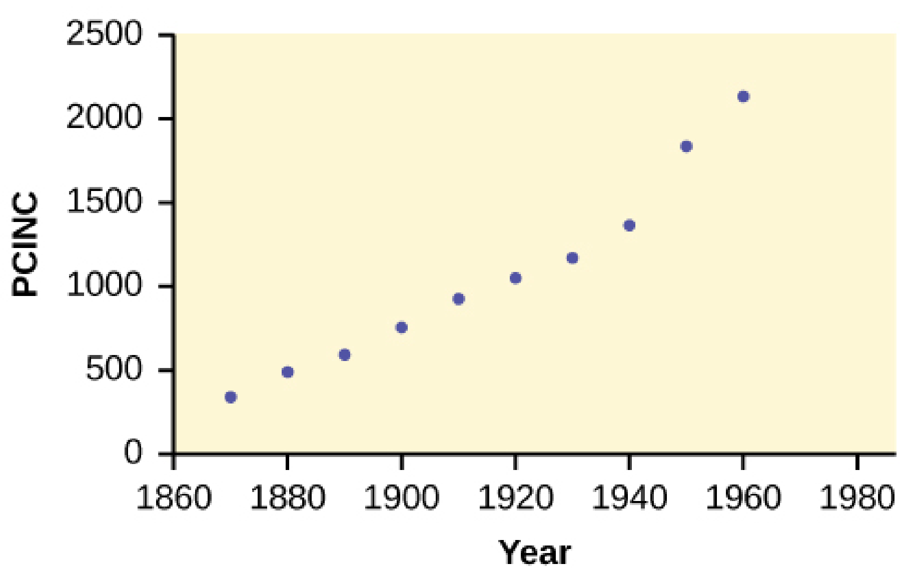


Figure 2.

c. $\hat{y} = 18.61x - 34574$; $r = 0.9732$

d. At $df = 8$, the critical value is 0.632. The r value is significant because it is greater than the critical value.

e. There does appear to be a linear relationship between the variables.

f. The coefficient of determination is 0.947, which means that 94.7% of the variation in PCINC is explained by the variation in the years.

g. and h. The slope of the regression equation is 18.61, and it means that per capita income increases by \$18.61 for each passing year. $\hat{y}$ = 785 when the year is 1900, and $\hat{y}$ = 2,646 when the year is 2000.

i. There do not appear to be any outliers.

Chapter 13. F Distribution and One-Way ANOVA

Section 2. The F Distribution and the F-Ratio

TRY IT 13.1

As part of an experiment to see how different types of soil cover would affect slicing tomato production, Marist College students grew tomato plants under different soil cover conditions. Groups of three plants each had one of the following treatments

- bare soil
- a commercial ground cover
- black plastic
- straw
- compost

All plants grew under the same conditions and were the same variety. Students recorded the weight (in grams) of tomatoes produced by each of the *n* = 15 plants:

Bare: n_1= 3	**Ground Cover: n_2= 3**	**Plastic: n_3= 3**	**Straw: n_4= 3**	**Compost: n_5= 3**
2,625	5,348	6,583	7,285	6,277
2,997	5,682	8,560	6,897	7,818
4,915	5,482	3,830	9,230	8,677

Table13.4

Create the one-way ANOVA table.
Solution:

Enter the data into lists L1, L2, L3, L4 and L5. Press STAT and arrow over to TESTS. Arrow down to ANOVA. Press ENTER and enter L1, L2, L3, L4, L5). Press ENTER. The table was filled in with the results from the calculator.

One-Way ANOVA table:

Source of Variation	Sum of Squares (*SS*)	Degrees of Freedom (*df*)	Mean Square (*MS*)	*F*
Factor (Between)	36,648,561	5 – 1 = 4	$\frac{36,648,561}{4}$ =9,162,140	$\frac{9,162,140}{2,044,672.6}$=4.4810
Error (Within)	20,446,726	15 – 5 = 10	$\frac{20,446,726}{10}$ =2,044,672.6	
Total	57,095,287	15 – 1 = 14		

Section 3. Facts About the F Distribution

TRY IT 13.2

MRSA, or *Staphylococcus aureus*, can cause a serious bacterial infections in hospital patients. Table 13.6 shows various colony counts from different patients who may or may not have MRSA. The data from the table is plotted in Figure 13.5.

Conc = 0.6	Conc = 0.8	Conc = 1.0	Conc = 1.2	Conc = 1.4
9	16	22	30	27
66	93	147	199	168
98	82	120	148	132

Table13.6

Plot of the data for the different concentrations:

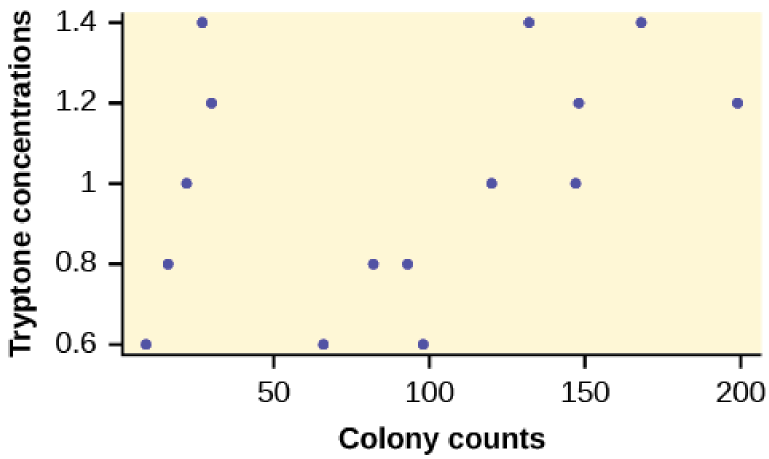


Figure 13.5

Test whether the mean number of colonies are the same or are different. Construct the ANOVA table (by hand or by using a TI-83, 83+, or 84+ calculator), find the *p*-value, and state your conclusion. Use a 5% significance level.

