

# Complete Solutions Manual to Accompany

---

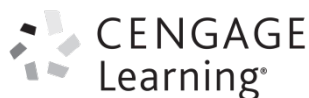
## **Introduction to Probability and Statistics** **15th Edition**

**William Mendenhall, III**  
1925-2009

**Robert J. Beaver**  
University of California, Riverside, Emeritus

**Barbara M. Beaver**  
University of California, Riverside, Emerita

Prepared by  
**Barbara M. Beaver**



---

Australia • Brazil • Mexico • Singapore • United Kingdom • United States

© 2020 Cengage Learning

ALL RIGHTS RESERVED. No part of this work covered by the copyright herein may be reproduced, transmitted, stored, or used in any form or by any means graphic, electronic, or mechanical, including but not limited to photocopying, recording, scanning, digitizing, taping, Web distribution, information networks, or information storage and retrieval systems, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the publisher except as may be permitted by the license terms below.

For product information and technology assistance, contact us at  
**Cengage Learning Customer & Sales Support,**  
**1-800-354-9706.**

For permission to use material from this text or product, submit  
all requests online at **[www.cengage.com/permissions](http://www.cengage.com/permissions)**  
Further permissions questions can be emailed to  
**[permissionrequest@cengage.com](mailto:permissionrequest@cengage.com)**.

ISBN-13: 978-1-337-55829-7  
ISBN-10: 1-337-55829-X

**Cengage Learning**  
20 Channel Center Street  
Boston, MA 02210  
USA

Cengage Learning is a leading provider of customized learning solutions with office locations around the globe, including Singapore, the United Kingdom, Australia, Mexico, Brazil, and Japan. Locate your local office at:  
**[www.cengage.com/global](http://www.cengage.com/global)**.

Cengage Learning products are represented in  
Canada by Nelson Education, Ltd.

To learn more about Cengage Learning Solutions  
or to purchase any of our products at our preferred  
online store, visit **[www.cengage.com](http://www.cengage.com)**.

**NOTE: UNDER NO CIRCUMSTANCES MAY THIS MATERIAL OR ANY PORTION THEREOF BE SOLD, LICENSED, AUCTIONED, OR OTHERWISE REDISTRIBUTED EXCEPT AS MAY BE PERMITTED BY THE LICENSE TERMS HEREIN.**

#### READ IMPORTANT LICENSE INFORMATION

Dear Professor or Other Supplement Recipient:

Cengage Learning has provided you with this product (the "Supplement") for your review and, to the extent that you adopt the associated textbook for use in connection with your course (the "Course"), you and your students who purchase the textbook may use the Supplement as described below. Cengage Learning has established these use limitations in response to concerns raised by authors, professors, and other users regarding the pedagogical problems stemming from unlimited distribution of Supplements.

Cengage Learning hereby grants you a nontransferable license to use the Supplement in connection with the Course, subject to the following conditions. The Supplement is for your personal, noncommercial use only and may not be reproduced, or distributed, except that portions of the Supplement may be provided to your students in connection with your instruction of the Course, so long as such students are advised that they may not copy or distribute any portion of the Supplement to any third party. Test banks, and other testing materials may be made available in the classroom and collected at the end of each class session, or posted electronically as described herein. Any

material posted electronically must be through a password-protected site, with all copy and download functionality disabled, and accessible solely by your students who have purchased the associated textbook for the Course. You may not sell, license, auction, or otherwise redistribute the Supplement in any form. We ask that you take reasonable steps to protect the Supplement from unauthorized use, reproduction, or distribution. Your use of the Supplement indicates your acceptance of the conditions set forth in this Agreement. If you do not accept these conditions, you must return the Supplement unused within 30 days of receipt.

All rights (including without limitation, copyrights, patents, and trade secrets) in the Supplement are and will remain the sole and exclusive property of Cengage Learning and/or its licensors. The Supplement is furnished by Cengage Learning on an "as is" basis without any warranties, express or implied. This Agreement will be governed by and construed pursuant to the laws of the State of New York, without regard to such State's conflict of law rules.

Thank you for your assistance in helping to safeguard the integrity of the content contained in this Supplement. We trust you find the Supplement a useful teaching tool.

# Contents

---

<b>Chapter 1:</b> Describing Data with Graphs.....	1
<b>Chapter 2:</b> Describing Data with Numerical Measures.....	30
<b>Chapter 3:</b> Describing Bivariate Data.....	68
<b>Chapter 4:</b> Probability.....	93
<b>Chapter 5:</b> Discrete Probability Distribution.....	121
<b>Chapter 6:</b> The Normal Probability Distribution.....	156
<b>Chapter 7:</b> Sampling Distributions.....	186
<b>Chapter 8:</b> Large-Sample Estimation.....	210
<b>Chapter 9:</b> Large-Sample Test of Hypotheses.....	240
<b>Chapter 10:</b> Inference from Small Samples.....	271
<b>Chapter 11:</b> The Analysis of Variance.....	324
<b>Chapter 12:</b> Linear Regression and Correlation.....	364
<b>Chapter 13:</b> Multiple Regression Analysis.....	415
<b>Chapter 14:</b> The Analysis of Categorical Data.....	439
<b>Chapter 15:</b> Nonparametric Statistics.....	469

# 1: Describing Data with Graphs

## Section 1.1

- 1.1.1 The experimental unit, the individual or object on which a variable is measured, is the student.
- 1.1.2 The experimental unit on which the number of errors is measured is the exam.
- 1.1.3 The experimental unit is the patient.
- 1.1.4 The experimental unit is the azalea plant.
- 1.1.5 The experimental unit is the car.
- 1.1.6 “Time to assemble” is a *quantitative* variable because a numerical quantity (1 hour, 1.5 hours, etc.) is measured.
- 1.1.7 “Number of students” is a *quantitative* variable because a numerical quantity (1, 2, etc.) is measured.
- 1.1.8 “Rating of a politician” is a *qualitative* variable since a quality (excellent, good, fair, poor) is measured.
- 1.1.9 “State of residence” is a *qualitative* variable since a quality (CA, MT, AL, etc.) is measured.
- 1.1.10 “Population” is a *discrete* variable because it can take on only integer values.
- 1.1.11 “Weight” is a *continuous* variable, taking on any values associated with an interval on the real line.
- 1.1.12 Number of claims is a *discrete* variable because it can take on only integer values.
- 1.1.13 “Number of consumers” is integer-valued and hence *discrete*.
- 1.1.14 “Number of boating accidents” is integer-valued and hence *discrete*.
- 1.1.15 “Time” is a *continuous* variable.
- 1.1.16 “Cost of a head of lettuce” is a *discrete* variable since money can be measured only in dollars and cents.
- 1.1.17 “Number of brothers and sisters” is integer-valued and hence *discrete*.
- 1.1.18 “Yield in bushels” is a *continuous* variable, taking on any values associated with an interval on the real line.
- 1.1.19 The statewide database contains a record of all drivers in the state of Michigan. The data collected represents the *population* of interest to the researcher.
- 1.1.20 The researcher is interested in the opinions of all citizens, not just the 1000 citizens that have been interviewed. The responses of these 1000 citizens represent a *sample*.
- 1.1.21 The researcher is interested in the weight gain of all animals that might be put on this diet, not just the twenty animals that have been observed. The responses of these twenty animals is a *sample*.
- 1.1.22 The data from the Internal Revenue Service contains the records of all wage earners in the United States. The data collected represents the *population* of interest to the researcher.
- 1.1.23
  - a The experimental unit, the item or object on which variables are measured, is the vehicle.
  - b Type (qualitative); make (qualitative); carpool or not? (qualitative); one-way commute distance (quantitative continuous); age of vehicle (quantitative continuous)
  - c Since five variables have been measured, this is *multivariate data*.
- 1.1.24
  - a The set of ages at death represents a population, because there have only been 38 different presidents in the United States history.
  - b The variable being measured is the continuous variable “age”.
  - c “Age” is a quantitative variable.

**1.1.25 a** The population of interest consists of voter opinions (for or against the candidate) at the time of the election for all persons voting in the election.

**b** Note that when a sample is taken (at some time prior or the election), we are not actually sampling from the population of interest. As time passes, voter opinions change. Hence, the population of voter opinions changes with time, and the sample may not be representative of the population of interest.

**1.1.26 a-b** The variable “survival time” is a quantitative continuous variable.

**c** The population of interest is the population of survival times for all patients having a particular type of cancer and having undergone a particular type of radiotherapy.

**d-e** Note that there is a problem with sampling in this situation. If we sample from all patients having cancer and radiotherapy, some may still be living and their survival time will not be measurable. Hence, we cannot sample directly from the population of interest, but must arrive at some reasonable alternate population from which to sample.

**1.1.27 a** The variable “reading score” is a quantitative variable, which is probably integer-valued and hence discrete.

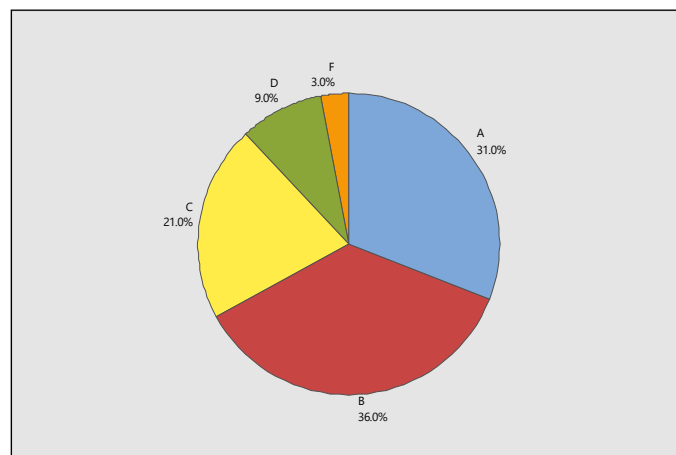
**b** The individual on which the variable is measured is the student.

**c** The population is hypothetical – it does not exist in fact – but consists of the reading scores for all students who could possibly be taught by this method.

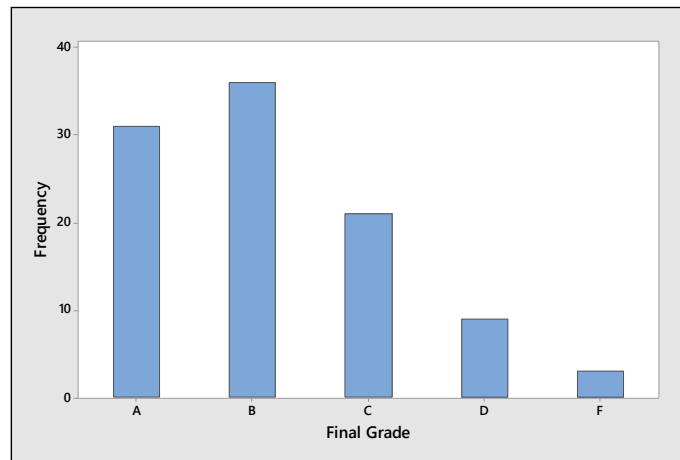
## Section 1.2

**1.2.1** The pie chart is constructed by partitioning the circle into five parts, according to the total contributed by each part. Since the total number of students is 100, the total number receiving a final grade of A represents  $31/100 = 0.31$  or 31% of the total. Thus, this category will be represented by a sector angle of  $0.31(360) = 111.6^\circ$ . The other sector angles are shown next, along with the pie chart.

Final Grade	Frequency	Fraction of Total	Sector Angle
A	31	.31	111.6
B	36	.36	129.6
C	21	.21	75.6
D	9	.09	32.4
F	3	.03	10.8

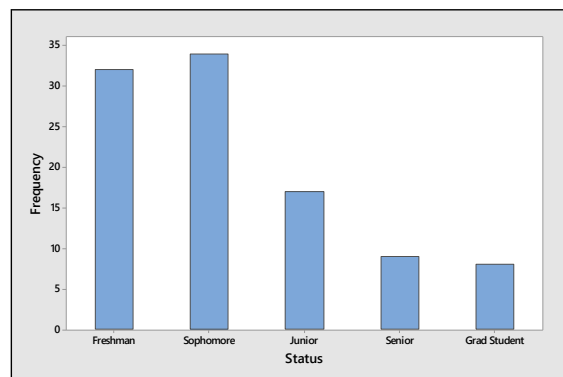
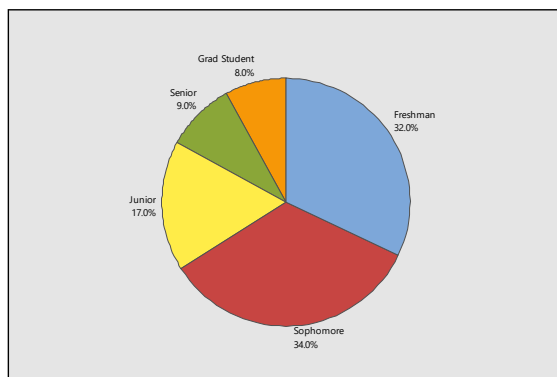


The bar chart represents each category as a bar with height equal to the frequency of occurrence of that category and is shown in the figure that follows.



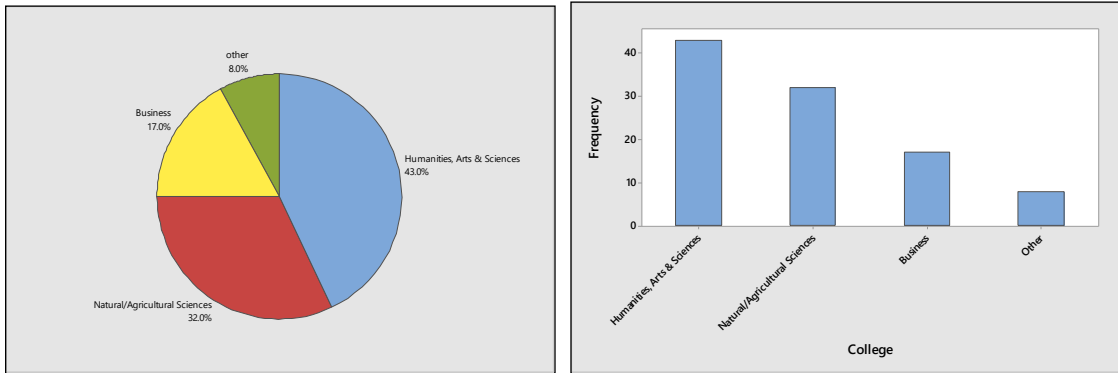
- 1.2.2** Construct a statistical table to summarize the data. The pie and bar charts are shown in the figures that follow.

Status	Frequency	Fraction of Total	Sector Angle
Freshman	32	.32	115.2
Sophomore	34	.34	122.4
Junior	17	.17	61.2
Senior	9	.09	32.4
Grad Student	8	.08	28.8



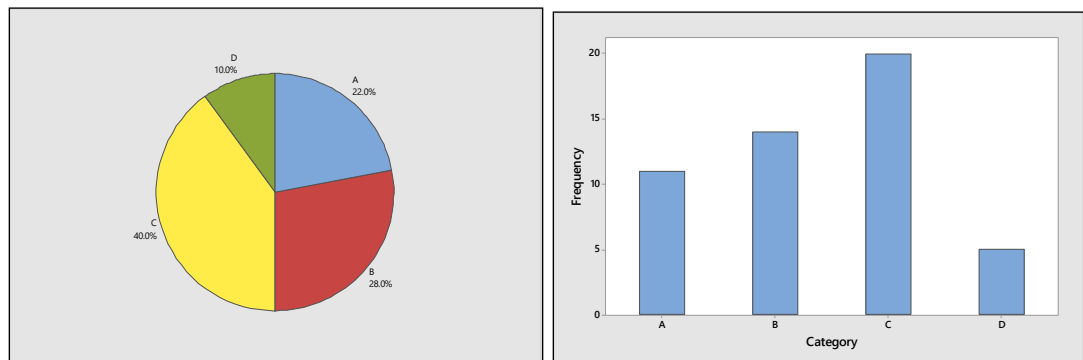
- 1.2.3** Construct a statistical table to summarize the data. The pie and bar charts are shown in the figures that follow.

Status	Frequency	Fraction of Total	Sector Angle
Humanities, Arts & Sciences	43	.43	154.8
Natural/Agricultural Sciences	32	.32	115.2
Business	17	.17	61.2
Other	8	.08	28.8



- 1.2.4 a** The pie chart is constructed by partitioning the circle into four parts, according to the total contributed by each part. Since the total number of people is 50, the total number in category A represents  $11/50 = 0.22$  or 22% of the total. Thus, this category will be represented by a sector angle of  $0.22(360) = 79.2^\circ$ . The other sector angles are shown below. The pie chart is shown in the figure that follows.

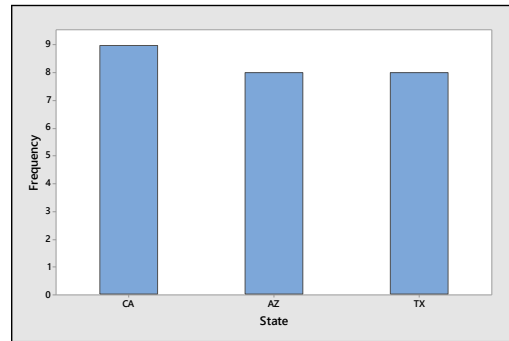
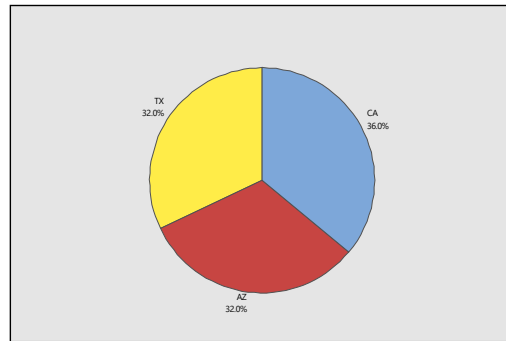
Category	Frequency	Fraction of Total	Sector Angle
A	11	.22	79.2
B	14	.28	100.8
C	20	.40	144.0
D	5	.10	36.0



- b** The bar chart represents each category as a bar with height equal to the frequency of occurrence of that category and is shown in the figure above.
- c** Yes, the shape will change depending on the order of presentation. The order is unimportant.
- d** The proportion of people in categories B, C, or D is found by summing the frequencies in those three categories, and dividing by  $n = 50$ . That is,  $(14 + 20 + 5)/50 = 0.78$ .
- e** Since there are 14 people in category B, there are  $50 - 14 = 36$  who are not, and the percentage is calculated as  $(36/50)100 = 72\%$ .

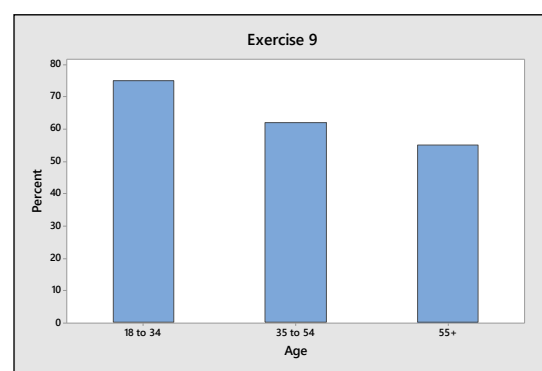
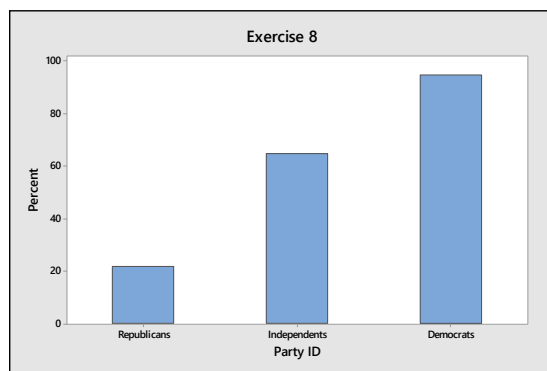
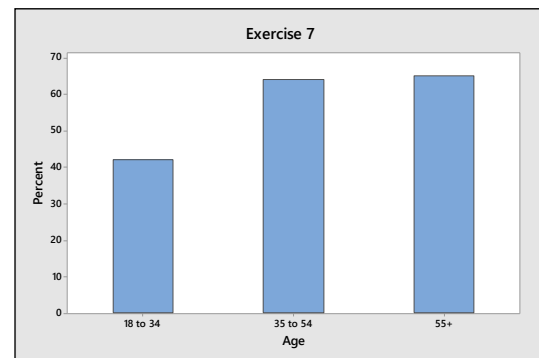
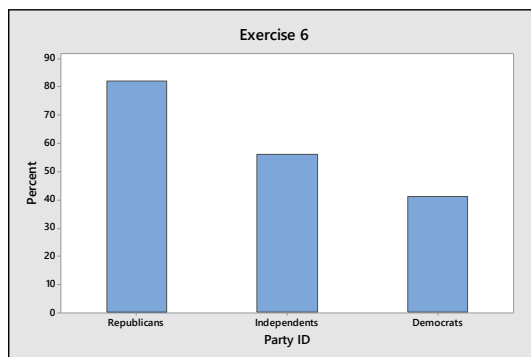
- 1.2.5 a-b** Construct a statistical table to summarize the data. The pie and bar charts are shown in the figures that follow.

State	Frequency	Fraction of Total	Sector Angle
CA	9	.36	129.6
AZ	8	.32	115.2
TX	8	.32	115.2



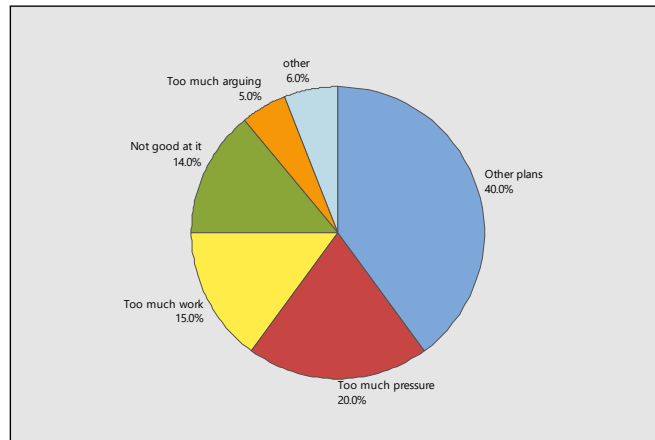
- c From the table or the chart, Texas produced  $8/25 = 0.32$  of the jeans.
- d The highest bar represents California, which produced the most pairs of jeans.
- e Since the bars and the sectors are almost equal in size, the three states produced roughly the same number of pairs of jeans.

**1.2.6-9** The bar charts represent each category as a bar with height equal to the frequency of occurrence of that category.



- 1.2.10** Answers will vary.
- 1.2.11 a** The percentages given in the exercise only add to 94%. We should add another category called “Other”, which will account for the other 6% of the responses.
- b** Either type of chart is appropriate. Since the data is already presented as percentages of the whole group, we choose to use a pie chart, shown in the figure that follows.

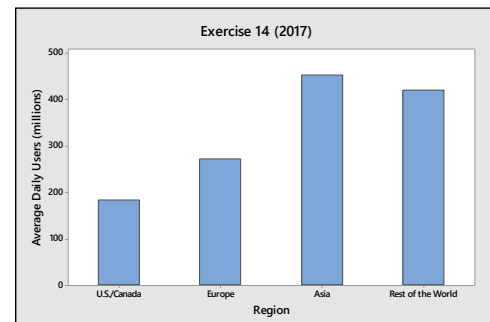
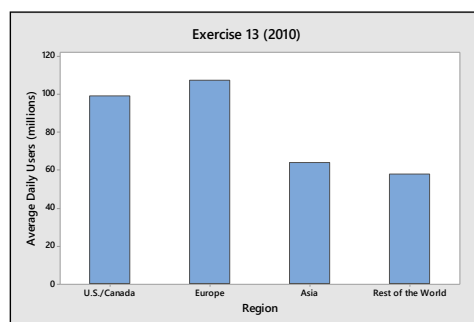
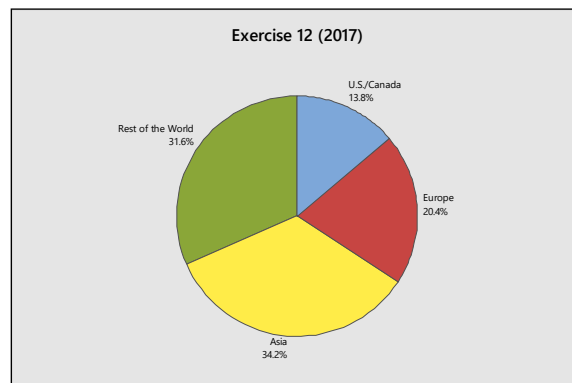




**c-d** Answers will vary.

**1.2.12-14** The percentages falling in each of the four categories in 2017 are shown next (in parentheses), and the pie chart for 2017 and bar charts for 2010 and 2017 follow.

Region	2010	2017
United States/Canada	99	183 (13.8%)
Europe	107	271 (20.4%)
Asia	64	453 (34.2%)
Rest of the World	58	419 (31.6%)
<b>Total</b>	<b>328</b>	<b>1326 (100%)</b>

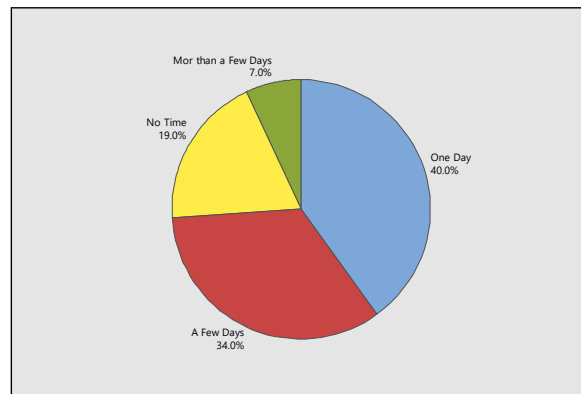


**1.2.15** Users in Asia and the rest of the world have increased more rapidly than those in the U.S., Canada or Europe over the seven-year period.

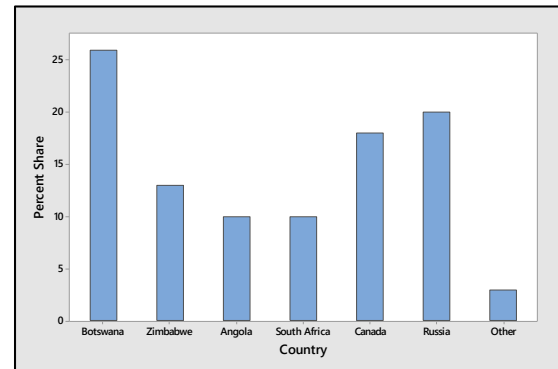
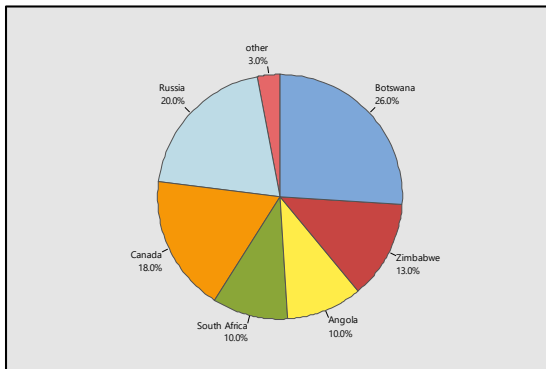
**1.2.16 a** The total percentage of responses given in the table is only  $(40 + 34 + 19)\% = 93\%$ . Hence there are 7% of the opinions not recorded, which should go into a category called “Other” or “More than a few days”.

**b** Yes. The bars are very close to the correct proportions.

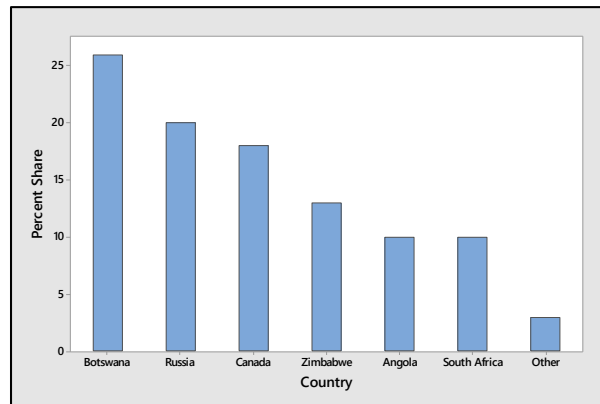
**c** Similar to previous exercises. The pie chart is shown next. The bar chart is probably more interesting to look at.



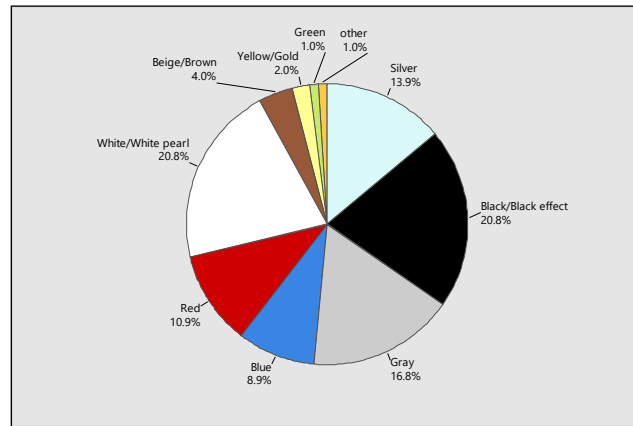
**1.2.17-18** Answers will vary from student to student. Since the graph gives a range of values for Zimbabwe’s share, we have chosen to use the 13% figure, and have used 3% in the “Other” category. The pie chart and bar charts are shown next.



**1.2.19-20** The Pareto chart is shown below. The Pareto chart is more effective than the bar chart or the pie chart.

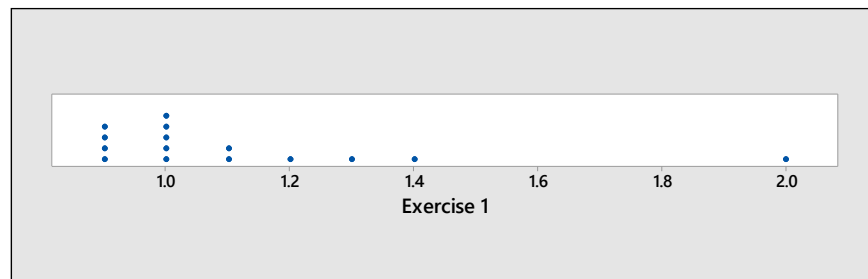


**1.2.21** The data should be displayed with either a bar chart or a pie chart. The pie chart is shown next.

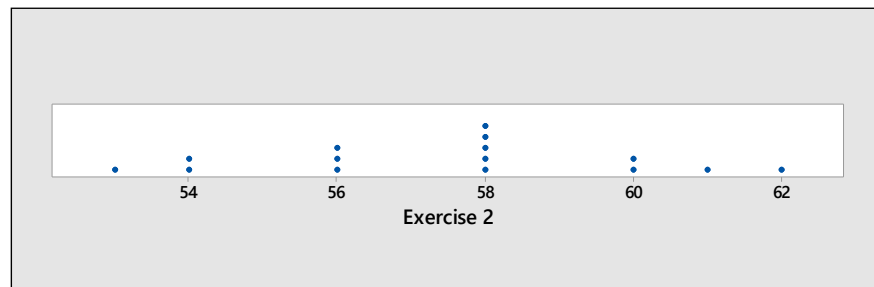


## Section 1.3

**1.3.1** The dotplot is shown next; the data is skewed right, with one outlier,  $x = 2.0$ .



**1.3.2** The dotplot is shown next; the data is relatively mound-shaped, with no outliers.



**1.3.3-5** The most obvious choice of a stem is to use the ones digit. The portion of the observation to the right of the ones digit constitutes the leaf. Observations are classified by row according to stem and also within each stem according to relative magnitude. The stem and leaf display is shown next.

```

1  6 8
2  1 2 5 5 5 7 8 8 9 9
3  1 1 4 5 5 6 6 6 7 7 7 7 8 9 9 9      leaf digit = 0.1
4  0 0 0 1 2 2 3 4 5 6 7 8 9 9 9      1 2 represents 1.2
5  1 1 6 6 7
6  1 2

```

- The stem and leaf display has a mound shaped distribution, with no outliers.
- From the stem and leaf display, the smallest observation is 1.6 (1 6).
- The eight and ninth largest observations are both 4.9 (4 9).

- 1.3.6** The stem is chosen as the ones digit, and the portion of the observation to the right of the ones digit is the leaf.

3 | 2 3 4 5 5 5 6 6 7 9 9 9 9

4 | 0 0 2 2 3 3 3 4 4 5 8      leaf digit = 0.1   1 2 represents 1.2

- 1.3.7-8** The stems are split, with the leaf digits 0 to 4 belonging to the first part of the stem and the leaf digits 5 to 9 belonging to the second. The stem and leaf display shown below improves the presentation of the data.

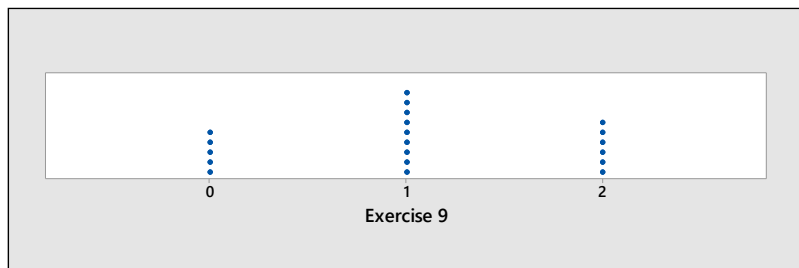
3 | 2 3 4

3 | 5 5 5 6 6 7 9 9 9 9      leaf digit = 0.1   1 2 represents 1.2

4 | 0 0 2 2 3 3 3 4 4

4 | 5 8

- 1.3.9** The scale is drawn on the horizontal axis and the measurements are represented by dots.



- 1.3.10** Since there is only one digit in each measurement, the ones digit must be the stem, and the leaf will be a zero digit for each measurement.

0 | 0 0 0 0 0

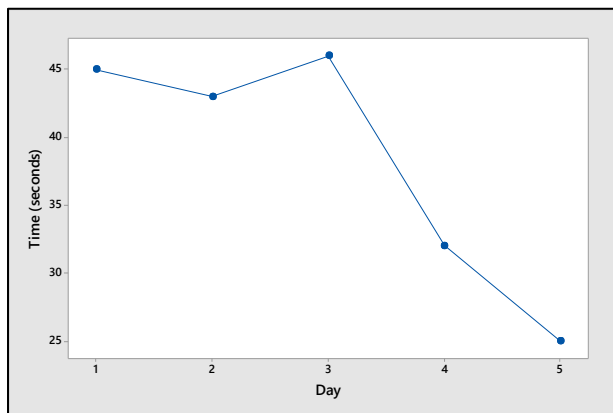
1 | 0 0 0 0 0 0 0 0 0

2 | 0 0 0 0 0 0

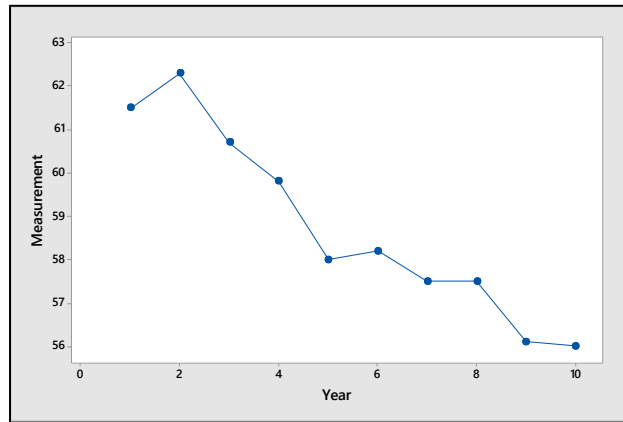
- 1.3.11** The distribution is relatively mound-shaped, with no outliers.

- 1.3.12** The two plots convey the same information if the stem and leaf plot is turned 90° and stretched to resemble the dotplot.

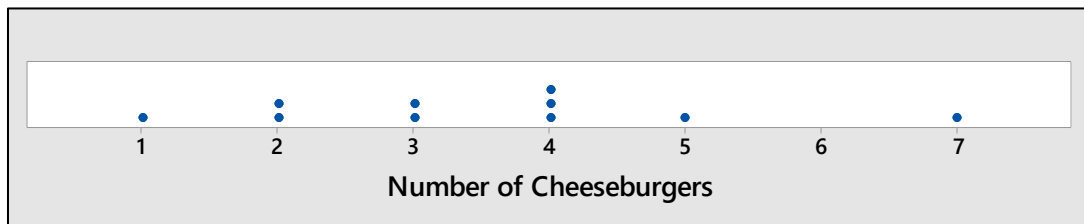
- 1.3.13** The line chart plots “day” on the horizontal axis and “time” on the vertical axis. The line chart shown next reveals that learning is taking place, since the time decreases each successive day.



- 1.3.14** The line graph is shown next. Notice the change in  $y$  as  $x$  increases. The measurements are decreasing over time.



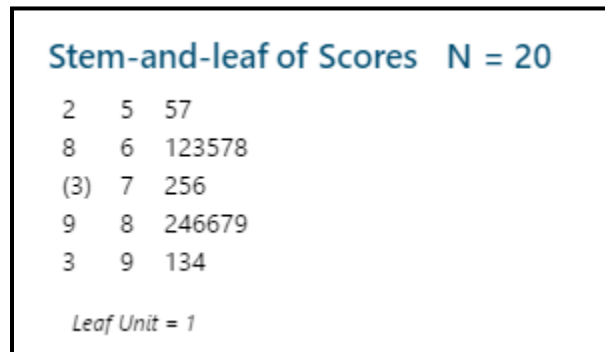
**1.3.15** The dotplot is shown next.



**a** The distribution is somewhat mound-shaped (as much as a small set can be); there are no outliers.

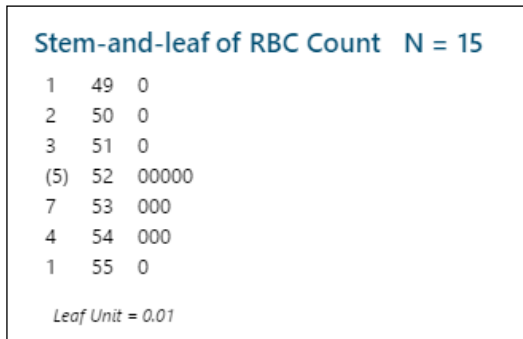
**b**  $2/10 = 0.2$

**1.3.16 a** The test scores are graphed using a stem and leaf plot generated by *Minitab*.



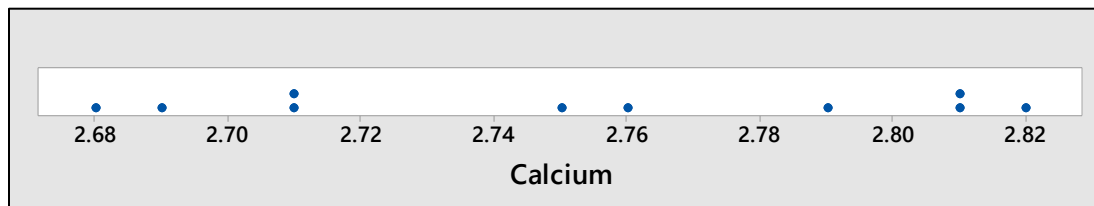
**b-c** The distribution is not mound-shaped, but is rather has two peaks centered around the scores 65 and 85. This might indicate that the students are divided into two groups – those who understand the material and do well on exams, and those who do not have a thorough command of the material.

**1.3.17 a** We choose a stem and leaf plot, using the ones and tenths place as the stem, and a zero digit as the leaf. The *Minitab* printout is shown next.



- b** The data set is relatively mound-shaped, centered at 5.2.
- c** The value  $x = 5.7$  does not fall within the range of the other cell counts, and would be considered somewhat unusual.

**1.3.18 a-b** The dotplot and the stem and leaf plot are drawn using *Minitab*.



**Stem-and-leaf of Calcium N = 10**

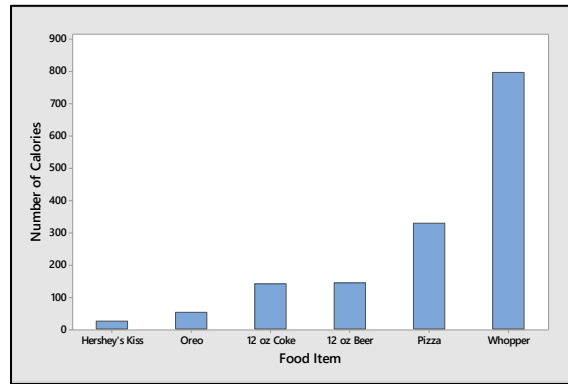
2	26	89
(5)	27	11569
3	28	112

Leaf Unit = 0.01

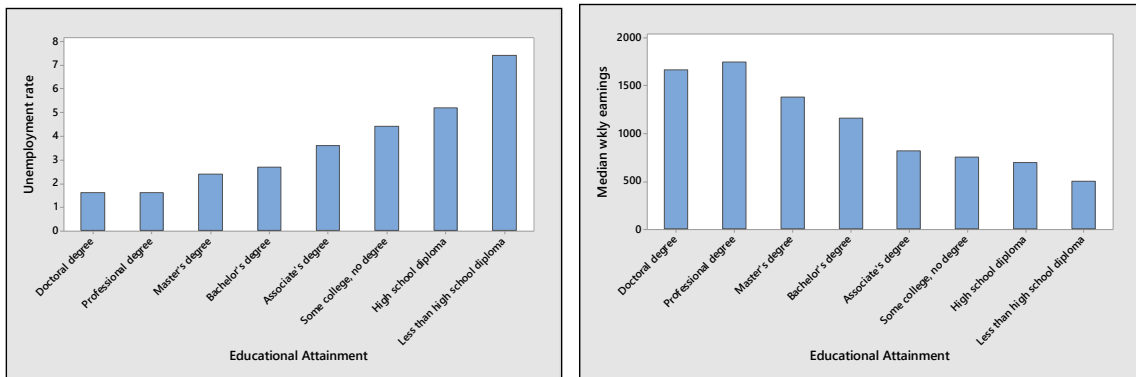
- c** The measurements all seem to be within the same range of variability. There do not appear to be any outliers.
- 1.3.19 a** Stem and leaf displays may vary from student to student. The most obvious choice is to use the tens digit as the stem and the ones digit as the leaf.

