ANSWERS TO TRY IT EXERCISES

Introductory Business Statistics

Chapter 1 Sampling and Data

Section 1 Definition of Statistics, Probability, and Key Terms

TRY IT 1.1:

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(categorical) data

TRY IT 1.9:

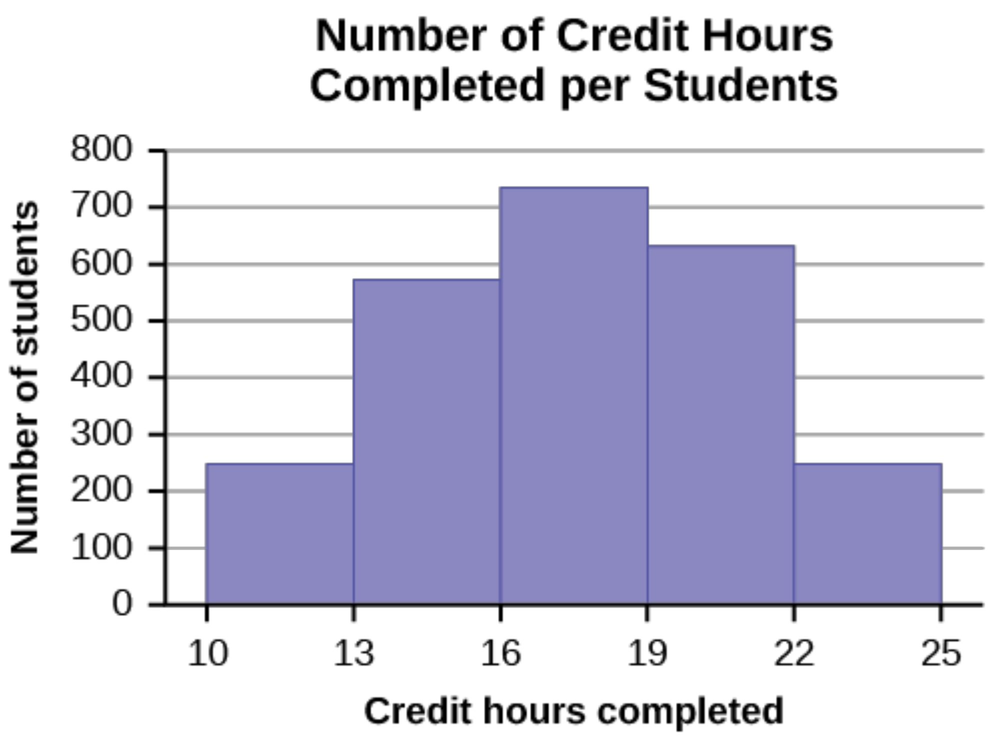
Determine the correct data type (quantitative or qualitative) for the number of cars in a parking lot. Indicate whether quantitative data are continuous or discrete.

Solution:

quantitative 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.



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 1.13:

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 Levels of Measurement

TRY IT 1.14:

Table 1.9 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	650650 = 0.12	0.12
4.97–6.99	7	750750 = 0.14	0.12 + 0.14 = 0.26
6.99–9.01	15	15501550 = 0.30	0.26 + 0.30 = 0.56
9.01–11.03	8	850850 = 0.16	0.56 + 0.16 = 0.72
11.03–13.05	9	950950 = 0.18	0.72 + 0.18 = 0.90
13.05–15.07	5	550550 = 0.10	0.90 + 0.10 = 1.00
	Total = 50	Total = 1.00	

Table1.9

From Table 1.9, find the percentage of rainfall that is less than 9.01 inches.

Solution:

0.56 or 56%

TRY IT 1.15:

From Table 1.9, 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 1.17:

Table 1.9 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:

$\frac{9}{50}$

TRY IT 1.18:

Table 1.12 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

Year	Total number of crashes	Year	Total number of crashes
2000	37,526	2010	30,296
2001	37,862	2011	29,757
2002	38,491	Total	653,782
2003	38,477		

Table1.12

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.

Chapter 2 Descriptive Statistics

Section 1 Display Data

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.

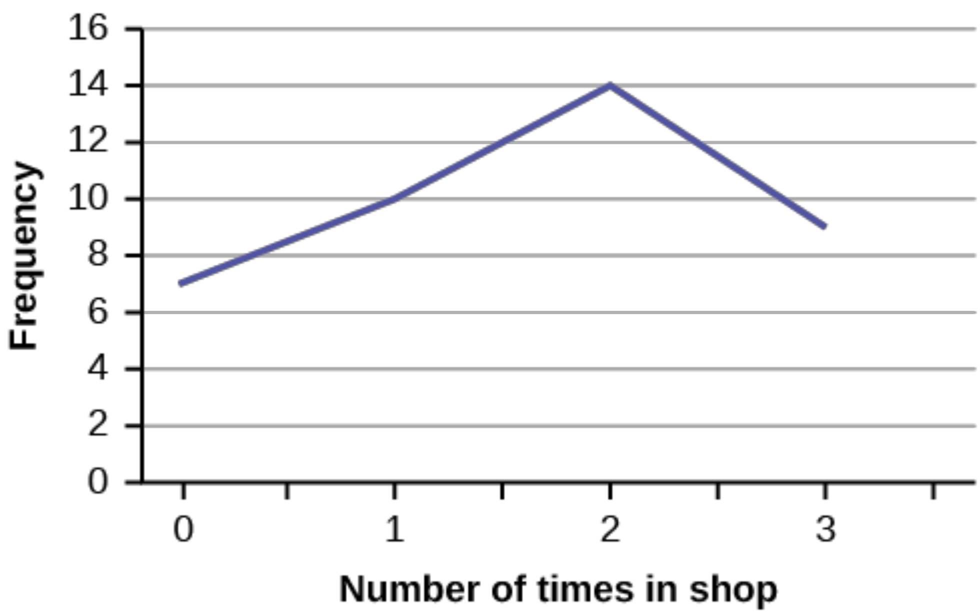
TRY IT 2.4:

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 2.7. Construct a line graph.