Solution:

While there are differences in the spreads between the groups (see Figure), the differences do not appear to be big enough to cause concern.

We test for the equality of mean number of colonies:

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

$H_a : \mu^i \neq \mu^j$ some $i \neq j$

The one-way ANOVA table results are shown in Table.

Source of Variation	**Sum of Squares (*SS*)**	**Degrees of Freedom (*df*)**	**Mean Square (*MS*)**	***F***
Factor (Between)	10,233	5 – 1 = 4	$\frac{10,233}{4}$ =2,558.25	$\frac{2,558.25}{4194.9}$=0.6099
Error (Within)	41,949	15 – 5 = 10	$\frac{41,949}{10}$ =4,194.9	
Total	52,182	15 – 1 = 14		

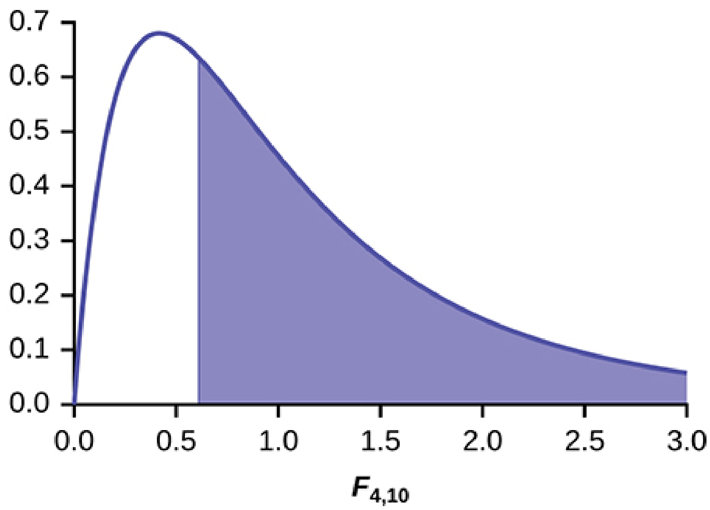


Figure 3.

Distribution for the test: $F_{4,10}$

Probability Statement: p-value = $P(F > 0.6099) = 0.6649$.

Compare α and the p-value: $\alpha = 0.05$, p-value = 0.669, $\alpha > p$-value

Make a decision: Since $\alpha > p$-value, we do not reject H_0.

Conclusion: At the 5% significance level, there is insufficient evidence from these data that different levels of tryptone will cause a significant difference in the mean number of bacterial colonies formed.

TRY IT 13.3

Four sports teams took a random sample of players regarding their GPAs for the last year. The results are shown in Table 13.8.

Basketball	Baseball	Hockey	Lacrosse
3.6	2.1	4.0	2.0
2.9	2.6	2.0	3.6
2.5	3.9	2.6	3.9
3.3	3.1	3.2	2.7

Basketball	Baseball	Hockey	Lacrosse
3.8	3.4	3.2	2.5

Table13.8 GPAs FOR FOUR SPORTS TEAMS

Use a significance level of 5%, and determine if there is a difference in GPA among the teams.

Solution:

With a p-value of 0.9271, we decline to reject the null hypothesis. There is not sufficient evidence to conclude that there is a difference among the GPAs for the sports teams.

TRY IT 13.4

Another fourth grader also grew bean plants, but this time in a jelly-like mass. The heights were (in inches) 24, 28, 25, 30, and 32. Do a one-way ANOVA test on the four groups. Are the heights of the bean plants different? Use the same method as shown in Example 13.4.

Solution:

- F = 0.9496
- p-value = 0.4402

From the sample data, the evidence is not sufficient to conclude that the mean heights of the bean plants are different.

Section 4. Test of Two Variances

TRY IT 13.5

The New York Choral Society divides male singers up into four categories from highest voices to lowest: Tenor1, Tenor2, Bass1, Bass2. In the table are heights of the men in the Tenor1 and Bass2 groups. One suspects that taller men will have lower voices, and that the variance of height may go up with the lower voices as well. Do we have good evidence that the variance of the heights of singers in each of these two groups (Tenor1 and Bass2) are different?

Tenor1	Bass2	Tenor 1	Bass 2	Tenor 1	Bass 2
69	72	67	72	68	67
72	75	70	74	67	70

Tenor1	Bass2	Tenor 1	Bass 2	Tenor 1	Bass 2
71	67	65	70	64	70
66	75	72	66		69
76	74	70	68		72
74	72	68	75		71
71	72	64	68		74
66	74	73	70		75
68	72	66	72		

Table 13.11

Solution:

The histograms are not as normal as one might like. Plot them to verify. However, we proceed with the test in any case.

Subscripts: T1= tenor1 and B2 = bass 2

The standard deviations of the samples are sT1 = 3.3302 and sB2 = 2.7208.

The hypotheses are

H0: $\sigma_{T1}^2 = \sigma_{B2}^2$ and H0: $\sigma_{T1}^2 \neq \sigma_{B2}^2$ (two tailed test)

The F statistic is 1.4894 with 20 and 25 degrees of freedom.

The p-value is 0.3430. If we assume alpha is 0.05, then we cannot reject the null hypothesis.

We have no good evidence from the data that the heights of Tenor1 and Bass2 singers have different variances (despite there being a significant difference in mean heights of about 2.5 inches.)