7		8 9
8		0 1 7
9		0 1 2 4 4 5 6 6 6 8 8
10		1 7 9
11		2

- b** The display is fairly mound-shaped, with a large peak in the middle.
- 1.3.20 a** The sizes and volumes of the food items do increase as the number of calories increase, but not in the correct proportion to the actual calories. The differences in calorie content are not accurately portrayed in the graph.
- b** The bar graph which accurately portrays the number of calories in the six food items is shown next.

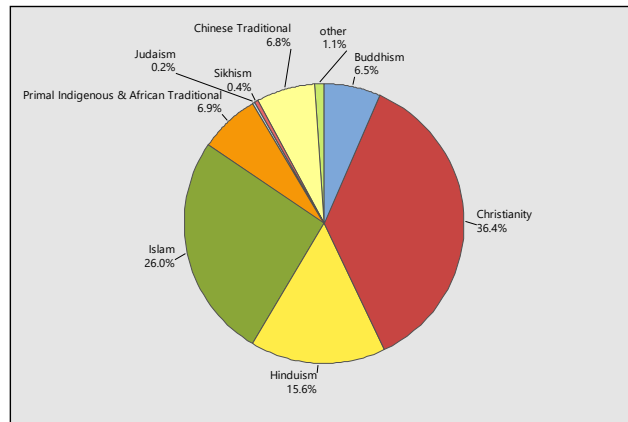


**1.3.21 a-b** The bar charts for the median weekly earnings and unemployment rates for eight different levels of education are shown next.

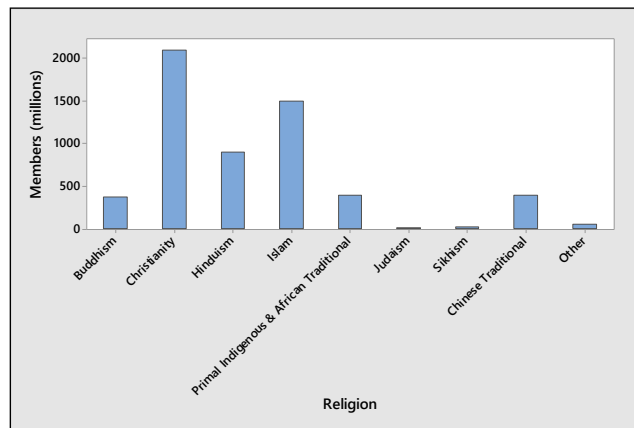


**c** The unemployment rate drops and the median weekly earnings rise as the level of educational attainment increases.

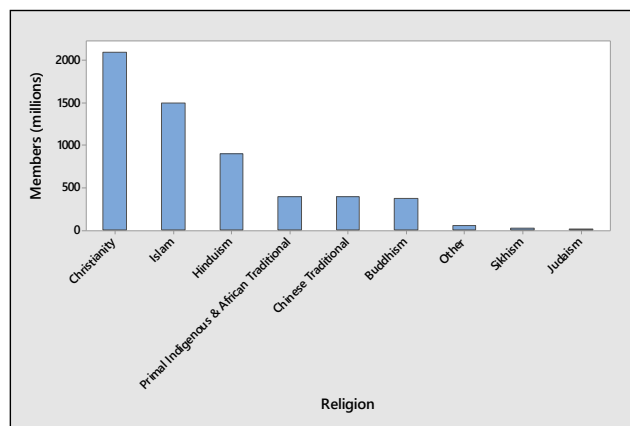
**1.3.22 a** Similar to previous exercises. The pie chart is shown next.



**b** The bar chart is shown next.



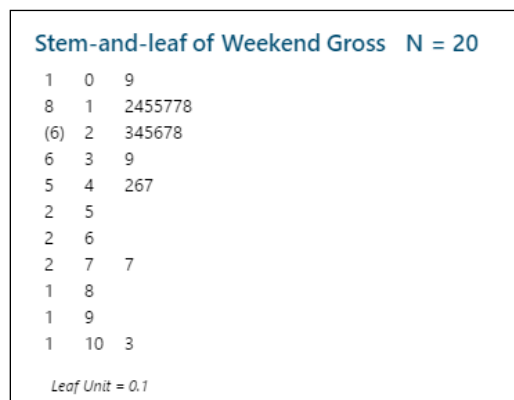
- c** The Pareto chart is a bar chart with the heights of the bars ordered from large to small. This display is more effective than the pie chart.



- 1.3.23 a** The distribution is skewed to the right, with a several unusually large measurements. The five states marked as HI are California, New Jersey, New York and Pennsylvania.

**b** Three of the four states are quite large in area, which might explain the large number of hazardous waste sites. However, New Jersey is relatively small, and other large states do not have unusually large number of waste sites. The pattern is not clear.

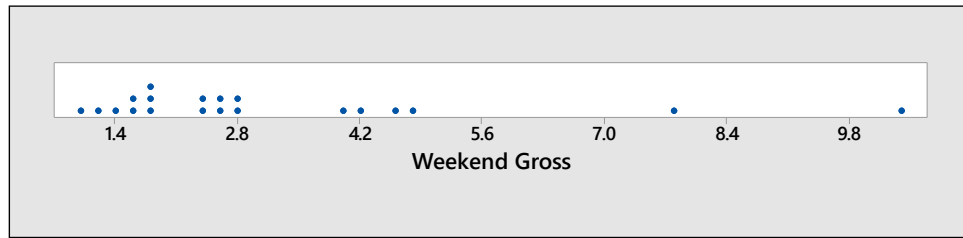
- 1.3.24 a**



The distribution is skewed to the right, with two outliers.

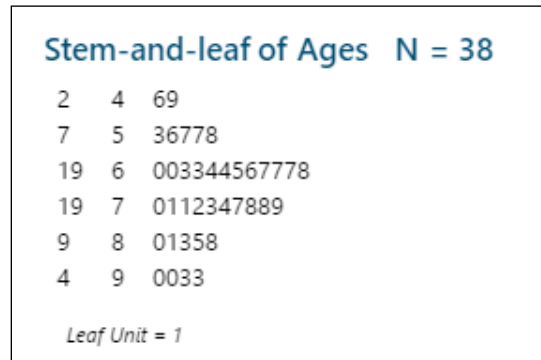
**b** The dotplot is shown next. It conveys nearly the same information, but the stem-and-leaf plot may be more informative.





**1.3.25 a** Answers will vary.

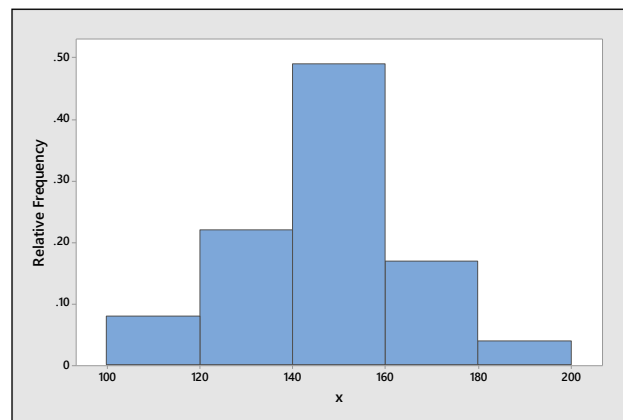
**b** The stem and leaf plot is constructed using the tens place as the stem and the ones place as the leaf. Notice that the distribution is roughly mound-shaped.



**c-d** Three of the five youngest presidents – Kennedy, Lincoln and Garfield – were assassinated while in office. This would explain the fact that their ages at death were in the lower tail of the distribution.

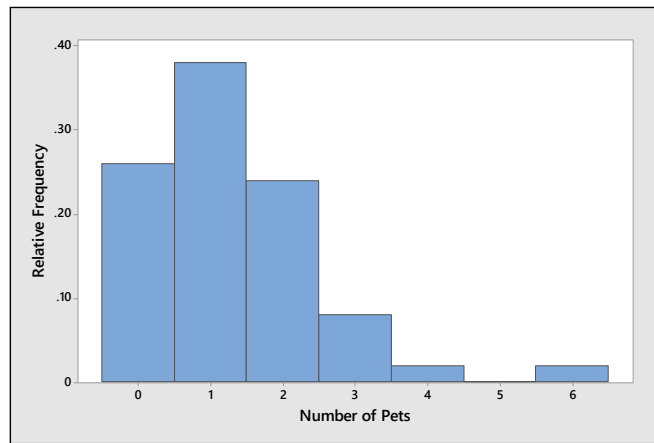
## Section 1.4

**1.4.1** The relative frequency histogram displays the relative frequency as the height of the bar over the appropriate class interval and is shown next. The distribution is relatively mound-shaped.



**1.4.2** Since the variable of interest can only take integer values, the classes can be chosen as the values 0, 1, 2, 3, 4, 5 and 6. The table containing the classes, their corresponding frequencies and their relative frequencies and the relative frequency histogram are shown next. The distribution is skewed to the right.

Number of Household Pets	Frequency	Relative Frequency
0	13	$13/50 = .26$
1	19	$19/50 = .38$
2	12	$12/50 = .24$
3	4	$4/50 = .08$
4	1	$1/50 = .02$
5	0	$0/50 = .00$
6	1	$1/50 = .02$
Total	50	$50/50 = 1.00$



**1.4.3-8** The proportion of measurements falling in each interval is equal to the sum of the heights of the bars over that interval. Remember that the lower class boundary is included, but not the upper class boundary.

3.  $.20 + .40 + .15 = .75$

4.  $.05 + .15 + .20 = .40$

5.  $.05$

6.  $.40 + .15 = .55$

7.  $.15$

8.  $.05 + .15 + .20 = .40$

**1.4.9** Answers will vary. The range of the data is  $110 - 10 = 90$  and we need to use seven classes. Calculate  $90/7 = 12.86$  which we choose to round up to 15. Convenient class boundaries are created, starting at 10: 10 to < 25, 25 to < 40, ..., 100 to < 115.

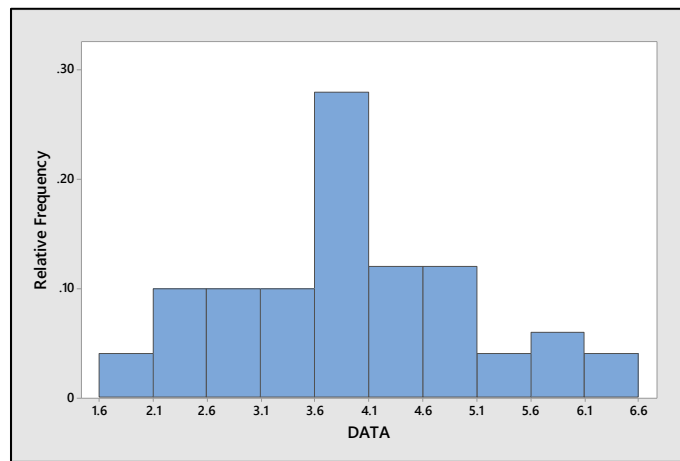
**1.4.10** Answers will vary. The range of the data is  $76.8 - 25.5 = 51.3$  and we need to use six classes. Calculate  $51.3/6 = 8.55$  which we choose to round up to 9. Convenient class boundaries are created, starting at 25: 25 to < 34, 34 to < 43, ..., 70 to < 79.

**1.4.11** Answers will vary. The range of the data is  $1.73 - .31 = 1.42$  and we need to use ten classes. Calculate  $1.42/10 = .142$  which we choose to round up to .15. Convenient class boundaries are created, starting at .30: .30 to < .45, .45 to < .60, ..., 1.65 to < 1.80.

**1.4.12** Answers will vary. The range of the data is  $192 - 0 = 192$  and we need to use eight classes. Calculate  $192/8 = 24$  which we choose to round up to 25. Convenient class boundaries are created, starting at 0: 0 to < 25, 25 to < 50, ..., 175 to < 200.

**1.4.13-16** The table containing the classes, their corresponding frequencies and their relative frequencies and the relative frequency histogram are shown next.

Class i	Class Boundaries	Tally	$f_i$	Relative frequency, $f_i/n$
1	1.6 to < 2.1	11	2	.04
2	2.1 to < 2.6	11111	5	.10
3	2.6 to < 3.1	11111	5	.10
4	3.1 to < 3.6	11111	5	.10
5	3.6 to < 4.1	11111 11111 1111	14	.28
6	4.1 to < 4.6	11111 11	7	.14
7	4.6 to < 5.1	11111	5	.10
8	5.1 to < 5.6	11	2	.04
9	5.6 to < 6.1	111	3	.06
10	6.1 to < 6.6	11	2	.04



13. The distribution is roughly mound-shaped.

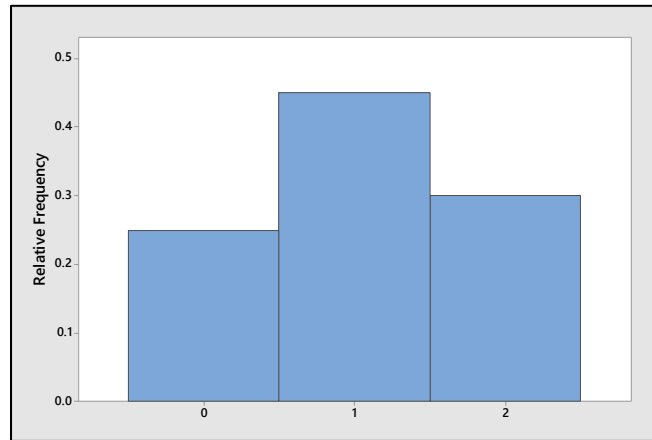
14. The fraction less than 5.1 is that fraction lying in classes 1-7, or  $(2+5+\cdots+7+5)/50 = 43/50 = 0.86$ .

15. The fraction larger than 3.6 lies in classes 5-10, or  $(14+7+\cdots+3+2)/50 = 33/50 = 0.66$ .

16. The fraction from 2.6 up to but not including 4.6 lies in classes 3-6, or  $(5+5+14+7)/50 = 31/50 = 0.62$ .

1.4.17-20 Since the variable of interest can only take the values 0, 1, or 2, the classes can be chosen as the integer values 0, 1, and 2. The table shows the classes, their corresponding frequencies and their relative frequencies. The relative frequency histogram follows the table.

Value	Frequency	Relative Frequency
0	5	.25
1	9	.45
2	6	.30



17. Using the table above, the proportion of measurements greater than 1 is the same as the proportion of “2”s, or 0.30.

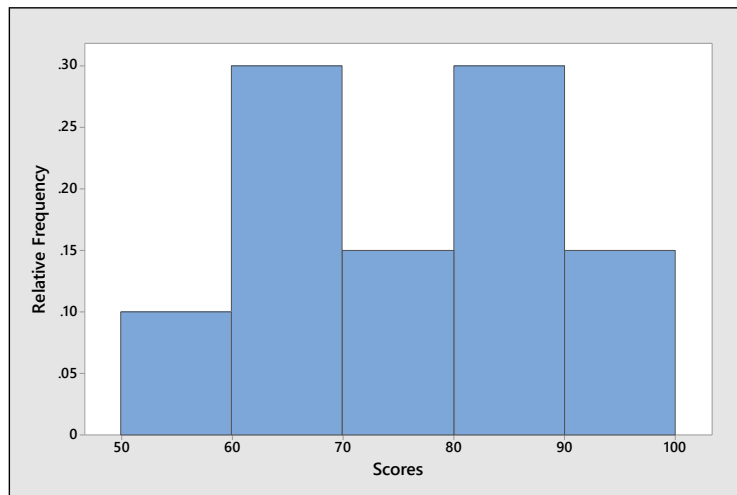
18. The proportion of measurements less than 2 is the same as the proportion of “0”s and “1”s, or  $0.25 + 0.45 = .70$ .

19. The probability of selecting a “2” in a random selection from these twenty measurements is  $6/20 = .30$ .

20. There are no outliers in this relatively symmetric, mound-shaped distribution.

1.4.21-23 Answers will vary. The range of the data is  $94 - 55 = 39$  and we choose to use 5 classes. Calculate  $39/5 = 7.8$  which we choose to round up to 10. Convenient class boundaries are created, starting at 50 and the table and relative frequency histogram are created.

Class Boundaries	Frequency	Relative Frequency
50 to < 60	2	.10
60 to < 70	6	.30
70 to < 80	3	.15
80 to < 90	6	.30
90 to < 100	3	.15



21. The distribution has two peaks at about 65 and 85. Depending on the way in which the student constructs the histogram, these peaks may or may not be clearly seen.

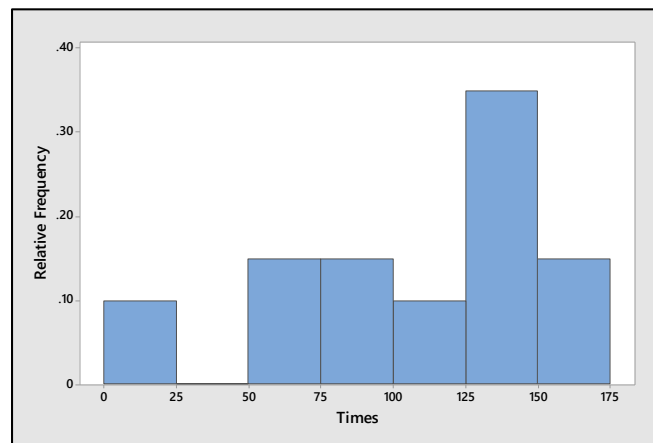
**22.** The shape is unusual. It might indicate that the students are divided into two groups – those who understand the material and do well on exams, and those who do not have a thorough command of the material.

**23.** The shapes are roughly the same, but this may not be the case if the student constructs the histogram using different class boundaries.

**1.4.24 a** There are a few extremely small numbers, indicating that the distribution is probably skewed to the left.

**b** The range of the data  $165 - 8 = 157$ . We choose to use seven class intervals of length 25, with subintervals  $0 < 25$ ,  $25 < 50$ ,  $50 < 75$ , and so on. The tally and relative frequency histogram are shown next.

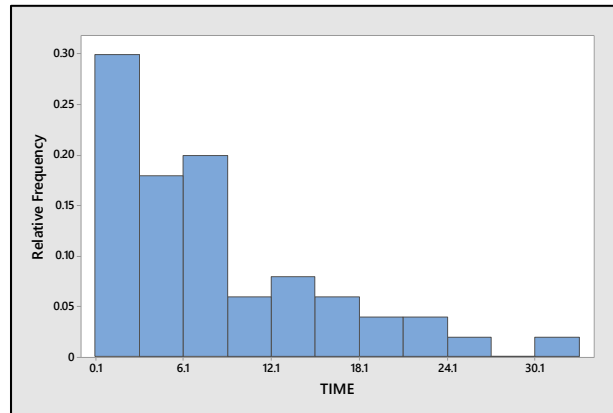
Class i	Class Boundaries	Tally	$f_i$	Relative frequency, $f_i/n$
1	0 to < 25	11	2	2/20
2	25 to < 50		0	0/20
3	50 to < 75	111	3	3/20
4	75 to < 100	111	3	3/20
5	100 to < 125	11	2	2/20
6	125 to < 150	11111 11	7	7/20
7	150 to < 175	111	3	3/20



**c** The distribution is indeed skewed left with two possible outliers:  $x = 8$  and  $x = 11$ .

**1.4.25 a** The range of the data  $32.3 - 0.2 = 32.1$ . We choose to use eleven class intervals of length 3 ( $32.1/11 = 2.9$ , which when rounded to the next largest integer is 3). The subintervals  $0.1 < 3.1$ ,  $3.1 < 6.1$ ,  $6.1 < 9.1$ , and so on, are convenient and the tally and relative frequency histogram are shown next.

Class i	Class Boundaries	Tally	$f_i$	Relative frequency, $f_i/n$
1	0.1 to < 3.1	11111 11111 11111	15	15/50
2	3.1 to < 6.1	11111 1111	9	9/50
3	6.1 to < 9.1	11111 11111	10	10/50
4	9.1 to < 12.1	111	3	3/50
5	12.1 to < 15.1	1111	4	4/50
6	15.1 to < 18.1	111	3	3/50
7	18.1 to < 21.1	11	2	2/50
8	21.1 to < 24.1	11	2	2/50
9	24.1 to < 27.1	1	1	1/50
10	27.1 to < 30.1		0	0/50
11	30.1 to < 33.1	1	1	1/50

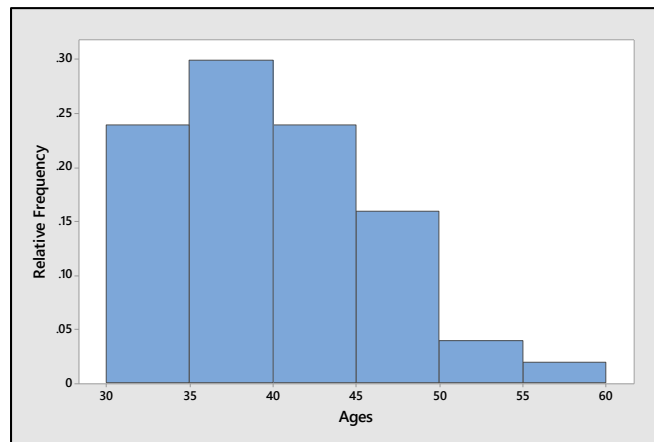


**b** The data is skewed to the right, with a few unusually large measurements.

**c** Looking at the data, we see that 36 patients had a disease recurrence within 10 months. Therefore, the fraction of recurrence times less than or equal to 10 is  $36/50 = 0.72$ .

**1.4.26 a** We use class intervals of length 5, beginning with the subinterval 30 to < 35. The tally and the relative frequency histogram are shown next.

Class i	Class Boundaries	Tally	$f_i$	Relative frequency, $f_i/n$
1	30 to < 35	11111 11111 11	12	12/50
2	35 to < 40	11111 11111 11111	15	15/50
3	40 to < 45	11111 11111 11	12	12/50
4	45 to < 50	11111 111	8	8/50
5	50 to < 55	11	2	2/50
6	55 to < 60	1	1	1/50



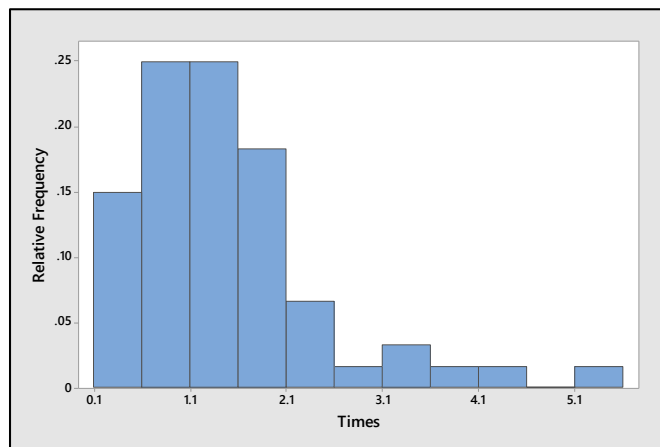
**b** Use the table or the relative frequency histogram. The proportion of children in the interval 35 to < 45 is  $(15 + 12)/50 = .54$ .

**c** The proportion of children aged less than 50 months is  $(12 + 15 + 12 + 8)/50 = .94$ .

**1.4.27 a** The data ranges from .2 to 5.2, or 5.0 units. Since the number of class intervals should be between five and twelve, we choose to use eleven class intervals, with each class interval having length 0.50 ( $5.0/11 = .45$ , which, rounded to the nearest convenient fraction, is .50). We must now select interval boundaries such that no measurement can fall on a boundary point. The subintervals .1 to < .6, .6 to < 1.1, and so on, are convenient and a tally is constructed.

Class i	Class Boundaries	Tally	$f_i$	Relative frequency, $f_i/n$
1	0.1 to < 0.6	11111 11111	10	.167
2	0.6 to < 1.1	11111 11111 11111	15	.250
3	1.1 to < 1.6	11111 11111 11111	15	.250
4	1.6 to < 2.1	11111 11111	10	.167
5	2.1 to < 2.6	1111	4	.067
6	2.6 to < 3.1	1	1	.017
7	3.1 to < 3.6	11	2	.033
8	3.6 to < 4.1	1	1	.017
9	4.1 to < 4.6	1	1	.017
10	4.6 to < 5.1		0	.000
11	5.1 to < 5.6	1	1	.017

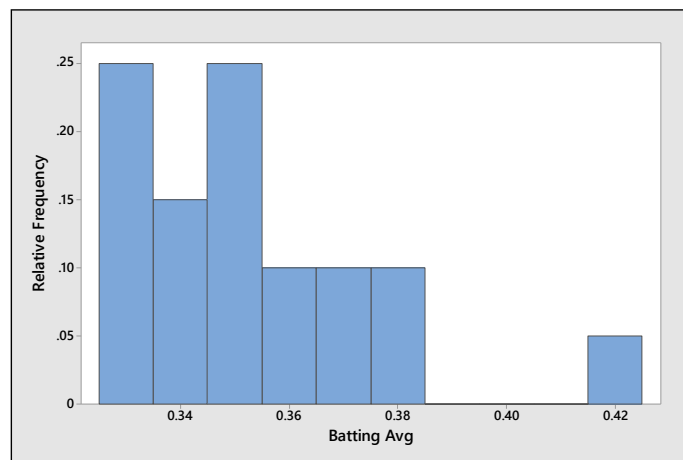
The relative frequency histogram is shown next.



**b** The distribution is skewed to the right, with several unusually large observations.

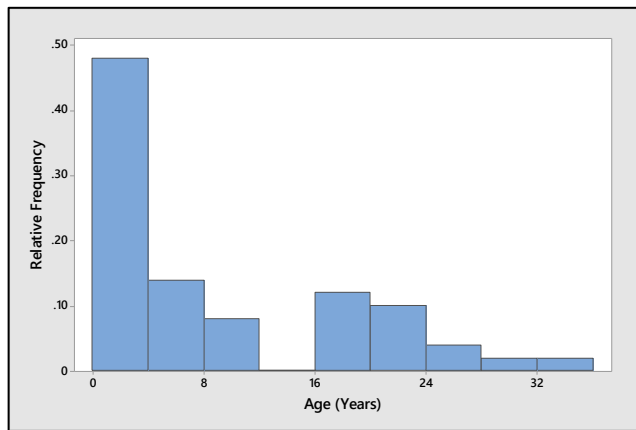
**c** For some reason, one person had to wait 5.2 minutes. Perhaps the supermarket was understaffed that day, or there may have been an unusually large number of customers in the store.

**1.4.28 a** Histograms will vary from student to student. A typical histogram generated by *Minitab* is shown next.

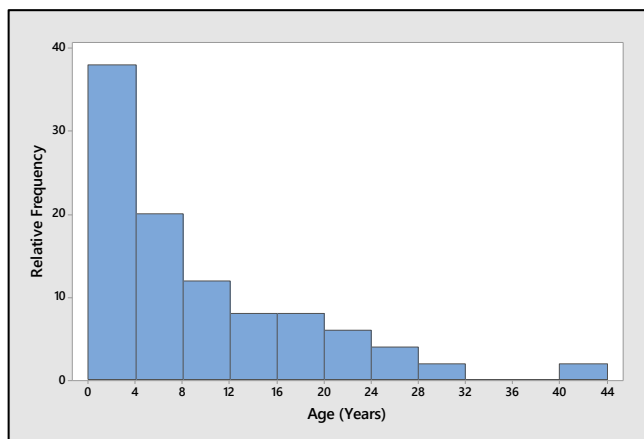


**b** Since 1 of the 20 players has an average above 0.400, the chance is 1 out of 20 or  $1/20 = 0.05$ .

**1.4.29 a-b** Answers will vary from student to student. The students should notice that the distribution is skewed to the right with a few pennies being unusually old. A typical histogram is shown next.



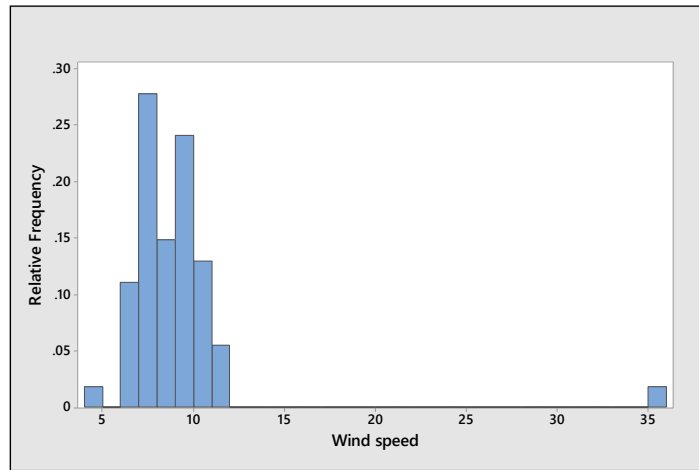
- 1.4.30 a** Answers will vary from student to student. A typical histogram is shown next. It looks very similar to the histogram from Exercise 1.4.29.



- b** There is one outlier,  $x = 41$ .
- 1.4.31 a** Answers will vary from student to student. The relative frequency histogram below was constructed using classes of length 1.0 starting at  $x = 4$ . The value  $x = 35.1$  is not shown in the table but appears on the graph shown next.

Class i	Class Boundaries	Tally	$f_i$	Relative frequency, $f_i/n$
1	4.0 to < 5.0	1	1	1/54
2	5.0 to < 6.0	0	0	0/54
3	6.0 to < 7.0	11111 1	6	6/54
4	7.0 to < 8.0	11111 11111 11111	15	15/54
5	8.0 to < 9.0	11111 111	8	8/54
6	9.0 to < 10.0	11111 11111 111	13	13/54
7	10.0 to < 11.0	11111 11	7	7/54
8	11.0 to < 12.0	111	3	3/54





**b** Since Mt. Washington is a very mountainous area, it is not unusual that the average wind speed would be very high.

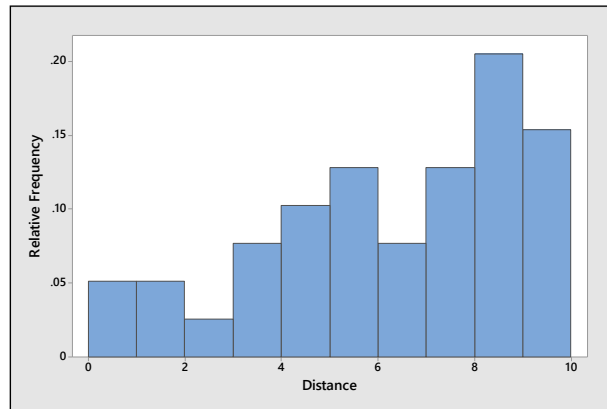
**c** The value  $x = 9.9$  does not lie far from the center of the distribution (excluding  $x = 35.1$ ). It would not be considered unusually high.

**1.4.32 a-b** The data is somewhat mound-shaped, but it appears to have two local peaks – high points from which the frequencies drop off on either side.

**c** Since these are student heights, the data can be divided into two groups – heights of males and heights of females. Both groups will have an approximate mound-shape, but the average female height will be lower than the average male height. When the two groups are combined into one data set, it causes a “mixture” of two mound-shaped distributions and produces the two peaks seen in the histogram.

**1.4.33 a** The relative frequency histogram below was constructed using classes of length 1.0 starting at  $x = 0.0$ .

Class i	Class Boundaries	Tally	$f_i$	Relative frequency, $f_i/n$
1	0.0 to < 1.0	11	2	$2/39$
2	1.0 to < 2.0	11	2	$2/39$
3	2.0 to < 3.0	1	1	$1/39$
4	3.0 to < 4.0	111	3	$3/39$
5	4.0 to < 5.0	111	4	$4/39$
6	5.0 to < 6.0	11111	5	$5/39$
7	6.0 to < 7.0	111	3	$3/39$
8	7.0 to < 8.0	11111	5	$5/39$
9	8.0 to < 9.0	11111 111	8	$8/39$
10	9.0 to < 10.0	11111 1	6	$6/39$

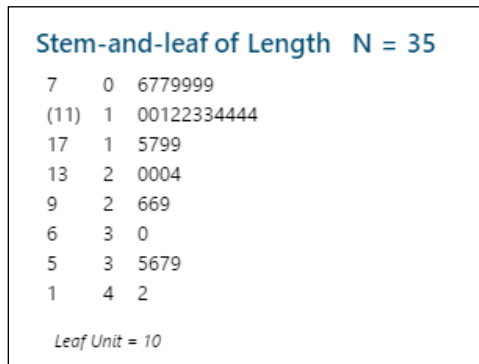


- a** The distribution is skewed to the left, with slightly higher frequency in the first two classes (within two miles of UCR).
- b** As the distance from UCR increases, each successive area increases in size, thus allowing for more Starbucks stores in that region.

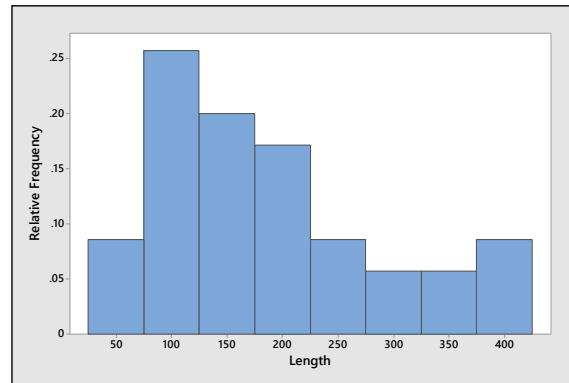
## Reviewing What You've Learned

- 1.R.1**
- a** “Ethnic origin” is a *qualitative variable* since a quality (ethnic origin) is measured.
  - b** “Score” is a *quantitative variable* since a numerical quantity (0-100) is measured.
  - c** “Type of establishment” is a *qualitative variable* since a category (Carl’s Jr., McDonald’s or Burger King) is measured.
  - d** “Mercury concentration” is a *quantitative variable* since a numerical quantity is measured.
- 1.R.2** To determine whether a distribution is likely to be skewed, look for the likelihood of observing extremely large or extremely small values of the variable of interest.
- a** The price of an 8-oz can of peas is not likely to contain unusually large or small values.
  - b** Not likely to be skewed.
  - c** If a package is dropped, it is likely that all the shells will be broken. Hence, a few large number of broken shells is possible. The distribution will be skewed.
  - d** If an animal has one tick, he is likely to have more than one. There will be some “0”s with uninfected rabbits, and then a larger number of large values. The distribution will not be symmetric.
- 1.R.3**
- a** The length of time between arrivals at an outpatient clinic is a continuous random variable, since it can be any of the infinite number of positive real values.
  - b** The time required to finish an examination is a continuous random variable as was the random variable described in part **a**.
  - c** Weight is continuous, taking any positive real value.
  - d** Body temperature is continuous, taking any real value.
  - e** Number of people is discrete, taking the values 0, 1, 2, ...
- 1.R.4**
- a** Number of properties is discrete, taking the values 0, 1, 2, ...
  - b** Depth is continuous, taking any non-negative real value.
  - c** Length of time is continuous, taking any non-negative real value.
  - d** Number of aircraft is discrete.

1.R.5 a



b

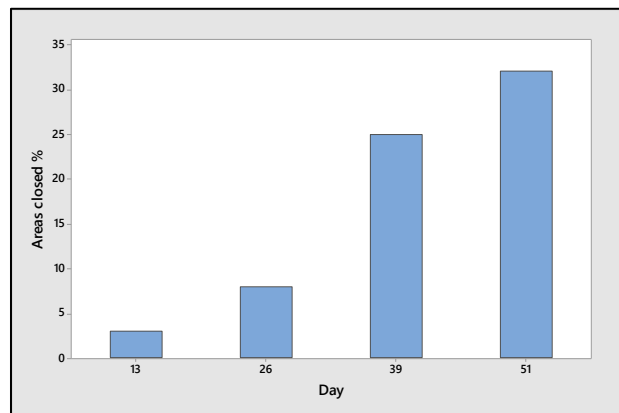
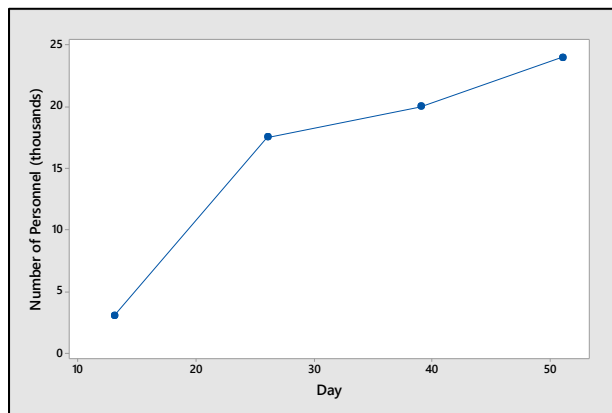


c These data are skewed right.

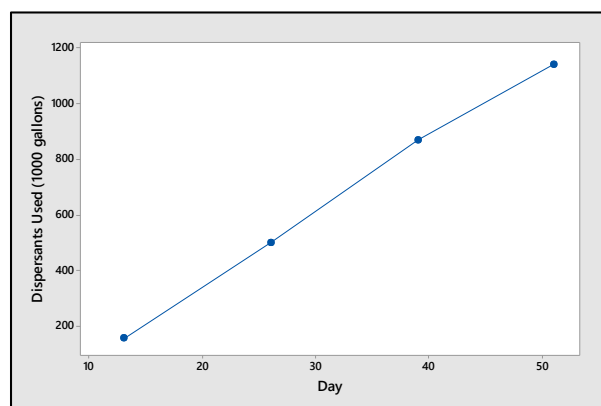
1.R.6 a The five quantitative variables are measured over time two months after the oil spill. Some sort of comparative bar charts (side-by-side or stacked) or a line chart should be used.

b As the time after the spill increases, the values of all five variables increase.

c-d The line chart for *number of personnel* and the bar chart for *fishing areas closed* are shown next.

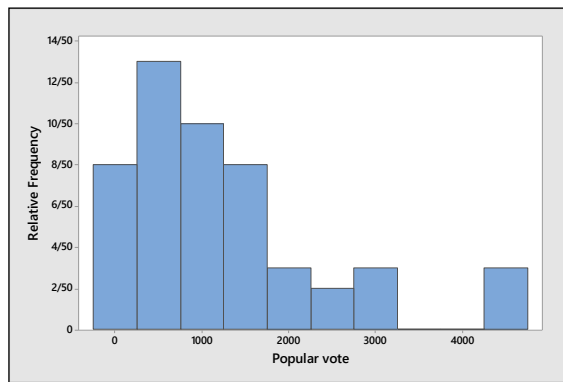


e The line chart for *amount of dispersants* is shown next. There appears to be a straight-line trend.

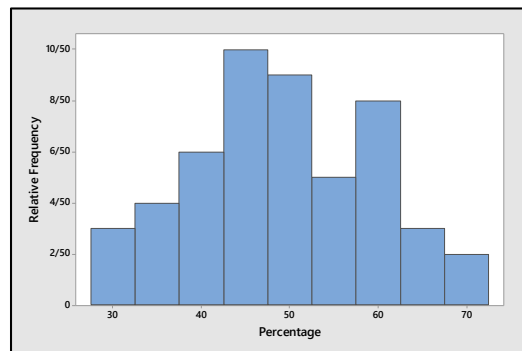


1.R.7 a The popular vote within each state should vary depending on the size of the state. Since there are several very large states (in population) in the United States, the distribution should be skewed to the right.

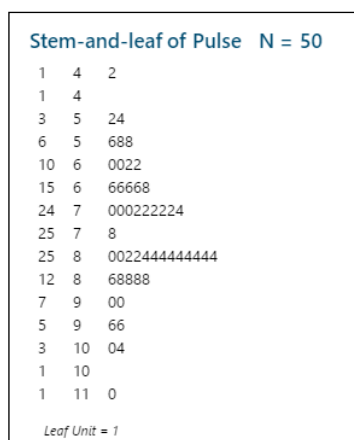
b-c Histograms will vary from student to student but should resemble the histogram generated by *Minitab* in the next figure. The distribution is indeed skewed to the right, with three “outliers” – California, Florida and Texas.



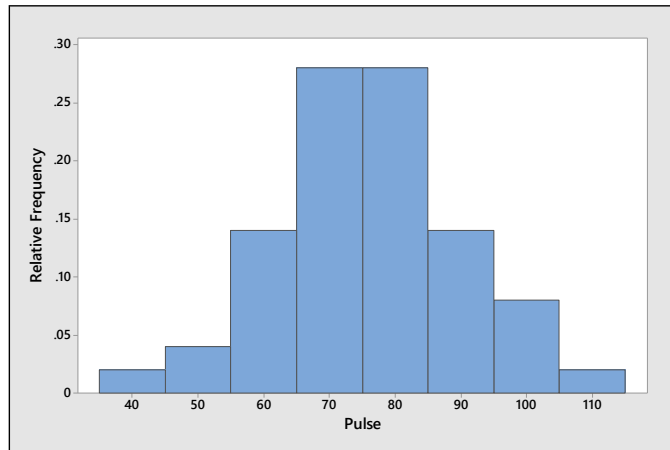
- 1.R.8 a-b** Once the size of the state is removed by calculating the percentage of the popular vote, the unusually large values in the Exercise 7 data set will disappear, and each state will be measured on an equal basis. Student histograms should resemble the histogram shown next. Notice the relatively mound-shape and the lack of any outliers.



- 1.R.9 a-b** Popular vote is skewed to the right while the percentage of popular vote is roughly mound-shaped. While the distribution of popular vote has outliers (California, Florida and Texas), there are no outliers in the distribution of percentage of popular vote. When the stem and leaf plots are turned 90°, the shapes are very similar to the histograms.
- c** Once the size of the state is removed by calculating the percentage of the popular vote, the unusually large values in the set of “popular votes” will disappear, and each state will be measured on an equal basis. The data then distribute themselves in a mound-shape around the average percentage of the popular vote.
- 1.R.10 a** The measurements are obtained by counting the number of beats for 30 seconds, and then multiplying by 2. Thus, the measurements should all be even numbers.
- b** The stem and leaf plot is shown next.