Number of times in shop	Frequency
0	7
1	10
2	14
3	9

Table2.7

Solution:



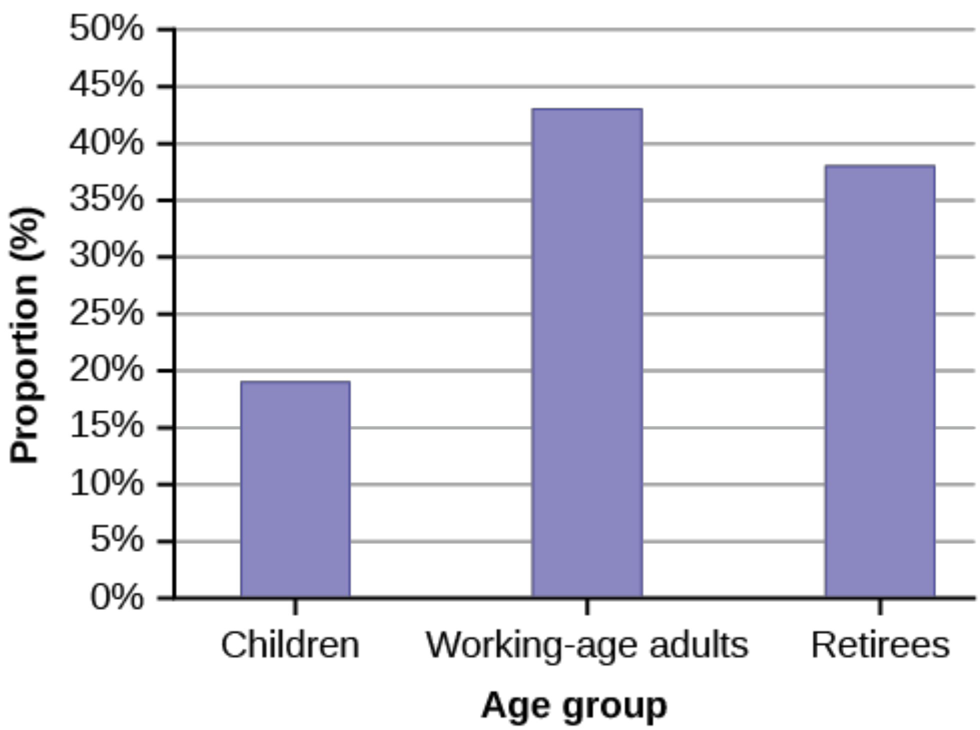
TRY IT 2.5:

The population in Park City is made up of children, working-age adults, and retirees. Table 2.9 shows 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%

Table2.9

Solution:



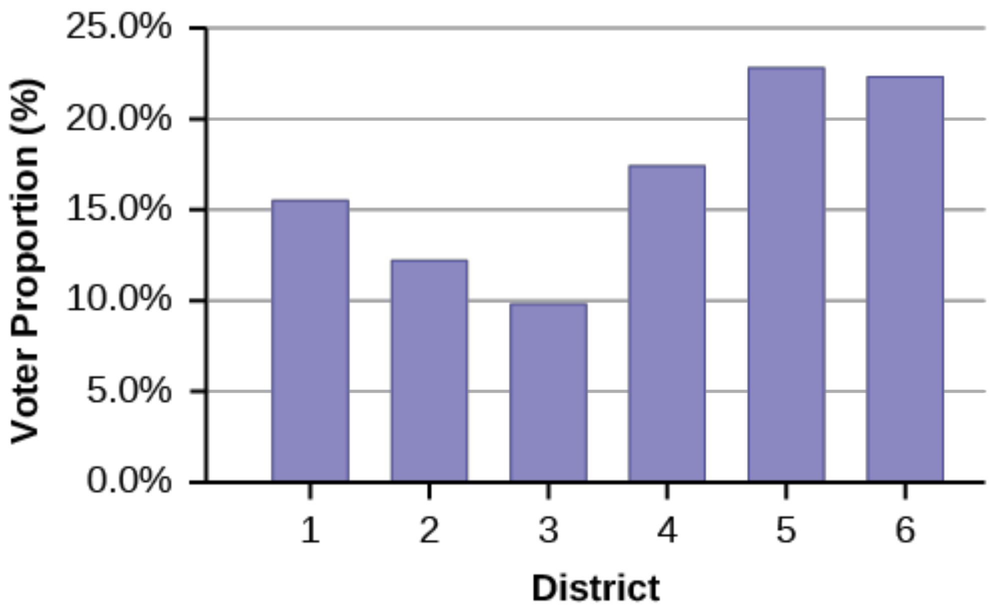
TRY IT 2.6:

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%

Table2.11

Solution:



TRY IT 2.8:

The following data are the shoe sizes of 50 male students. The sizes are continuous data since shoe size is measured. 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

$\frac{14.05-8.95}{6}$ = 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 2.10:

Using this data set, construct a histogram.

Number of hours my classmates spent playing video games on weekends

9.95	10	2.25	16.75	0
19.5	22.5	7.5	15	12.75
5.5	11	10	20.75	17.5
23	21.9	24	23.75	18
20	15	22.9	18.8	20.5

Table2.13

Solution:

Some values in this data set fall on boundaries for the class intervals. A value is counted in a class interval if it falls on the left boundary, but not if it falls on the right boundary. Different researchers may set up histograms for the same data in different ways. There is more than one correct way to set up a histogram.

TRY IT 2.11:

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

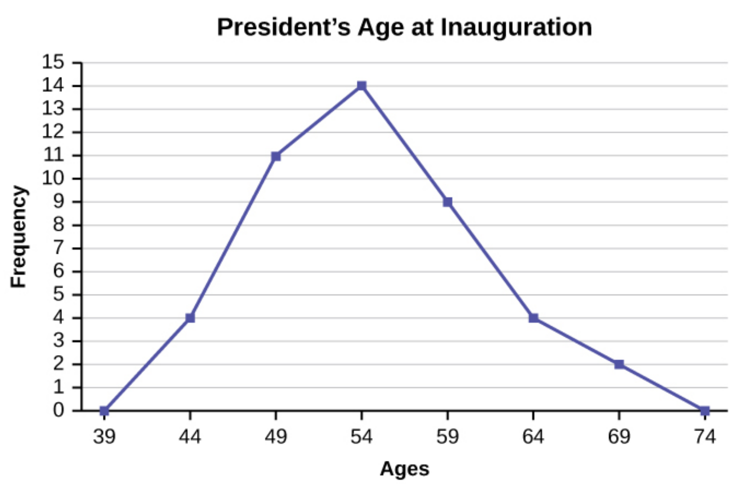
Age at inauguration	**Frequency**
41.5–46.5	4
46.5–51.5	11
51.5–56.5	14
56.5–61.5	9

Age at inauguration	Frequency
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.



TRY IT 2.13:

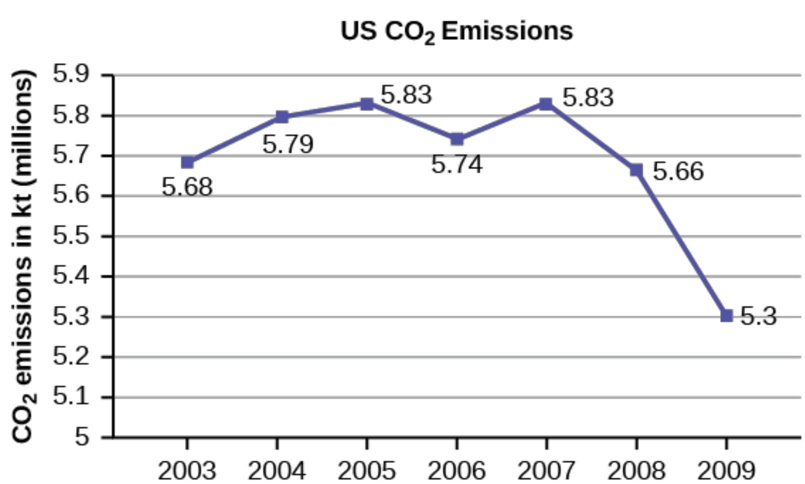
The following table is a portion of a data set from www.worldbank.org. Use the table to construct a time series graph for CO_2 emissions for the United States.

CO_2 emissions

Year	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:



Section 2 Measures of the Location of the Data

TRY IT 2.16:

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
4	10	0.25	0.90
5	4	0.10	1.00

Solution:

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

The 65th percentile is 3.5.

TRY IT 2.18:

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 = $\frac{k}{100}(n+1)$ = $\frac{20}{100}(29+1)$ = 6. The age in the sixth position is 27. The 20th percentile is 27 years.

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

TRY IT 2.21:

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.

Section 3 Measures of the Center of the Data

TRY IT 2.28:

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
11.5–15.5	7
15.5–19.5	9

Table2.26

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 = (1.75)(3) + (5.5)(7) + (9.5)(12) + (13.5)(7) + (17.5)(9) = 409.75/38 = 10.78

Section 7 Measures of the Spread of the Data

TRY IT 2.32:

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

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. $A \cap B$ = _____________________, $A \cup B$ = _________________

e. $P(A \cap B)$ = __________, $P(A \cup B)$ = ______________

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. $A \cap B$ = {(2,2), (2,4), (3,1), (3,3)}

$A \cup B$ = {(1,1), (1,3), (2,1), (2,2), (2,3), (2,4), (3,1), (3,2), (3,3), (3,4)}

e. $P(A \cap B) = \frac{1}{3}$, $P(A \cup B) = \frac{5}{6}$

f. B' = {(1,1), (1,2), (1,3), (1,4)}, $P(B') = \frac{1}{3}$

g. $P(B) + P(B') = 1$

h. $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{2}$, $P(B|A) = \frac{P(A \cap B)}{P(B)} = \frac{2}{3}$, 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: a. Possible; b. Impossible, c. Possible

with replacement: a. Possible; b. Possible, c. 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 $A \cap B$ = learning Spanish and German. Suppose P(A) = 0.4 and P(B) = 0.2P(B) = 0.2. $P(A \cap B) = 0.08$. Are events *A* and *B* independent? Hint: You must show ONE of the following:

- $P(A|B) = P(A)$
- $P(B|A) = P(B)$
- $P(A \cap B) = P(A)P(B)$

Solution:

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

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 $S = \{R1, R2, R3, R4, R5, R6, G1, G2, G3, G4\}$.