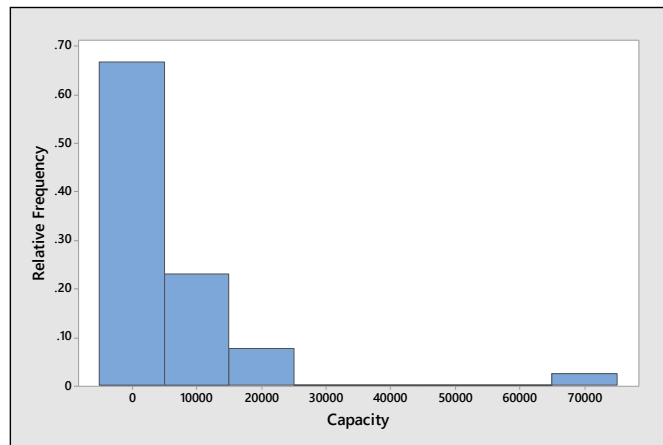


**c** Answers will vary. A typical histogram, generated by *Minitab*, is shown next.



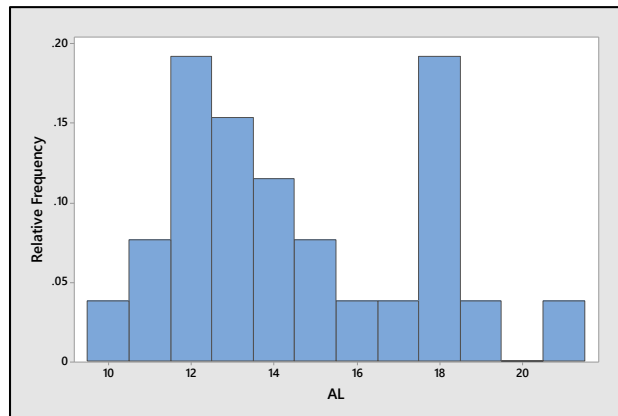
**d** The distribution of pulse rates is mound-shaped and relatively symmetric around a central location of 75 beats per minute. There are no outliers.

**1.R.11 a-b** Answers will vary from student to student. A typical histogram is shown next—the distribution is skewed to the right, with an extreme outlier (Texas).

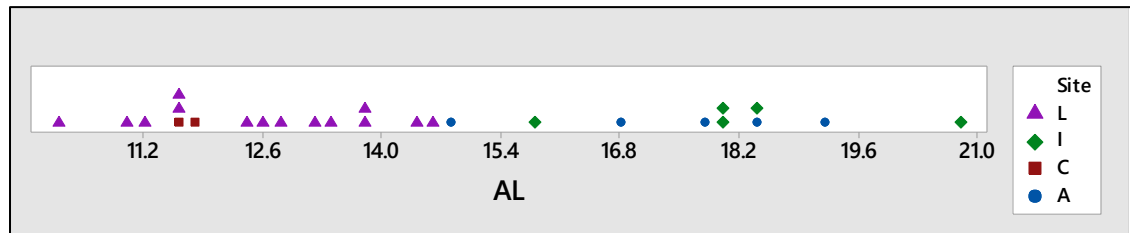


**c** Answers will vary.

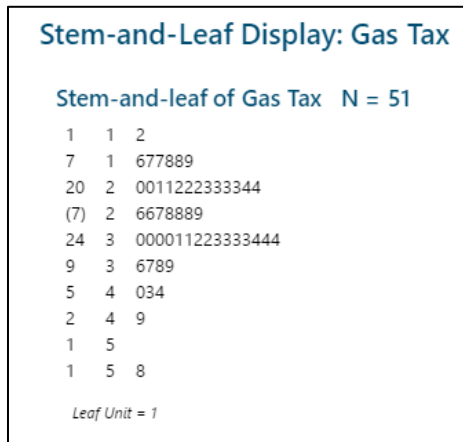
**1.R.12 a-b** Answers will vary. A typical histogram is shown next. Notice the gaps and the bimodal nature of the histogram, probably due to the fact that the samples were collected at different locations.



c The dotplot is shown as follows. The locations are indeed responsible for the unusual gaps and peaks in the relative frequency histogram.

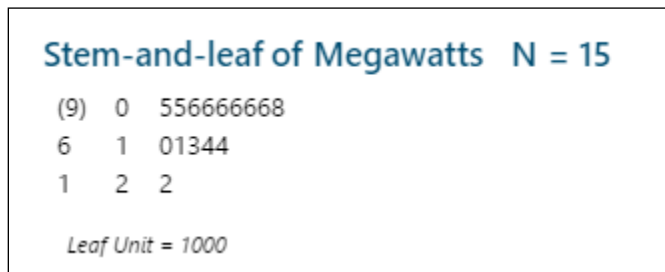


1.R.13 a-b The *Minitab* stem and leaf plot is shown next. The distribution is slightly skewed to the right.



c Pennsylvania (58.20) has an unusually high gas tax.

1.R.14 a-b Answers will vary. The *Minitab* stem and leaf plot is shown next. The distribution is skewed to the right.

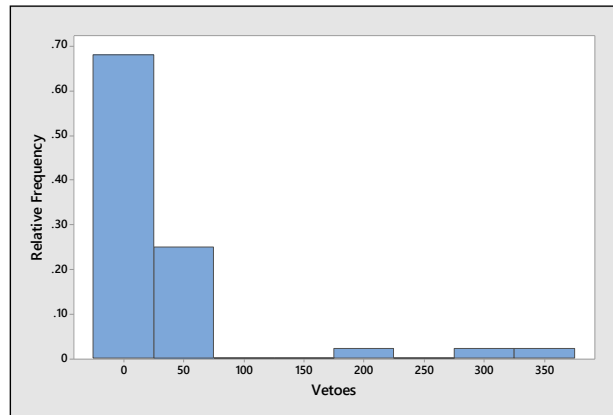


1.R.15 a-b The distribution is approximately mound-shaped, with one unusual measurement, in the class with midpoint at 100.8°. Perhaps the person whose temperature was 100.8 has some sort of illness coming on?

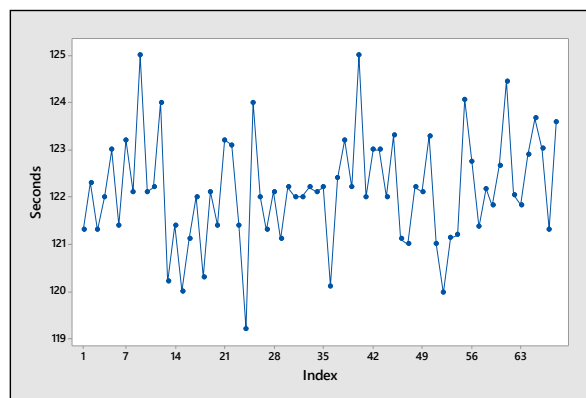
c The value 98.6° is slightly to the right of center.

### On Your Own

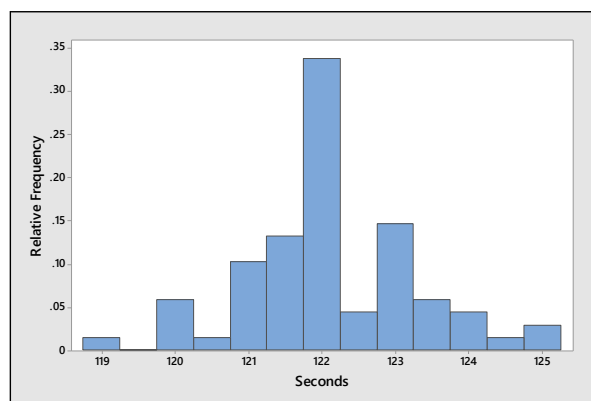
1.R.16 Answers will vary from student to student. The students should notice that the distribution is skewed to the right with a few presidents (Truman, Cleveland, and F.D. Roosevelt) casting an unusually large number of vetoes.



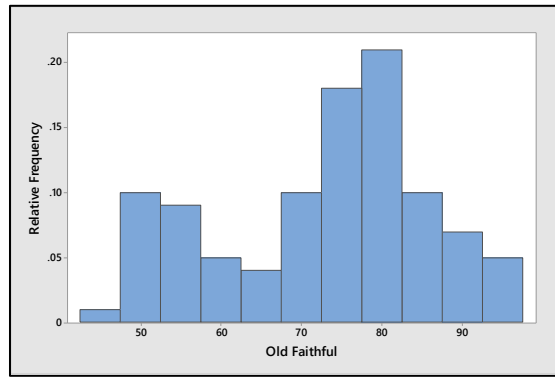
- 1.R.17 a** The line chart is shown next. The year in which a horse raced does not appear to have an effect on his winning time.



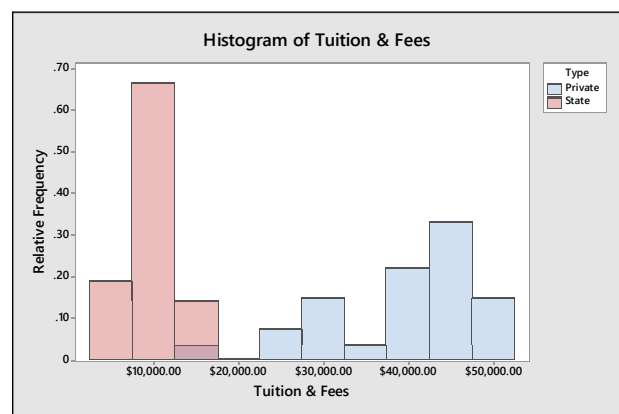
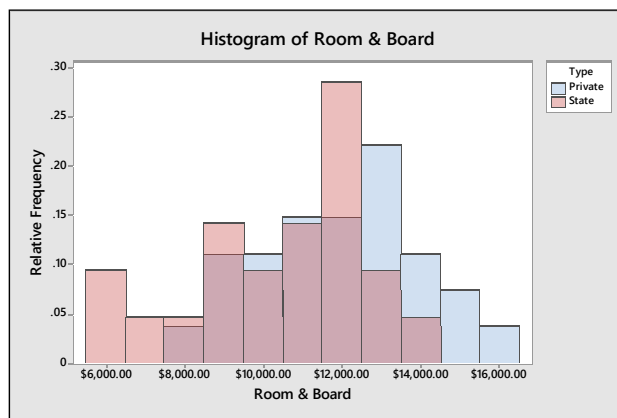
- b** Since the year of the race is not important in describing the data set, the distribution can be described using a relative frequency histogram. The distribution that follows is roughly mound-shaped with an unusually fast ( $x = 119.2$ ) race times the year that *Secretariat* won the derby.



- 1.R.18** Answers will vary from student to student. Students should notice that both distributions are skewed left. The higher peak with a low bar to its left in the laptop group may indicate that students who would generally receive average scores (65-75) are scoring higher than usual. This may or may not be *caused* by the fact that they used laptop computers.
- 1.R.19** Answers will vary. A typical relative frequency histogram is shown next. There is an unusual bimodal feature.



**1.R.20** Answers will vary. The student should notice that there is a clear difference in tuition and fees between private and state schools. Both distributions are roughly mound-shaped. The distribution of room and board costs are also roughly mound-shaped, but there is less of a difference between private and state schools.



## CASE STUDY: How is Your Blood Pressure?

1. The following variables have been measured on the participants in this study: sex (qualitative); age in years (quantitative discrete); diastolic blood pressure (quantitative continuous, but measured to an integer value) and systolic blood pressure (quantitative continuous, but measured to an integer value). For each person, both systolic and diastolic readings are taken, making the data bivariate.
2. The important variables in this study are diastolic and systolic blood pressure, which can be described singly with histograms in various categories (male vs. female or by age categories). Further, the relationship between systolic and diastolic blood pressure can be displayed together using a scatterplot or a bivariate histogram.
3. Answers will vary from student to student, depending on the choice of class boundaries or the software package which is used. The histograms should look fairly mound-shaped.
4. Answers will vary from student to student.
5. In determining how a student's blood pressure compares to those in a comparable sex and age group, female students (ages 15-20) must compare to the population of females, while male students (ages 15-20) must compare to the population of males. The student should use his or her blood pressure and compare it to the scatterplot generated in part 4.



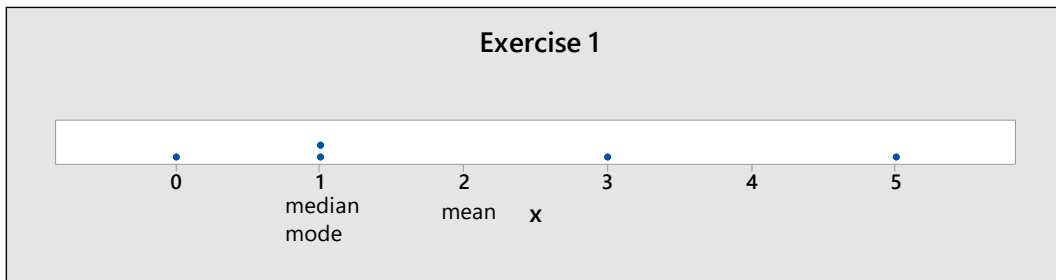
## 2: Describing Data with Numerical Measures

### Section 2.1

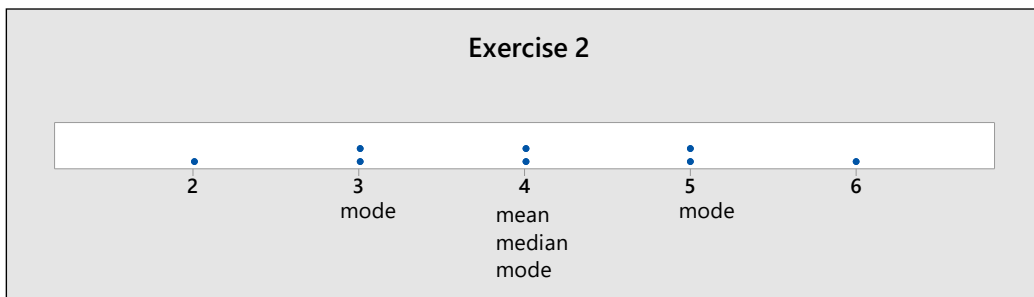
**2.1.1** The mean is the sum of the measurements divided by the number of measurements, or

$$\bar{x} = \frac{\sum x_i}{n} = \frac{0+5+1+1+3}{5} = \frac{10}{5} = 2$$

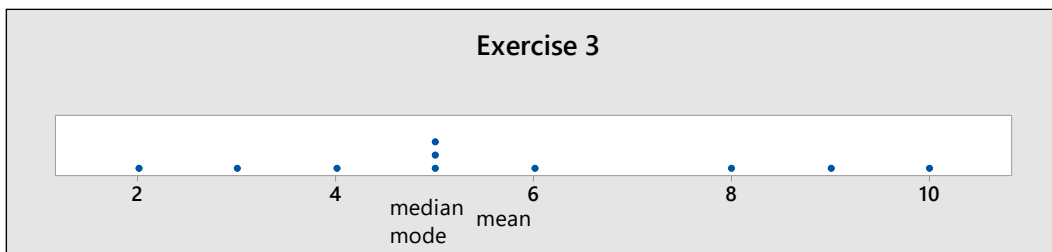
To calculate the median, the observations are first ranked from smallest to largest: 0, 1, 1, 3, 5. Then since  $n=5$ , the position of the median is  $0.5(n+1) = 3$ , and the median is the 3<sup>rd</sup> ranked measurement, or  $m = 1$ . The mode is the measurement occurring most frequently, or mode = 1. The three measures are located on the dotplot that follows.



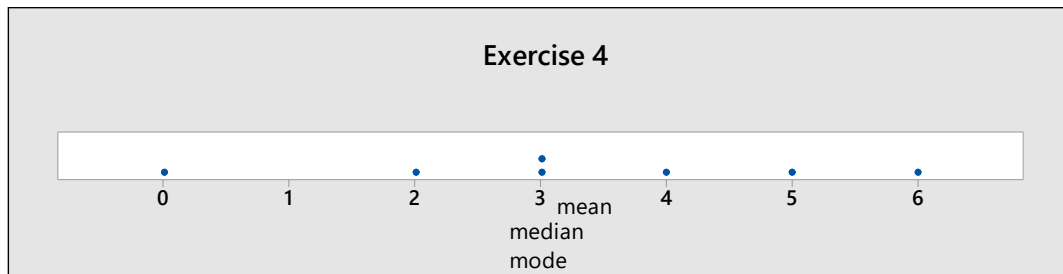
**2.1.2** The mean is  $\bar{x} = \frac{\sum x_i}{n} = \frac{3+2+\dots+5}{8} = \frac{32}{8} = 4$ . To calculate the median, the observations are first ranked from smallest to largest: 2, 3, 3, 4, 4, 5, 5, 6. Since  $n=8$  is even, the position of the median is  $0.5(n+1) = 4.5$ , and the median is the average of the 4<sup>th</sup> and 5<sup>th</sup> measurements, or  $m = (4+4)/2 = 4$ . Since three measurements occur twice, there are three modes: 3, 4, and 5. The dotplot is shown next.



**2.1.3** The mean is  $\bar{x} = \frac{\sum x_i}{n} = \frac{57}{10} = 5.7$ . The ranked observations are: 2, 3, 4, 5, 5, 5, 6, 8, 9, 10. Since  $n=10$ , the median is halfway between the 5<sup>th</sup> and 6<sup>th</sup> ordered observations, or  $m = (5+5)/2 = 5$ . The measurement  $x = 5$  occurs three times. Since this is the highest frequency of occurrence for the data set, the *mode* is 5. The dotplot is shown next.



- 2.1.4** The mean is  $\bar{x} = \frac{\sum x_i}{n} = \frac{23}{7} = 3.286$ . The ranked observations are: 0, 2, 3, 3, 4, 5, 6. Since  $n = 7$ , the median is the 4<sup>th</sup> ordered observation, or  $m = 3$ . The measurement  $x = 3$  occurs twice. Since this is the highest frequency of occurrence for the data set, the *mode* is 3. The dotplot is shown next.



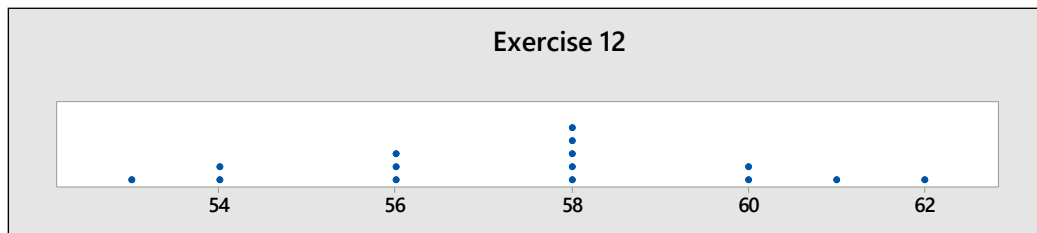
- 2.1.5** Since the mean is less than the median, the data set is skewed left.
- 2.1.6** Since the mean and the median are nearly equal, the data set is roughly symmetric.
- 2.1.7** Since the mean is greater than the median, the data set is skewed right.
- 2.1.8** Since the mean is less than the median, the data set is skewed left.
- 2.1.9** The mean is

$$\bar{x} = \frac{\sum x_i}{n} = \frac{94.3}{24} = 3.929$$

The data set is ordered from smallest to largest below. The position of the median is  $.5(n + 1) = 12.5$  and the median is the average of the 12<sup>th</sup> and 13<sup>th</sup> measurements, or  $(3.9 + 3.9)/2 = 3.9$ .

3.2	3.5	3.7	3.9	4.2	4.4
3.3	3.5	3.9	4.0	4.3	4.4
3.4	3.6	3.9	4.0	4.3	4.5
3.5	3.6	3.9	4.2	4.3	4.8

- 2.1.10** Since the mean and the median are nearly equal, the data set is roughly symmetric.
- 2.1.11** The mean is  $\bar{x} = \frac{\sum x_i}{n} = \frac{862}{15} = 57.467$ . The ranked observations are: 53, 54, 54, 56, 56, 56, 58, 58, 58, 58, 58, 60, 60, 61, 62. Since  $n = 15$ , the median is the  $.5(15 + 1) = 8^{\text{th}}$  ordered observation, or  $m = 58$ . The measurement  $x = 58$  occurs five times. Since this is the highest frequency of occurrence for the data set, the *mode* is 58.
- 2.1.12** Since the mean and the median are nearly equal, the data set is roughly symmetric. The dotplot is shown next, confirming our result.



- 2.1.13** Similar to previous exercises. The mean is

$$\bar{x} = \frac{\sum x_i}{n} = \frac{0.99 + 1.92 + \cdots + 0.66}{14} = \frac{12.55}{14} = 0.896$$

To calculate the median, rank the observations from smallest to largest. The position of the median is  $0.5(n + 1) = 7.5$ , and the median is the average of the 7<sup>th</sup> and 8<sup>th</sup> ranked measurement or

$m = (0.67 + 0.69)/2 = 0.68$ . The mode is .60, which occurs twice. Since the mean is slightly larger than the median, the distribution is slightly skewed to the right.

**2.1.14** Similar to previous exercises. The mean is

$$\bar{x} = \frac{\sum x_i}{n} = \frac{125 + 45 + \cdots + 80}{15} = \frac{1305}{15} = 87$$

To calculate the median, rank the observations from smallest to largest. The position of the median is  $0.5(n+1) = 8$ , and the median is the 8<sup>th</sup> ranked measurement or  $m = 105$ . There are two modes—110 and 115—which both occur twice. Since the mean is smaller than the median, the distribution is skewed to the left.

**2.1.15** Similar to previous exercises. The mean is

$$\bar{x} = \frac{\sum x_i}{n} = \frac{2150}{10} = 215$$

The ranked observations are: 175, 185, 190, 190, 200, 225, 230, 240, 250, 265. The position of the median is  $0.5(n+1) = 5.5$  and the median is the average of the 5<sup>th</sup> and 6<sup>th</sup> observation or

$$\frac{200 + 225}{2} = 212.5$$

and the mode is 190 which occurs twice. The mean is slightly larger than the median, and the data may be slightly skewed right.

**2.1.16**   **a**    $\bar{x} = \frac{\sum x_i}{n} = \frac{13904}{4} = 3476$                       **b**    $\bar{x} = \frac{\sum x_i}{n} = \frac{12674}{4} = 3168.50$

**c**   The average premium cost in several different cities is not as important to the consumer as the cost for your particular characteristics, including the city in which you live.

**2.1.17**   **a**   Although there may be a few households who own more than one DVR, the majority should own either 0 or 1. The distribution should be slightly skewed to the right.

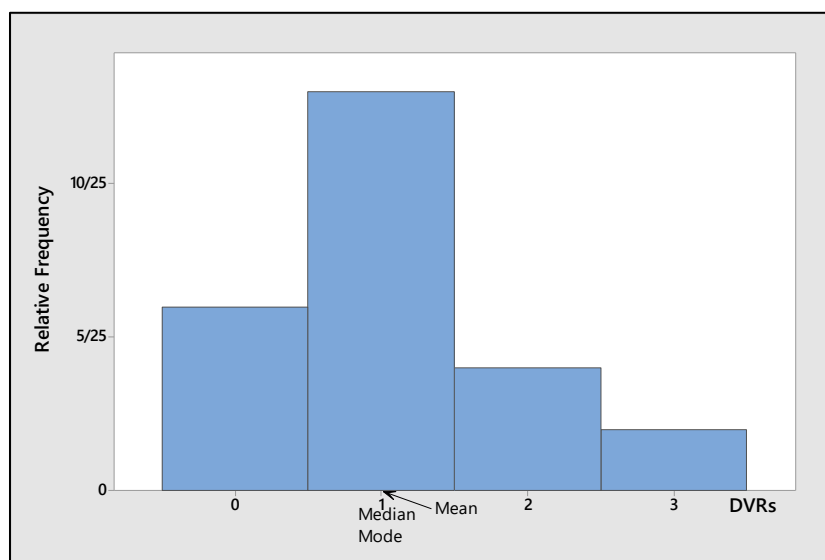
**b**   Since most households will have only one DVR, we guess that the mode is 1.

**c**   The mean is

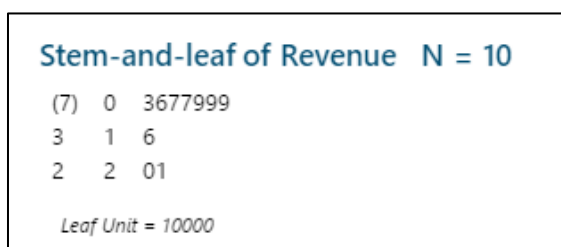
$$\bar{x} = \frac{\sum x_i}{n} = \frac{1 + 0 + \cdots + 1}{25} = \frac{27}{25} = 1.08$$

To calculate the median, the observations are first ranked from smallest to largest: There are six 0s, thirteen 1s, four 2s, and two 3s. Then since  $n = 25$ , the position of the median is  $0.5(n+1) = 13$ , which is the 13<sup>th</sup> ranked measurement, or  $m = 1$ . The mode is the measurement occurring most frequently, or mode = 1.

**d**   The relative frequency histogram is shown next, with the three measures superimposed. Notice that the mean falls slightly to the right of the median and mode, indicating that the measurements are slightly skewed to the right.



**2.1.18 a** The stem and leaf plot that follows was generated by *Minitab*. It is skewed to the right.



**b** The mean is

$$\bar{x} = \frac{\sum x_i}{n} = \frac{166380 + 79919 + \dots + 205004}{10} = \frac{1129104}{10} = 112,910.40$$

To calculate the median, rank the observations from smallest to largest.

37,949	69,495	71,890	79,919	93,662
94,571	94,595	166,380	205,004	215,639

Then since  $n = 10$ , the position of the median is  $0.5(n+1) = 5.5$ , the average of the 5<sup>th</sup> and 6<sup>th</sup> ranked measurements or  $m = (93662 + 94571)/2 = 94,116.5$ .

**c** Since the mean is strongly affected by outliers, the median would be a better measure of center for this data set.

**2.1.19** It is obvious that any one family cannot have 2.5 children, since the number of children per family is a quantitative discrete variable. The researcher is referring to the average number of children per family calculated for all families in the United States during the 1930s. The average does not necessarily have to be integer-valued.

**2.1.20** The distribution of sports salaries will be skewed to the right, because of the very high salaries of some sports figures. Hence, the median salary would be a better measure of center than the mean.

**2.1.21 a** The mean is

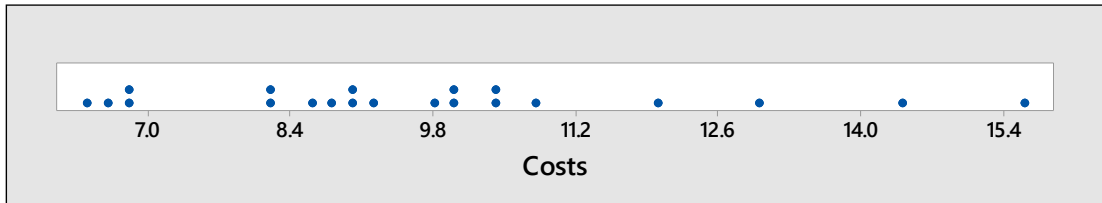
$$\bar{x} = \frac{\sum x_i}{n} = \frac{10.4 + 10.7 + \dots + 8.8}{21} = \frac{203.5}{21} = 9.690$$

To calculate the median, rank the observations from smallest to largest.

6.4   6.5   6.8   6.8   8.2   8.2  
 8.6   8.8   8.9   9.0   9.2  
 9.7   10.0   10.0   10.4   10.4  
 10.7   12.0   13.0   14.3   15.6

Then since  $n = 21$ , the position of the median is  $0.5(n+1) = 11$ , the 11<sup>th</sup> ranked measurement or  $m = 9.2$ . There are four modes—6.8, 8.2, 10.0 and 10.4—all of which occur twice.

**b-c** Since the mean is larger than the median, the data set is skewed right. The dotplot is shown next.



**2.1.22 a** Similar to previous exercises.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{743.87}{16} = 46.492$$

**b** The ranked observations are:

39.99   39.99   39.99   39.99  
 39.99   41.46   42.46   43.99  
 46.49   47.19   47.49   47.90  
 51.97   54.99   59.99   59.99

The position of the median is  $0.5(n+1) = 8.5$ , and the median is the average of the 8<sup>th</sup> and 9<sup>th</sup> observations or  $m = (43.99 + 46.49)/2 = 45.24$ . The mode is 39.99, which occurs 5 times.

**c** Answers will vary. Ease of purchase, speed of delivery, delivery costs, and return policy might be important.

## Section 2.2

**2.2.1** Calculate  $\bar{x} = \frac{\sum x_i}{n} = \frac{12}{5} = 2.4$ . Then create a table of differences,  $(x_i - \bar{x})$  and their squares,  $(x_i - \bar{x})^2$ .

$x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
2	-0.4	0.16
1	-1.4	1.96
1	-1.4	1.96
3	0.6	0.36
5	2.6	6.76
Total	0	11.20

Then, using the definition formula,

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{(2-2.4)^2 + \dots + (5-2.4)^2}{4} = \frac{11.20}{4} = 2.8$$

To use the computing formula, calculate  $\sum x_i^2 = 2^2 + 1^2 + \dots + 5^2 = 40$ . Then

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{40 - \frac{(12)^2}{5}}{4} = \frac{11.2}{4} = 2.8$$

The sample standard deviation is the positive square root of the variance or

$$s = \sqrt{s^2} = \sqrt{2.8} = 1.673$$

**2.2.2** Calculate  $\bar{x} = \frac{\sum x_i}{n} = \frac{17}{8} = 2.125$ . Then create a table of differences,  $(x_i - \bar{x})$  and their squares,  $(x_i - \bar{x})^2$ .

$x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
4	1.875	3.515625
1	-1.125	1.265625
3	.875	.765625
1	-1.125	1.265625
3	.875	.765625
1	-1.125	1.265625
2	-.125	.015625
2	-.125	.015625
Total	0	8.875

Then, using the definition formula,

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{(4-2.125)^2 + \dots + (2-2.125)^2}{7} = \frac{8.875}{7} = 1.2679$$

To use the computing formula, calculate  $\sum x_i^2 = 4^2 + 1^2 + \dots + 2^2 = 45$ . Then

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{45 - \frac{(17)^2}{8}}{7} = \frac{8.875}{7} = 1.2679 \text{ and } s = \sqrt{s^2} = \sqrt{1.2679} = 1.126.$$

**2.2.3** Calculate  $\bar{x} = \frac{\sum x_i}{n} = \frac{31}{8} = 3.875$ . Then create a table of differences,  $(x_i - \bar{x})$  and their squares,  $(x_i - \bar{x})^2$ .

$x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
3	-.875	.765625	4	.125	.015625
1	-2.875	8.265625	4	.125	.015625
5	1.125	1.265625	3	-.875	.765625
6	2.125	4.515625	5	1.125	1.265625

Then, using the definition formula,

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{(3-3.875)^2 + \dots + (5-3.875)^2}{7} = \frac{16.875}{7} = 2.4107$$

To use the computing formula, calculate  $\sum x_i^2 = 3^2 + 1^2 + \dots + 5^2 = 137$ . Then

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{137 - \frac{(31)^2}{8}}{7} = \frac{16.875}{7} = 2.4107 \text{ and } s = \sqrt{s^2} = \sqrt{2.4107} = 1.553.$$

**2.2.4** Use the computing formula, with  $\sum x_i^2 = 374.65$  and  $\sum x_i = 94.3$ . Then  $\bar{x} = \frac{\sum x_i}{n} = \frac{94.3}{24} = 3.929$ ,

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{374.65 - \frac{(94.3)^2}{24}}{23} = \frac{4.129533}{23} = .179547$$

$$\text{and } s = \sqrt{s^2} = \sqrt{.179547} = .4237.$$

The range is calculated as  $R = 4.8 - 3.2 = 1.6$  and  $R/s = 1.6/.4237 = 3.78$ , so that the range is approximately 4 standard deviations.

**2.2.5** Use the computing formula, with  $\sum x_i^2 = 49,634$  and  $\sum x_i = 862$ . Then  $\bar{x} = \frac{\sum x_i}{n} = \frac{862}{15} = 57.467$ ,

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{49634 - \frac{(862)^2}{15}}{14} = \frac{97.73333}{14} = 6.98095$$

$$\text{and } s = \sqrt{s^2} = \sqrt{6.98095} = 2.642.$$

The range is calculated as  $R = 62 - 53 = 9$  and  $R/s = 9/2.642 = 3.41$ , so that the range is approximately 3.5 standard deviations.

**2.2.6** Use the computing formula, with  $\sum x_i^2 = 130$  and  $\sum x_i = 32$ . Then  $\bar{x} = \frac{\sum x_i}{n} = \frac{32}{10} = 3.2$ ,

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{130 - \frac{(32)^2}{10}}{9} = \frac{27.60}{9} = 3.066667 \text{ and } s = \sqrt{s^2} = \sqrt{3.066667} = 1.75119.$$

The range,  $R = 6 - 1 = 5$ , is  $5/1.75119 = 2.855$  standard deviations.

**2.2.7** The range is  $R = 1.92 - .53 = 1.39$ . Use the computing formula, with  $\sum x_i^2 = 13.3253$  and  $\sum x_i = 12.55$ . Then,

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{13.3253 - \frac{(12.55)^2}{14}}{13} = \frac{2.07512}{13} = .15962 \text{ and } s = \sqrt{s^2} = \sqrt{.15962} = .3995.$$

**2.2.8** The range is  $R = 125 - 25 = 100$ . Use the computing formula, with  $\sum x_i^2 = 131,125$  and  $\sum x_i = 1305$ . Then,

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{131,125 - \frac{(1305)^2}{15}}{14} = \frac{17,590}{14} = 1256.42857 \text{ and } s = \sqrt{s^2} = \sqrt{1256.42857} = 35.446.$$

**2.2.9** The range is  $R = 265 - 175 = 90$ . Use the computing formula, with  $\sum x_i^2 = 470,900$  and  $\sum x_i = 2150$ . Then,

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{470,900 - \frac{(2150)^2}{10}}{9} = \frac{8650}{9} = 961.11111 \text{ and } s = \sqrt{s^2} = \sqrt{961.11111} = 31.002.$$

**2.2.10 a** The range is  $R = 2.39 - 1.28 = 1.11$ .

**b** Calculate  $\sum x_i^2 = 1.28^2 + 2.39^2 + \dots + 1.51^2 = 15.415$ . Then

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{15.415 - \frac{(8.56)^2}{5}}{4} = \frac{.76028}{4} = .19007$$

and  $s = \sqrt{s^2} = \sqrt{.19007} = .436$

**c** The range,  $R = 1.11$ , is  $1.11/.436 = 2.5$  standard deviations.

**2.2.11 a** The range is  $R = 459.21 - 233.97 = 225.24$ . **b**  $\bar{x} = \frac{\sum x_i}{n} = \frac{3777.30}{12} = 314.775$

**c** Calculate  $\sum x_i^2 = 243.92^2 + 233.97^2 + \dots + 286.41^2 = 1,267,488.3340$ . Then

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{1267488.334 - \frac{(3777.30)^2}{12}}{11} = 7135.338773$$

and  $s = \sqrt{s^2} = \sqrt{7135.338773} = 84.471$ .

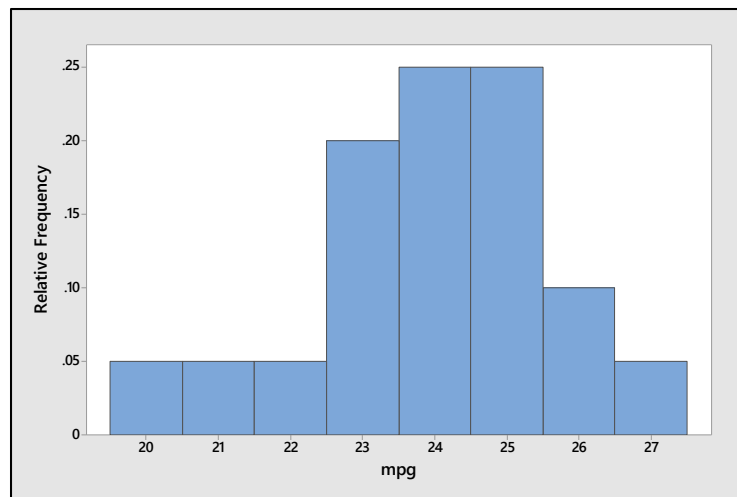
**2.2.12 a** Calculate  $n = 10$ ,  $\sum x_i = 68.5$ ,  $\sum x_i^2 = 478.375$ . Then  $\bar{x} = \frac{\sum x_i}{n} = \frac{68.5}{10} = 6.85$  and

$$s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{478.375 - \frac{(68.5)^2}{10}}{9}} = 1.008$$

**b** The most frequently recorded measurement is the mode or  $x = 7$  hours of sleep.

**2.2.13 a** Max = 27, Min = 20.2 and the range is  $R = 27 - 20.2 = 6.8$ .

**b** Answers will vary. A typical histogram is shown next. The distribution is slightly skewed to the left.



**c** Calculate  $n = 20$ ,  $\sum x_i = 479.2$ ,  $\sum x_i^2 = 11532.82$ . Then  $\bar{x} = \frac{\sum x_i}{n} = \frac{479.2}{20} = 23.96$

$$s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{11532.82 - \frac{(479.2)^2}{20}}{19}} = \sqrt{2.694} = 1.641$$

**2.2.14 a** The range is  $R = 71 - 40 = 31$ .



- b** Calculate  $n = 10$ ,  $\sum x_i = 592$ ,  $\sum x_i^2 = 36014$ . Then  $\bar{x} = \frac{\sum x_i}{n} = \frac{592}{10} = 59.2$

$$s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{36014 - \frac{(592)^2}{10}}{9}} = \sqrt{107.5111} = 10.369$$

- c** Calculate  $R/s = 31/10.369 = 3.00$ , so that the range is approximately 3 standard deviations.

**2.2.15** They are probably referring to the average number of times that men and women go camping per year.

## Section 2.3

- 2.3.1** The range of the data is  $R = 6 - 1 = 5$  and the range approximation with  $n = 10$  is  $s \approx \frac{R}{3} = 1.67$ . The standard deviation of the sample is

$$s = \sqrt{s^2} = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{130 - \frac{(32)^2}{10}}{9}} = \sqrt{3.0667} = 1.751$$

which is very close to the estimate.

- 2.3.2** The range of the data is  $R = 30 - 22 = 8$  and the range approximation with  $n = 28$  is  $s \approx \frac{R}{4} = 2$ . The standard deviation of the sample is

$$s = \sqrt{s^2} = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{19285 - \frac{(733)^2}{28}}{27}} = \sqrt{3.55952} = 1.887$$

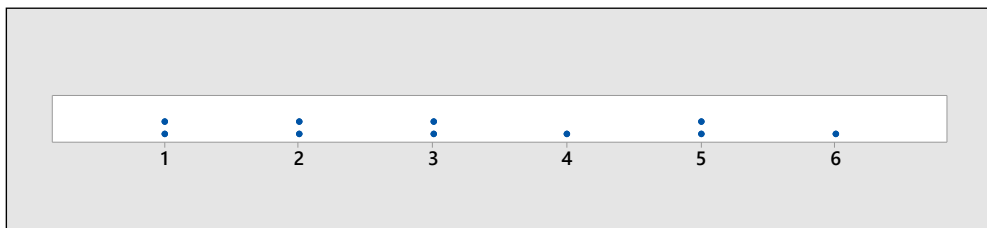
which is very close to the estimate.

- 2.3.3** The range of the data is  $R = 7.1 - 4.0 = 3.1$  and the range approximation with  $n = 15$  is  $s \approx \frac{R}{3} = 1.033$ . The standard deviation of the sample is

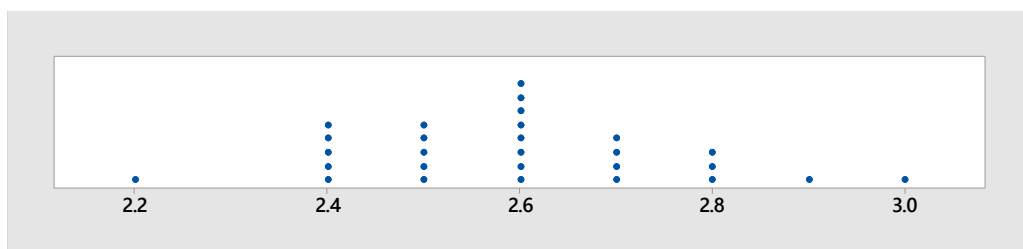
$$s = \sqrt{s^2} = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{531.17 - \frac{(88.1)^2}{15}}{14}} = \sqrt{.98067} = .990$$

which is very close to the estimate.

- 2.3.4** From the dotplot on the next page, you can see that the data set is not mound-shaped. Hence you can use Tchebysheff's Theorem, but not the Empirical Rule to describe the data.

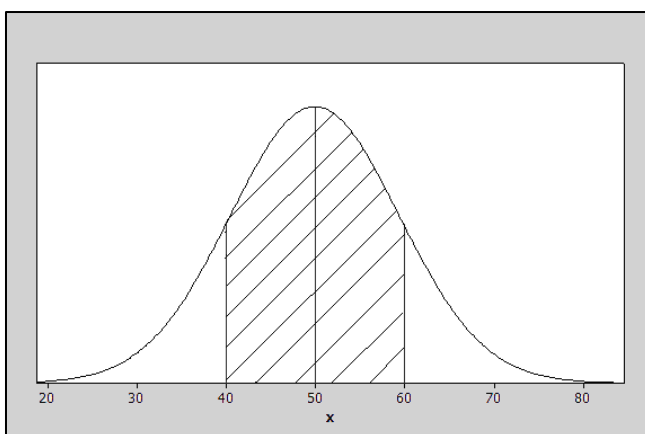


- 2.3.5** From the dotplot below, you can see that the data set is relatively mound-shaped. Hence you can use both Tchebysheff's Theorem and the Empirical Rule to describe the data.