S has ten outcomes. What is $P(G \cap O)$?

Solution:

Event G and $O = \{G1, G3\}$

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

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) = 0.40P(B) = 0.40, P(D) = 0.30P(D) = 0.30 and P(B∩D) = 0.20P(B∩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 $P(B \cap A) = 0.20$, 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 $P(F) = 0.56$
- $P(I \cap F) = 0$ because Mark will take only one route to work.

What is the probability of $P(I \cup F)$?

Solution:

Because $P(I \cap F) = 0$,

$P(I \cup F) = P(I)+P(F)-P(I \cap F) = 0.44+0.56-0 = 1$

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) = 14

b. P(T|F) = 12

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:

$P(D|C) = 0.85$

$P(C \cap D) = P(D \cap C)$

$P(D \cap C) = P(D|C)P(C) = (0.85)(0.75) = 0.6375$
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 $P(B \cap D)$.

b. Find $P(B \cup D)$.

Solution:

a. $P(B \cap D) = P(D|B)P(B) = (0.5)(0.4) = 0.20$.

b. $P(B \cup D) = P(B) + P(D) - P(B \cap D) = 0.40 + 0.30 - 0.20 = 0.50$

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.

$P(B) = \frac{140}{200}$

$P(B|A) = \frac{50}{140}$

$P(A \cap B) = P(B|A)P(A)$

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

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 $P(D \cap B)$.

c. Find $P(B|D)$.

d. Find $P(D \cap B')$.

e. Find $P(D|B')$.

Solution:

a. $P(B') = 0.60$

b. $P(D \cap B) = P(D|B)P(B) = 0.20$

c. $P(B|D) = \frac{P(B \cap D)}{P(D)} = \frac{(0.20)}{(0.30)} = 0.66$

d. $P(D \cap B') = P(D) - P(D \cap B) = 0.30 - 0.20 = 0.10$

e. $P(D|B') = P(D \cap B')P(B') = (P(D) - P(D \cap B))(0.60) = (0.10)(0.60) = 0.06$

Section 4 Contingency Tables and Probability Trees

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) = $\frac{250}{800}$ = 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. $P(H|M) = \frac{52}{90} = 0.5778$
b. For M and H to be independent, show $P(H|M) = P(H)$

   $P(H|M) = 0.5778$, $P(H) = \frac{90}{200} = 0.45$

   $P(H|M)$ 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	
Normal	20	51	28	

Weight/height	Tall	Medium	Short	Totals
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:

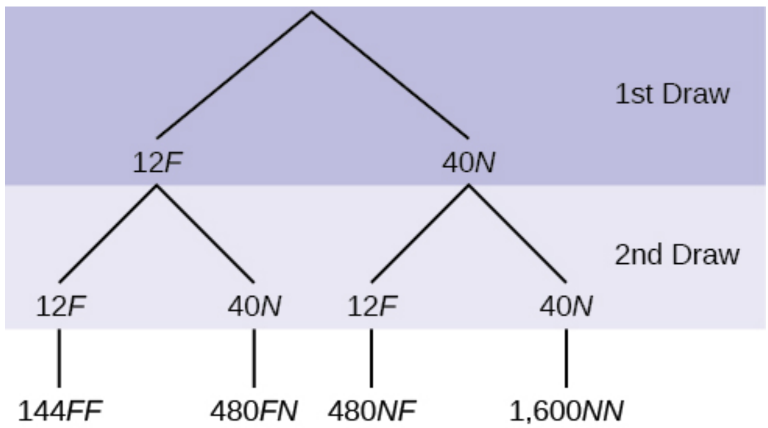
Weight/height	Tall	Medium	Short	Totals
Obese	18	28	14	60
Normal	20	51	28	99
Underweight	12	25	9	46
Totals	50	104	51	205

a. Row Totals: 60, 99, 46. Column totals: 50, 104, 51.
b. $P(\text{Tall}) = \frac{50}{205} = 0.244$
c. $P(\text{Obese} \cap \text{Tall}) = \frac{18}{205} = 0.088$
d. $P(\text{Tall} | \text{Obese}) = \frac{18}{60} = 0.3$
e. $P(\text{Obese} | \text{Tall}) = \frac{18}{50} = 0.36$
f. $P(\text{Tall} \cap \text{Underweight}) = \frac{12}{205} = 0.0585$

g. No. P(Tall) does not equal P(Tall||Obese).

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 $P(FF)$.



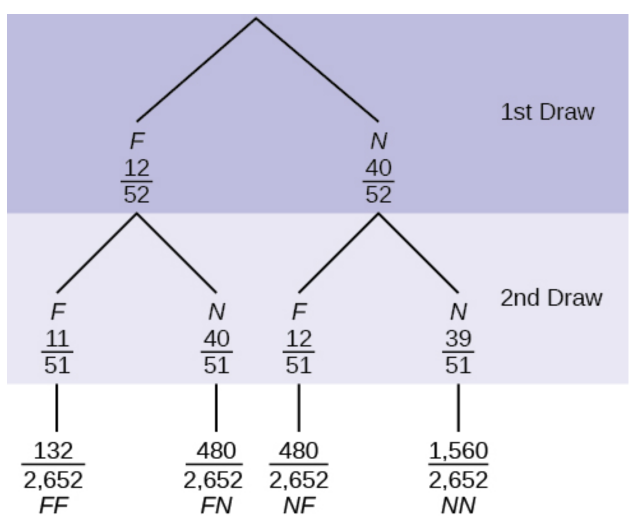
Solutions:

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.



a. Find P(FN∪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∪NF) = $\frac{480}{2{,}652} + \frac{280}{2{,}652} = \frac{960}{2{,}652} = \frac{80}{221}$

b. $P(N|F) = \frac{40}{51}$

c. *P*(at most one face card) = $\frac{(480+480+1{,}560)}{2{,}652} = \frac{2{,}520}{2{,}652}$

d. *P*(at least one face card) = $\frac{(132+480+480)}{2{,}652} = \frac{1{,}092}{2{,}652}$

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:

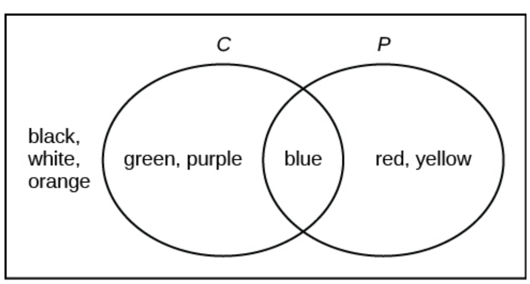
$\left(\frac{4}{7}\right)\left(\frac{3}{6}\right) + \left(\frac{3}{7}\right)\left(\frac{4}{6}\right)$

Section 5 Venn Diagrams

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∩P = {blue} and C∪P = {green, blue, purple, red, yellow}. Draw a Venn diagram representing this situation.

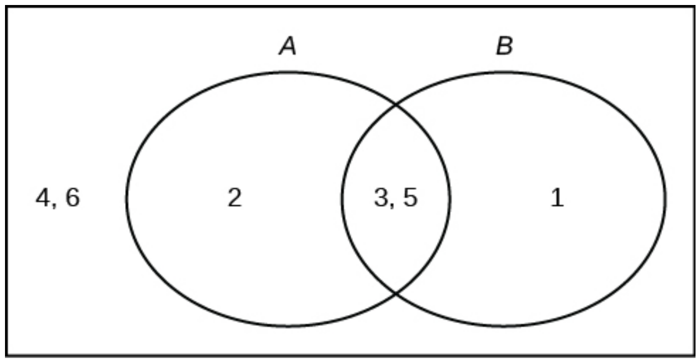
Solution:



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∩B = {3, 5}. A∪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.

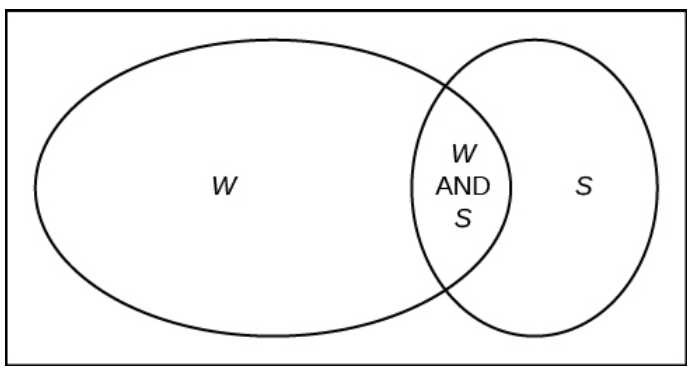
Solution:



TRY IT 3.30:

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:



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.

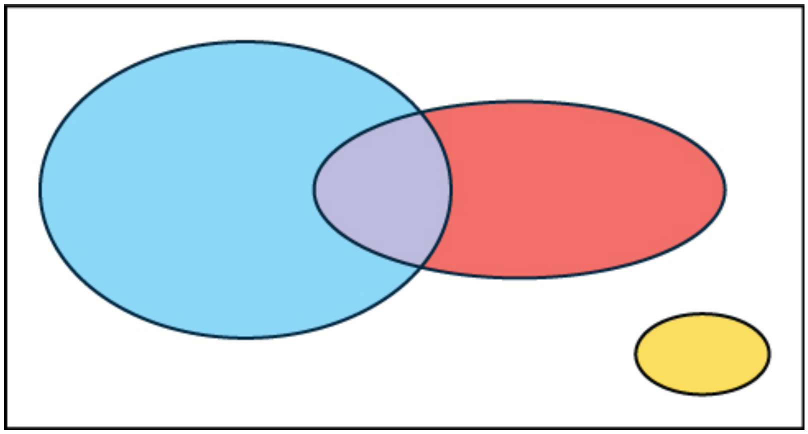
b. Find the probability that the customer buys either a novel or a non-fiction book.

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.



b. P(novel or non-fiction) = P(Blue ∪ Red) = P(Blue) + P(Red) - P(Blue ∩ 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 Hypergeometric Distribution

TRY IT 4.1:

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.

Section 2 Binomial Distribution

TRY IT 4.2:

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. Find the P(X=12) using the binomial Pdf.