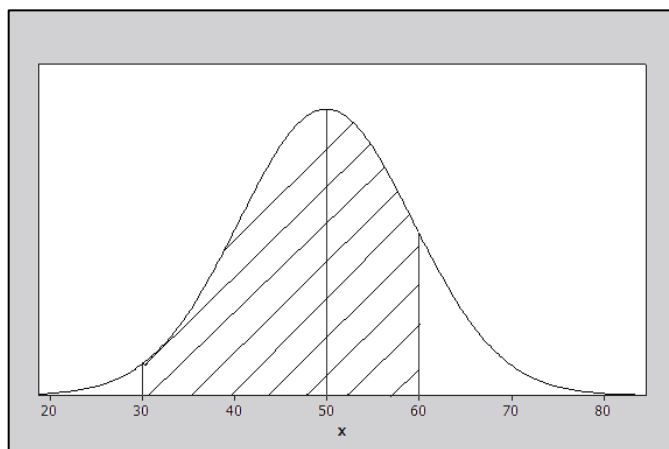
**2.3.6-11** Since the distribution is relatively mound-shaped, use the Empirical Rule.

**6.** The interval from 40 to 60 represents  $\mu \pm \sigma = 50 \pm 10$ . Since the distribution is relatively mound-shaped, the proportion of measurements between 40 and 60 is .68 (or 68%) according to the Empirical Rule and is shown next.



**7.** Refer to the figure above. Again, using the Empirical Rule, the interval  $\mu \pm 2\sigma = 50 \pm 2(10)$  or between 30 and 70 contains approximately .95 (or 95%) of the measurements.

**8.** Refer to the figure that follows.



Since approximately 68% of the measurements are between 40 and 60, the symmetry of the distribution implies that 34% of the measurements are between 50 and 60. Similarly, since 95% of the measurements are between 30 and 70, approximately 47.5% are between 30 and 50. Thus, the proportion of measurements between 30 and 60 is  $0.34 + 0.475 = 0.815$ .

**9.** From the figure in Exercise 6, the proportion of the measurements between 50 and 60 is 0.34 and the proportion of the measurements which are greater than 50 is 0.50. Therefore, the proportion that is greater than 60 must be  $0.5 - 0.34 = 0.16$ .

**10.** Refer to Exercise 9. The proportion of the measurements that are greater than 60 is .16. Since the total area under the curve is 1 or 100%, the proportion less than 60 must be  $1 - .16 = .84$ .

**11.** From the figure in Exercise 6, the proportion of the measurements between 40 and 50 is 0.34 and the proportion of the measurements which are greater than 50 is 0.50. Therefore, the proportion that is 40 or more must be  $0.5 + 0.34 = 0.84$ .

**2.3.12-14** Since nothing is known about the shape of the data distribution, you must use Tchebysheff's Theorem to describe the data.

**12.** The interval from 60 to 90 represents  $\mu \pm 3\sigma$  which will contain at least 8/9 of the measurements.

**13.** The interval from 65 to 85 represents  $\mu \pm 2\sigma$  which will contain at least 3/4 of the measurements.

**14.** The interval from 62.5 to 87.5 represents  $\mu \pm 2.5\sigma$  which will contain at least  $1 - \frac{1}{k^2} = 1 - \frac{1}{2.5^2} = .84$  or 84% of the measurements.

**2.3.15** The range of the data is  $R = 1.1 - 0.5 = 0.6$  and the approximate value of  $s$  is

$$s \approx \frac{R}{3} = 0.2$$

Calculate  $\sum x_i = 7.6$  and  $\sum x_i^2 = 6.02$ . The sample variance is

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{6.02 - \frac{(7.6)^2}{10}}{9} = \frac{0.244}{9} = .027111$$

and the standard deviation of the sample is  $s = \sqrt{.027111} = 0.165$  which is very close to the range approximation.

**2.3.16** The range of the data is  $R = 1.92 - .53 = 1.39$  and the approximate value of  $s$  is

$$s \approx \frac{R}{3} = .4633$$

Calculate  $\sum x_i = 12.55$  and  $\sum x_i^2 = 13.3253$ . The sample variance is

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{13.3253 - \frac{(12.55)^2}{14}}{13} = .159625$$

and the standard deviation of the sample is  $s = \sqrt{.159625} = 0.3995$  which is very close to the range approximation.

**2.3.17** The range is  $R = 125 - 25 = 100$  and the approximate value of  $s$  is

$$s \approx \frac{R}{3} = 33.33$$

Use the computing formula, with  $\sum x_i^2 = 131,125$  and  $\sum x_i = 1305$ . Then,

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{131,125 - \frac{(1305)^2}{15}}{14} = \frac{17,590}{14} = 1256.42857 \text{ and } s = \sqrt{s^2} = \sqrt{1256.42857} = 35.446.$$

which is very close to the range approximation.

**2.3.18** The range of the data is  $R = 2.39 - 1.28 = 1.11$  and the approximate value of  $s$  is

$$s \approx \frac{R}{2.5} = 0.444$$

Calculate  $\sum x_i = 8.56$  and  $\sum x_i^2 = 15.4150$ . The sample variance is

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{15.4150 - \frac{(8.56)^2}{5}}{4} = .19007$$

and the standard deviation of the sample is  $s = \sqrt{.19007} = 0.436$  which is very close to the range approximation.

- 2.3.19 a** The stem and leaf plot generated by *Minitab* shows that the data is roughly mound-shaped. Note however the gap in the center of the distribution and the two measurements in the upper tail.

Stem-and-leaf of Weight N = 27		
1	7	5
2	8	3
6	8	7999
8	9	23
13	9	66789
13	10	
(3)	10	688
11	11	2244
7	11	788
4	12	4
3	12	8
2	13	
2	13	8
1	14	1

Leaf Unit = 0.01

- b** Calculate  $\sum x_i = 28.41$  and  $\sum x_i^2 = 30.6071$ , the sample mean is  $\bar{x} = \frac{\sum x_i}{n} = \frac{28.41}{27} = 1.052$  and the standard deviation of the sample is

$$s = \sqrt{s^2} = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{30.6071 - \frac{(28.41)^2}{27}}{26}} = 0.166$$

- c** The following table gives the actual percentage of measurements falling in the intervals  $\bar{x} \pm ks$  for  $k = 1, 2, 3$ .

$k$	$\bar{x} \pm ks$	Interval	Number in Interval	Percentage
1	$1.052 \pm 0.166$	0.866 to 1.218	21	78%
2	$1.052 \pm 0.332$	0.720 to 1.384	26	96%
3	$1.052 \pm 0.498$	0.554 to 1.550	27	100%

- d** The percentages in part **c** do not agree too closely with those given by the Empirical Rule, especially in the one standard deviation range. This is caused by the lack of rounding (indicated by the gap) in the center of the distribution.

- e** The lack of any one-pound packages is probably a marketing technique intentionally used by the supermarket. People who buy slightly less than one-pound would be drawn by the slightly lower price, while those who need exactly one-pound of meat for their recipe might tend to opt for the larger package, increasing the store's profit.

- 2.3.20** According to the Empirical Rule, if a distribution of measurements is approximately mound-shaped,

- a** approximately 68% or 0.68 of the measurements fall in the interval  $\mu \pm \sigma = 12 \pm 2.3$  or 9.7 to 14.3
- b** approximately 95% or 0.95 of the measurements fall in the interval  $\mu \pm 2\sigma = 12 \pm 4.6$  or 7.4 to 16.6
- c** approximately 99.7% or 0.997 of the measurements fall in the interval  $\mu \pm 3\sigma = 12 \pm 6.9$  or 5.1 to 18.9

Therefore, approximately 0.3% or 0.003 will fall outside this interval.

- 2.3.21 a** The stem and leaf plots are shown next. The second set has a slightly higher location and spread.

Stem-and-leaf of Titanium content			Method = 1	N = 10
1	10	0		
3	11	00		
4	12	0		
(4)	13	0000		
2	14	0		
1	15	0		
Leaf Unit = 0.0001				

Stem-and-leaf of Titanium content			Method = 2	N = 10
1	11	0		
3	12	00		
5	13	00		
5	14	0		
4	15	00		
2	16	0		
1	17	0		
Leaf Unit = 0.0001				

- b Method 1:** Calculate  $\sum x_i = 0.125$  and  $\sum x_i^2 = 0.001583$ . Then  $\bar{x} = \frac{\sum x_i}{n} = 0.0125$  and

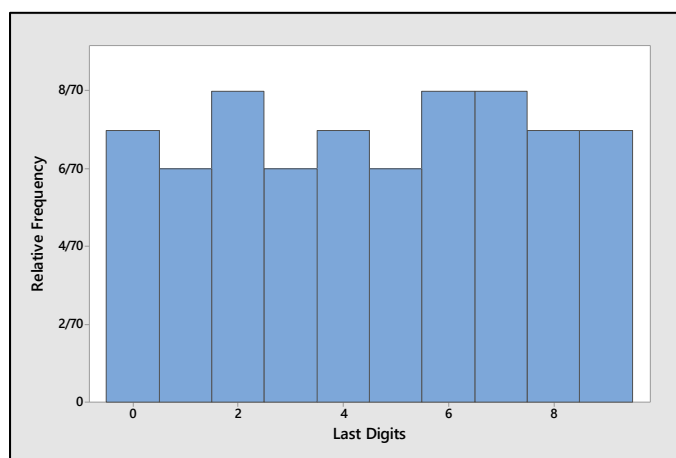
$$s = \sqrt{s^2} = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{0.001583 - \frac{(0.125)^2}{10}}{9}} = 0.00151$$

- Method 2:** Calculate  $\sum x_i = 0.138$  and  $\sum x_i^2 = 0.001938$ . Then  $\bar{x} = \frac{\sum x_i}{n} = 0.0138$  and

$$s = \sqrt{s^2} = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{0.001938 - \frac{(0.138)^2}{10}}{9}} = 0.00193$$

The results confirm the conclusions of part **a**.

- 2.3.22 a** The relative frequency histogram is shown next.



The distribution is relatively “flat” or “uniformly distributed” between 0 and 9. Hence, the center of the distribution should be approximately halfway between 0 and 9 or  $(0+9)/2 = 4.5$ .

**b** The range of the data is  $R = 9 - 0 = 9$ . Using the range approximation,  $s \approx R/4 = 9/4 = 2.25$ .

**c** Using the data entry method the students should find  $\bar{x} = 4.586$  and  $s = 2.892$ , which are fairly close to our approximations.

**2.3.23 a** Similar to previous exercises. The intervals, counts and percentages are shown in the table.

$k$	$\bar{x} \pm ks$	Interval	Number in Interval	Percentage
1	$4.586 \pm 2.892$	1.694 to 7.478	43	61%
2	$4.586 \pm 5.784$	-1.198 to 10.370	70	100%
3	$4.586 \pm 8.676$	-4.090 to 13.262	70	100%

**b** The percentages in part **a** do not agree with those given by the Empirical Rule. This is because the shape of the distribution is not mound-shaped, but flat.

**2.3.24 a** Although most of the animals will die at around 32 days, there may be a few animals that survive a very long time, even with the infection. The distribution will probably be skewed right.

**b** Using Tchebysheff’s Theorem, at least  $3/4$  of the measurements should be in the interval  $\mu \pm \sigma \Rightarrow 32 \pm 72$  or 0 to 104 days.

**2.3.25 a** The value of  $x$  is  $\mu - \sigma = 32 - 36 = -4$ .

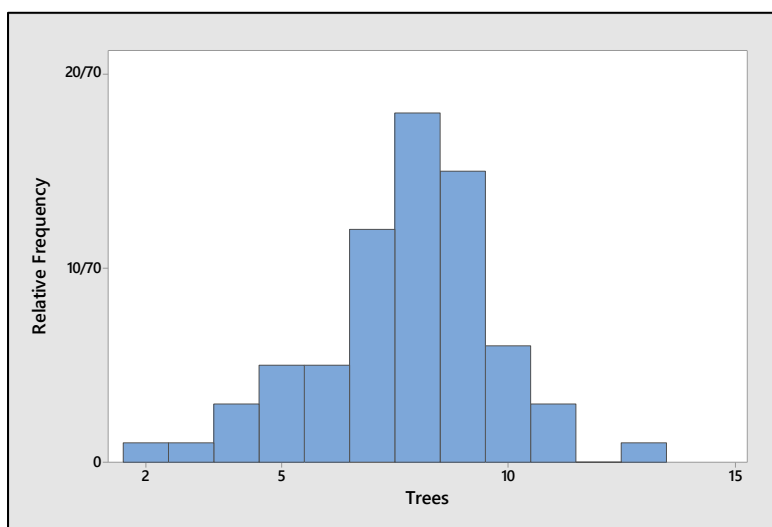
**b** The interval  $\mu \pm \sigma$  is  $32 \pm 36$  or  $-4$  to 68. This interval should contain approximately 68% of the survival times, so that 16% will be longer than 68 days and 16% will be less than  $-4$  days.

**c** The latter is clearly impossible. Therefore, the approximate values given by the Empirical Rule are not accurate, indicating that the distribution cannot be mound-shaped.

**2.3.26 a** We choose to use 12 classes of length 1.0. The tally and the relative frequency histogram follow.

Class i	Class Boundaries	Tally	$f_i$	Relative frequency, $f_i/n$
1	2 to < 3	1	1	1/70
2	3 to < 4	1	1	1/70
3	4 to < 5	111	3	3/70
4	5 to < 6	11111	5	5/70
5	6 to < 7	11111	5	5/70
6	7 to < 8	11111 11111 11	12	12/70
7	8 to < 9	11111 11111 11111 111	18	18/70
8	9 to < 10	11111 11111 11111	15	15/70
9	10 to < 11	11111 1	6	6/70

10	11 to < 12	111	3	3/70
11	12 to < 13		0	0
12	13 to < 14	1	1	1/70



**b** Calculate  $n = 70$ ,  $\sum x_i = 541$  and  $\sum x_i^2 = 4453$ . Then  $\bar{x} = \frac{\sum x_i}{n} = \frac{541}{70} = 7.729$  is an estimate of  $\mu$ .

**c** The sample standard deviation is

$$s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{4453 - \frac{(541)^2}{70}}{69}} = \sqrt{3.9398} = 1.985$$

The three intervals,  $\bar{x} \pm ks$  for  $k = 1, 2, 3$  are calculated below. The table shows the actual percentage of measurements falling in a particular interval as well as the percentage predicted by Tchebysheff's Theorem and the Empirical Rule. Note that the Empirical Rule should be fairly accurate, as indicated by the mound-shape of the histogram in part **a**.

$k$	$\bar{x} \pm ks$	Interval	Fraction in Interval	Tchebysheff	Empirical Rule
1	$7.729 \pm 1.985$	5.744 to 9.774	$50/70 = 0.71$	at least 0	$\approx 0.68$
2	$7.729 \pm 3.970$	3.759 to 11.699	$67/70 = 0.96$	at least 0.75	$\approx 0.95$
3	$7.729 \pm 5.955$	1.774 to 13.684	$70/70 = 1.00$	at least 0.89	$\approx 0.997$

**2.3.27 a-b** Calculate  $R = 93 - 51 = 42$  so that  $s \approx R/4 = 42/4 = 10.5$ .

**c** Calculate  $n = 30$ ,  $\sum x_i = 2145$  and  $\sum x_i^2 = 158,345$ . Then

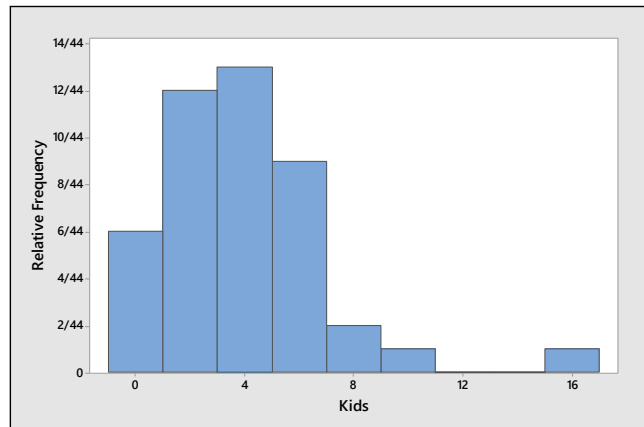
$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{158,345 - \frac{(2145)^2}{30}}{29} = 171.6379 \text{ and } s = \sqrt{171.6379} = 13.101$$

which is fairly close to the approximate value of  $s$  from part **b**.

**d** The two intervals are calculated below. The proportions agree with Tchebysheff's Theorem but are not too close to the percentages given by the Empirical Rule. (This is because the distribution is not quite mound-shaped.)

$k$	$\bar{x} \pm ks$	Interval	Fraction in Interval	Tchebysheff	Empirical Rule
2	$71.5 \pm 26.20$	45.3 to 97.7	$30/30 = 1.00$	at least 0.75	$\approx 0.95$
3	$71.5 \pm 39.30$	32.2 to 110.80	$30/30 = 1.00$	at least 0.89	$\approx 0.997$

**2.3.28 a** Answers will vary. A typical histogram is shown next. The distribution is skewed to the right.



**b** Calculate  $n = 44$ ,  $\sum x_i = 158$  and  $\sum x_i^2 = 926$ . Then

$$\bar{x} = \frac{\sum x_i}{n} = \frac{158}{44} = 3.591,$$

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{926 - \frac{(158)^2}{44}}{43} = 8.34038$$

and  $s = \sqrt{8.34038} = 2.888$

**c** The three intervals,  $\bar{x} \pm ks$  for  $k = 1, 2, 3$  are calculated below. The table shows the actual percentage of measurements falling in a particular interval as well as the percentage predicted by Tchebysheff's Theorem and the Empirical Rule. Note that the Empirical Rule is not very accurate for the first interval, since the histogram in part **a** is skewed.

$k$	$\bar{x} \pm ks$	Interval	Fraction in Interval	Tchebysheff	Empirical Rule
1	$3.591 \pm 2.888$	.703 to 6.419	$34/44 = .77$	at least 0	$\approx 0.68$
2	$3.591 \pm 5.776$	-2.785 to 9.367	$42/44 = .95$	at least 0.75	$\approx 0.95$
3	$3.591 \pm 8.664$	-5.073 to 12.255	$43/44 = .98$	at least 0.89	$\approx 0.997$

**2.3.29 a** Answers will vary. A typical stem and leaf plot is generated by *Minitab*.



**b** Calculate  $n = 16$ ,  $\sum x_i = 386$  and  $\sum x_i^2 = 9476$ . Then  $\bar{x} = \frac{\sum x_i}{n} = \frac{386}{16} = 24.125$ ,



$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{9476 - \frac{(386)^2}{16}}{15} = 10.91667$$

and  $s = \sqrt{s^2} = \sqrt{10.91667} = 3.304$ .

**c** Calculate  $\bar{x} \pm 2s \Rightarrow 24.125 \pm 6.608$  or 17.517 to 30.733. From the original data set, 16 of the 16 measurements, or 100% fall in this interval.

**2.3.30** The following information is available:

$$n = 400, \bar{x} = 600, s^2 = 4900$$

The standard deviation of these scores is then 70, and the results of Tchebysheff's Theorem follow:

$k$	$\bar{x} \pm ks$	Interval	Tchebysheff
1	$600 \pm 70$	530 to 670	at least 0
2	$600 \pm 140$	460 to 740	at least 0.75
3	$600 \pm 210$	390 to 810	at least 0.89

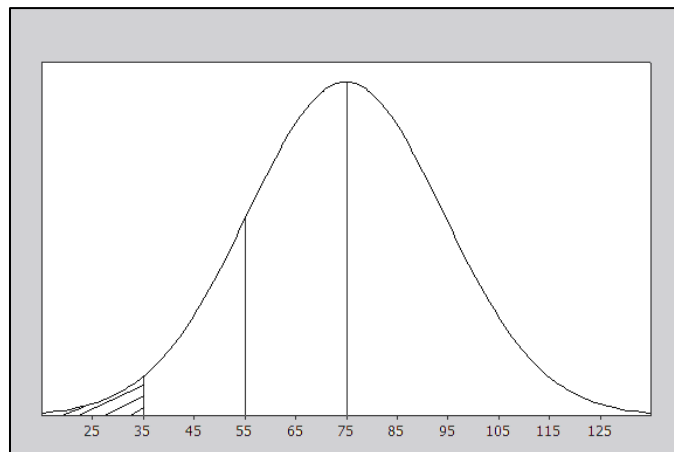
If the distribution of scores is mound-shaped, we use the Empirical Rule, and conclude that approximately 68% of the scores would lie in the interval 530 to 670 (which is  $\bar{x} \pm s$ ). Approximately 95% of the scores would lie in the interval 460 to 740.

**2.3.31** The results of the Empirical Rule follow:

$k$	$\bar{x} \pm ks$	Interval	Empirical Rule
1	$420 \pm 5$	415 to 425	approximately 0.68
2	$420 \pm 10$	410 to 430	approximately 0.95
3	$420 \pm 15$	405 to 435	approximately 0.997

Notice that we are assuming that attendance follows a mound-shaped distribution and hence that the Empirical Rule is appropriate.

**2.3.32 a** It is known that duration times are approximately normal, with mean 75 and standard deviation 20. In order to determine the probability that a commercial lasts less than 35 seconds, we must determine the fraction of the curve which lies within the shaded area in the figure that follows. Using the Empirical Rule, the fraction of the area between 35 and 75 is half of 0.95 or 0.475. Hence, the fraction below 35 would be  $0.5 - 0.475 = 0.025$ .



**b** The fraction of the curve area that lies above the 55 second mark may again be determined by using the Empirical Rule. Refer to the figure in part **a**. The fraction between 55 and 75 is 0.34 and the fraction above 75 is 0.5. Hence, the probability that a commercial lasts longer than 55 seconds is  $0.5 + 0.34 = 0.84$ .

## Section 2.4

**2.4.1** The ordered data are: 0, 1, 3, 4, 4, 5, 6, 6, 6, 7, 7, 8

Calculate  $n=12$ ,  $\sum x_i = 57$  and  $\sum x_i^2 = 337$ . Then  $\bar{x} = \frac{\sum x_i}{n} = \frac{57}{12} = 4.75$  and the sample standard deviation is

$$s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{337 - \frac{(57)^2}{12}}{11}} = \sqrt{6.022727} = 2.454$$

For the smallest observation,  $x=0$ ,  $z\text{-score} = \frac{x-\bar{x}}{s} = \frac{0-4.75}{2.454} = -1.94$  and for the largest observation,

$x=8$ ,  $z\text{-score} = \frac{x-\bar{x}}{s} = \frac{8-4.75}{2.454} = 1.32$ . Since neither  $z$ -score exceeds 2 in absolute value, none of the observations are unusually small or large.

**2.4.2** The ordered data are: 1.0, 1.7, 2.0, 2.1, 2.3, 2.8, 2.9, 4.4, 5.1, 6.5, 7.8

Calculate  $n=11$ ,  $\sum x_i = 38.6$  and  $\sum x_i^2 = 182.3$ . Then  $\bar{x} = \frac{\sum x_i}{n} = \frac{38.6}{11} = 3.509$  and the sample standard deviation is

$$s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{182.30 - \frac{(38.6)^2}{11}}{10}} = \sqrt{4.68491} = 2.164$$

For the smallest observation,  $x=1.0$ ,  $z\text{-score} = \frac{x-\bar{x}}{s} = \frac{1.0-3.509}{2.164} = -1.16$  and for the largest observation,

$x=7.8$ ,  $z\text{-score} = \frac{x-\bar{x}}{s} = \frac{7.8-3.509}{2.164} = 1.98$ . Since neither  $z$ -score exceeds 2 in absolute value, none of the observations are unusually small or large.

**2.4.3** The ordered data are: 2, 3, 4, 5, 6, 6, 6, 7, 8, 9, 9, 10, 25

Calculate  $n=13$ ,  $\sum x_i = 100$  and  $\sum x_i^2 = 1162$ . Then  $\bar{x} = \frac{\sum x_i}{n} = \frac{100}{13} = 7.692$  and the sample standard deviation is

$$s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{1162 - \frac{(100)^2}{13}}{12}} = \sqrt{32.73077} = 5.721$$

For the smallest observation,  $x=2$ ,  $z\text{-score} = \frac{x-\bar{x}}{s} = \frac{2-7.692}{5.721} = -.99$  and for the largest observation,

$x=25$ ,  $z\text{-score} = \frac{x-\bar{x}}{s} = \frac{25-7.692}{5.721} = 3.03$ . The value  $x=25$  is an outlier.

**2.4.4** The ordered data are: 2, 3, 4, 6, 7, 8, 10. With  $n=7$ , the median is in position  $0.5(n+1)=4$ , the lower quartile is in position  $0.25(n+1)=2$  and the upper quartile is in position  $0.75(n+1)=6$ . Hence,  $m=6$ ,  $Q_1=3$  and  $Q_3=8$ .

**2.4.5** The ordered data are: 0, 1, 2, 3, 5, 5, 6, 6, 7. With  $n=9$ , the median is in position  $0.5(n+1)=5$ , the lower quartile is in position  $0.25(n+1)=2.5$  (halfway between the 2<sup>nd</sup> and 3<sup>rd</sup> ordered observations) and the upper

quartile is in position  $0.75(n+1) = 7.5$  (halfway between the 7<sup>th</sup> and 8<sup>th</sup> ordered observations). Hence,  $m = 5$ ,  $Q_1 = (1+2)/2 = 1.5$  and  $Q_3 = (6+6)/2 = 6$ .

- 2.4.6** The ordered data are: 1, 2, 4, 5, 7, 9. With  $n = 6$ , the median is in position  $0.5(n+1) = 3.5$  (halfway between the 3<sup>rd</sup> and 4<sup>th</sup> ordered observations), the lower quartile is in position  $0.25(n+1) = 1.75$  (3/4 of the way between the 1<sup>st</sup> and 2<sup>nd</sup> ordered observations) and the upper quartile is in position  $0.75(n+1) = 5.25$  (1/4 of the way between the 5<sup>th</sup> and 6<sup>th</sup> ordered observations). Hence,  $m = (4+5)/2 = 4.5$ ,  $Q_1 = 1 + .75(2-1) = 1.75$  and  $Q_3 = 7 + .25(9-7) = 7.5$ .

- 2.4.7** The ordered data are: 0, 1, 5, 6, 7, 8, 9, 10, 12, 12, 13, 14, 16, 19, 19. With  $n = 15$ , the median is in position  $0.5(n+1) = 8$ , so that  $m = 10$ . The lower quartile is in position  $0.25(n+1) = 4$  so that  $Q_1 = 6$  and the upper quartile is in position  $0.75(n+1) = 12$  so that  $Q_3 = 14$ . Then the five-number summary is

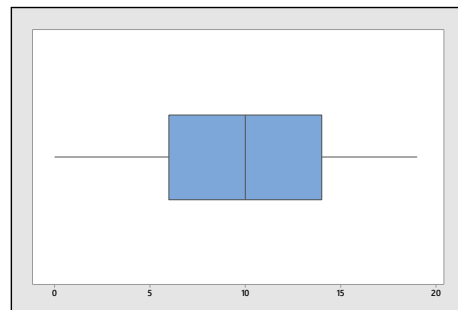
Min	$Q_1$	Median	$Q_3$	Max
0	6	10	14	19

and  $IQR = Q_3 - Q_1 = 14 - 6 = 8$ . The lower and upper fences are:

$$Q_1 - 1.5IQR = 6 - 12 = -6$$

$$Q_3 + 1.5IQR = 14 + 12 = 26$$

There are no data points that lie outside the fences. The box plot is shown next. The lower whisker connects the box to the smallest value that is not an outlier,  $x = 0$ . The upper whisker connects the box to the largest value that is not an outlier or  $x = 19$ .



- 2.4.8** Notice that the data is already ranked from smallest to largest. With  $n = 8$ , the median is in position  $0.5(n+1) = 4.5$  (halfway between the 4<sup>th</sup> and 5<sup>th</sup> observations), the lower quartile is in position  $0.25(n+1) = 2.25$  (one-fourth of the way between the 2<sup>nd</sup> and 3<sup>rd</sup> observations) and the upper quartile is in position  $0.75(n+1) = 6.75$  (three-fourths of the way between the 6<sup>th</sup> and 7<sup>th</sup> observations). Hence,  $m = (.41 + .56)/2 = .485$ ,  $Q_1 = .30 + 0.25(.35 - .30) = .3125$  and  $Q_3 = .58 + 0.75(.76 - .58) = .7150$ . The five-number summary is

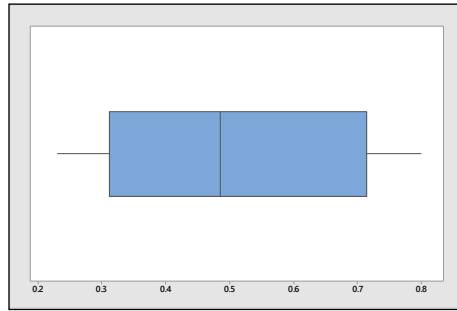
Min	$Q_1$	Median	$Q_3$	Max
.23	.3125	.485	.7150	.80

and  $IQR = Q_3 - Q_1 = .7150 - .3125 = .4025$ . The upper and lower fences are then calculated as:

$$\text{Lower fence} = Q_1 - 1.5IQR = .3125 - 1.5(.4025) = -.29125$$

$$\text{Upper fence} = Q_3 + 1.5IQR = .7150 + 1.5(.4025) = 1.31875$$

There are no data points that lie outside these fences. The box plot is shown next. The lower whisker connects the box to the smallest value that is not an outlier,  $x = .23$ . The upper whisker connects the box to the largest value that is not an outlier or  $x = .80$ .



**2.4.9** The ordered data are: 12, 18, 22, 23, 24, 25, 25, 26, 26, 27, 28. For  $n = 11$ , the position of the median is  $0.5(n+1) = 0.5(11+1) = 6$  and  $m = 25$ . The positions of the quartiles are  $0.25(n+1) = 3$  and  $0.75(n+1) = 9$ , so that  $Q_1 = 22$ ,  $Q_3 = 26$ , and  $IQR = 26 - 22 = 4$ . The five-number summary is

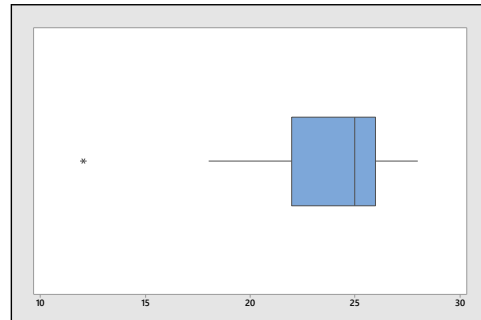
Min	$Q_1$	Median	$Q_3$	Max
12	22	25	26	28

The lower and upper fences are:

$$Q_1 - 1.5IQR = 22 - 6 = 16$$

$$Q_3 + 1.5IQR = 26 + 6 = 32$$

The only observation falling outside the fences is  $x = 12$  which is identified as an outlier. The box plot is shown next. The lower whisker connects the box to the smallest value that is not an outlier,  $x = 18$ . The upper whisker connects the box to the largest value that is not an outlier or  $x = 28$ .



**2.4.10 Mean, standard deviation and z-scores:** The ordered data are: 4, 8, 9, 10, 10, 11, 11, 12, 13, 13, 14, 15, 17, 19, 29. Calculate  $n = 15$ ,  $\sum x_i = 195$  and  $\sum x_i^2 = 2997$ . Then  $\bar{x} = \frac{\sum x_i}{n} = \frac{195}{15} = 13.00$  and the sample standard deviation is

$$s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{2997 - \frac{(195)^2}{15}}{14}} = 5.7446$$

For the smallest observation,  $x = 4$ ,  $z\text{-score} = \frac{x - \bar{x}}{s} = \frac{4 - 13}{5.7446} = -1.57$  and for the largest observation,

$x = 29$ ,  $z\text{-score} = \frac{x - \bar{x}}{s} = \frac{29 - 13}{5.7446} = 2.79$ . The value  $x = 29$  is a suspect outlier.

**Five-number summary and box plot:** For  $n = 15$ , the position of the median is  $0.5(n+1) = 0.5(15+1) = 8$  and  $m = 12$ . The positions of the quartiles are  $0.25(n+1) = 4$  and  $0.75(n+1) = 12$ , so that  $Q_1 = 10$ ,  $Q_3 = 15$ , and  $IQR = 15 - 10 = 5$ . The five-number summary is

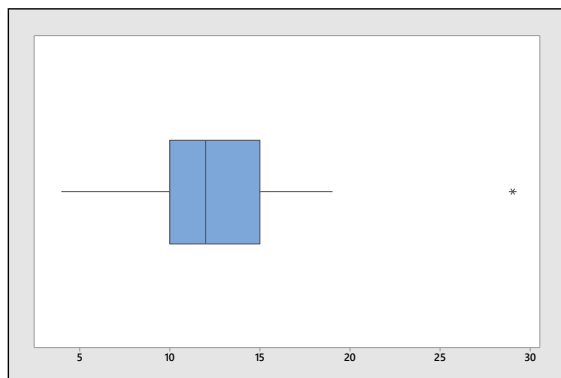
Min	Q <sub>1</sub>	Median	Q <sub>3</sub>	Max
4	10	12	15	29

The lower and upper fences are:

$$Q_1 - 1.5IQR = 10 - 7.5 = 2.5$$

$$Q_3 + 1.5IQR = 15 + 7.5 = 22.5$$

The only observation falling outside the fences is  $x = 29$  which is identified as an outlier. The box plot is shown next. The lower whisker connects the box to the smallest value that is not an outlier,  $x = 4$ . The upper whisker connects the box to the largest value that is not an outlier or  $x = 19$ .



The results of the two different methods for identifying outliers are similar.

**2.4.11 Mean, standard deviation and z-scores:** From Exercise 3, the ordered data are: 2, 3, 4, 5, 6, 6, 6, 7, 8, 9, 9, 10, 25. Calculate  $n = 13$ ,  $\sum x_i = 100$  and  $\sum x_i^2 = 1162$ . Then  $\bar{x} = \frac{\sum x_i}{n} = \frac{100}{13} = 7.692$  and the sample standard deviation is

$$s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{1162 - \frac{(100)^2}{13}}{12}} = \sqrt{32.73077} = 5.721$$

For the smallest observation,  $x = 2$ ,  $z\text{-score} = \frac{x - \bar{x}}{s} = \frac{2 - 7.692}{5.721} = -0.99$  and for the largest observation,

$x = 25$ ,  $z\text{-score} = \frac{x - \bar{x}}{s} = \frac{25 - 7.692}{5.721} = 3.03$ . The value  $x = 25$  is an outlier.

**Five-number summary and box plot:** For  $n = 13$ , the position of the median is  $0.5(n+1) = 0.5(13+1) = 7$  and  $m = 6$ . The positions of the quartiles are  $0.25(n+1) = 3.5$  and  $0.75(n+1) = 10.5$ , so that  $Q_1 = 4.5$ ,  $Q_3 = 9$ , and  $IQR = 9 - 4.5 = 4.5$ . The five-number summary is

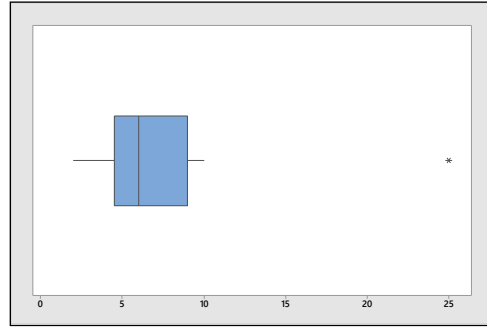
Min	Q <sub>1</sub>	Median	Q <sub>3</sub>	Max
2	4.5	6	9	25

The lower and upper fences are:

$$Q_1 - 1.5IQR = 4.5 - 6.75 = -2.25$$

$$Q_3 + 1.5IQR = 9 + 6.75 = 15.75$$

The only observation falling outside the fences is  $x = 25$  which is identified as an outlier. The box plot is shown next. The lower whisker connects the box to the smallest value that is not an outlier,  $x = 2$ . The upper whisker connects the box to the largest value that is not an outlier or  $x = 10$ .



The results of the two different methods for identifying outliers are identical.

**2.4.12** The 69<sup>th</sup> percentile ( $x = 78$ ) implies that 69% of all students scored below 78, and only 31% scored higher.

**2.4.13** Since 14.5% of U.S. men are six feet or taller,  $(100 - 14.5)\% = 85.5\%$  are less than six feet tall. Six feet is the 85.5<sup>th</sup> percentile.

**2.4.14** Ninety percent of all students scored lower than 513, while 10% scored higher.

**2.4.15** Since 46% of all 19-year-old females in a certain height-weight category have a BMI greater than 21.9,  $(100 - 46)\% = 54\%$  have a BMI less than or equal to 21.9. The value 21.9 is the 54<sup>th</sup> percentile.

**2.4.16** Calculate  $n = 25$ ,  $\sum x_i = 155.5$  and  $\sum x_i^2 = 1260.75$ . Then  $\bar{x} = \frac{\sum x_i}{n} = \frac{155.5}{25} = 6.22$  and

$$s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{1260.75 - \frac{(155.5)^2}{25}}{24}} = 3.497$$

For the smallest observation,  $x = 1.0$ ,  $z\text{-score} = \frac{x - \bar{x}}{s} = \frac{1 - 6.22}{3.497} = -1.49$  and for the largest observation,

$x = 15$ ,  $z\text{-score} = \frac{x - \bar{x}}{s} = \frac{15 - 6.22}{3.497} = 2.51$ . The value  $x = 15$  is a suspect outlier.

**2.4.17** Calculate  $\sum x_i = 28.41$  and  $\sum x_i^2 = 30.6071$ , the sample mean is  $\bar{x} = \frac{\sum x_i}{n} = \frac{28.41}{27} = 1.052$  and the standard deviation of the sample is

$$s = \sqrt{s^2} = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{30.6071 - \frac{(28.41)^2}{27}}{26}} = 0.166$$

For the smallest observation,  $x = .75$ ,  $z\text{-score} = \frac{x - \bar{x}}{s} = \frac{.75 - 1.052}{.166} = -1.82$  and for the largest observation,

$x = 1.41$ ,  $z\text{-score} = \frac{x - \bar{x}}{s} = \frac{1.41 - 1.052}{.166} = 2.16$ . The value  $x = 1.41$  is a suspect outlier.

**2.4.18** For  $n = 14$ , the position of the median is  $0.5(n+1) = 7.5$  and the positions of the quartiles are  $0.25(n+1) = 3.75$  and  $0.75(n+1) = 11.25$ . The lower quartile is  $\frac{3}{4}$  the way between the 3<sup>rd</sup> and 4<sup>th</sup> measurements or  $Q_1 = 0.60 + 0.75(0.63 - 0.60) = 0.6225$  and the upper quartile is  $\frac{1}{4}$  the way between the 11<sup>th</sup> and 12<sup>th</sup> measurements or  $Q_3 = 1.12 + 0.25(1.23 - 1.12) = 1.1475$ . Then the five-number summary is

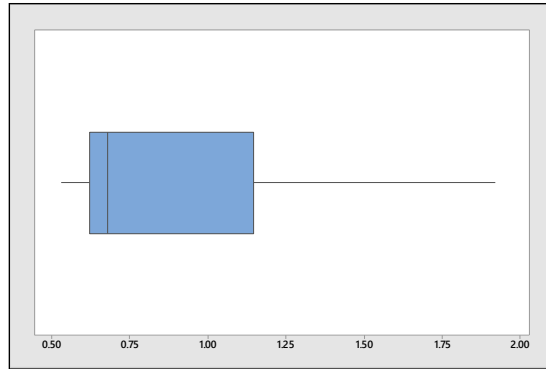
Min	Q <sub>1</sub>	Median	Q <sub>3</sub>	Max
0.53	0.6225	0.68	1.1475	1.92

Calculate  $IQR = Q_3 - Q_1 = 1.1475 - 0.6225 = 0.5250$ . Then the *lower and upper fences* are:

$$Q_1 - 1.5IQR = 0.6225 - 0.7875 = -0.165$$

$$Q_3 + 1.5IQR = 1.1475 + 0.7875 = 1.935$$

The box plot is shown next. Since there are no outliers, the whiskers connect the box to the minimum and maximum values in the ordered set.



**2.4.19** The ordered set is: 40, 49, 52, 54, 59, 61, 67, 69, 70, 71. Since  $n = 10$ , the positions of  $m$ ,  $Q_1$ , and  $Q_3$  are 5.5, 2.75 and 8.25 respectively, and  $m = (59 + 61)/2 = 60$ ,  $Q_1 = 49 + 0.75(52 - 49) = 51.25$ ,  $Q_3 = 69 + .25(70 - 69) = 69.25$  and  $IQR = 69.25 - 51.25 = 18$ . Then the five-number summary is

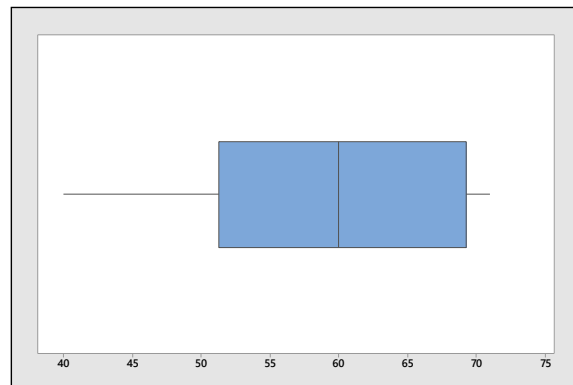
Min	$Q_1$	Median	$Q_3$	Max
40	51.25	60	69.25	71

The *lower and upper fences* are:

$$Q_1 - 1.5IQR = 51.25 - 27 = 24.25$$

$$Q_3 + 1.5IQR = 69.25 + 27 = 96.25$$

and the box plot is shown next. There are no outliers and the data set is slightly skewed left.



**2.4.20 a** The ordered data are shown below:

1.70	101.00	209.00	264.00	316.00	445.00
1.72	118.00	218.00	278.00	318.00	481.00
5.90	168.00	221.00	286.00	329.00	485.00
8.80	180.00	241.00	314.00	397.00	
85.40	183.00	252.00	315.00	406.00	

For  $n = 28$ , the position of the median is  $0.5(n+1) = 14.5$  and the positions of the quartiles are  $0.25(n+1) = 7.25$  and  $0.75(n+1) = 21.75$ . The lower quartile is  $\frac{1}{4}$  the way between the 7<sup>th</sup> and 8<sup>th</sup> measurements or  $Q_1 = 118 + 0.25(168 - 118) = 130.5$  and the upper quartile is  $\frac{3}{4}$  the way between the 21<sup>st</sup> and 22<sup>nd</sup> measurements or  $Q_3 = 316 + 0.75(318 - 316) = 317.5$ . Then the five-number summary is

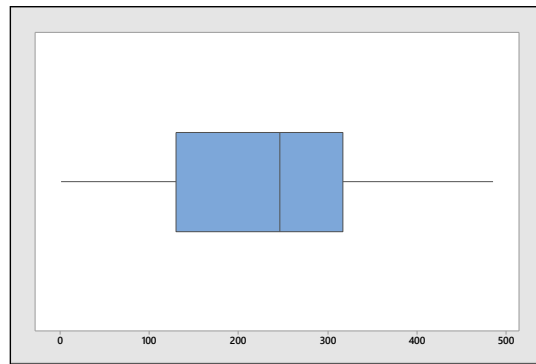
Min	$Q_1$	Median	$Q_3$	Max
1.70	130.5	246.5	317.5	485

**b** Calculate  $IQR = Q_3 - Q_1 = 317.5 - 130.5 = 187$ . Then the *lower and upper fences* are:

$$Q_1 - 1.5IQR = 130.5 - 280.5 = -150$$

$$Q_3 + 1.5IQR = 317.5 + 280.5 = 598$$

The box plot is shown next. Since there are no outliers, the whiskers connect the box to the minimum and maximum values in the ordered set.



**c-d** The boxplot does not identify any of the measurements as outliers, mainly because the large variation in the measurements cause the IQR to be large. However, the student should notice the extreme difference in the magnitude of the first four observations taken on young dolphins. These animals have not been alive long enough to accumulate a large amount of mercury in their bodies.

**2.4.21 a** For  $n = 15$ , the position of the median is  $0.5(n+1) = 8$  and the positions of the quartiles are  $0.25(n+1) = 4$  and  $0.75(n+1) = 12$ , while for  $n = 16$ , the position of the median is  $0.5(n+1) = 8.5$  and the positions of the quartiles are  $0.25(n+1) = 4.25$  and  $0.75(n+1) = 12.75$ . The sorted measurements are shown below.

**Alex Smith:** 14, 16, 19, 19, 20, 21, 23, 23, 25, 25, 25, 27, 27, 28, 29

**Joe Flacco:** 8, 9, 10, 19, 20, 20, 22, 23, 24, 25, 25, 26, 27, 29, 31, 34

For Alex Smith,

$$m = 23, Q_1 = 19 \text{ and } Q_3 = 27.$$

For Joe Flacco,

$$m = (23 + 24)/2 = 23.5, Q_1 = 19 + 0.25(20 - 19) = 19.25 \text{ and } Q_3 = 26 + 0.75(27 - 26) = 26.75.$$

Then the five-number summaries are

	Min	$Q_1$	Median	$Q_3$	Max
Smith	14	19	23	27	29
Flacco	8	19.25	23.5	26.75	34



**b** For Alex Smith, calculate  $IQR = Q_3 - Q_1 = 27 - 19 = 8$ . Then the *lower and upper fences* are:

$$Q_1 - 1.5IQR = 19 - 12 = 7$$

$$Q_3 + 1.5IQR = 27 + 12 = 39$$

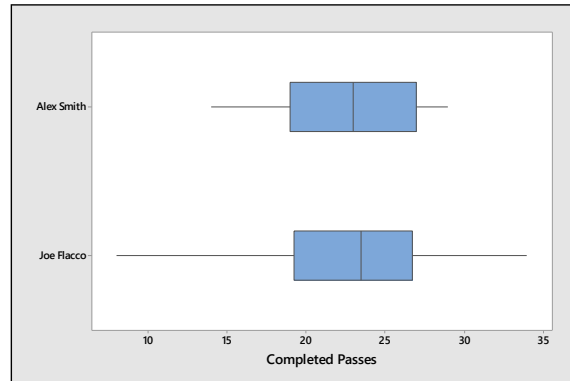
and there are no outliers.

For Joe Flacco, calculate  $IQR = Q_3 - Q_1 = 26.75 - 19.25 = 7.5$ . Then the *lower and upper fences* are:

$$Q_1 - 1.5IQR = 19.25 - 11.25 = 8$$

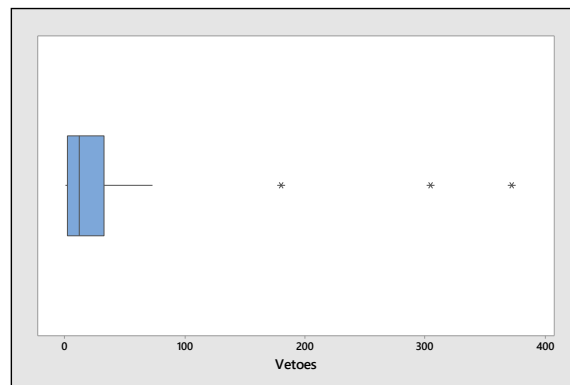
$$Q_3 + 1.5IQR = 26.75 + 11.25 = 38$$

and there are no outliers. The box plots are shown next.



**c** Answers will vary. Both distributions are relatively symmetric and somewhat mound-shaped. The Flacco distribution is more variable, but Flacco has a slightly higher median number of completed passes.

**2.4.22** Answers will vary from student to student. The distribution is skewed to the right with three outliers (Truman, Cleveland and F. Roosevelt). The box plot is shown next.



**2.4.23 a** Just by scanning through the 20 measurements, it seems that there are a few unusually small measurements, which would indicate a distribution that is skewed to the left.

**b** The position of the median is  $0.5(n+1) = 0.5(25+1) = 10.5$  and  $m = (120+127)/2 = 123.5$ . The mean is

$$\bar{x} = \frac{\sum x_i}{n} = \frac{2163}{20} = 108.15$$

which is smaller than the median, indicate a distribution skewed to the left.

**c** The positions of the quartiles are  $0.25(n+1) = 5.25$  and  $0.75(n+1) = 15.75$ , so that

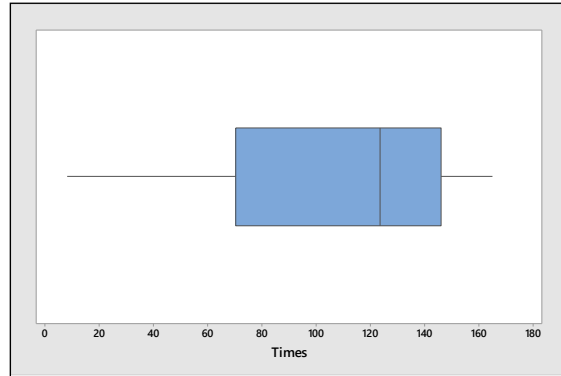
$$Q_1 = 65 - .25(87 - 65) = 70.5, Q_3 = 144 + .75(147 - 144) = 146.25, \text{ and } IQR = 146.25 - 70.5 = 75.75.$$

The *lower and upper fences* are:

$$Q_1 - 1.5IQR = 70.5 - 113.625 = -43.125$$

$$Q_3 + 1.5IQR = 146.25 + 113.625 = 259.875$$

The box plot is shown next. There are no outliers. The long left whisker and the median line located to the right of the center of the box indicates that the distribution that is skewed to the left.



**2.4.24 a** The sorted data is:

233.97, 243.92, 247.34, 252.44, 255.40, 273.80

286.35, 286.41, 383.68, 408.48, 446.30, 459.21

The positions of the median and the quartiles are  $0.5(n+1) = 6.5$ ,  $0.25(n+1) = 3.25$  and  $0.75(n+1) = 9.75$ ,

$$m = (273.80 + 286.35) / 2 = 280.075$$

so that  $Q_1 = 247.34 + .25(252.44 - 247.34) = 248.615$

$$Q_3 = 383.68 + .75(408.48 - 383.68) = 402.28$$

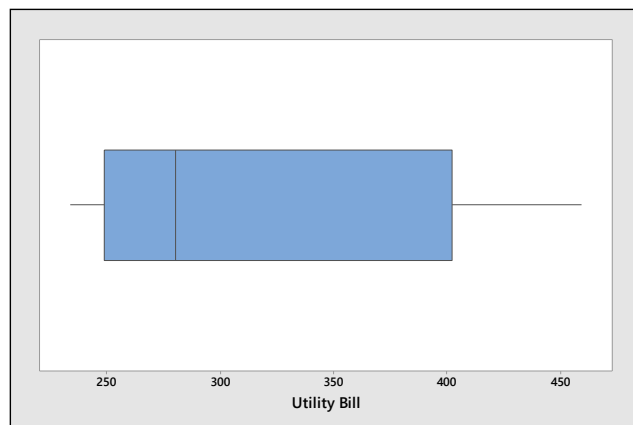
and  $IQR = 402.28 - 248.615 = 153.665$ .

The *lower and upper fences* are:

$$Q_1 - 1.5IQR = 248.615 - 230.4975 = 18.1175$$

$$Q_3 + 1.5IQR = 402.28 + 230.4975 = 632.7775$$

There are no outliers, and the box plot is shown next.



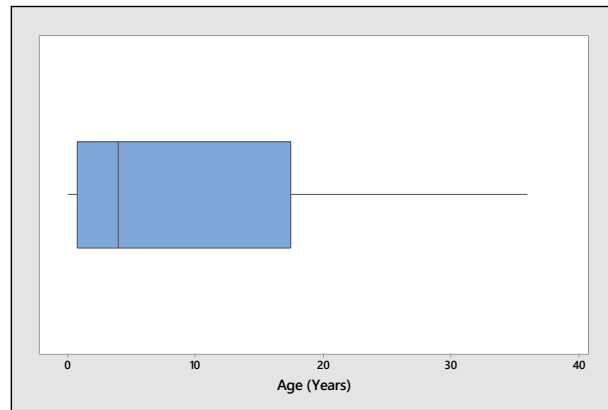
**b** Because of the slightly longer right whisker, there are a few unusually large bills (probably in the summer due to air conditioning). A large portion of the bills are in the \$240 to \$285 range.

**2.4.25 a** Calculate  $n = 50$ ,  $\sum x_i = 418$ , so that  $\bar{x} = \frac{\sum x_i}{n} = \frac{418}{50} = 8.36$ .

- b** The position of the median is  $.5(n+1) = 25.5$  and  $m = (4 + 4)/2 = 4$ .
- c** Since the mean is larger than the median, the distribution is skewed to the right.
- d** Since  $n = 50$ , the positions of  $Q_1$  and  $Q_3$  are  $.25(51) = 12.75$  and  $.75(51) = 38.25$ , respectively. Then  $Q_1 = 0 + 0.75(1 - 0) = 12.75$ ,  $Q_3 = 17 + .25(19 - 17) = 17.5$  and  $IQR = 17.5 - .75 = 16.75$ .

The lower and upper fences are:  $Q_1 - 1.5IQR = .75 - 25.125 = -24.375$   
 $Q_3 + 1.5IQR = 17.5 + 25.125 = 42.625$

and the box plot is shown next. There are no outliers and the data is skewed to the right.



**2.4.26** Each bulleted statement produces a percentile.

- $x$  = hours per week spent volunteering. The value  $x = 5$  is the  $(100 - 12) = 88^{\text{th}}$  percentile.
- $x$  = age in years of currently operated cars. The value  $x = 8$  is the  $(100 - 58) = 42^{\text{nd}}$  percentile.
- $x$  = amount paid for a ticket to a top concert tour. The value  $x = 75$  is the  $(100 - 22) = 78^{\text{th}}$  percentile.

## Reviewing What You've Learned

**2.R.1 a** Calculate  $n = 14$ ,  $\sum x_i = 367$  and  $\sum x_i^2 = 9641$ . Then  $\bar{x} = \frac{\sum x_i}{n} = \frac{367}{14} = 26.214$  and

$$s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{9641 - \frac{(367)^2}{14}}{13}} = 1.251$$

**b** Calculate  $n = 14$ ,  $\sum x_i = 366$  and  $\sum x_i^2 = 9644$ . Then  $\bar{x} = \frac{\sum x_i}{n} = \frac{366}{14} = 26.143$  and

$$s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{9644 - \frac{(366)^2}{14}}{13}} = 2.413$$

**c** The centers are roughly the same; the Sunmaid raisins appear slightly more variable.

**2.R.2 a** The ordered sets are shown as follows:

Generic					Sunmaid				
24	25	25	25	26	22	24	24	24	24

26	26	26	26	27		25	25	27	28	28
27	28	28	28			28	28	29	30	

For  $n = 14$ , the position of the median is  $0.5(n+1) = 0.5(14+1) = 7.5$  and the positions of the quartiles are  $0.25(n+1) = 3.75$  and  $0.75(n+1) = 11.25$ , so that

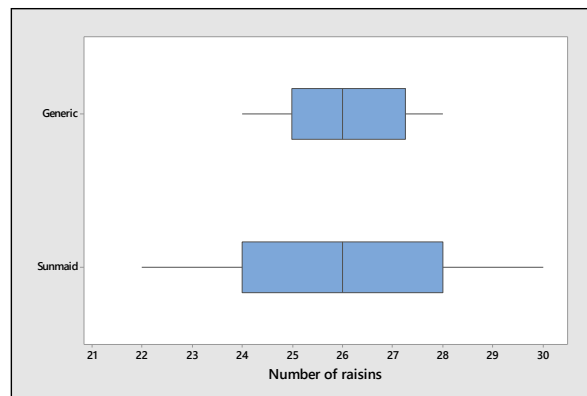
**Generic:**  $m = 26$ ,  $Q_1 = 25$ ,  $Q_3 = 27.25$ , and  $IQR = 27.25 - 25 = 2.25$

**Sunmaid:**  $m = 26$ ,  $Q_1 = 24$ ,  $Q_3 = 28$ , and  $IQR = 28 - 24 = 4$

**b Generic:** Lower and upper fences are:  $Q_1 - 1.5IQR = 25 - 3.375 = 21.625$   
 $Q_3 + 1.5IQR = 27.25 + 3.375 = 30.625$

**Sunmaid:** Lower and upper fences are:  $Q_1 - 1.5IQR = 24 - 6 = 18$   
 $Q_3 + 1.5IQR = 28 + 6 = 34$

The box plots are shown next. There are no outliers.



**c** The two stem and leaf plots, generated by *Minitab* are shown next. If the plots were rotated left by  $90^\circ$ , they would reflect the same shape as shown in the box plots.

**Stem-and-leaf of Generic N = 14**

1	24	0
4	25	000
(5)	26	00000
5	27	00
3	28	000

Leaf Unit = 0.1

**Stem-and-leaf of Sunmaid N = 14**

1	22	0
1	23	
5	24	0000
7	25	00
7	26	
7	27	0
6	28	0000
2	29	0
1	30	0

Leaf Unit = 0.1

**d** If the boxes are not being underfilled, the average size of the raisins is roughly the same for the two brands. However, since the number of raisins is more variable for the Sunmaid brand, it would appear that some of the Sunmaid raisins are large while others are small. The individual sizes of the generic raisins are not as variable.

**2.R.3 a** Calculate  $R = 499.9 - 219.9 = 280$  so that  $s \approx R/4 = 280/4 = 70$ .

**b** Calculate  $n = 25$ ,  $\sum x_i = 7996.3$  and  $\sum x_i^2 = 2634119.27$ . Then  $\bar{x} = \frac{7996.3}{25} = 319.852$

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{2,634,119.27 - \frac{(7996.3)^2}{25}}{24} = 3186.946767 \text{ and } s = \sqrt{3186.946767} = 56.453$$

which is of the same order of magnitude as the approximate value of  $s$  from part **a**.

**c** Calculate  $\bar{x} \pm 2s \Rightarrow 319.852 \pm 112.906$  or 206.946 to 432.758. The proportion of measurements in this interval is  $24/25 = .96$ , close to the proportion given by the Empirical Rule.

**2.R.4 a** The largest observation found in the data is 32.3, while the smallest is 0.2. Therefore, the range is  $R = 32.3 - 0.2 = 32.1$ .

**b** Using the range, the approximate value for  $s$  is:  $s \approx R/4 = 32.1/4 = 8.025$ .

**c** Calculate  $n = 50$ ,  $\sum x_i = 418.4$  and  $\sum x_i^2 = 6384.34$ . Then

$$s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{6384.34 - \frac{(418.4)^2}{50}}{49}} = 7.671$$

**2.R.5 a** Refer to Exercise 4. Since  $\sum x_i = 418.4$ , the sample mean is  $\bar{x} = \frac{\sum x_i}{n} = \frac{418.4}{50} = 8.368$ .

The three intervals of interest is shown in the table, along with the number of observations which fall in each interval.

$k$	$\bar{x} \pm ks$	Interval	Number in Interval	Percentage
1	$8.368 \pm 7.671$	0.697 to 16.039	37	74%
2	$8.368 \pm 15.342$	-6.974 to 23.710	47	94%
3	$8.368 \pm 23.013$	-14.645 to 31.381	49	98%

**b** The percentages falling in the intervals do agree with Tchebysheff's Theorem. At least 0 fall in the first interval, at least  $3/4 = 0.75$  fall in the second interval, and at least  $8/9 = 0.89$  fall in the third. The percentages are not too close to the percentages described by the Empirical Rule (68%, 95%, and 99.7%).

**c** The Empirical Rule may be unsuitable for describing these data. The data distribution does not have a strong mound-shape (see the relative frequency histogram in the solution to Exercise 25, Section 1.4), but is skewed to the right.

**2.R.6** The ordered data are shown below.

0.2	2.0	4.3	8.2	14.7
0.2	2.1	4.4	8.3	16.7
0.3	2.4	5.6	8.7	18.0
0.4	2.4	5.8	9.0	18.0
1.0	2.7	6.1	9.6	18.4
1.2	3.3	6.6	9.9	19.2
1.3	3.5	6.9	11.4	23.1
1.4	3.7	7.4	12.6	24.0
1.6	3.9	7.4	13.5	26.7
1.6	4.1	8.2	14.1	32.3

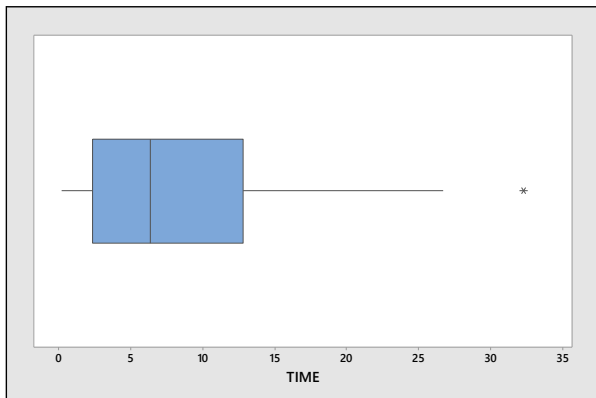
Since  $n = 50$ , the position of the median is  $0.5(n+1) = 25.5$  and the positions of the lower and upper quartiles are  $0.25(n+1) = 12.75$  and  $0.75(n+1) = 38.25$ . Then

$$m = (6.1 + 6.6)/2 = 6.35, Q_1 = 2.1 + 0.75(2.4 - 2.1) = 2.325$$

$$Q_3 = 12.6 + 0.25(13.5 - 12.6) = 12.825 \text{ and } IQR = 12.825 - 2.325 = 10.5.$$

The lower and upper fences are:  $Q_1 - 1.5IQR = 2.325 - 15.75 = -13.425$   
 $Q_3 + 1.5IQR = 12.825 + 15.75 = 28.575$

and the box plot is shown next. There is one outlier,  $x = 32.3$ . The distribution is skewed to the right.



**2.R.7** First calculate the intervals:

$$\bar{x} \pm s = 0.17 \pm 0.01 \quad \text{or } 0.16 \text{ to } 0.18$$

$$\bar{x} \pm 2s = 0.17 \pm 0.02 \quad \text{or } 0.15 \text{ to } 0.19$$

$$\bar{x} \pm 3s = 0.17 \pm 0.03 \quad \text{or } 0.14 \text{ to } 0.20$$

**a** If no prior information as to the shape of the distribution is available, we use Tchebysheff's Theorem. We would expect at least  $(1 - 1/1^2) = 0$  of the measurements to fall in the interval 0.16 to 0.18; at least  $(1 - 1/2^2) = 3/4$  of the measurements to fall in the interval 0.15 to 0.19; at least  $(1 - 1/3^2) = 8/9$  of the measurements to fall in the interval 0.14 to 0.20.

**b** According to the Empirical Rule, approximately 68% of the measurements will fall in the interval 0.16 to 0.18; approximately 95% of the measurements will fall between 0.15 to 0.19; approximately 99.7% of the measurements will fall between 0.14 and 0.20. Since mound-shaped distributions are so frequent, if we do have a sample size of 30 or greater, we expect the sample distribution to be mound-shaped. Therefore, in this exercise, we would expect the Empirical Rule to be suitable for describing the set of data.

**c** If the chemist had used a sample size of four for this experiment, the distribution would not be mound-shaped. Any possible histogram we could construct would be non-mound-shaped. We can use at most 4 classes, each with frequency 1, and we will not obtain a histogram that is even close to mound-shaped. Therefore, the Empirical Rule would not be suitable for describing  $n = 4$  measurements.

**2.R.8** Since it is obvious that the distribution of amount of chloroform per liter of water in various water sources is not mound-shaped (one standard deviation below the mean is negative, which is impossible since chloroform amounts are always zero or greater), we cannot make this assumption. Tchebysheff's Theorem can be used, however, and the necessary intervals and fractions falling in these intervals are given in the table.

$k$	$\bar{x} \pm ks$	Interval	Tchebysheff
1	$34 \pm 53$	-19 to 87	at least 0
2	$34 \pm 106$	-72 to 140	at least 0.75
3	$34 \pm 159$	-125 to 193	at least 0.89

- 2.R.9 a** For  $n=10$ , the position of the median is  $0.5(n+1)=5.5$  and the positions of the quartiles are  $0.25(n+1)=2.75$  and  $0.75(n+1)=8.25$ . The sorted data are: 5, 6, 6, 6.75, 7, 7, 7, 7.25, 8, 8.5

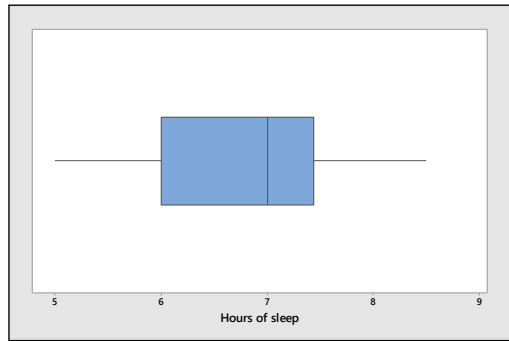
Then  $m = (7+7)/2 = 7$ ,  $Q_1 = 6 + 0.75(6-6) = 6$  and  $Q_3 = 7.25 + 0.25(8-7.25) = 7.4375$ .

Calculate  $IQR = 7.4375 - 6 = 1.4375$  and the *lower and upper fences*:

$$Q_1 - 1.5IQR = 6 - 2.15625 = 3.84$$

$$Q_3 + 1.5IQR = 7.4375 + 2.15625 = 9.59$$

The box plot is shown next.



- b** There are no outliers, not even  $x = 8.5$ !
- 2.R.10 a** Calculate  $n = 20$ ,  $\sum x_i = 479.2$ ,  $\sum x_i^2 = 11532.82$ . Then

$$\bar{x} = \frac{\sum x_i}{n} = \frac{479.2}{20} = 23.96$$

$$s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{11532.82 - \frac{(479.2)^2}{20}}{19}} = \sqrt{2.694} = 1.641$$

The sorted data is shown below:

20.2	21.3	22.2	22.7	22.9
23.1	23.2	23.6	23.7	24.2
24.4	24.4	24.6	24.7	24.7
24.9	25.3	25.9	26.2	27.0

The  $z$ -scores for  $x = 20.2$  and  $x = 27$  are

$$z = \frac{x - \bar{x}}{s} = \frac{20.2 - 23.96}{1.641} = -2.29 \text{ and } z = \frac{x - \bar{x}}{s} = \frac{27 - 23.96}{1.641} = 1.85$$

Since neither of the  $z$ -scores are greater than 3 in absolute value, the measurements are not judged to be outliers.

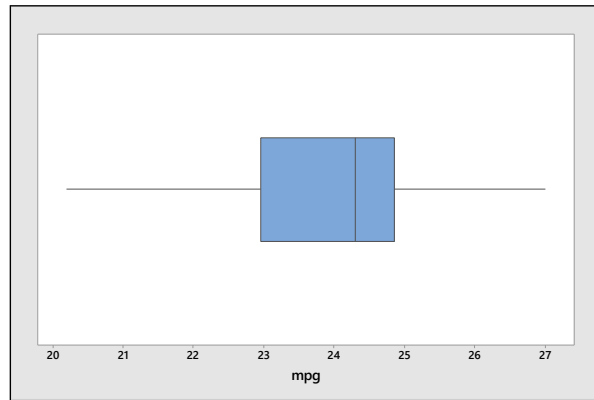
- b** The position of the median is  $0.5(n+1)=10.5$  and the median is  $m = (24.2 + 24.4)/2 = 24.3$ . The positions of the quartiles are  $0.25(n+1)=5.25$  and  $0.75(n+1)=15.75$ . Then  $Q_1 = 22.9 + 0.25(23.1 - 22.9) = 22.95$  and  $Q_3 = 24.7 + 0.75(24.9 - 24.7) = 24.85$ .

- c** Calculate  $IQR = 24.85 - 22.95 = 1.9$ . The *lower and upper fences* are:

$$Q_1 - 1.5IQR = 22.95 - 2.85 = 20.10$$

$$Q_3 + 1.5IQR = 24.85 + 2.85 = 27.70$$

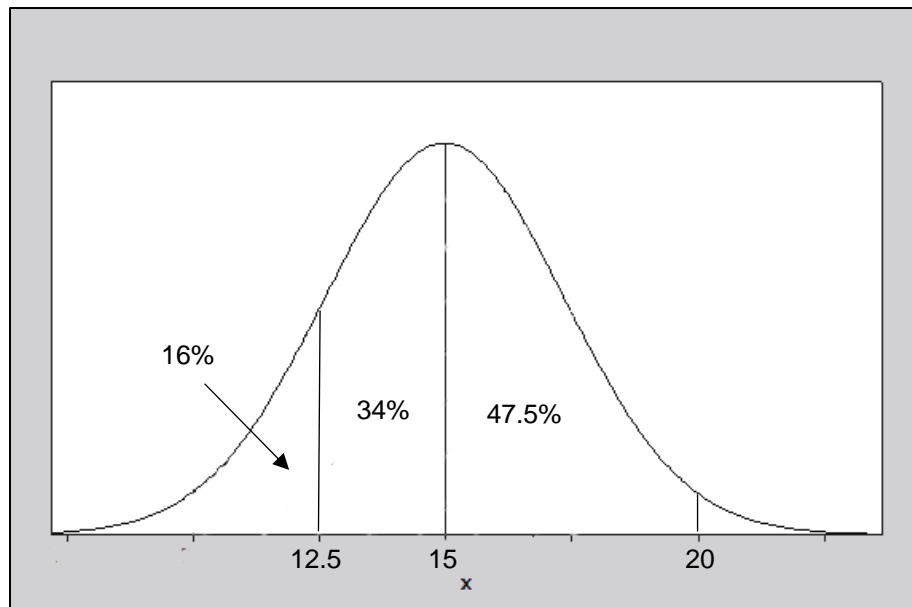
There are no outliers, which confirms the conclusion in part a. The box plot is shown next.



**2.R.11** If the distribution is mound-shaped, then almost all of the measurements will fall in the interval  $\mu \pm 3\sigma$ , which is an interval  $6\sigma$  in length. That is, the range of the measurements should be approximately  $6\sigma$ . In this case, the range is  $800 - 200 = 600$ , so that  $\sigma \approx 600/6 = 100$ .

**2.R.12** The stem lengths are approximately normal with mean 15 and standard deviation 2.5.

**a** In order to determine the percentage of roses with length less than 12.5, we must determine the proportion of the curve which lies within the shaded area in the figure that follows. Using the Empirical Rule, the proportion of the area between 12.5 and 15 is half of 0.68 or 0.34. Hence, the fraction below 12.5 would be  $0.5 - 0.34 = 0.16$  or 16%.



**b** Refer to the figure shown in part a. Again we use the Empirical Rule. The proportion of the area between 12.5 and 15 is half of 0.68 or 0.34, while the proportion of the area between 15 and 20 is half of 0.95 or 0.475. The total area between 12.5 and 20 is then  $0.34 + 0.475 = .815$  or 81.5%.

**2.R.13 a** The range is  $R = 172 - 108 = 64$  and the range approximation is

$$s \approx R/4 = 64/4 = 16$$

**b** Calculate  $n = 15$ ,  $\sum x_i = 2041$ ,  $\sum x_i^2 = 281,807$ . Then  $\bar{x} = \frac{\sum x_i}{n} = \frac{2041}{15} = 136.07$



$$s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{281,807 - \frac{(2041)^2}{15}}{14}} = \sqrt{292.495238} = 17.102$$

**c** According to Tchebysheff's Theorem, with  $k = 2$ , at least  $3/4$  or 75% of the measurements will lie within  $k = 2$  standard deviations of the mean. For this data, the two values,  $a$  and  $b$ , are calculated as

$$\bar{x} \pm 2s \Rightarrow 136.07 \pm 2(17.10) \Rightarrow 137.07 \pm 34.20 \text{ or } a = 101.87 \text{ and } b = 170.27.$$

**2.R.14** The diameters of the trees are approximately mound-shaped with mean 14 and standard deviation 2.8.

**a** The value  $x = 8.4$  lies two standard deviations below the mean, while the value  $x = 22.4$  is three standard deviations above the mean. Use the Empirical Rule. The fraction of trees with diameters between 8.4 and 14 is half of 0.95 or 0.475, while the fraction of trees with diameters between 14 and 22.4 is half of 0.997 or 0.4985. The total fraction of trees with diameters between 8.4 and 22.4 is  $0.475 + 0.4985 = .9735$

**b** The value  $x = 16.8$  lies one standard deviation above the mean. Using the Empirical Rule, the fraction of trees with diameters between 14 and 16.8 is half of 0.68 or 0.34, and the fraction of trees with diameters greater than 16.8 is  $0.5 - 0.34 = .16$

**2.R.15 a** The range is  $R = 19 - 4 = 15$  and the range approximation is  $s \approx R/4 = 15/4 = 3.75$ .

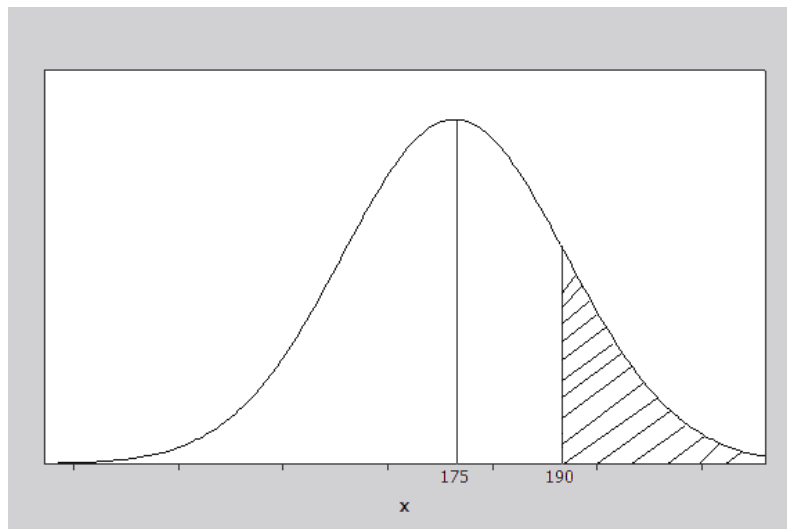
**b** Calculate  $n = 15$ ,  $\sum x_i = 175$ ,  $\sum x_i^2 = 2237$ . Then  $\bar{x} = \frac{\sum x_i}{n} = \frac{175}{15} = 11.67$

$$s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{2237 - \frac{(175)^2}{15}}{14}} = \sqrt{13.95238} = 3.735$$

**c** Calculate the interval  $\bar{x} \pm 2s \Rightarrow 11.67 \pm 2(3.735) \Rightarrow 11.67 \pm 7.47$  or 4.20 to 19.14. Referring to the original data set, the fraction of measurements in this interval is  $14/15 = .93$ .

**2.R.16 a** The percentage of colleges that have between 145 and 205 teachers corresponds to the fraction of measurements expected to lie within two standard deviations of the mean. Tchebysheff's Theorem states that this fraction will be at least  $3/4$  or 75%.

**b** If the population is normally distributed, the Empirical Rule is appropriate and the desired fraction is calculated. Referring to the normal distribution shown below, the fraction of area lying between 175 and 190 is 0.34, so that the fraction of colleges having more than 190 teachers is  $0.5 - 0.34 = 0.16$ .



**2.R.17** We must estimate  $s$  and compare with the student's value of 0.263. In this case,  $n = 20$  and the range is  $R = 17.4 - 16.9 = 0.5$ . The estimated value for  $s$  is then  $s \approx R/4 = 0.5/4 = 0.125$  which is less than 0.263. It is important to consider the magnitude of the difference between the "rule of thumb" and the calculated value. For example, if we were working with a standard deviation of 100, a difference of 0.142 would not be great. However, the student's calculation is twice as large as the estimated value. Moreover, two standard deviations, or  $2(0.263) = 0.526$ , already exceeds the range. Thus, the value  $s = 0.263$  is probably incorrect. The correct value of  $s$  is

$$s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{5851.95 - \frac{117032.41}{20}}{19}} = \sqrt{0.0173} = 0.132$$

**2.R.18** Answers will vary. Students should notice that the distribution of baseline measurements is relatively mound-shaped. Therefore, the Empirical Rule will provide a very good description of the data. A measurement which is further than two or three standard deviations from the mean would be considered unusual.

**2.R.19 a** Calculate  $n = 25$ ,  $\sum x_i = 104.9$ ,  $\sum x_i^2 = 454.810$ . Then  $\bar{x} = \frac{\sum x_i}{n} = \frac{104.9}{25} = 4.196$

$$s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{454.810 - \frac{(104.9)^2}{25}}{24}} = \sqrt{.610} = .781$$

**b** The ordered data set is shown below:

2.5	3.0	3.1	3.3	3.6
3.7	3.8	3.8	3.9	3.9
4.1	4.2	4.2	4.2	4.3
4.3	4.4	4.7	4.7	4.8
4.8	5.2	5.3	5.4	5.7

**c** The  $z$ -scores for  $x = 2.5$  and  $x = 5.7$  are

$$z = \frac{x - \bar{x}}{s} = \frac{2.5 - 4.196}{.781} = -2.17 \text{ and } z = \frac{x - \bar{x}}{s} = \frac{5.7 - 4.196}{.781} = 1.93$$

Since neither of the  $z$ -scores are greater than 3 in absolute value, the measurements are not judged to be extremely unlikely; however, the minimum value,  $x = -2.17$  is somewhat unlikely since it is greater than 2 standard deviations from the mean.

**2.R.20 a** For  $n = 25$ , the position of the median is  $0.5(n+1) = 13$  and the positions of the quartiles are  $0.25(n+1) = 6.5$  and  $0.75(n+1) = 19.5$ . Then  $m = 4.2$ ,  $Q_1 = (3.7 + 3.8)/2 = 3.75$  and  $Q_3 = (4.7 + 4.8)/2 = 4.75$ . Then the five-number summary is

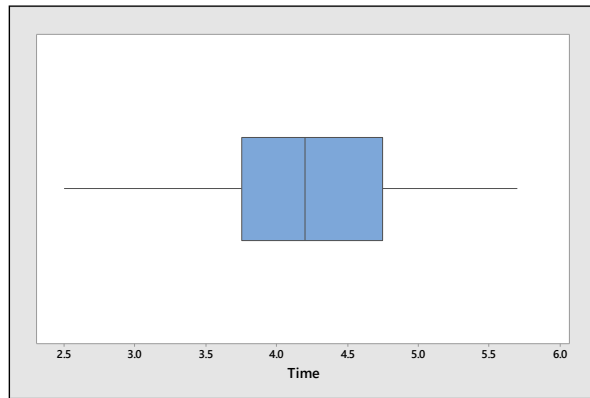
Min	$Q_1$	Median	$Q_3$	Max
2.5	3.75	4.2	4.75	5.7

**b-c** Calculate  $IQR = Q_3 - Q_1 = 4.75 - 3.75 = 1$ . Then the *lower and upper fences* are:

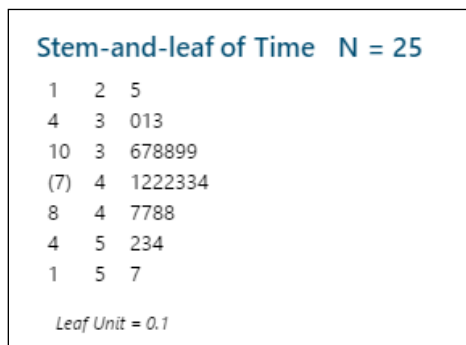
$$Q_1 - 1.5IQR = 3.75 - 1.5 = 2.25$$

$$Q_3 + 1.5IQR = 4.75 + 1.5 = 6.25$$

There are no unusual measurements, and the box plot is shown below.



**d** Answers will vary. A stem and leaf plot, generated by *Minitab*, is shown below. The data is roughly mound-shaped.



### On Your Own

**2.R.21 a** Calculate  $n = 15$ ,  $\sum x_i = 21$  and  $\sum x_i^2 = 49$ . Then  $\bar{x} = \frac{\sum x_i}{n} = \frac{21}{15} = 1.4$  and

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{49 - \frac{(21)^2}{15}}{14} = 1.4$$

**b** Using the frequency table and the grouped formulas, calculate

$$\sum x_i f_i = 0(4) + 1(5) + 2(2) + 3(4) = 21$$

$$\sum x_i^2 f_i = 0^2(4) + 1^2(5) + 2^2(2) + 3^2(4) = 49$$

Then, as in part **a**,  $\bar{x} = \frac{\sum x_i f_i}{n} = \frac{21}{15} = 1.4$

$$s^2 = \frac{\sum x_i^2 f_i - \frac{(\sum x_i f_i)^2}{n}}{n-1} = \frac{49 - \frac{(21)^2}{15}}{14} = 1.4$$

**2.R.22** Use the formulas for grouped data given in Exercise 21. Calculate  $n = 17$ ,  $\sum x_i f_i = 79$ , and  $\sum x_i^2 f_i = 393$ .

Then,  $\bar{x} = \frac{\sum x_i f_i}{n} = \frac{79}{17} = 4.65$

$$s^2 = \frac{\sum x_i^2 f_i - \frac{(\sum x_i f_i)^2}{n}}{n-1} = \frac{393 - \frac{(79)^2}{17}}{16} = 1.6176 \text{ and } s = \sqrt{1.6176} = 1.27$$

**2.R.23 a** The data in this exercise have been arranged in a frequency table.

$x_i$	0	1	2	3	4	5	6	7	8	9	10
$f_i$	10	5	3	2	1	1	1	0	0	1	1

Using the frequency table and the grouped formulas, calculate

$$\sum x_i f_i = 0(10) + 1(5) + \cdots + 10(1) = 51$$

$$\sum x_i^2 f_i = 0^2(10) + 1^2(5) + \cdots + 10^2(1) = 293$$

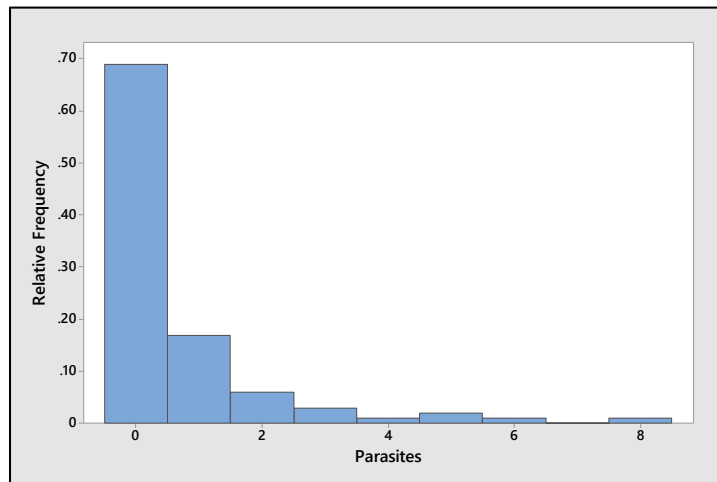
Then  $\bar{x} = \frac{\sum x_i f_i}{n} = \frac{51}{25} = 2.04$

$$s^2 = \frac{\sum x_i^2 f_i - \frac{(\sum x_i f_i)^2}{n}}{n-1} = \frac{293 - \frac{(51)^2}{25}}{24} = 7.873 \text{ and } s = \sqrt{7.873} = 2.806.$$

**b-c** The three intervals  $\bar{x} \pm ks$  for  $k=1,2,3$  are calculated in the table along with the actual proportion of measurements falling in the intervals. Tchebysheff's Theorem is satisfied and the approximation given by the Empirical Rule are fairly close for  $k=2$  and  $k=3$ .