Solution:

$P(x = 12)$

TRY IT 4.4:

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 4.4:

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. 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 = $\sqrt{\text{npq}} = \sqrt{80(0.613)(0.387)} \approx 3.564$

c. $P(x = 60) = 0.0036$

d. $P(x > 50) = 1 – P(x \leq 50) = 1 – 0.6282 = 0.3718$

Section 3 Geometric Distribution

TRY IT 4.5:

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 4.6:

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 4.9:

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 The United Colonies of Independence is 12%. Let X = the number of 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?

Solution:

a. $X \sim G(0.12)$

b. $P(x = 5) = 0.0720$

c. $P(x = 10) = 0.0380$

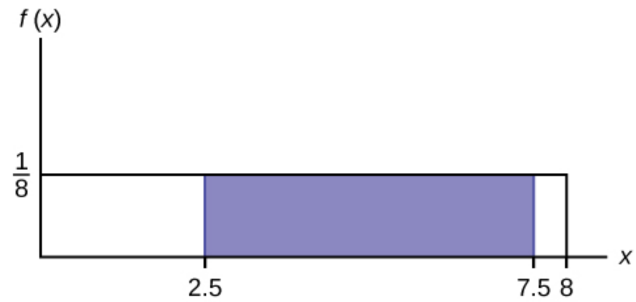
Chapter 5 Continuous Random Variables

Section 1 Properties of Continuous Probability Density Functions

TRY IT 5.1:

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

Solution:



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

Section 2 The Uniform Distribution

TRY IT 5.1:

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

Solution:

a is zero; *b* is 14; $X \sim U(0, 14)$; μ = 7 passengers; σ = 4.04 passengers

TRY IT 5.2:

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.

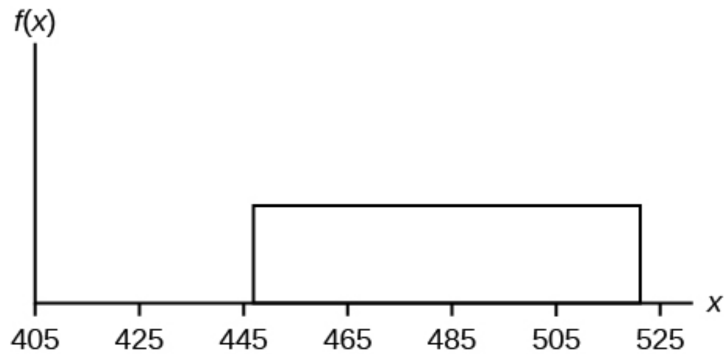
d. What is the probability that the duration of games for a team for the 2011 season is between 480 and 500 hours?

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. $P(480 < x < 500) = 0.2703$

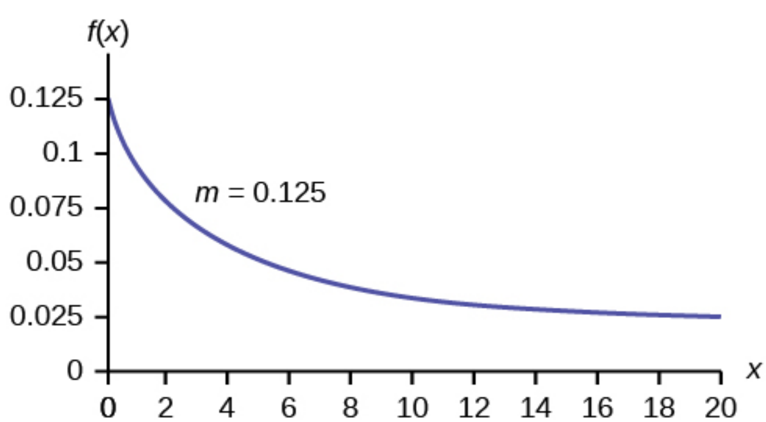
Section 3 The Exponential Distribution

TRY IT 5.3:

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:

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



TRY IT 5.4:

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 5.5:

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

Chapter 6 The Normal Distribution

Section 2 Using the Normal Distribution

TRY IT 6.3:

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 6.4:

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

Chapter 8 Confidence Intervals

Section 2 A Confidence Interval for a Population Standard Deviation Unknown, Small Sample Case

TRY IT 8.5:

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)

Section 3 A Confidence Interval for A Population Proportion

TRY IT 8.6:

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 8.8:

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. 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 $\alpha = 1 - 0.97 = 0.03$ and $\frac{\alpha}{2} = 0.015$.

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

$$EPB = \left(z\frac{\alpha}{2}\right)\sqrt{\frac{p'q'}{n}} = 2.17\sqrt{\frac{0.68(0.32)}{300}} \approx 0.0584$$

p' – EPB = 0.68 – 0.0584 = 0.0584
p' + EPB = 0.68 + 0.0584 = 0.0584
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.

Section 4 Calculating the Sample Size n: Continuous and Binary Random Variables

TRY IT 8.9:

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. Check the Real Estate section in your local

---

Chapter 9 Hypothesis Testing with One Sample

Section 2 Outcomes and the Type I and Type II Errors

TRY IT 9.5:

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 9.6:

“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 be H_0: the mean level of toxins is at most 800 μg, $H_0 : \mu_0 \leq 800\ \mu$g.

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.

Section 4 Full Hypothesis Test Examples

TRY IT 9.8:

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.