$k$	$\bar{x} \pm ks$	Interval	Fraction in Interval	Tchebysheff	Empirical Rule
1	$2.04 \pm 2.806$	-0.766 to 4.846	$21/25 = 0.84$	at least 0	$\approx 0.68$
2	$2.04 \pm 5.612$	-3.572 to 7.652	$23/25 = 0.92$	at least 0.75	$\approx 0.95$
3	$2.04 \pm 8.418$	-6.378 to 10.458	$25/25 = 1.00$	at least 0.89	$\approx 0.997$

**2.R.24 a** The relative frequency histogram for these data is shown next.



**b** Refer to the formulas given in Exercise 21. Using the frequency table and the grouped formulas,

calculate  $n = 100$ ,  $\sum x_i f_i = 66$ ,  $\sum x_i^2 f_i = 234$ . Then  $\bar{x} = \frac{\sum x_i f_i}{n} = \frac{66}{100} = 0.66$

$$s^2 = \frac{\sum x_i^2 f_i - \frac{(\sum x_i f_i)^2}{n}}{n-1} = \frac{234 - \frac{(66)^2}{100}}{99} = 1.9236 \text{ and } s = \sqrt{1.9236} = 1.39.$$

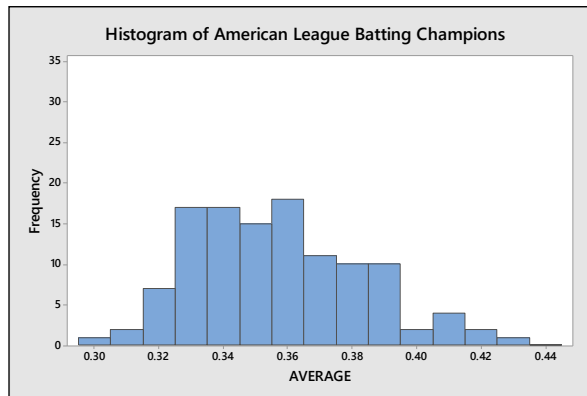
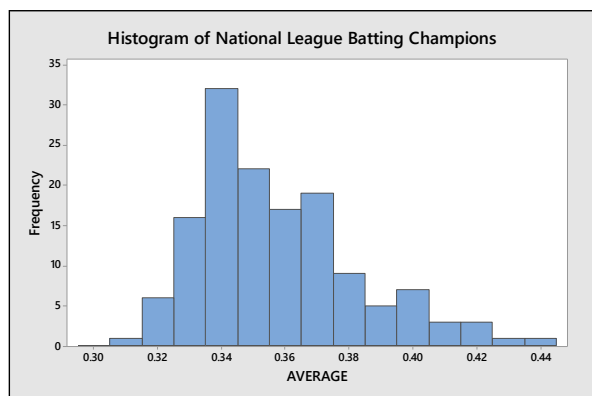
c The three intervals,  $\bar{x} \pm ks$  for  $k = 2, 3$  are calculated in the table along with the actual proportion of measurements falling in the intervals. Tchebysheff's Theorem is satisfied and the approximation given by the Empirical Rule are fairly close for  $k = 2$  and  $k = 3$ .

$k$	$\bar{x} \pm ks$	Interval	Fraction in Interval	Tchebysheff	Empirical Rule
2	$0.66 \pm 2.78$	-2.12 to 3.44	$95/100 = 0.95$	at least 0.75	$\approx 0.95$
3	$0.66 \pm 4.17$	-3.51 to 4.83	$96/100 = 0.96$	at least 0.89	$\approx 0.997$

**2.R.25** Answers will vary. The student should notice the outliers in the female group, and that the median female temperature is higher than the median male temperature.

## Case Study: The Boys of Summer

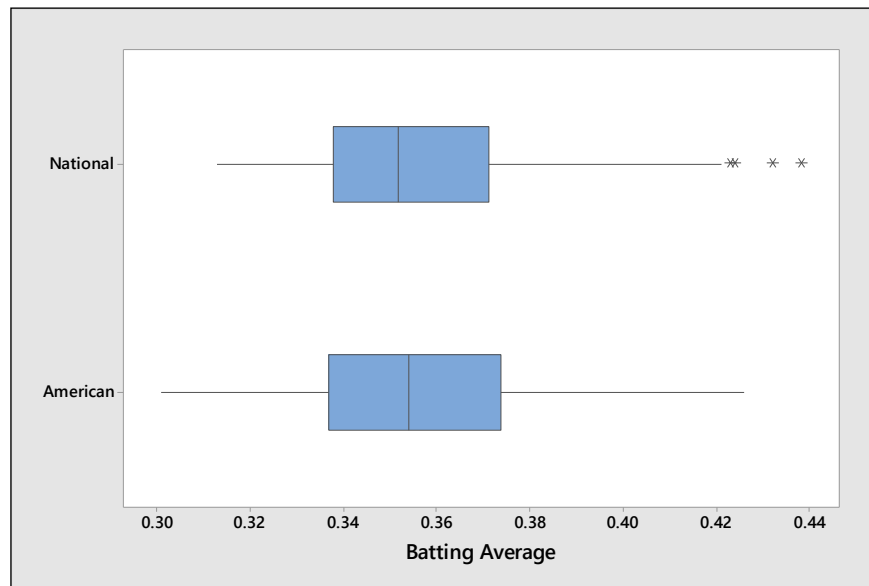
1 The *Minitab* computer package was used to analyze the data. In the printout below, various descriptive statistics as well as histograms and box plots are shown.



### Statistics

Variable	LEAGUE	N	N*	Mean	SE Mean	StDev	Variance	Sum	Sum of Squares	Minimum
AVERAGE	0	142	0	0.35718	0.00214	0.02552	0.00065	50.71900	18.20745	0.31300
	1	117	0	0.35643	0.00243	0.02631	0.00069	41.70200	14.94402	0.30100
Variable	LEAGUE	Q1	Median	Q3	Maximum	Range				
AVERAGE	0	0.33800	0.35200	0.37125	0.43800	0.12500				
	1	0.33700	0.35400	0.37400	0.42600	0.12500				

2 Notice that the mean percentage of hits is almost the same for the two leagues, but that the American League (1) is slightly more variable.

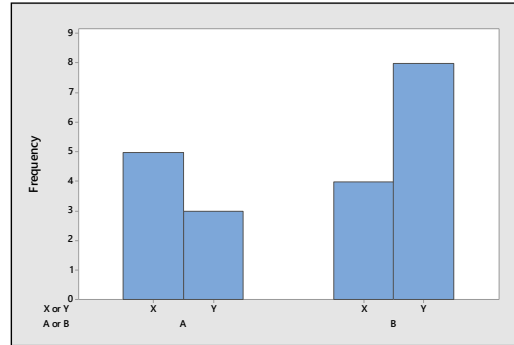


- 3 The box plot shows that there are four outliers in the National League data set.
- 4 In summary, except for the four outliers, there is very little difference between the two leagues.

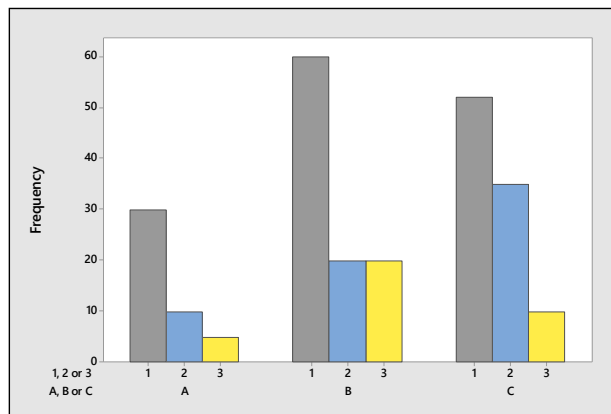
### 3: Describing Bivariate Data

#### Section 3.1

- 3.1.1** The side-by-side bar measures the frequency of occurrence for each of the two categories, A and B. A separate bar is used for each of the other two categories—X and Y. The student may choose to group by X and Y first, which will change the look of the bar chart.

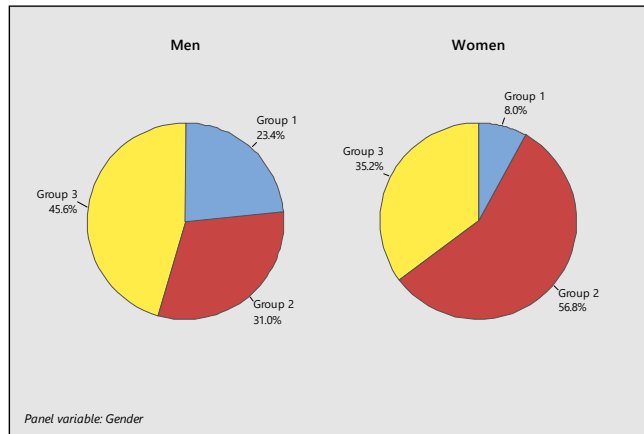


- 3.1.2** Similar to Exercise 1. Students may choose the categories to group together—the side-by-side bar chart below shows the distribution of categories 1, 2 and 3 as they vary for categories A, B and C.



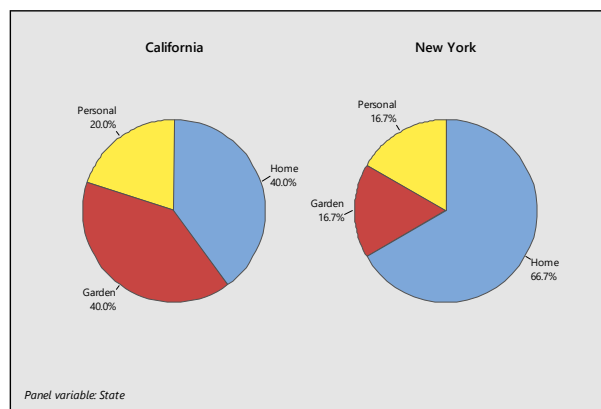
- 3.1.3** The side-by-side pie charts are constructed as in Chapter 1 for each of the two groups (men and women) and are displayed next using the percentages shown in the table below.

	Group 1	Group 2	Group 3	Total
<b>Men</b>	23%	31%	46%	100%
<b>Women</b>	8%	57%	35%	100%

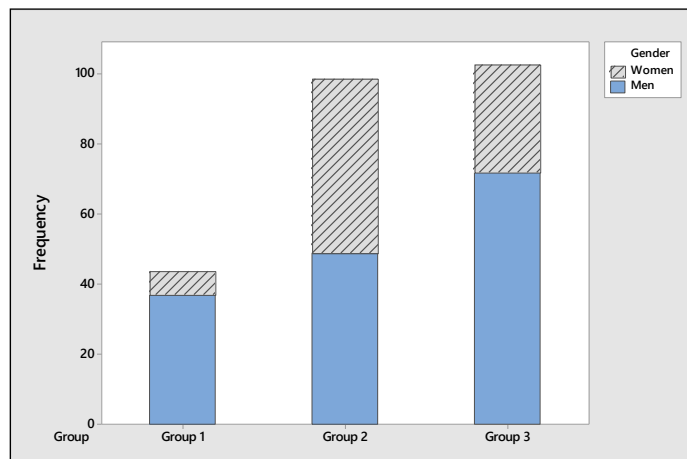


**3.1.4** Similar to Exercise 3. Pie Charts are constructed for each of the two states and are displayed next using the proportions shown in the table.

	Home	Garden	Personal	Total
<b>New York</b>	20/30	5/30	5/30	30/30
<b>California</b>	10/25	10/25	5/25	25/25



**3.1.5** Use the data from Exercise 3. The stacked bar chart shows the frequency of occurrence for each of the three groups. A portion of each bar is used for men and women.

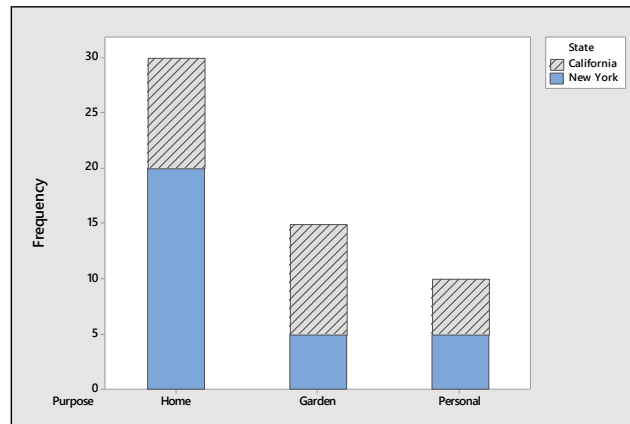


The differences in the proportions of men and women in the three groups is most graphically portrayed by the pie charts, since the unequal number of men and women tend to confuse the interpretation of the bar



charts. However, the bar chart is useful in retaining the actual frequencies of occurrence in each group, which is lost in the pie chart.

- 3.1.6** Use the data from Exercise 4. The stacked bar chart) shows each of the three categories, with the number of items for New York and California (each shaded differently) stacked on top of each other.



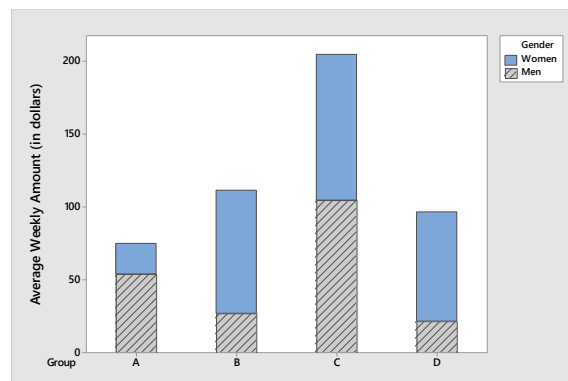
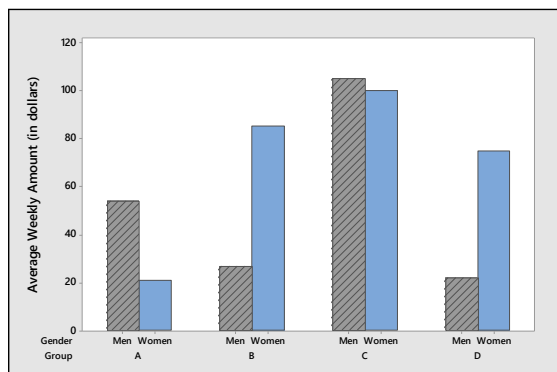
The differences in the proportions clothing items in the three groups is most graphically portrayed by the pie charts, since the unequal number of items in each state tend to confuse the interpretation of the bar charts. However, the bar chart is useful in retaining the actual frequencies of occurrence in each group, which is lost in the pie chart.

- 3.1.7-8** The conditional distribution in each of the groups given that the person was either male or female is shown in the table that follows. These are the values that were used to construct the pie charts in Exercise 3.

	Group 1	Group 2	Group 3	Total
Men	23%	31%	46%	100%
Women	8%	57%	35%	100%

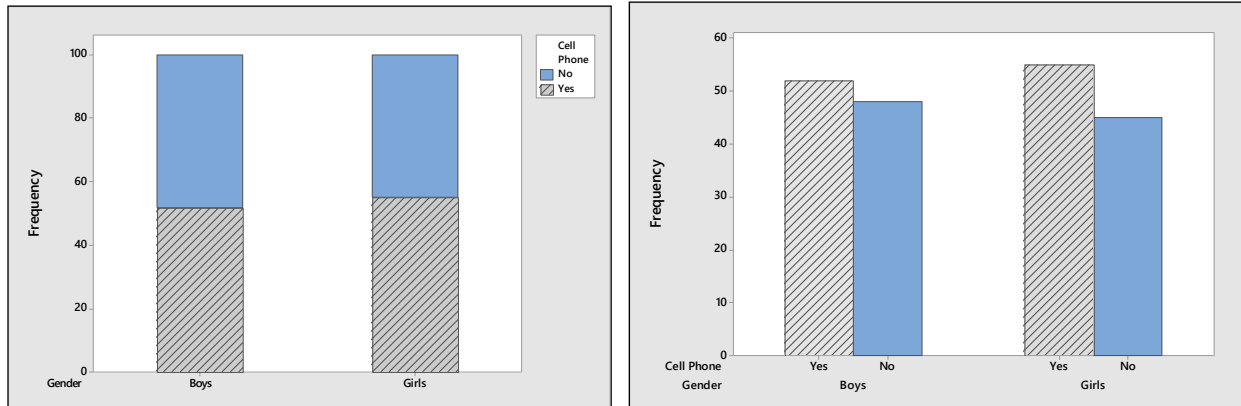
- 3.1.9 a** Any of the comparative charts (side-by-side pie charts, stacked or side-by-side bar charts) can be used.

**b-c** The two types of comparative bar charts are shown next. The amounts spent in each of the four categories seem to be quite different for men and women, except in category C. In category C which involves the largest dollar amount of purchase, there is little difference between the genders.

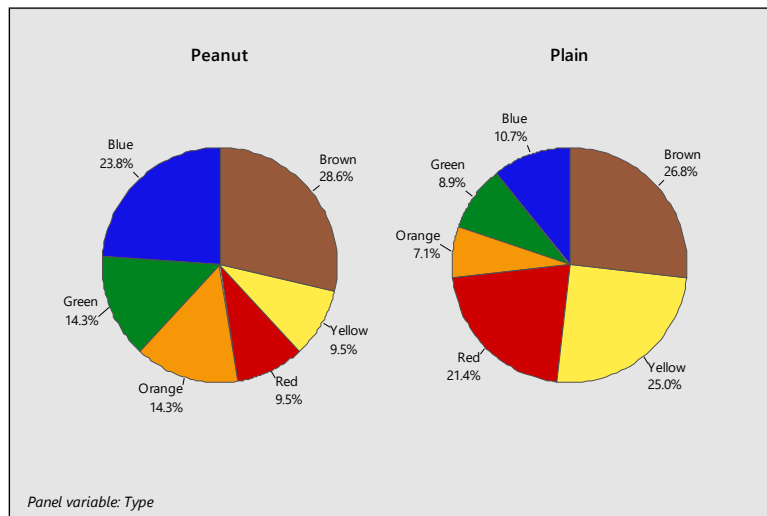


- d** Although it is really a matter of preference, the only advantage to the stacked chart is that the reader can easily see the total dollar amount for each category. For comparison purposes, the side-by-side chart may be better.

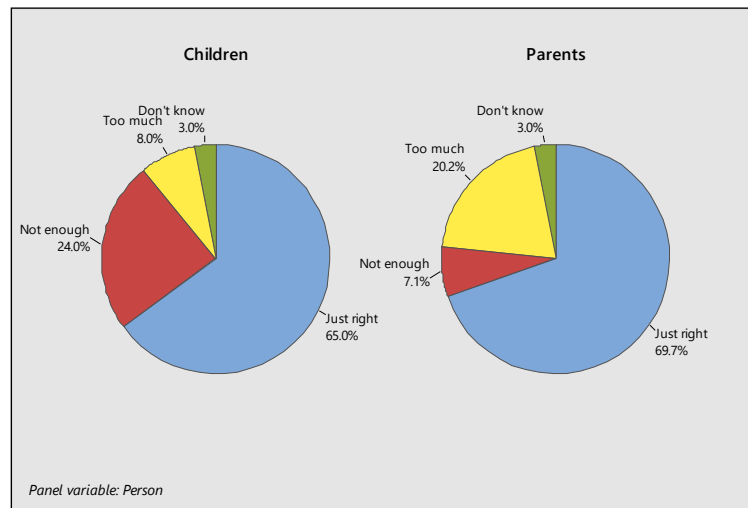
- 3.1.10 a-c** The stacked and side-by-side bar charts are shown next. The number of eighth grade boys and girls with cell phones is nearly the same; the girls in this group of eighth graders had slightly more cell phones than the boys.



- 3.1.11** It is clear that there are more plain M&Ms than peanut M&Ms, since the peanut M&Ms are larger. We choose to ignore the differences in total numbers by using side-by-side pie charts (below). There is a distinct difference in the distribution of colors.

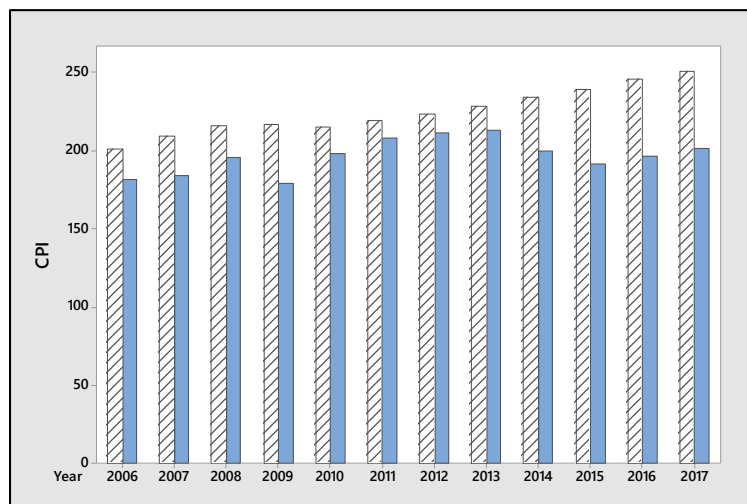


- 3.1.12 a** The population of interest is the population of responses to the question about free time for all parents and children in the United States. The sample is the set of responses generated for the 198 parents and 200 children in the survey.
- b** The data can be considered bivariate if, for each person interviewed, we record the person's relationship (Parent or Child) and their response to the question (just the right amount, not enough, too much, don't know). Since the measurements are not numerical in nature, the variables are qualitative.
- c** The entry in a cell represents the number of people who fell into that relationship-opinion category.
- d** A pie chart is created for both the "parent" and the "children" categories. The size of each sector angle is proportional to the fraction of measurements falling into that category.

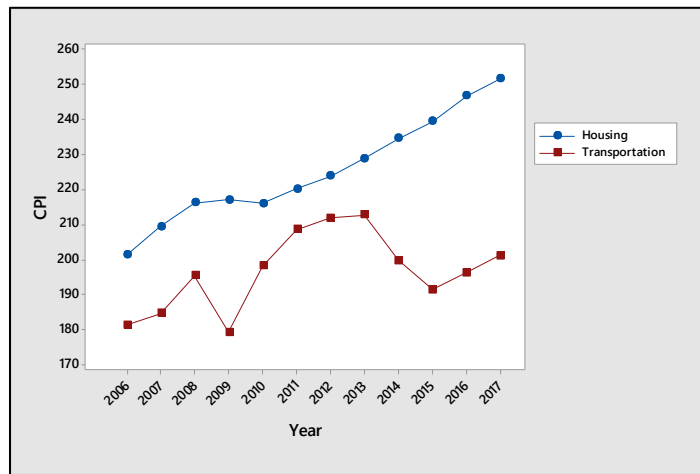


**e** Either stacked or comparative bar charts could be used, but since the height of the bar represents the frequency of occurrence (and hence is tied to the sample size), this type of chart would be misleading. The comparative pie charts are the best choice.

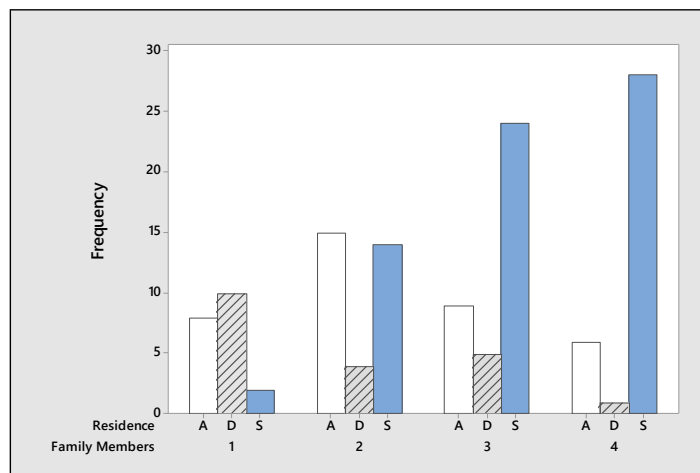
**3.1.13 a** The side-by-side comparative bar charts that follow measure the CPIs on the vertical axis, while the clusters of bars represent the grouping into housing (stripes) and transportation (solid) over the years.



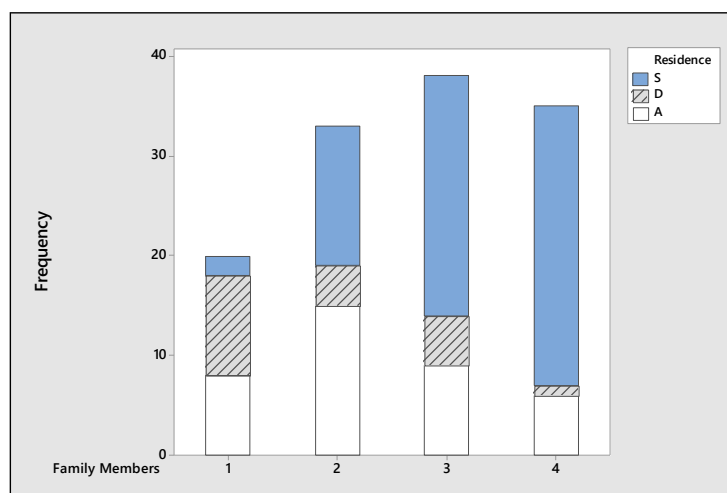
**b-c** The comparative line chart is shown next. The line chart is more effective, showing that the housing CPI has increased steadily, except for a small dip around 2009-2010, but that the CPI for transportation has been a lot more variable, with increases and decreases over the years.



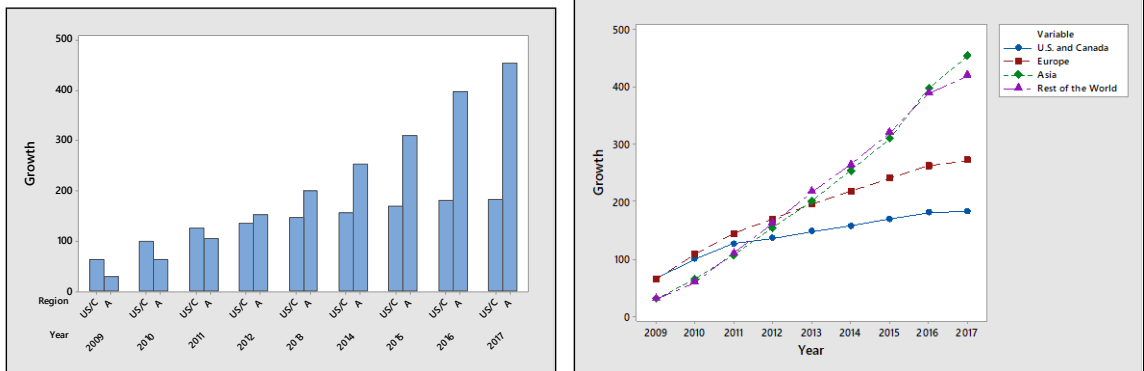
**3.1.14 a** The side-by-side bar chart is shown next.



**b-c** The stacked bar chart is shown next. Both charts indicate that the more family members there are, the more likely it is that the family lives in a duplex or a single residence. The fewer the number of family members, the more likely it is that the family lives in an apartment.

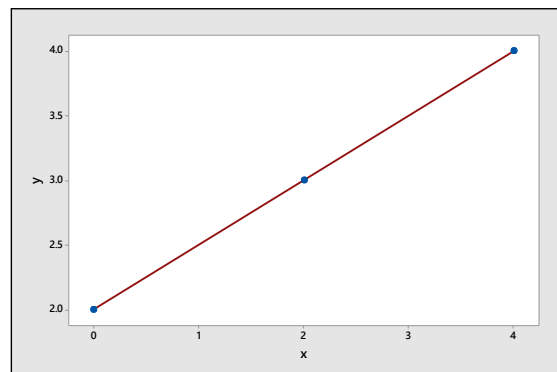


- 3.1.15** The side-by-side bar chart and the comparative line charts are shown next. Conclusions will vary from student to student, but it is clear that markets in Asia and the rest of the world are increasing at a faster rate than Europe, the United States and Canada.

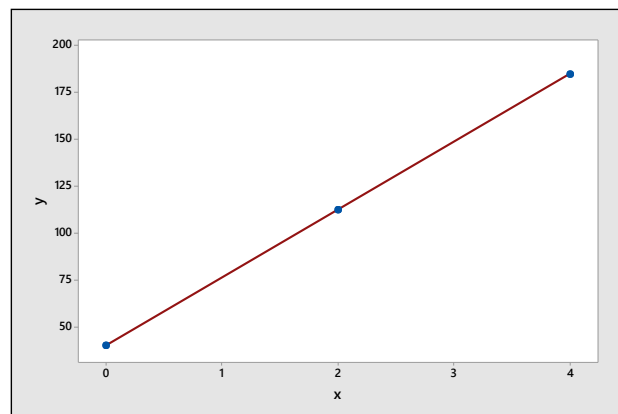


## Section 3.2

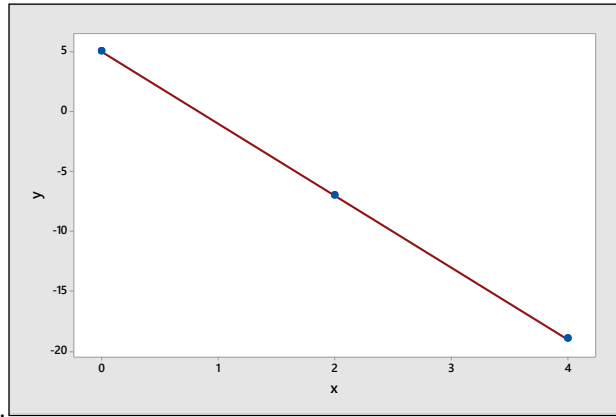
- 3.2.1** The line is plotted next. The change in  $y$  for a one-unit change in  $x$  is defined as the *slope* of the line—the coefficient of the  $x$  variable. For this line, the slope is  $b = 0.5$ . The point at which the line crosses the  $y$ -axis is called the  $y$ -intercept and is the constant term in the equation of the line. For this line, the  $y$ -intercept is  $a = 2.0$ . When  $x = 2.5$ , the predicted value of  $y$  is  $y = 2.0 + 0.5(2.5) = 3.25$ .



- 3.2.2** Similar to Exercise 1. The line is plotted next. The *slope* of the line is  $b = 36.2$  and the  $y$ -intercept is  $a = 40$ . When  $x = 2.5$ , the predicted value of  $y$  is  $y = 40 + 36.2(2.5) = 130.5$ .



- 3.2.3** Similar to Exercise 1. The line is plotted next. The *slope* of the line is  $b = -6$  and the  $y$ -intercept is  $a = 5$ . When  $x = 2.5$ , the predicted value of  $y$  is  $y = 5 - 6(2.5) = -10$ .



**3.2.4** The pattern is linear, but not very strong.

**3.2.5** The pattern is strong linear, with two outliers.

**3.2.6** The pattern is strong quadratic.

**3.2.7** The pattern is strong linear, with two distinct clusters.

**3.2.8** Calculate the sums, sums of square sand sum of cross products for the pairs  $(x_i, y_i)$ .

$$\sum x_i = 19; \sum y_i = 37; \sum x_i^2 = 71; \sum y_i^2 = 237; \sum x_i y_i = 126$$

Then the covariance is 
$$s_{xy} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{n-1} = \frac{126 - \frac{(19)(37)}{6}}{5} = 1.76667$$

and the sample standard deviations are

$$s_x = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{71 - \frac{(19)^2}{6}}{5}} = 1.472 \quad \text{and} \quad s_y = \sqrt{\frac{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}{n-1}} = \sqrt{\frac{237 - \frac{(37)^2}{6}}{5}} = 1.329$$

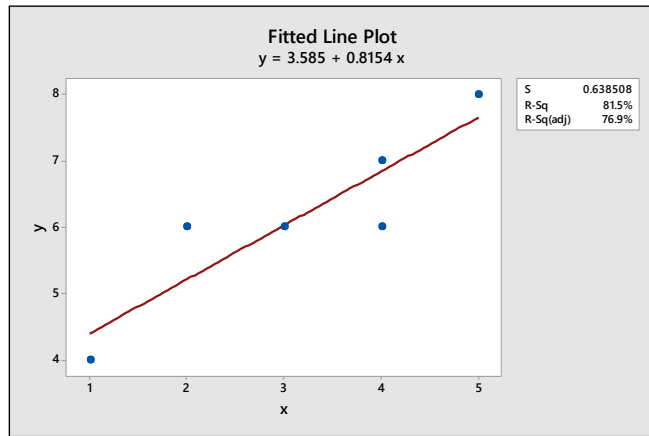
The correlation coefficient is 
$$r = \frac{s_{xy}}{s_x s_y} = \frac{1.76667}{(1.472)(1.329)} = 0.902986 \approx 0.903$$

The slope and y-intercept of the regression line are

$$b = r \frac{s_y}{s_x} = 0.902986 \left( \frac{1.329}{1.472} \right) = 0.81526 \quad \text{and} \quad a = \bar{y} - b\bar{x} = \frac{37}{6} - 0.81526 \left( \frac{19}{6} \right) = 3.58$$

and the equation of the regression line is  $\hat{y} = 3.58 + 0.815x$ .

The graph of the data points and the best fitting line is shown next. The line fits through the data points.



**3.2.9** Calculate the sums, sums of square sand sum of cross products for the pairs  $(x_i, y_i)$ .

$$\sum x_i = 6; \sum y_i = 12; \sum x_i^2 = 14; \sum y_i^2 = 56; \sum x_i y_i = 20$$

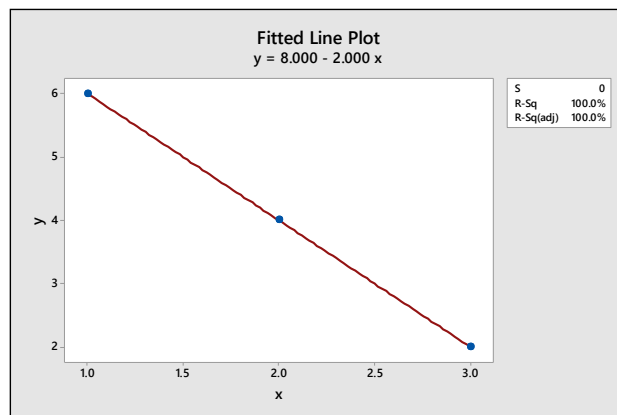
Then the covariance is  $s_{xy} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{n-1} = \frac{20 - \frac{(6)(12)}{5}}{5-1} = -2$ . The sample standard deviations are

$$s_x = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{14 - \frac{(6)^2}{5}}{5-1}} = 1 \text{ and } s_y = \sqrt{\frac{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}{n-1}} = \sqrt{\frac{56 - \frac{(12)^2}{5}}{5-1}} = 2$$

and the correlation coefficient is  $r = \frac{s_{xy}}{s_x s_y} = \frac{-2}{(1)(2)} = -1$ . The slope and y-intercept of the regression line

$$\text{are } b = r \frac{s_y}{s_x} = -1 \left( \frac{2}{1} \right) = -2 \text{ and } a = \bar{y} - b\bar{x} = \frac{12}{5} - (-2) \left( \frac{6}{5} \right) = 8$$

and the equation of the regression line is  $\hat{y} = 8 - 2x$ . The graph of the data points and the best fitting line is shown next. Notice that, since  $r = -1$ , the points all fall exactly on the straight line, which has a negative slope.



**3.2.10** Calculate the sums, sums of square sand sum of cross products for the pairs  $(x_i, y_i)$ .

$$\sum x_i = 21; \sum y_i = 24.3; \sum x_i^2 = 91; \sum y_i^2 = 103.99; \sum x_i y_i = 75.3$$

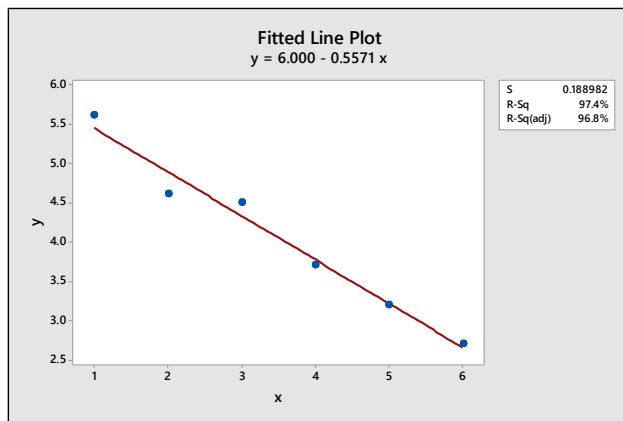
Then the covariance is  $s_{xy} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{n-1} = \frac{75.3 - \frac{(21)(24.3)}{6}}{5} = -1.95$  and the sample standard deviations are

$$s_x = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{91 - \frac{(21)^2}{6}}{5}} = 1.8708 \text{ and } s_y = \sqrt{\frac{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}{n-1}} = \sqrt{\frac{103.99 - \frac{(24.3)^2}{6}}{5}} = 1.0559$$

The correlation coefficient is  $r = \frac{s_{xy}}{s_x s_y} = \frac{-1.95}{(1.8708)(1.0559)} = -0.987$ . The slope and y-intercept of the

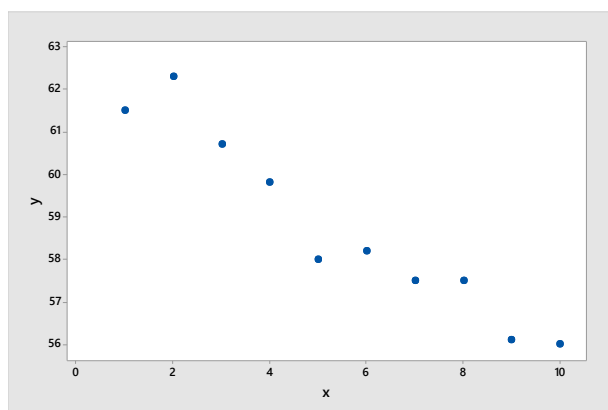
regression line are  $b = r \frac{s_y}{s_x} = -0.987 \left( \frac{1.0559}{1.8708} \right) = -.557$  and  $a = \bar{y} - b\bar{x} = \frac{24.3}{6} - (-.557) \left( \frac{21}{6} \right) = 6.00$

and the equation of the regression line is  $\hat{y} = 6 - .557x$ . The graph of the data points and the best fitting line is shown next. The line fits through the data points.



- 3.2.11-13** Answers will vary, depending on the type of calculator the student is using. The results should agree with the results of Exercises 8-10.
- 3.2.14** The grade on a history test *depends on* the number of hours spent studying. Grade,  $y$ , is the dependent variable and number of hours,  $x$ , is the independent variable.
- 3.2.15** The number of calories burned *depends on* the number of minutes running on the treadmill. Number of calories,  $y$ , is the dependent variable and number of minutes running on the treadmill,  $x$ , is the independent variable.
- 3.2.16** The amount of electricity generated *depends on* the speed of the wind turbine. Amount of electricity,  $y$ , is the dependent variable and the speed of the turbine,  $x$ , is the independent variable.
- 3.2.17** The number of ice cream cones sold *depends on* the temperature on a given day. Number of cones sold,  $y$ , is the dependent variable and daily temperature,  $x$ , is the independent variable.
- 3.2.18** The weight of a newborn puppy *depends on* the litter size. Weight,  $y$ , is the dependent variable and number of puppies in the litter,  $x$ , is the independent variable.
- 3.2.19 a-b** The scatterplot is shown next. Notice that there is a negative relationship between  $y$  and  $x$ . The measurements are decreasing over time.





**c** From the Minitab printout,  $s_{xy} = -6.42222$ ,  $s_x = \sqrt{9.16667}$  and  $s_y = \sqrt{4.84933}$  so that

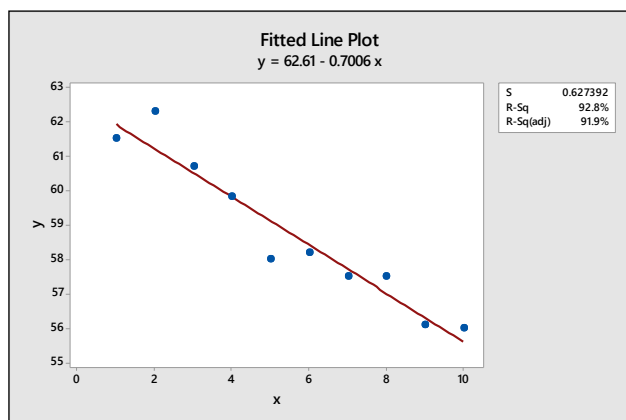
$$r = \frac{s_{xy}}{s_x s_y} = \frac{-6.42222}{\sqrt{(9.16667)(4.84933)}} = -0.9632$$

**d** The slope and y-intercept of the regression line are

$$b = r \frac{s_y}{s_x} = -0.9632 \left( \frac{2.2021}{3.0277} \right) = -0.700 \quad \text{and} \quad a = \bar{y} - b\bar{x} = \frac{587.6}{10} - (-0.700) \left( \frac{55}{10} \right) = 62.61$$

and the equation of the regression line is  $\hat{y} = 62.61 - 0.70x$ .

**e** The graph of the data points and the best fitting line is shown below. The line fits through the points very well.



**3.2.20 a** Use your scientific calculator. You can verify that  $\sum x_i = 18$ ;  $\sum y_i = 2644$ ;

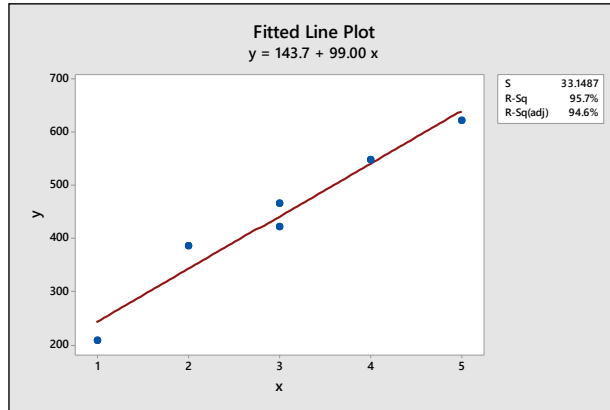
$\sum x_i^2 = 64$ ;  $\sum y_i^2 = 1,267,528$ ;  $\sum x_i y_i = 8922$ . Then the covariance is

$$s_{xy} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{n-1} = \frac{8922 - \frac{18(2644)}{6}}{5} = 198$$

The sample standard deviations are  $s_x = 1.4142136$  and  $s_y = 143.112077$  so that  $r = 0.9783$ . Then

$$b = r \frac{s_y}{s_x} = 99.00 \quad \text{and} \quad a = \bar{y} - b\bar{x} = 440.6667 - 99(3) = 143.6667$$

and the equation of the regression line is  $\hat{y} = 143.67 + 99.00x$ . The graph of the data points and the best fitting line is shown below. The line provides a good description of the data.



**3.2.21** Since BMI is a function of income, BMI,  $y$ , will *depend on* income,  $x$ . Use your scientific calculator. You can verify that  $\sum x_i = 240.5$ ;  $\sum y_i = 162$ ;

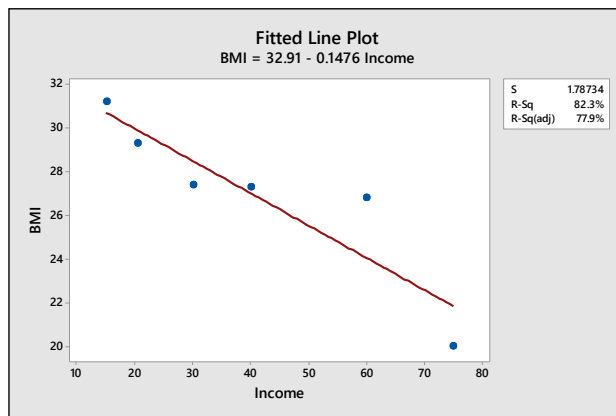
$\sum x_i^2 = 12,370.25$ ;  $\sum y_i^2 = 4446.22$ ;  $\sum x_i y_i = 6090.65$ . Then the covariance is

$$s_{xy} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{n-1} = \frac{6090.65 - \frac{240.5(162)}{6}}{5} = -80.57$$

The sample standard deviations are  $s_x = 23.3675$  and  $s_y = 3.8005$  so that  $r = -.9072$ . Then

$$b = r \frac{s_y}{s_x} = -.14755 \text{ and } a = \bar{y} - b\bar{x} = 27 - (-.14755)(40.08333) = 32.914$$

and the equation of the regression line is  $\hat{y} = 32.914 - .148x$ . The graph of the data points and the best fitting line is shown next. The line provides a good description of the data.



**3.2.22** The percent that are rearrested may *depend on* the age of the prisoners. Therefore,  $y$  = percent arrested and  $x$  = group median age. Use your scientific calculator. You can verify that  $\sum x_i = 259$ ;  $\sum y_i = 331.2$ ;

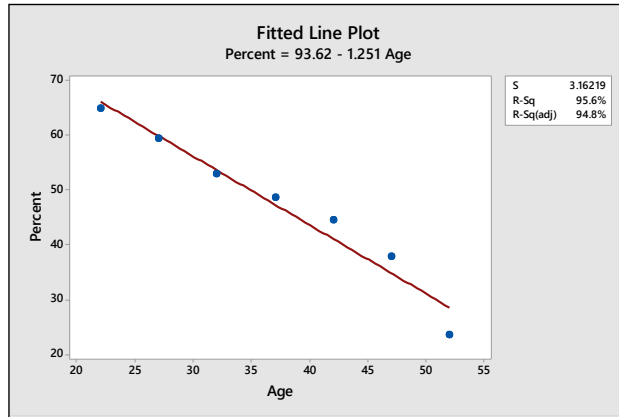
$\sum x_i^2 = 10,283$ ;  $\sum y_i^2 = 16,816.74$ ;  $\sum x_i y_i = 11,378.4$ . Then the covariance is

$$s_{xy} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{n-1} = \frac{11,378.4 - \frac{259(331.2)}{7}}{6} = -146$$

The sample standard deviations are  $s_x = 10.8012345$  and  $s_y = 13.821774$  so that  $r = -.9779$ . Then

$$b = r \frac{s_y}{s_x} = -1.2514 \text{ and } a = \bar{y} - b\bar{x} = 47.314286 - (-1.2514)(37) = 93.62$$

and the equation of the regression line is  $\hat{y} = 93.62 - 1.251x$ . The graph of the data points and the best fitting line is shown next. The line provides a good description of the data.



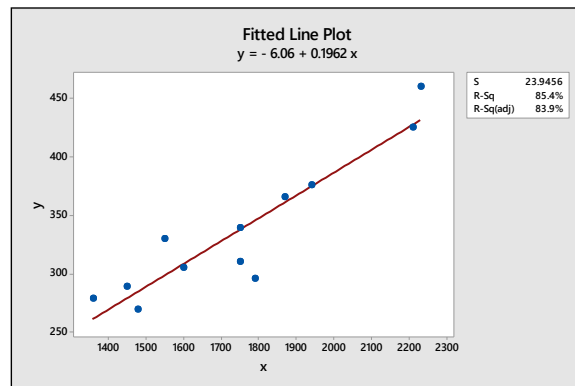
- 3.2.23 a** Similar to previous exercises. Calculate  $n = 12$ ;  $\sum x_i = 20,980$ ;  $\sum y_i = 4043.5$ ;  $\sum x_i^2 = 37,551,600$ ;  $\sum y_i^2 = 1,401,773.75$ ;  $\sum x_i y_i = 7,240,383$ . Then the covariance is

$$s_{xy} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{n-1} = 15,545.19697$$

The sample standard deviations are  $s_x = 281.48416$  and  $s_y = 59.75916$  so that  $r = 0.9241$ . Then

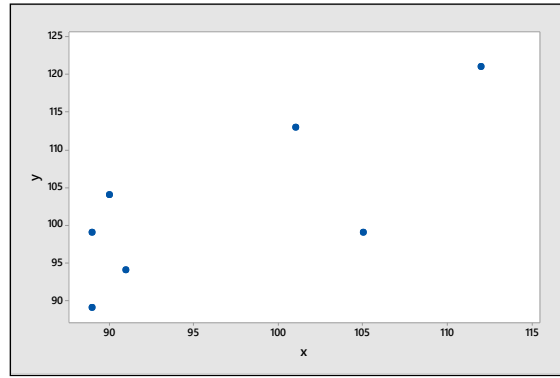
$$b = r \frac{s_y}{s_x} = 0.19620 \text{ and } a = \bar{y} - b\bar{x} = 336.95833 - 0.19620(1748.333) = -6.06$$

and the equation of the regression line is  $y = -6.06 + 0.1962x$ . The graph of the data points and the best fitting line is shown next.



- b** The fitted line describes the data quite well.

- 3.2.24 a** The scatterplot is shown below. There is a slight positive trend between pre- and post-test scores, but the trend is not too pronounced.



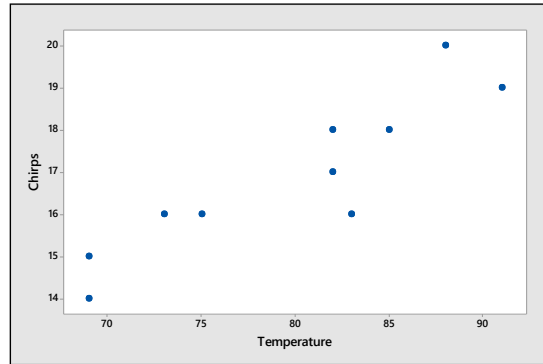
**b** Calculate  $n = 7$ ;  $\sum x_i = 677$ ;  $\sum y_i = 719$ ;  $\sum x_i^2 = 65,993$ ;  $\sum y_i^2 = 74,585$ ;  $\sum x_i y_i = 70,006$ . Then the covariance is

$$s_{xy} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{n-1} = 78.071429$$

The sample standard deviations are  $s_x = 9.286447$  and  $s_y = 11.056134$  so that  $r = 0.760$ . This is a relatively strong positive correlation, confirming the interpretation of the scatterplot.

**3.2.25 a** Since we would be interested in predicting the number of chirps based on temperature, the number of chirps is the dependent variable ( $y$ ) and temperature is the independent variable ( $x$ ).

**b** The scatterplot is shown next. There appears to be a positive linear trend, as expected.



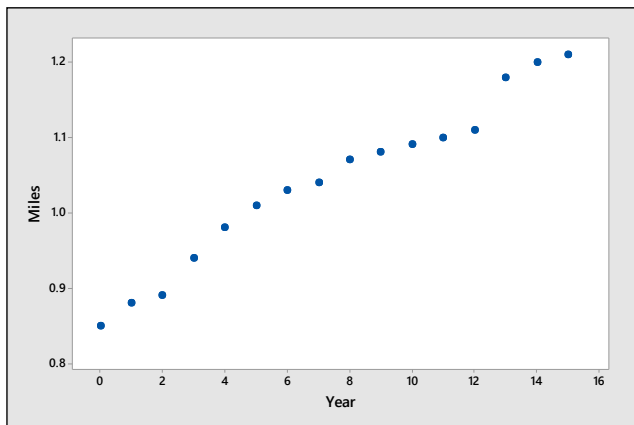
**c** Calculate  $n = 10$ ;  $\sum x_i = 797$ ;  $\sum y_i = 169$ ;  $\sum x_i^2 = 64,063$ ;  $\sum y_i^2 = 2887$ ;  $\sum x_i y_i = 13,586$ . Then the covariance is

$$s_{xy} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{n-1} = 12.966667$$

The sample standard deviations are  $s_x = 7.761014$  and  $s_y = 1.852926$  so that  $r = 0.9017$ . Then (using full accuracy),  $b = r \frac{s_y}{s_x} = .21527$  and  $a = \bar{y} - b\bar{x} = 16.9 - .21527(79.7) = -.257$  and the equation of the regression line is  $\hat{y} = -.257 + .215x$ .

**d** When  $x = 80$ , the predicted number of chirps is  $y = -.257 + .215(80) = 16.943$ , or approximately 17 chirps.

- 3.2.26 a** The scatterplot is shown next. It appears that there might be a slight curvature to the pattern.



- b** Calculate  $n = 16$ ;  $\sum x_i = 120$ ;  $\sum y_i = 16.66$ ;  $\sum x_i^2 = 1240$ ;  $\sum y_i^2 = 17.5336$ ;  $\sum x_i y_i = 13,586$ . Then the covariance is

$$s_{xy} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{n-1} = .523333$$

The sample standard deviations are  $s_x = 4.7609523$  and  $s_y = .1114675$  so that  $r = 0.9861348$ . Then

$b = r \frac{s_y}{s_x} = .023088$  and  $a = \bar{y} - b\bar{x} = 1.04125 - .023088(7.5) = .86809$  and the equation of the regression line is  $\hat{y} = .868 + .023x$ .

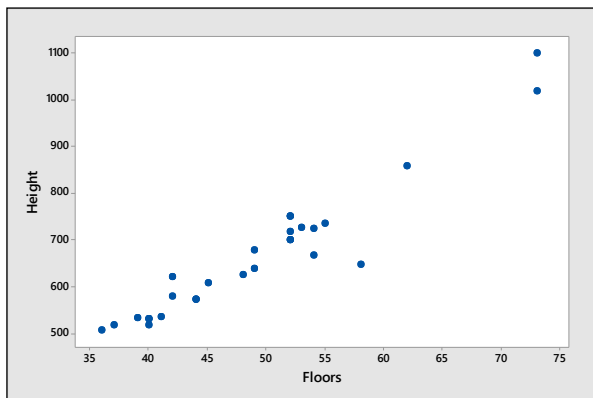
- 3.2.27 a** The height of the building will in part *depend upon* how many floors there are. Then  $y$  = height and  $x$  = number of floors.

- b** Calculate  $n = 28$ ;  $\sum x_i = 1378$ ;  $\sum y_i = 18,637$ ;  $\sum x_i^2 = 70,250$ ;  $\sum y_i^2 = 12,960,309$ ;  $\sum x_i y_i = 952,387$ . Then the covariance is

$$s_{xy} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{n-1} = 1302.97619$$

The sample standard deviations are  $s_x = 9.4921$  and  $s_y = 143.4223$  so that  $r = .9571$ . There is a strong correlation between the number of floors and the height of the building.

- c** The scatterplot is shown next. There is a strong linear relationship with no outliers.



**d** The slope of the line is  $b = r \frac{s_y}{s_x} = 14.4614$  and  $a = \bar{y} - b\bar{x} = 665.6071 - 14.4614(49.2143) = -46.10$  and the equation of the regression line is  $\hat{y} = -46.10 + 14.46x$ . If  $x = 48$ , we predict that the height of the building will be  $\hat{y} = -46.10 + 14.46(48) = 647.98$  or approximately 648 feet.

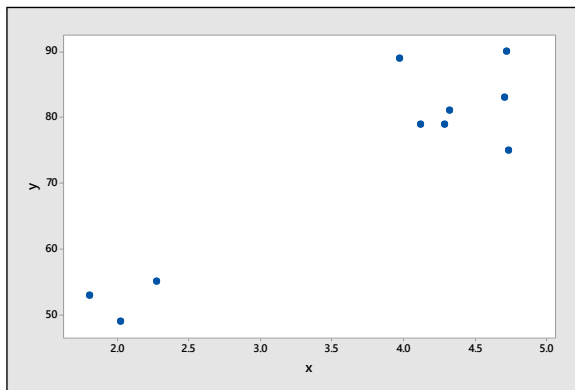
**3.2.28 a** Calculate  $n = 10$ ;  $\sum x_i = 36.917$ ;  $\sum y_i = 733$ ;  $\sum x_i^2 = 148.846879$ ;  $\sum y_i^2 = 55,813$ ;  $\sum x_i y_i = 2854.773$ . Then the covariance is

$$s_{xy} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{n-1} = 16.5285444$$

The sample standard deviations are  $s_x = 1.1813547$  and  $s_y = 15.2173147$  so that  $r = 0.9194249$ . Then

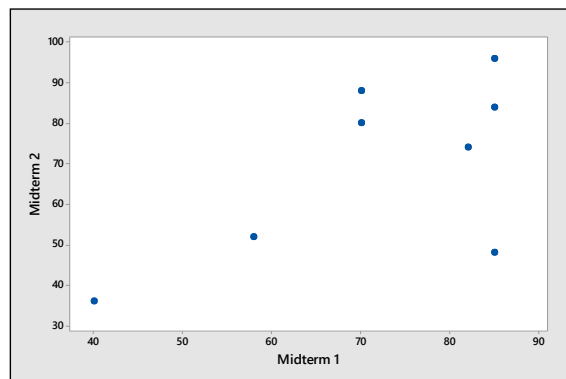
$b = r \frac{s_y}{s_x} = 11.84333$  and  $a = \bar{y} - b\bar{x} = 73.3 - 11.84333(3.6917) = 29.578$  and the equation of the regression line is  $\hat{y} = 29.578 + 11.843x$ .

**b** The scatterplot is shown next. Notice the two distinct clusters in the data. If the eruption lasts only a short amount of time, the waiting time until the next eruption is also short, indicating that the pressure buildup within the geyser was not totally dissipated during the eruption.



## Reviewing What You've Learned

**3.R.1 a-b** The scatterplot is shown next. There is a fairly strong positive linear relationship between  $x$  and  $y$ , with one possible outlier in the lower right corner of the plot. Perhaps that student forgot to study for midterm 2!



- 3.R.2 a** Calculate  $n = 8$ ;  $\sum x_i = 575$ ;  $\sum y_i = 558$ ;  $\sum x_i^2 = 43,163$ ;  $\sum y_i^2 = 42,196$ ;  $\sum x_i y_i = 41,664$ . Then the covariance is

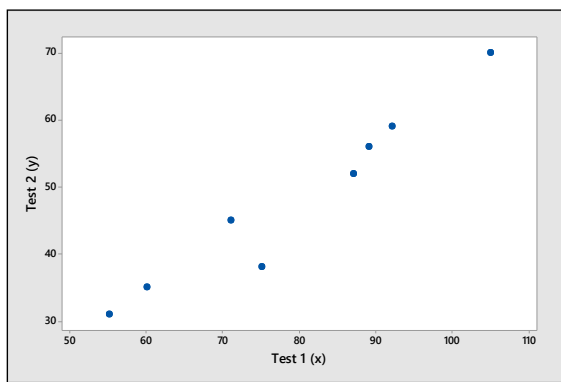
$$s_{xy} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{n-1} = 222.535714$$

The sample standard deviations are  $s_x = 16.190275$  and  $s_y = 21.631657$  so that  $r = 0.635$ .

- b** Calculate (using full accuracy)  $b = r \frac{s_y}{s_x} = .848968$  and  $a = \bar{y} - b\bar{x} = 69.75 - .848968(71.875) = 8.730$  and the equation of the regression line is  $\hat{y} = 8.730 + .849x$ .

- c** When  $x = 85$ , the predicted midterm 2 score is  $\hat{y} = 8.730 + .849(85) = 80.895$ .

- 3.R.3 a-b** The scatterplot is shown next. There is a strong positive linear relationship between  $x$  and  $y$ .



- 3.R.4 a** Calculate  $n = 8$ ;  $\sum x_i = 634$ ;  $\sum y_i = 386$ ;  $\sum x_i^2 = 52270$ ;  $\sum y_i^2 = 19876$ ;  $\sum x_i y_i = 32136$ .

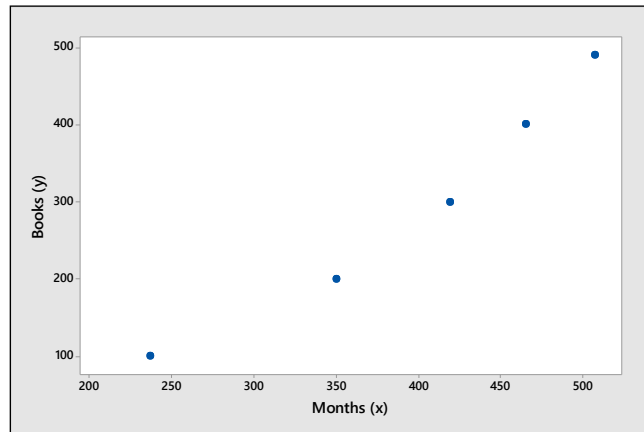
Then the covariance is

$$s_{xy} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{n-1} = 220.78571$$

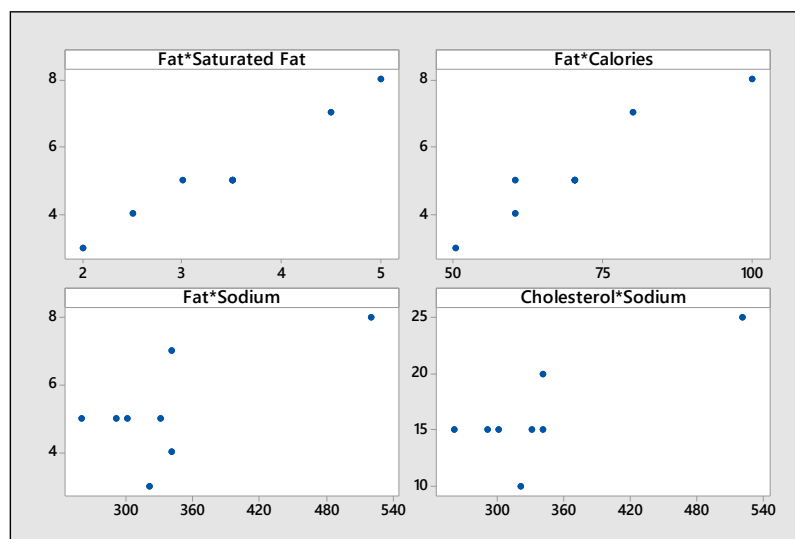
The sample standard deviations are  $s_x = 17.010501$  and  $s_y = 13.3710775$  so that  $r = 0.971$ .

- b** Since the correlation coefficient is so close to 1, the strong correlation indicates that the second and quicker test could be used in place of the longer test-interview.

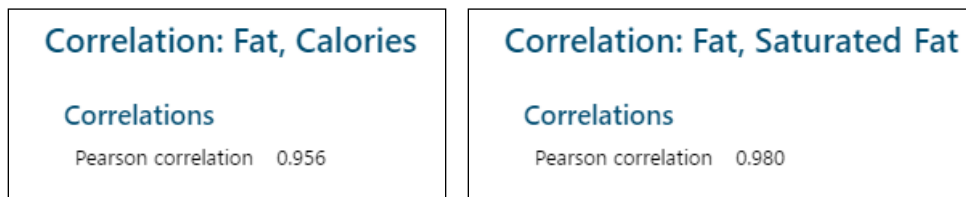
- 3.R.5 a-b** The data is graphed as a scatterplot in the figure that follows, with the time in months plotted on the horizontal axis and the number of books on the vertical axis. The data points are then connected to form a line graph. There is a very distinct pattern in this data, with the number of books increasing with time, a response which might be modeled by a quadratic equation. The professor's productivity appears to increase, with less time required to write later books.



**3.R.6 a-c** Answers will vary. The four scatterplots are generated by *Minitab* in the figures that follow.

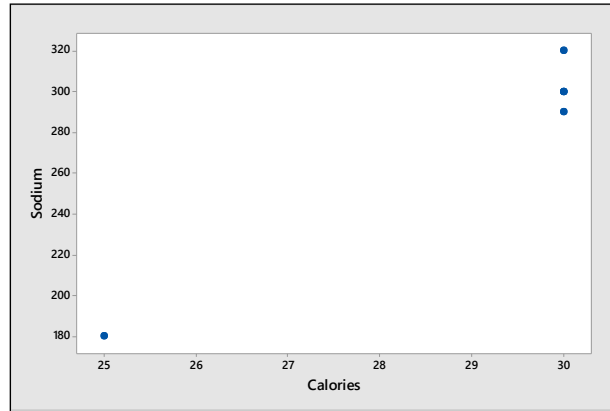


**d-e** The first two pairs of variables (parts b and c) show strong relationships. The correlation coefficients for these two pairs are given in the *Minitab* printout. The second two pairs are not strongly related and, except for one unusual measurement in the upper right corner of the graph, exhibit a relatively random scatter. Notice the similarity in the patterns of the top row of figures and the bottom row of figures. This probably is because fat and cholesterol are highly correlated, as are saturated fat and calories.



**3.R.7 a-b** The scatterplot is shown next. Four of the measurements are clustered together, and the other (Smart Beat American) is an outlier containing much less sodium and calories. There seems to be very little pattern to the plot.



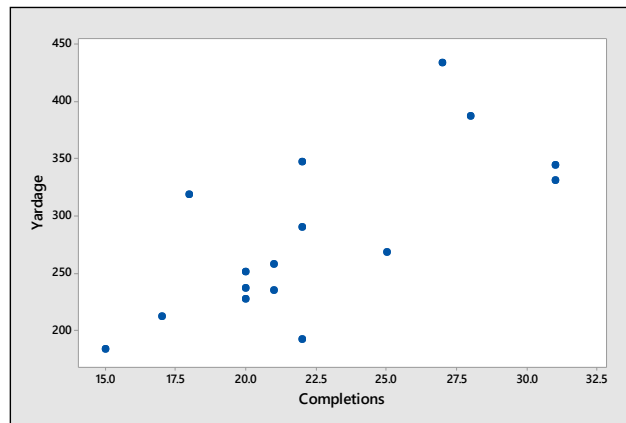


**b** Based only on these two variables, *Smart Beat* would be the brand of choice. However, it might taste poorly, or have some other undesirable qualities (cholesterol, price, melting ability, etc.).

**3.R.8 a** For each student in the class, the following variables are recorded: gender (qualitative) and height (quantitative).

**b** Males are typically taller than females. If you look at the two groups separately, the box plots show two relatively symmetric distributions, probably mound-shaped, with one unusually tall female. When the two groups are mixed together, it causes the two local peaks seen in the histogram.

**3.R.9 a** The scatterplot is shown next. The relationship is positive linear.



**b** There are no outliers to the linear trend; however, in Weeks 9 and 11, Rivers threw an unusually low number of passes.

**c**  $n = 16$ ;  $\sum x_i = 360$ ;  $\sum y_i = 4515$ ;  $\sum x_i^2 = 8432$ ;  $\sum y_i^2 = 1,351,901$ ;  $\sum x_i y_i = 105,161$ .

Then the covariance is

$$s_{xy} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{n-1} = 238.233333$$

The sample standard deviations are  $s_x = 4.704608$  and  $s_y = 72.0298260$  so that  $r = 0.70302$ .

**d** Calculate (using full accuracy)

$$b = r \frac{s_y}{s_x} = 10.76355 \text{ and } a = \bar{y} - b\bar{x} = 282.1875 - 10.76355(22.5) = 40.008$$

and the equation of the regression line is  $\hat{y} = 40.01 + 10.76x$ .

**e** Use the regression line from part **d**. When  $x = 20$ ,  $\hat{y} = 40.01 + 10.76(20) = 255.21$ .

**3.R.10 a** The variables measured are amount of aluminum oxide (quantitative), and archeological site (qualitative).

**b** Looking at the box plots, you can see that there are higher levels of aluminum oxide at the Ashley Rails and Island Thorns sites. The variability is about the same at the three sites.

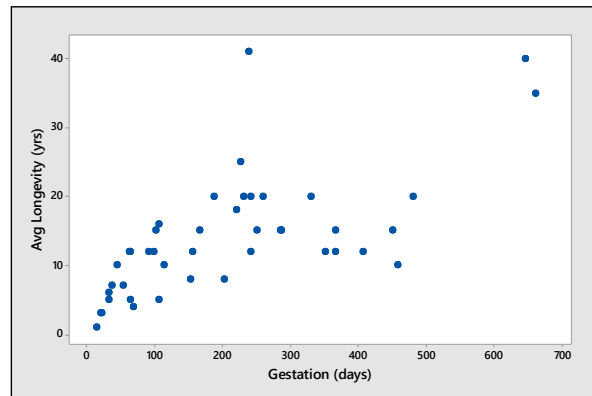
**3.R.11 a** The calculations for the correlation coefficients are the same as used in previous exercises. The *Minitab* printout shows the three correlation coefficients in a 2 x 2 matrix.

Correlations		
	Al	Fe
Fe	-0.617	
Mg	-0.189	0.626

Cell Contents  
Pearson correlation

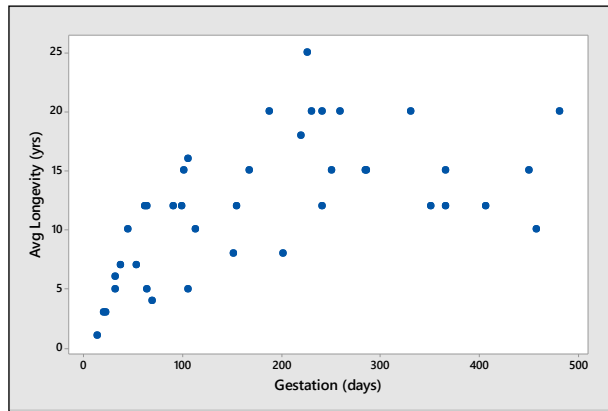
**b** There is a strong positive relationship between iron and magnesium oxide, while there is a strong negative relationship between aluminum and iron oxide. There is very little relationship between aluminum and magnesium oxide.

**3.R.12 a-b** The scatterplot is shown next. The relationship appears to be positive linear, but fairly weak.

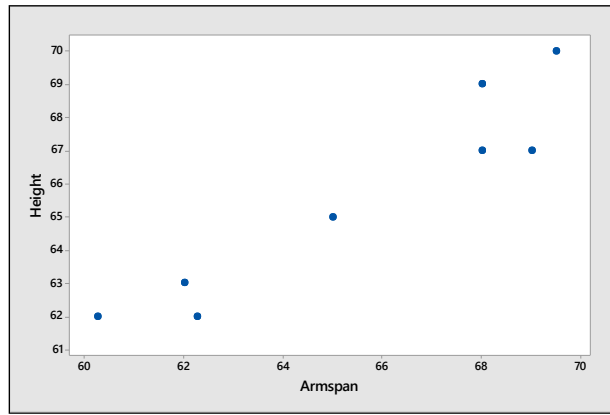


**c** There is one outlier to the linear trend [hippo-(238, 41)] and two data points that are unusually large [African elephant-(660, 35) and Asian elephant-(645, 40)].

**d** When the three data points from part **c** are removed, the scatterplot is as shown on the next page. It appears that the relationship may be more curvilinear than linear.



**3.R.13 a** The scatterplot is shown next. There is a positive linear relationship between armspan and height.



**b** Calculate  $n = 8$ ;  $\sum x_i = 524$ ;  $\sum y_i = 525$ ;  $\sum x_i^2 = 34413.375$ ;  $\sum y_i^2 = 34521$ ;  $\sum x_i y_i = 34462$ .

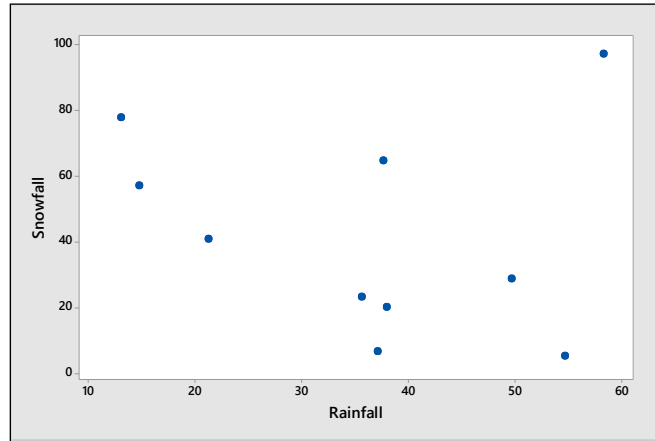
Then the covariance is  $s_{xy} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{n-1} = 10.64286$ . The sample standard deviations are  $s_x = 3.61297$  and  $s_y = 3.1139089$  so that  $r = 0.946$ .

**c** Since DaVinci indicated that a person's armspan is roughly equal to his height, the slope of the line should be approximately equal to 1.

**d** Calculate (using full accuracy)  $b = r \frac{s_y}{s_x} = .815$  and  $a = \bar{y} - b\bar{x} = 65.625 - .815(65.5) = 12.22$  and the equation of the regression line is  $\hat{y} = 12.22 + .815x$ .

**e** Use the regression line from part **d**. When  $x = 62$ ,  $\hat{y} = 12.22 + .815(62) = 62.75$  inches.

**3.R.14 a** The scatterplot is shown next.



**b** Calculate

$$n = 10; \sum x_i = 359.87; \sum y_i = 420.30; \sum x_i^2 = 15,189.9963; \sum y_i^2 = 26,344.81; \sum x_i y_i = 14,323.507.$$

Then the covariance is  $s_{xy} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{n-1} = -89.092122$ . The sample standard deviations are  $s_x = 15.7739399$  and  $s_y = 31.054792$  so that  $r = -0.182$ . There is a weak negative relationship between rainfall and snowfall.

**c** There is one outlier in the scatterplot—corresponding to Juneau, Alaska. Juneau has a lot of rain **and** a lot of snow!

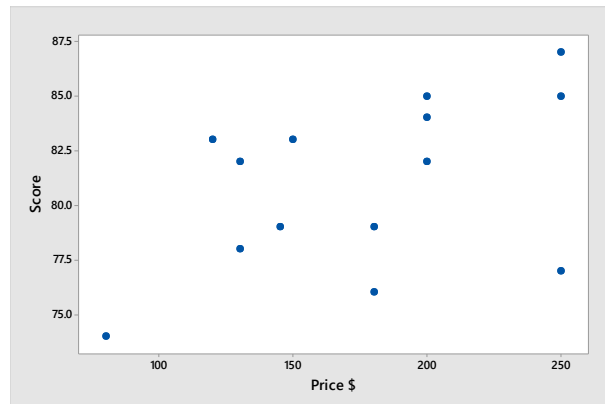
**d** With the outlier removed, calculate

$$n = 9; \sum x_i = 301.54; \sum y_i = 323.3; \sum x_i^2 = 11,787.3074; \sum y_i^2 = 16,935.81; \sum x_i y_i = 8665.497.$$

Then the covariance is  $s_{xy} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{n-1} = -270.811236$ . The sample standard deviations are  $s_x = 14.511535$  and  $s_y = 25.7928177$  so that  $r = -0.724$ .

The correlation between rainfall and snowfall now shows a fairly strong linear relationship.

**3.R.15 a** The scatterplot is shown next. The relationship is positive linear, but very weak.



**b** Calculate  $n = 14; \sum x_i = 2465; \sum y_i = 1134; \sum x_i^2 = 470,425; \sum y_i^2 = 92,048; \sum x_i y_i = 200,935$ . Then

the covariance is  $s_{xy} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{n-1} = 97.6923077$ . The sample standard deviations are

$s_x = 52.921516$  and  $s_y = 3.8630399$  so that  $r = .47786$ . Calculate (using full accuracy)  $b = r \frac{s_y}{s_x} = .034882$

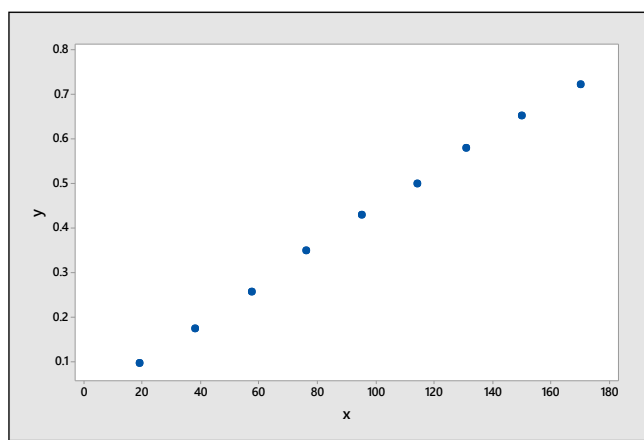
and  $a = \bar{y} - b\bar{x} = 81 - .034882(176.07143) = 74.858$  and the equation of the regression line is

$$\hat{y} = 74.858 + .035x.$$

**c** Because the correlation between  $x$  and  $y$  is so weak, the fit is not good, and the regression line will not be effective.

### On Your Own

**3.R.16** The scatterplot for these two quantitative variables is shown next. Notice the almost perfect positive correlation. The correlation coefficient and best fitting line can be calculated for descriptive purposes.



Calculate  $n = 9$ ;  $\sum x_i = 850.8$ ;  $\sum y_i = 3.755$ ;  $\sum x_i^2 = 101,495.78$ ;  $\sum y_i^2 = 1.941467$ ;  $\sum x_i y_i = 443.7727$ .

Then the covariance is  $s_{xy} = \frac{\sum x_i y_i - (\sum x_i)(\sum y_i)}{n - 1} = 11.10000417$ . The sample standard deviations are

$s_x = 51.31620$  and  $s_y = 0.216448$  so that  $r = 0.999$ . Then

$$b = r \frac{s_y}{s_x} = 0.004215 \text{ and } a = \bar{y} - b\bar{x} = 0.417222 - 0.004215(94.5333) = 0.0187 \text{ and the equation of}$$

the regression line is  $\hat{y} = 0.0187 + 0.0042x$ .

**R.17-18** Answers will vary. Here are the correlations for the six variables. There are strong positive correlations between total gross and number of weeks in the theatre, and between weekend gross and average gross.

Correlations					
	Weekend Gross	Theater Count	Average	Total Gross	Budget
Theater Count	0.433				
Average	0.948	0.141			
Total Gross	-0.116	0.210	-0.185		
Budget	0.087	0.397	-0.057	0.064	
No. of Weeks	-0.459	-0.047	-0.468	0.737	0.133
Cell Contents					
Pearson correlation					

**3.R.19 a** There seems to be a large cluster of points in the lower left hand corner, but the data points show no apparent relationship between the variables.

**b** The pattern described in parts **a** and **c** would indicate a weak correlation:

$$r = \frac{s_{xy}}{s_x s_y} = \frac{-67.770}{\sqrt{(9346.603)(705.363)}} = -0.026$$

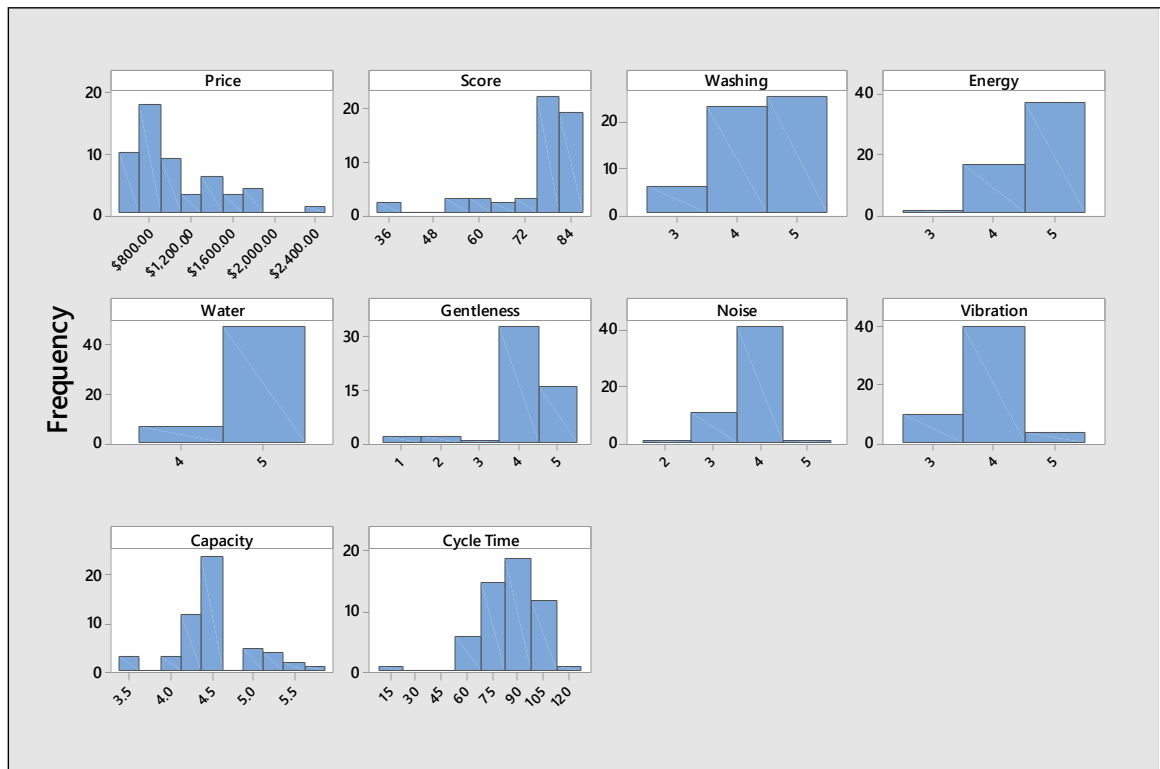
**d** Number of waste sites is only slightly affected by the size of the state. Some other possible explanatory variables might be local environmental regulations, population per square mile, or geographical region in the United States.

**3.R.20** Answers will vary. Students should notice the two distinct clusters in the data and a positive linear relationship between the two variables. If the eruption lasts only a short amount of time, the waiting time until the next eruption is also short, indicating that the pressure buildup within the geyser was not totally dissipated during the eruption.

## Case Study: Are Your Clothes are *Really* Clean?

**1** We have used *Minitab* to describe the five variables, and the printout is shown next.

Histograms for the variables show a variety of shapes, some skewed, some relatively symmetric, and some which cannot be described with a well-behaved shape. “Price” and “Score” are definitely skewed; “Cycle Time” would be fairly mound-shaped except for an outlier in the left tail.



**2-3** The pairs of variables can be compared by looking at the pairwise correlations for each pair.

### Correlations

	Price	Score	Washing	Energy	Water	Gentleness
Score	0.232					
Washing	0.063	0.408				
Energy	0.011	0.453	-0.036			
Water	0.063	0.465	-0.208	0.397		
Gentleness	0.364	0.677	-0.086	0.478	0.661	
Noise	0.074	0.426	0.234	0.366	0.049	0.257
Vibration	0.349	0.614	0.117	0.366	0.247	0.443
Capacity	0.076	0.257	0.158	0.333	-0.193	0.012
Cycle Time	0.041	0.190	0.202	0.228	-0.044	0.171
	Noise	Vibration	Capacity			
Vibration	0.350					
Capacity	0.589	0.408				
Cycle Time	0.417	0.281	0.549			

Cell Contents

Pearson correlation

In examining the correlations with Price, only Gentleness and Vibration have a positive correlation that is not close to zero. However, Score has fairly large correlations with the factors Washing, Energy, Water, Gentleness, Noise and Vibration. It appears that the score is a better indicator of quality than price, since a high score is positively correlated with almost all the factors included in the investigation.