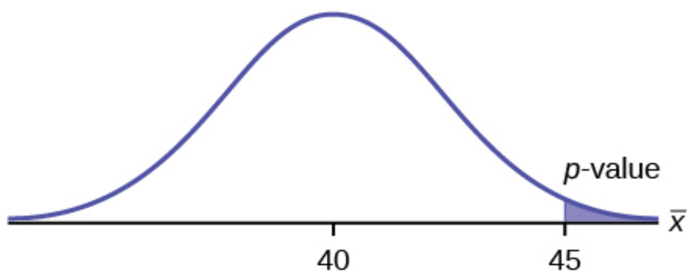
Solution:

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

$H_0 : \mu = 40$

$H_a : \mu > 40$

$p = 0.0062$



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 9.9:

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, state your conclusion, and identify the Type I 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 9.11:

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.

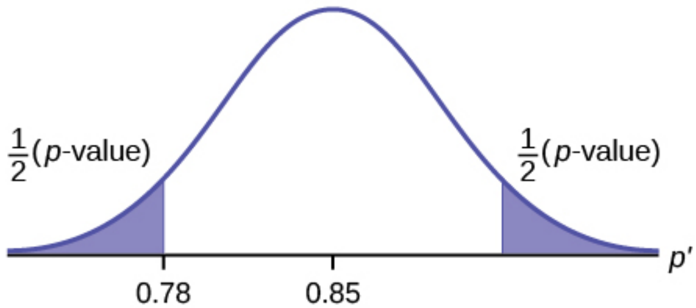
Solution:

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

H_0 : $p = 0.85$

H_a: $p \neq 0.85$

$p = 0.7554$



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%.

Chapter 10 Hypothesis Testing with Two Samples

Section 4 Comparing Two Independent Population Proportions

TRY IT 10.6:

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.

Section 5 Two Population Means with Known Standard Deviations

TRY IT 10.7:

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 10.4 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

Engine	Sample mean number of RPM	Population standard deviation
2	1,600	60

Table10.4

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.

Chapter 11 The Chi-Square Distribution

Section 2 Test of a Single Variance

TRY IT 11.1:

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: $\sigma^2 < 3^2$

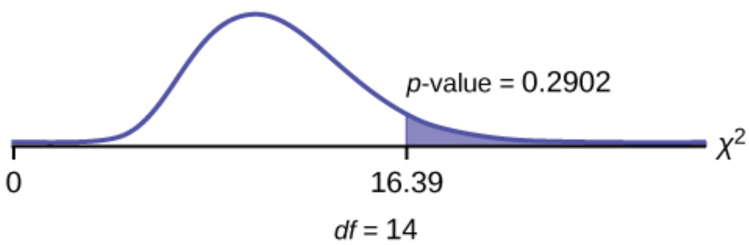
TRY IT 11.3:

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 distribution and mark the area associated with the level of confidence, 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.

Section 3 Goodness-of-Fit Test

TRY IT 11.4:

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 11.5:

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	Saturd
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 11.6:

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

Number of pets	Frequency
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?

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.

Section 4 Test of Independence

TRY IT 11.8:

A sample of 300 students is taken. Of the students surveyed, 50 were music students, while 250 were not. Ninety-seven of the 300 surveyed 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 11.9:

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 11.16 shows the results:

Industry sector	2000	2010	2020	Tota
Nonagriculture wage and salary	13,243	13,044	15,018	41,3
Goods-producing, excluding agriculture	2,457	1,771	1,950	6,17
Services-providing	10,786	11,273	13,068	35,1
Agriculture, forestry, fishing, and hunting	240	214	201	655
Nonagriculture self-employed and unpaid family worker	931	894	972	2,79
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	86,5

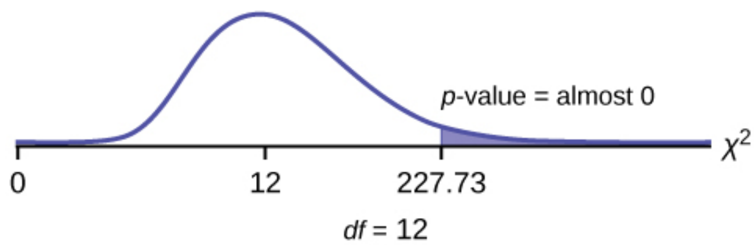
Table11.16

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 **STAT** and 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**.

Section 5 Test for Homogeneity

TRY IT 11.11:

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 11.19. 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

Table11.19

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 11.11:

Ivy League schools receive many applications, but only some can be accepted. At the schools listed in Table 11.20, 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	1,24
Early decision	577	627	1,228	444	1,195	761

Table11.20

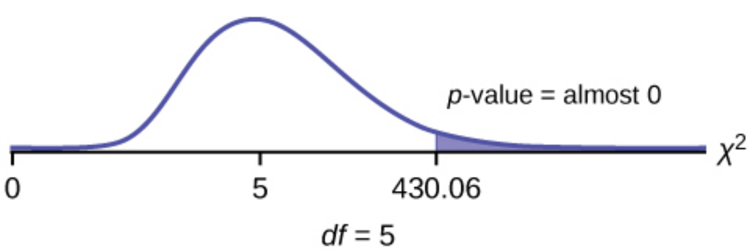
We want to know if the number of regular applications accepted follows the same distribution as the number of early applications accepted. State the null and alternative hypotheses, the degrees of freedom and the test statistic, sketch the graph of the χ^2 distribution and show the critical value and the calculated value of the test statistic, and draw a conclusion about the test of homogeneity.

Solution:

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 χ^2 distribution and show the critical value and the calculated value of the test statistic, and draw a conclusion about the test of homogeneity.

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



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 to **C: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**.

Chapter 12 F Distribution and One-Way ANOVA

Section 1 Test of Two Variances

TRY IT 12.1:

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

Table12.2

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 s_{T1} = 3.3302 and s_{B2} = 2.7208.

The hypotheses are

H_0: $\sigma_{T1}^2 = \sigma_{B2}^2$ and H_0: $\sigma_{T1}^2 \neq \sigma_{B2}^2$ (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.)

Section 3 The F Distribution and the F-Ratio

TRY IT 12.2:

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

Bare: n_1 = 3	Ground Cover: n_2 = 3	Plastic: n_3 = 3	Straw: n_4 = 3	Compost: n_5 = 3
4,915	5,482	3,830	9,230	8,677

Table12.6

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		

TRY IT 12.3:

MRSA, or *Staphylococcus aureus*, can cause a serious bacterial infections in hospital patients. Table 12.8 shows various colony counts from different patients who may or may not have MRSA. The data from the table is plotted in Figure 12.5.

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

Table12.8

Plot of the data for the different concentrations:

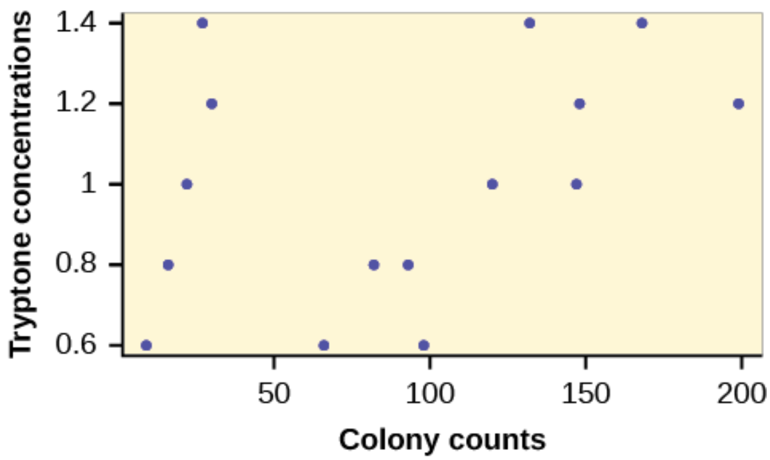


Figure 12.5

Test whether the mean number of colonies are the same or are different. Construct the ANOVA table, 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 2), the differences do not appear to be big enough to cause concern.

We test for the equality of mean number of colonies:

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

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

The one-way ANOVA table results are shown.

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

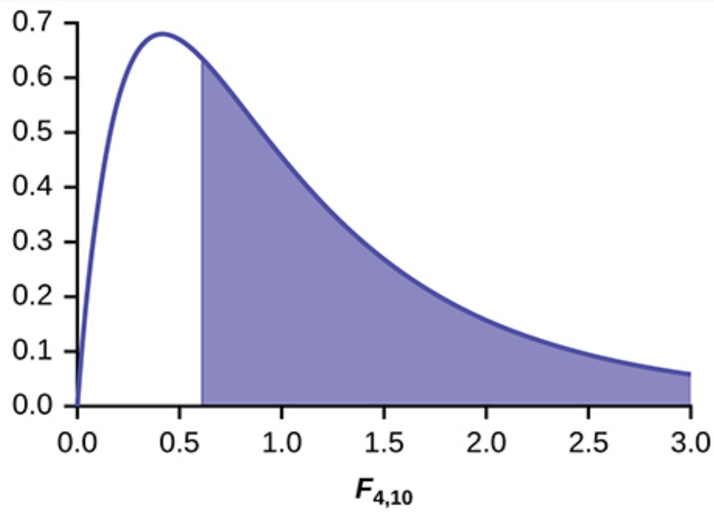


Figure 2

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 12.4:

Four sports teams took a random sample of players regarding their GPAs for the last year. The results are shown in Table 12.10.

Basketball	Baseball	Hockey	Lacrosse
3.6	2.1	4.0	2.0

Basketball	Baseball	Hockey	Lacrosse
2.9	2.6	2.0	3.6
2.5	3.9	2.6	3.9
3.3	3.1	3.2	2.7
3.8	3.4	3.2	2.5

Table12.10 GPAs for four sports teams

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

Solution:

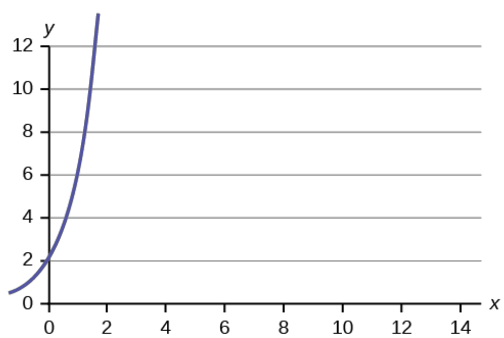
With a *p*-value of 0.9271, we decline to reject the null hypothesis. There is not sufficient evidence to conclude that there is a difference among the GPAs for the sports teams.

Chapter 13 Linear Regression and Correlation

Section 3 Linear Equations

TRY IT 13.2:

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



Solution:

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