## 4: Probability

### Section 4.1

**4.1.1-6** This experiment involves tossing a single die and observing the outcome. The sample space for this experiment consists of the following simple events:

$E_1$ : Observe a 1	$E_4$ : Observe a 4
$E_2$ : Observe a 2	$E_5$ : Observe a 5
$E_3$ : Observe a 3	$E_6$ : Observe a 6

Events A through F are compound events and are composed in the following manner:

1. A:  $\{2\}$
2. B:  $\{2, 4, 6\}$
3. C:  $\{3, 4, 5, 6\}$
4. D:  $\{1, 2, 3, 4\}$
5. E:  $\{2, 4, 6\}$
6. F: contains no simple events

**4.1.7** Both A and B:  $\{E_3\}$

**4.1.8** A or B or both:  $\{E_1, E_3, E_4, E_5, E_6, E_7\}$

**4.1.9** B or C or both:  $\{E_1, E_2, E_3, E_4, E_5, E_7\}$

**4.1.10** Both A and C:  $\{E_4\}$

**4.1.11** A or C or both:  $\{E_2, E_3, E_4, E_6\}$

**4.1.12** Not A:  $\{E_1, E_2, E_5, E_7\}$

**4.1.13** Both A and B:  $\{E_2\}$ ; A or B or both:  $\{E_1, E_2, E_3\}$ ; Not B:  $\{E_1, E_4\}$

**4.1.14** Both A and B:  $\{E_2, E_4\}$ ; A or B or both:  $\{E_2, E_3, E_4\}$ ; Not B:  $\{E_1\}$

**4.1.15** Both A and B:  $\{E_1\}$ ; A or B or both:  $\{E_1, E_3, E_4\}$ ; Not B:  $\{E_2\}$

**4.1.16** There are 36 pairs generated when the die is tossed twice.

(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

**4.1.17** There are four simple events: (H, H), (H, T), (T, H), (T, T)

**4.1.18** There are only four possible grades: Freshman, Sophomore, Junior, Senior

**4.1.19** Each of the three children can be either male (M) or female (F), generating eight possible simple events:

(M, M, M)	(M, M, F)	(M, F, M)	(F, M, M)
(F, F, M)	(F, M, F)	(M, F, F)	(F, F, F)

**4.1.20** There are 52 possible outcomes when one card is drawn from a deck of 52:

**Hearts:** Ace, King, Queen, Jack, 10, 9, 8, 7, 6, 5, 4, 3, 2

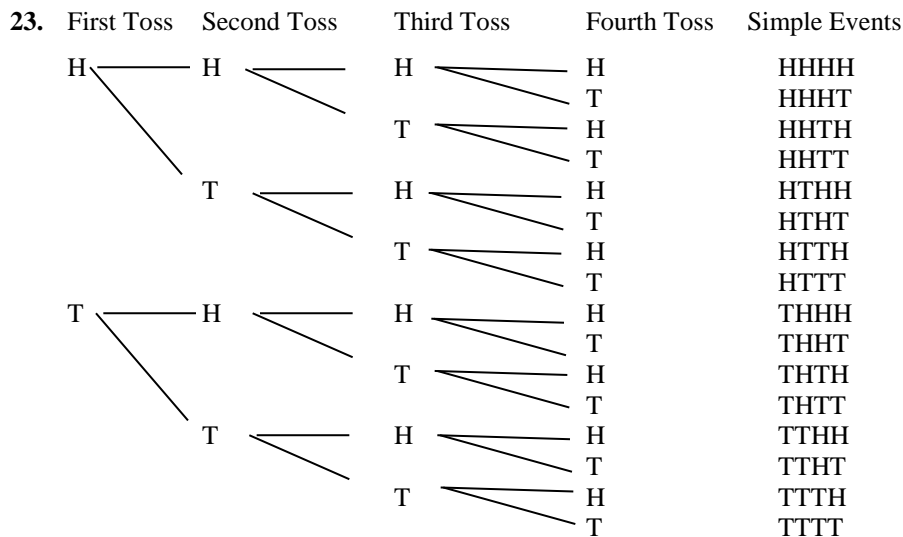
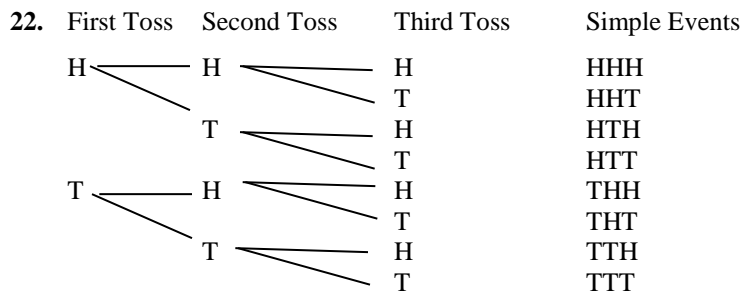
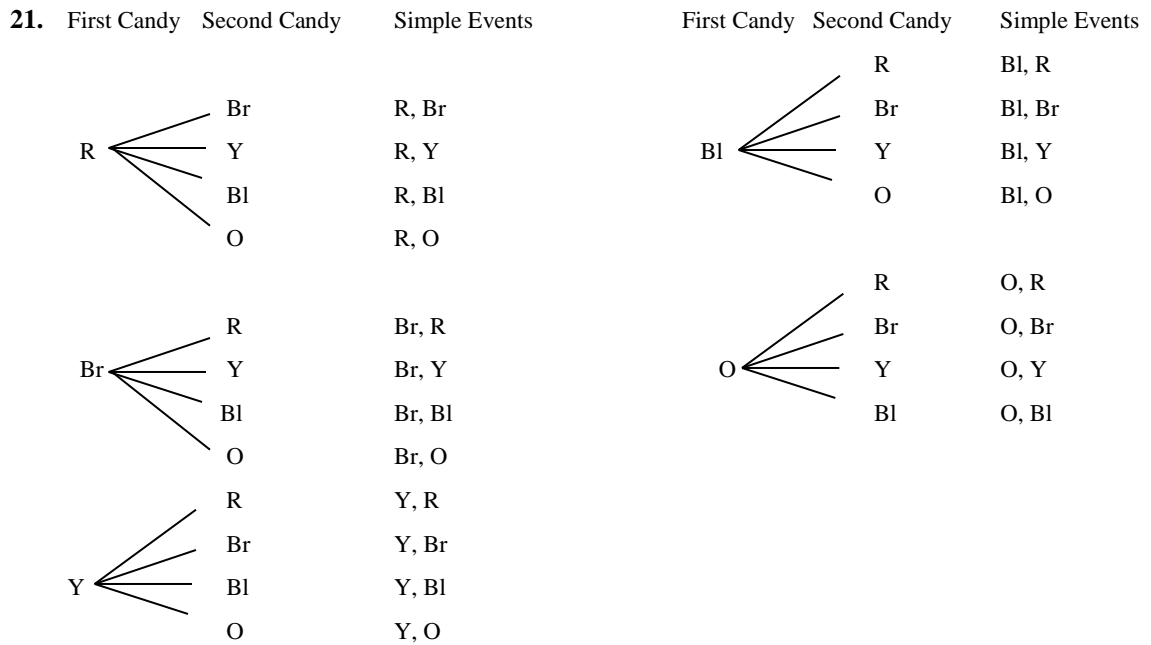
**Diamonds:** Ace, King, Queen, Jack, 10, 9, 8, 7, 6, 5, 4, 3, 2

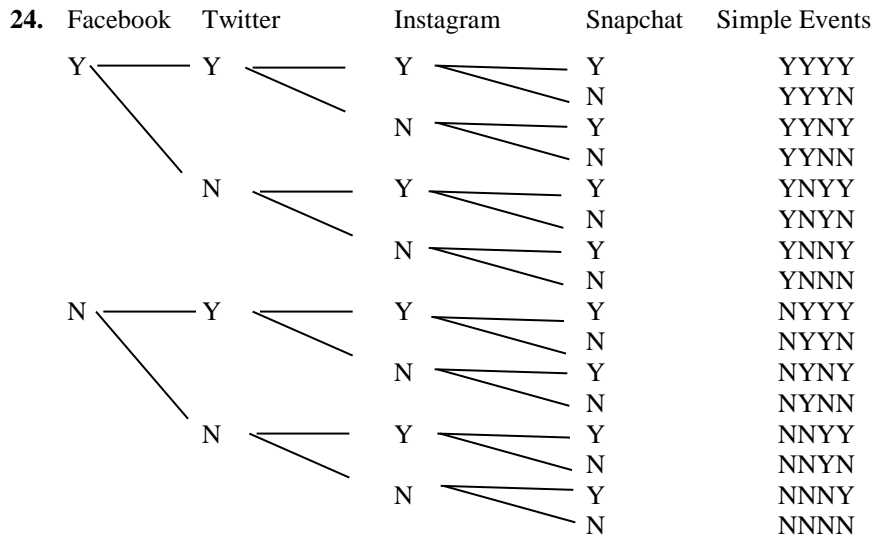
**Clubs:** Ace, King, Queen, Jack, 10, 9, 8, 7, 6, 5, 4, 3, 2

**Spades:** Ace, King, Queen, Jack, 10, 9, 8, 7, 6, 5, 4, 3, 2



**4.1.21-24** Tree diagrams can be used to identify simple events when the experiment is accomplished in stages. Each successive branching (column) corresponds to one of the stages necessary to generate the final outcome.





**4.1.25-27** A table of outcomes can be used to identify simple events when the experiment is accomplished in two stages. The columns and rows correspond to the outcomes on the two stages necessary to generate the final outcome.

**25.**

	Homeschooled (H)	Not homeschooled (NH)
Male (M)	H, M	NH, M
Female (F)	H, F	NH, F

**26.**

	Club	Heart	Diamond	Spade
Face card (FC)	C, FC	H, FC	D, FC	S, FC
Not face card (NFC)	C, NFC	H, NFC	D, NFC	S, NFC

**27.**

	Blonde (BL)	Brown (BR)	Black (BLK)	Red (R)	Other (O)
Male (M)	BL, M	BR, M	BLK, M	R, M	O, M
Female (F)	BL, F	BR, F	BLK, F	R, F	O, F

**4.1.28** Label the five balls as  $R_1$ ,  $R_2$ ,  $R_3$ ,  $Y_1$  and  $Y_2$ . The selection of two balls is accomplished in two stages to produce the simple events in the tree diagram that follows.

First Ball	Second Ball	Simple Events	First Ball	Second Ball	Simple Events
R <sub>1</sub>	R <sub>2</sub>	R <sub>1</sub> R <sub>2</sub>	Y <sub>1</sub>	R <sub>1</sub>	Y <sub>1</sub> R <sub>1</sub>
	R <sub>3</sub>	R <sub>1</sub> R <sub>3</sub>		R <sub>2</sub>	Y <sub>1</sub> R <sub>2</sub>
	Y <sub>1</sub>	R <sub>1</sub> Y <sub>1</sub>		R <sub>3</sub>	Y <sub>1</sub> R <sub>3</sub>
R <sub>2</sub>	Y <sub>2</sub>	R <sub>1</sub> Y <sub>2</sub>	Y <sub>2</sub>	Y <sub>2</sub>	Y <sub>1</sub> Y <sub>2</sub>
	R <sub>1</sub>	R <sub>2</sub> R <sub>1</sub>		R <sub>1</sub>	Y <sub>2</sub> R <sub>1</sub>
	R <sub>3</sub>	R <sub>2</sub> R <sub>3</sub>		R <sub>2</sub>	Y <sub>2</sub> R <sub>2</sub>
R <sub>3</sub>	Y <sub>1</sub>	R <sub>2</sub> Y <sub>1</sub>		R <sub>3</sub>	Y <sub>2</sub> R <sub>3</sub>
	Y <sub>2</sub>	R <sub>2</sub> Y <sub>2</sub>		Y <sub>1</sub>	Y <sub>2</sub> Y <sub>1</sub>
	R <sub>1</sub>	R <sub>3</sub> R <sub>1</sub>			
	R <sub>2</sub>	R <sub>3</sub> R <sub>2</sub>			
	Y <sub>1</sub>	R <sub>3</sub> Y <sub>1</sub>			
	Y <sub>2</sub>	R <sub>3</sub> Y <sub>2</sub>			

**4.1.29** When the first ball is replaced before the second ball is drawn, five additional simple events become possible:

$R_1R_1, R_2R_2, R_3R_3, Y_1Y_1,$  and  $Y_2Y_2$

## Section 4.2

**4.2.1-6** From Exercises 1-6 (Section 4.1), the simple events are:

$E_1$ : Observe a 1

$E_2$ : Observe a 2

$E_3$ : Observe a 3

$E_4$ : Observe a 4

$E_5$ : Observe a 5

$E_6$ : Observe a 6

Since the simple events  $E_i, i = 1, 2, 3, \dots, 6$  are equally likely,  $P(E_i) = 1/6$ . To find the probability of an event, we sum the probabilities assigned to the simple events in that event.

$$1. \quad P(A) = P(E_2) = \frac{1}{6}$$

$$2. \quad P(B) = P(E_2) + P(E_4) + P(E_6) = \frac{3}{6} = \frac{1}{2}$$

$$3. \quad P(C) = \frac{4}{6} = \frac{2}{3}$$

$$4. \quad P(D) = P(E_1) + P(E_2) + P(E_3) + P(E_4) = \frac{4}{6} = \frac{2}{3}$$

$$5. \quad P(E) = P(E_2) + P(E_4) + P(E_6) = \frac{3}{6} = \frac{1}{2}$$

6. Since event  $F$  contains no simple events,  $P(F) = 0$ .

**4.2.7-10** It is given that events  $E_1$  through  $E_6$  are equally likely, but that event  $E_7$  is twice as likely as the others. Define  $P(E)$  to be the probability of one of the first six events. Since all seven probabilities must add to 1, we can write

$$6P(E) + 2P(E) = 1 \Rightarrow 8P(E) = 1 \text{ or } P(E) = 1/8 \text{ for events } E_1 \text{ through } E_6 \text{ and } P(E_7) = 2\left(\frac{1}{8}\right) = \frac{1}{4}.$$

$$7. \quad P(A) = P(E_3) + P(E_4) + P(E_6) = \frac{3}{8}$$

$$8. \quad P(B) = P(E_1) + P(E_3) + P(E_5) + P(E_7) = \frac{3}{8} + \frac{2}{8} = \frac{5}{8}$$

$$9. \quad P(C) = P(E_2) + P(E_4) = \frac{2}{8} = \frac{1}{4}$$

10. Since there are no events in common to all three events,  $P(A \text{ and } B \text{ and } C) = 0$

4.2.11 It is given that  $P(E_1) = P(E_2) = .15$  and  $P(E_3) = .40$ . Since  $\sum_s P(E_i) = 1$ , we know that

$$P(E_4) + P(E_5) = 1 - .15 - .15 - .40 = .30 \quad (i)$$

Also, it is given that

$$P(E_4) = 2P(E_5) \quad (ii)$$

We have two equations in two unknowns which can be solved simultaneously for  $P(E_4)$  and  $P(E_5)$ . Substituting equation (ii) into equation (i), we have

$$2P(E_5) + P(E_5) = .3$$

$$3P(E_5) = .3 \text{ so that } P(E_5) = .1$$

$$\text{Then from (i), } P(E_4) + .1 = .3 \text{ and } P(E_4) = .2.$$

4.2.12-13 Use the results of Exercise 11. To find the necessary probabilities, sum the probabilities of the simple events:

$$P(A) = P(E_1) + P(E_3) + P(E_4) = .15 + .4 + .2 = .75$$

$$P(B) = P(E_2) + P(E_3) = .15 + .4 = .55$$

4.2.14 Use the results of Exercise 12-13. The following events are in either A or B or both:  $\{E_1, E_2, E_3, E_4\}$ .

$$P(A \text{ or } B \text{ or both}) = P(E_1) + P(E_2) + P(E_3) + P(E_4) = .15 + .15 + .4 + .2 = .90$$

4.2.15 The events that are not in event A are:  $\{E_2, E_5\}$  so that  $P(\text{not } A) = .15 + .1 = .25$ .

4.2.16 It is given that  $P(E_1) = .45$  and that  $3P(E_2) = .45$ , so that  $P(E_2) = .15$ . Since  $\sum_s P(E_i) = 1$ , the remaining 8 simple events must have probabilities whose sum is  $P(E_3) + P(E_4) + \dots + P(E_{10}) = 1 - .45 - .15 = .4$ . Since it is given that they are equiprobable,

$$P(E_i) = \frac{.4}{8} = .05 \text{ for } i = 3, 4, \dots, 10$$

4.2.17 The simple events are listed in Exercise 20 (Section 4.1). Since each card is equally likely to be drawn, the probability of each simple event is  $1/52$  and the probability of an ace is  $4/52 = 1/13$ .

4.2.18 The simple events are listed in Exercise 20 (Section 4.1). Since each card is equally likely to be drawn, the probability of each simple event is  $1/52$ . There are 12 cards with numbers less than 5 (2, 3 and 4 in each of four suits) and the probability  $12/52 = 3/13$ .

4.2.19 The simple events are shown in Exercise 19 (Section 4.1). Each simple event is equally likely, with probability  $1/8$ .

(M, M, M)	(M, M, F)	(M, F, M)	(F, M, M)
(F, F, M)	(F, M, F)	(M, F, F)	(F, F, F)

The probability of two boys and a girl is  $P(MMF) + P(MFM) + P(FMM) = 3/8$

4.2.20 Refer to Exercise 19. The probability of three girls is  $P(FFF) = 1/8$ .

4.2.21 The simple events are shown in Exercise 21 (Section 4.1). Each simple event is equally likely, with probability  $1/20$ . If we designate the first candy in the pair to be your candy, the event of interest consists of three simple events--(O, BR), (O, Y) and (O, BL)—and the probability is  $3/20$ .

- 4.2.22** Refer to Exercise 21. The event of interest consist of four simple events—(BL, R), (BL, BR), (O, R) and (O, BR)—and the probability is  $4/20 = 1/5$ .
- 4.2.23** The simple events are shown in Exercise 16 (Section 4.1). Each simple event is equally likely, with probability  $1/36$ . The event of interest consists of four simple events—(6, 3), (6, 4), (6, 5) and (6, 6)—and the probability is  $4/36 = 1/9$ .
- 4.2.24** Refer to Exercise 23 and Exercise 16 (Section 4.1). The event of interest consists of two simple events—(6, 5) and (5, 6)—and the probability is  $2/36 = 1/18$ .
- 4.2.25** There are 38 simple events, each corresponding to a single outcome of the wheel's spin. The 38 simple events are indicated next.

$E_1$ : Observe a 1       $E_2$ : Observe a 2    ...       $E_{36}$ : Observe a 36  
 $E_{37}$ : Observe a 0       $E_{38}$ : Observe a 00

Since any pocket is just as likely as any other,  $P(E_i) = 1/38$ . The probability of observing either a 0 or a 00 contains two simple events,  $E_{37}$  and  $E_{38}$ , so that the probability is  $P(E_{37}) + P(E_{38}) = 2/38 = 1/19$ .

- 4.2.26** Refer to Exercise 25. Define event A as the event that you win on a single spin. Since you have bet on the numbers 1 through 18, event A contains 18 simple events,  $E_1, E_2, \dots, E_{18}$ . Then

$$P(A) = P(E_1) + P(E_2) + \dots + P(E_{18}) = 18/38 = 9/19$$

- 4.2.27 a** It is required that  $\sum_s P(E_i) = 1$ . Hence,  $P(E_2) = 1 - .49 - .21 - .09 = .21$

**b** The player will hit on at least one of the two free throws if he hits on the first, the second, or both. The associated simple events are  $E_1, E_2$ , and  $E_3$  and  $P(\text{hits on at least one}) = P(E_1) + P(E_2) + P(E_3) = .91$ .

- 4.2.28 a** The experiment consists of choosing three coins at random from four. The order in which the coins are drawn is unimportant. Hence, each simple event consists of a triplet, indicating the three coins drawn. Using the letters N, D, Q, and H to represent the nickel, dime, quarter, and half-dollar, respectively, the four possible simple events are listed next.

$E_1$ : (NDQ)       $E_2$ : (NDH)       $E_3$ : (NQH)       $E_4$ : (DQH)

The event that a half-dollar is chosen is associated with the simple events  $E_2, E_3$ , and  $E_4$ . Hence,

$$P[\text{choose a half-dollar}] = P(E_2) + P(E_3) + P(E_4) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$$

since each simple event is equally likely.

- b** The simple events along with their monetary values follow:

$E_1$  NDQ \$0.40       $E_2$  NDH 0.65  
 $E_3$  NQH 0.80       $E_4$  DQH 0.85

Hence,  $P[\text{total amount is \$0.60 or more}] = P(E_2) + P(E_3) + P(E_4) = 3/4$ .

- 4.2.29 a** The experiment is accomplished in two stages, as shown in the tree diagram shown next.

<u>Gender</u>	<u>Preschool</u>	<u>Simple Events</u>	<u>Probability</u>
Male	Yes	$E_1$ : Male, Preschool	8/25
	No	$E_2$ : Male, No preschool	6/25
Female	Yes	$E_3$ : Female, Preschool	9/25
	No	$E_4$ : Female, No preschool	2/25

**b** Since each of the 25 students are equally likely to be chosen, the probabilities will be proportional to the number of students in each of the four gender-preschool categories. These probabilities are shown in the last column of the tree diagram above.

**c** 
$$P(\text{Male}) = P(E_1) + P(E_2) = \frac{8}{25} + \frac{6}{25} = \frac{14}{25}$$

**d** 
$$P(\text{Female and no preschool}) = P(E_4) = \frac{2}{25}$$

**4.2.30** The four possible outcomes of the experiment, or simple events, are represented as the cells of a  $2 \times 2$  table, and have probabilities as given in the table.

- a**  $P[\text{adult judged to need glasses}] = .44 + .14 = .58$   
**b**  $P[\text{adult needs glasses but does not use them}] = .14$   
**c**  $P[\text{adult uses glasses}] = .44 + .02 = .46$   
**d**  $P[\text{adult used glasses when they didn't need them}] = .02$

**4.2.31 a** This experiment consists of two patients, each swallowing one of four tablets (two cold and two aspirin). There are four tablets to choose from, call them  $C_1$ ,  $C_2$ ,  $A_1$  and  $A_2$ . The resulting simple events are then all possible ordered pairs which can be formed from the four choices. There are 12 simple events, each with equal probability  $1/12$ .

$(C_1C_2) \quad (C_2C_1) \quad (A_1C_1) \quad (A_2C_1)$   
 $(C_1A_1) \quad (C_2A_1) \quad (A_1C_2) \quad (A_2C_2)$   
 $(C_1A_2) \quad (C_2A_2) \quad (A_1A_2) \quad (A_2A_1)$

Notice that it is important to consider the order in which the tablets are chosen, since it makes a difference, for example, which patient (1 or 2) swallows the cold tablet.

- b**  $A = \{(C_1C_2), (C_1A_1), (C_1A_2), (C_2A_1), (C_2C_1), (C_2A_2)\}$  and the probability is  $P(A) = 6/12 = 1/2$   
**c**  $B = \{(C_1A_1), (C_1A_2), (C_2A_1), (C_2A_2), (A_1C_1), (A_1C_2), (A_2C_1), (A_2C_2)\}$  and the probability is  $P(B) = 8/12 = 2/3$ .  
**d**  $C = \{(A_1A_2), (A_2A_1)\}$  and the probability is  $P(C) = 2/12 = 1/6$ .

**4.2.32** Refer to Exercise 31. Again, there are 12 simple events, each with equal probability  $1/12$ . By summing the probabilities of simple events in the events of interest we have

- a**  $P(A \text{ or } B \text{ or both}) = 10/12 = 5/6$       **b**  $P(\text{both } A \text{ and } B) = 4/12 = 1/3$   
**c**  $P(A \text{ or } C \text{ or both}) = 8/12 = 2/3$       **d**  $P(\text{both } A \text{ and } C) = 0$

**4.2.33 a** *Experiment:* Select three people and record their gender (M or F).

**b** The simple events are the same as those shown in Exercise 19 (Section 4.1):

$\begin{array}{cccc} \text{FFF} & \text{FMM} & \text{MFM} & \text{MMF} \\ \text{MFF} & \text{FMF} & \text{FFM} & \text{MMM} \end{array}$

- c** Since there are  $N = 8$  equally likely simple events, each is assigned probability,  $P(E_i) = 1/N = 1/8$ .

**d-e** Sum the probabilities of the appropriate simple events:

$$P(\text{only one man}) = P(MFF) + P(FMF) + P(FFM) = 3\left(\frac{1}{8}\right) = \frac{3}{8}$$

$$P(\text{all three are women}) = P(FFF) = \frac{1}{8}$$

**4.2.34 a** Refer to Exercise 33. List the six jurors as  $M_1, M_2, M_3, M_4, F_1$  and  $F_2$ . The simple events can be found using a tree diagram and are shown next:

$M_1M_2$	$M_1M_3$	$M_1M_4$	$M_1F_1$	$M_1F_2$
$M_2M_3$	$M_2M_4$	$M_2F_1$	$M_2F_2$	$M_3M_4$
$M_3F_1$	$M_3F_2$	$M_4F_1$	$M_4F_2$	$F_1F_2$

**b** Since each simple event is equally likely with probability  $1/15$ ,

$$P(\text{both jurors are women}) = P(F_1F_2) = \frac{1}{15}$$

**4.2.35** A taster tastes and ranks three varieties of tea A, B, and C, according to preference and the simple events in  $S$  are in triplet form, each with probability  $1/6$ :

$E_1 : (A, B, C)$	$E_4 : (B, C, A)$
$E_2 : (A, C, B)$	$E_5 : (C, B, A)$
$E_3 : (B, A, C)$	$E_6 : (C, A, B)$

Here the most desirable is in the first position, the next most desirable is in the second position, and the least desirable is in third position. Define the events

D: variety A is ranked first and F: variety A is ranked third

Then  $P(D) = P(E_1) + P(E_2) = 1/6 + 1/6 = 1/3$  and  $P(F) = P(E_5) + P(E_6) = 1/6 + 1/6 = 1/3$

**4.2.36** Visualize a tree diagram with four stages – selecting the runner who places first, second, third and fourth, respectively. There are four choices at the first stage, three choices (branches) at the second stage, two choices (branches) at the third stage, and only one choice (branch) for the last runner. The total number of simple events is  $4(3)(2)(1) = 24$ . The simple events are listed next:

JBED	JEBD	JEDB	JBDE	JDEB	JDBE
BJED	BEJD	BEDJ	BJDE	BDEJ	BDJE
EBJD	EJBD	EJDB	EBDJ	EDJB	EDBJ
DBEJ	DEBJ	DEJB	DBJE	DJEB	DJBE

**a** If all runners are equally qualified, the simple events will be equally likely with probability

$$P(E_i) = 1/24. \text{ Sum the probabilities of the appropriate simple events to find } P(\text{Dave wins}) = \frac{6}{24} = \frac{1}{4}.$$

$$\text{b } P(\text{Dave is first, John is second}) = \frac{2}{24} = \frac{1}{12} \quad \text{c } P(\text{Ed is last}) = \frac{6}{24} = \frac{1}{4}$$

**4.2.37** Similar to Exercise 30. The four possible outcomes of the experiment, or simple events, are represented as the cells of a  $2 \times 2$  table and have probabilities (when divided by 300) as given in the table.

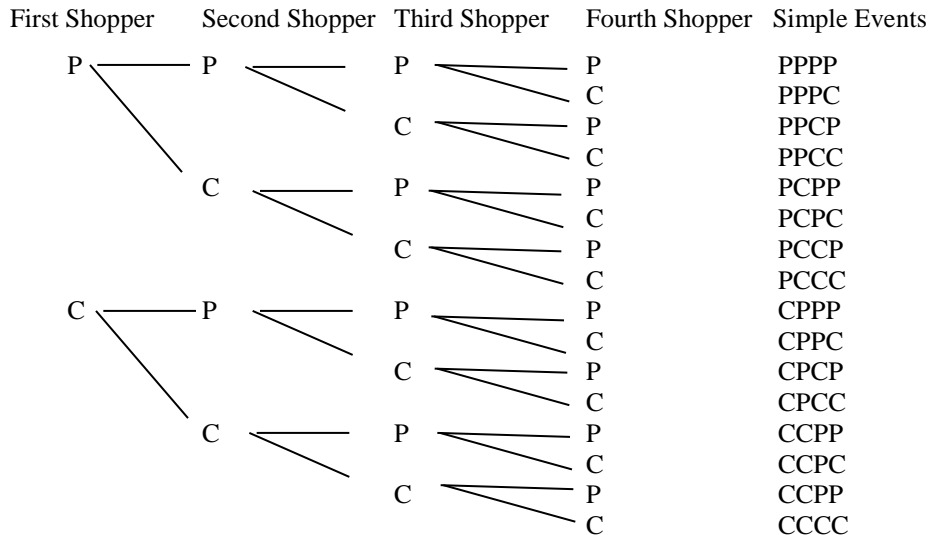
**a**  $P(\text{normal eyes and normal wing size}) = 140/300 = .467$

**b**  $P(\text{vermillion eyes}) = (3 + 151)/300 = 154/300 = .513$

**c**  $P(\text{either vermillion eyes or miniature wings or both}) = (3 + 151 + 6)/300 = 160/300 = .533$

**4.2.38** Since there are four possible shapes for each of the three slots, there are a total of  $N = (4)(4)(4) = 64$  possible configurations for the three slots. Of these, only  $n_A = 4$  (CCC, LLL, SSS, or BBB) will result in a win. Hence, the probability of winning is  $P(\text{win}) = 4/64 = 1/16$ .

**4.2.39 a** Define  $P$ : shopper prefers Pepsi and  $C$ : shopper prefers Coke. There are 16 simple events in the experiment, shown in the tree diagram shown next.



Then if there is actually no difference in the taste, each simple event is equally likely, with probability  $1/16$ , and  $P(\text{all four prefer Pepsi}) = P(PPPP) = \frac{1}{16} = .0625$ .

**b**  $P(\text{exactly one prefers Pepsi}) = P(PCCC) + P(CPCC) + P(CCPC) + P(CCCP) = \frac{4}{16} = .25$

**4.2.40** The necessary probabilities can be found by summing the necessary cells in the probability table and dividing by 220, the total number of firms.

**a**  $P(\text{city } A) = \frac{114}{220} = .5182$

**b**  $P(\text{city } B \text{ and flextime}) = \frac{25}{220} = .1136$

**c**  $P(\text{no flextime}) = \frac{156}{220} = .7091$

### Section 4.3

**4.3.1** Use the  $mn$  Rule. There are  $10(8) = 80$  possible pairs.

**4.3.2** Use the extended  $mn$  Rule. There are  $4(7)(3) = 84$  possible triplets.

**4.3.3** Use the extended  $mn$  Rule. Each coin can fall in one of  $n_i = 2$  ways, and the total number of simple events is  $2(2)(2)(2) = 2^4 = 16$ .

**4.3.4** Use the extended  $mn$  Rule. The first die can fall in one of 6 ways, and the second and third die can each fall in one of 6 ways. The total number of simple events is  $6(6)(6) = 216$ .



$$4.3.5 \quad P_3^5 = \frac{5!}{2!} = 5(4)(3) = 60$$

$$4.3.6 \quad P_9^{10} = \frac{10!}{1!} = 3,628,800$$

$$4.3.7 \quad P_6^6 = \frac{6!}{0!} = 6! = 720$$

$$4.3.8 \quad P_1^{20} = \frac{20!}{19!} = 20$$

$$4.3.9 \quad C_3^5 = \frac{5!}{3!2!} = \frac{5(4)}{2(1)} = 10$$

$$4.3.10 \quad C_9^{10} = \frac{10!}{9!1!} = 10$$

$$4.3.11 \quad C_6^6 = \frac{6!}{6!0!} = 1$$

$$4.3.12 \quad C_1^{20} = \frac{20!}{1!19!} = 20$$

$$4.3.13 \quad \text{Since order is important, you use } \textit{permutations} \text{ and } P_5^8 = \frac{8!}{3!} = 8(7)(6)(5)(4) = 6720 .$$

$$4.3.14 \quad \text{Since order is unimportant, you use } \textit{combinations} \text{ and } C_2^{20} = \frac{20!}{2!18!} = \frac{20(19)}{2(1)} = 190 .$$

$$4.3.15 \quad \text{Since order is unimportant, you use } \textit{combinations} \text{ and } C_3^{10} = \frac{10!}{3!17!} = \frac{10(9)(8)}{3(2)(1)} = 120 .$$

4.3.16 Use the extended *mn* Rule. There are  $m = 4$  pairs of jeans,  $n = 12$  T-shirts and  $t = 4$  pairs of sneakers. The total number of outfits is  $mnt = 4(12)(4) = 192$  .

4.3.17 This exercise involves the arrangement of 6 different cities in all possible orders. Each city will be visited once and only once. Hence, order is important and elements are being chosen from a single set. Permutations are used and the number of arrangements is

$$P_6^6 = \frac{6!}{0!} = 6! = 6(5)(4)(3)(2)(1) = 720$$

4.3.18 Use the extended *mn* Rule. The total number of options are  $4(5)(3) = 60$

4.3.19 **a** Each student has a choice of 52 cards, since the cards are replaced between selections. The *mn* Rule allows you to find the total number of configurations for three students as  $52(52)(52) = 140,608$  .

**b** Now each student must pick a different card. That is, the first student has 52 choices, but the second and third students have only 51 and 50 choices, respectively. The total number of configurations is found using the *mn* Rule or the rule for permutations:

$$mnt = 52(51)(50) = 132,600 \quad \text{or} \quad P_3^{52} = \frac{52!}{49!} = 132,600 .$$

**c** Let  $A$  be the event of interest. Since there are 52 different cards in the deck, there are 52 configurations in which all three students pick the same card (one for each card). That is, there are  $n_A = 52$  ways for the event  $A$  to occur, out of a total of  $N = 140,608$  possible configurations from part **a**. The probability of interest is  $P(A) = \frac{n_A}{N} = \frac{52}{140,608} = .00037$ .

**d** Again, let  $A$  be the event of interest. There are  $n_A = 132,600$  ways (from part **b**) for the event  $A$  to occur, out of a total of  $N = 140,608$  possible configurations from part **a**, and the probability of interest is

$$P(A) = \frac{n_A}{N} = \frac{132,600}{140,608} = .943$$

**4.3.20** Use the extended *mn* Rule. The total number of options are  $2(2)(2) = 8$ .

**4.3.21 a** Since the order of selection for the five-card hand is unimportant, use *combinations* to find the number of possible hands as  $N = C_5^{52} = \frac{52!}{5!47!} = \frac{52(51)(50)(49)(48)}{5(4)(3)(2)(1)} = 2,598,960$ .

**b** Since there are only four different suits, there are  $n_A = 4$  ways to get a royal flush.

**c** From parts **a** and **b**,  $P(\text{royal flush}) = \frac{n_A}{N} = \frac{4}{2,598,960} = .000001539$

**4.3.22 a** From Exercise 21a,  $N = 2,598,960$ .

**b** The count is accomplished in two stages. There are 13 different ways to get four cards of the same face value (four Aces, four 'twos', and so on). Once these four cards are chosen, there are 48 cards left from which to choose the fifth card. Using the *mn* Rule, the total number of ways is  $13(48) = 624$ .

**c** From parts **a** and **b**,  $P(\text{four of a kind}) = \frac{n_A}{N} = \frac{624}{2,598,960} = .00024$

**4.3.23** Notice that a sample of 10 nurses will be the same no matter in which order they were selected. Hence, order is unimportant and combinations are used. The number of samples of 10 selected from a total of 90 is

$$C_{10}^{90} = \frac{90!}{10!80!} = \frac{2.0759076(10^{19})}{3.6288(10^6)} = 5.720645(10^{12})$$

**4.3.24 a** Since the order of selection for the two-person subcommittee is unimportant, use *combinations* to find the number of possible subcommittees as  $N = C_2^5 = \frac{5!}{2!3!} = \frac{5(4)}{(2)(1)} = 10$ .

**b** Since there is only  $n_A = 1$  subcommittee made up of Smith and Jones, the probability of this choice is  $1/10$ .

**4.3.25 a** Use the *mn* Rule. The Western conference team can be chosen in one of  $m = 6$  ways, and there are 6 ways to choose the Eastern conference team, for a total of  $N = mn = 6(6) = 36$  possible pairings.

**b** You must choose Los Angeles from the first group and New York from the second group, so that  $n_A = (1)(1) = 1$  and the probability is  $n_A/N = 1/36$ .

**c** Since there is only one California team in the Western conference, there are five choices for the first team and six choices for the second team. Hence  $n_A = 5(6) = 30$  and the probability is  $n_A/N = 30/36 = 5/6$ .

**4.3.26 a** There are  $P_4^4 = 4(3)(2)(1) = 24$  possible orders of finish.

**b** There is only one choice for the first position (the winner must be Dave) and, since the remaining runners may occupy any of the remaining three positions, there are  $P_3^3 = 3(2)(1) = 6$  possible arrangements. The probability that Dave wins is  $n_A/N = 6/24 = 1/4$ .

**c** There is only one choice for the first position (Dave), one choice for the second position (John), and  $P_2^2 = 2$  ways to arrange the other two runners. The probability that Dave wins and John finishes second is then  $n_A/N = (1 \times 1 \times 2)/24 = 1/12$ .

**d** There is only choice for the last position (Ed) and, since the remaining runners may occupy any of the remaining three positions there are  $P_3^3 = 3(2)(1) = 6$  possible arrangements. The probability that Ed is last is  $n_A/N = 6/24 = 1/4$ .

**4.3.27** The situation presented here is analogous to drawing 5 items from a jar (the five members voting in favor of the plaintiff). If the jar contains 5 red and 3 white items (5 females and 3 males), what is the probability that all five items are red? That is, if there is no sex bias, five of the eight members are randomly chosen to be those voting for the plaintiff. What is the probability that all five are female? There are

$$N = C_5^8 = \frac{8!}{5!3!} = 56$$

simple events in the experiment, only one of which results in choosing 5 females. Hence,

$$P(\text{five females}) = \frac{1}{56}.$$

**4.3.28** The instructor can select five out of the list of ten questions in  $C_5^{10}$  ways. If the student is to be able to answer all five questions, the instructor must choose five questions out of the six which the student can answer, and none of the four questions which the student cannot answer. The number of ways in which this event can occur is  $C_5^6 C_0^4$ . Hence the probability that the student can answer all five questions is

$$\frac{C_5^6 C_0^4}{C_5^{10}} = \frac{6}{252} = \frac{1}{42}$$

**4.3.29** The monkey can place the twelve blocks in any order. Each arrangement will yield a simple event and hence the total number of simple events (arrangements) is  $P_{12}^{12} = 12!$ . It is necessary to determine the number of simple events in the event of interest (that he draws three of each kind, in order). First, he may draw the four different types of blocks in any order. Thus we need the number of ways of arranging these four items, which is  $P_4^4 = 4!$ . Once this order has been chosen, the three squares can be arranged in  $P_3^3 = 3!$  ways, the three triangles can be arranged in  $P_3^3 = 3!$  ways, and so on. Thus, the total number of simple events in the event of interest is  $P_4^4 (P_3^3)^4$  and the associated probability is

$$\frac{P_4^4 (P_3^3)^4}{P_{12}^{12}} = \frac{4!(3!)^4}{12!}$$

**4.3.30** Would it be unlikely, if three houses were chosen at random, to choose three adjacent houses in a row of 12? The number of choices for the three houses is

$$C_3^{12} = \frac{12(11)(10)}{3(2)(1)} = 220$$

There are 10 choices of three in which the houses are adjacent; that is,  $\{(123), (234), (456), \dots, (10\ 11\ 12)\}$

so that the probability of choosing three adjacent houses is  $\frac{10}{220} = .045$  which is an unusual occurrence.

## Section 4.4

**4.4.1** Each simple event is equally likely, with probability  $1/5$ .

$$\mathbf{a} \quad A^c = \{E_2, E_4, E_5\} \quad P(A^c) = 3/5$$

$$\mathbf{b} \quad A \cap B = \{E_1\} \quad P(A \cap B) = 1/5$$

$$\mathbf{c} \quad B \cap C = \{E_4\} \quad P(B \cap C) = 1/5$$

$$\mathbf{d} \quad A \cup B = S = \{E_1, E_2, E_3, E_4, E_5\} \quad P(A \cup B) = 1$$

$$\mathbf{e} \quad B | C = \{E_4\} \text{ from among } \{E_3, E_4\} \quad P(B | C) = 1/2$$

$$\mathbf{f} \quad A | B = \{E_1\} \text{ from among } \{E_1, E_2, E_4, E_5\} \quad P(A | B) = 1/4$$

$$\mathbf{g} \quad A \cup B \cup C = S \quad P(A \cup B \cup C) = 1$$

$$\mathbf{h} \quad (A \cap B)^c = \{E_2, E_3, E_4, E_5\} \quad P(A \cap B)^c = 4/5$$

$$\mathbf{4.4.2} \quad \mathbf{a} \quad P(A^c) = 1 - P(A) = 1 - \frac{2}{5} = \frac{3}{5} \quad \mathbf{b} \quad P(A \cap B)^c = 1 - P(A \cup B) = 1 - \frac{1}{5} = \frac{4}{5}$$

$$\mathbf{4.4.3} \quad \mathbf{a} \quad P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{1/5}{4/5} = \frac{1}{4} \quad \mathbf{b} \quad P(B | C) = \frac{P(B \cap C)}{P(C)} = \frac{1/5}{2/5} = \frac{1}{2}$$

$$\mathbf{4.4.4} \quad \mathbf{a} \quad P(A \cup B) = P(A) + P(B) - P(A \cap B) = 2/5 + 4/5 - 1/5 = 5/5 = 1$$

$$\mathbf{b} \quad P(A \cap B) = P(A | B)P(B) = (1/4)(4/5) = 1/5$$

$$\mathbf{c} \quad P(B \cap C) = P(B | C)P(C) = (1/2)(2/5) = 1/5$$

**4.4.5** From Exercise 3,  $P(A | B) = 1/4$  while  $P(A) = 2/5$ . Therefore, A and B are not independent.

**4.4.6** From Exercise 4,  $P(A \cap B) = 1/5$ . Since  $P(A \cap B) \neq 0$ , A and B are not mutually exclusive.

$$\mathbf{4.4.7} \quad P(A \cap B) = P(B)P(A | B) = .5(.1) = .05$$

**4.4.8** Since  $P(A) = P(A | B) = .1$ , events A and B are independent.

**4.4.9** Since  $P(A \cap B) = 0$  and  $P(A)P(B) = .1(.5) = .05$ ,  $P(A \cap B) \neq P(A)P(B)$ . Hence, events A and B are not independent.

**4.4.10** Using the Addition Rule, we have

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$.65 = .6 - P(A \cap B) \text{ so that } P(A \cap B) = .05.$$

Since  $P(A \cap B) \neq 0$ , events A and B are not mutually exclusive.

**4.4.11 a** The event A will occur whether event B occurs or not ( $B^c$ ). Hence,

$$P(A) = P(A \cap B) + P(A \cap B^c) = .34 + .15 = .49$$

**b** Similar to part **a**.  $P(B) = P(A \cap B) + P(A^c \cap B) = .34 + .46 = .80$ .

**c** The contents of the cell in the first row and first column is  $P(A \cap B) = .34$ .

**d**  $P(A \cup B) = P(A) + P(B) - P(A \cap B) = .49 + .80 - .34 = .95$ .

**e** Use the definition of conditional probability to find  $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{.34}{.80} = .425$ .

**f** Similar to part **e**.  $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{.34}{.49} = .694$

**4.4.12** From Exercise 11, since  $P(A \cap B) = .34$ , the two events are not mutually exclusive.

**4.4.13** From Exercise 11,  $P(A|B) = .425$  and  $P(A) = .49$ . The two events are not independent.

**4.4.14 a** Since  $P(A)P(B) = .3(.4) = .12$  and  $P(A \cap B) = .12$ , events A and B are independent.

**b** Use the Addition Rule:  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ . Substituting,

$$.7 = .3 + .4 - P(A \cap B) \text{ so that } P(A \cap B) = 0.$$

**c** If A and B are independent, then  $P(A|B) = P(A) = .3$ .

**d** If A and B are mutually exclusive, then  $P(A \cap B) = 0$ , so that

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0}{P(B)} = 0$$

**4.4.15 a** Since A and B are independent,  $P(A \cap B) = P(A)P(B) = .4(.2) = .08$ .

**b**  $P(A \cup B) = P(A) + P(B) - P(A \cap B) = .4 + .2 - (.4)(.2) = .52$

**4.4.16 a** Since A and B are mutually exclusive,  $P(A \cap B) = 0$ .

**b**  $P(A \cup B) = P(A) + P(B) - P(A \cap B) = .3 + .5 - 0 = .8$

**4.4.17 a** Use the definition of conditional probability to find

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{.12}{.4} = .3$$

**b** Since  $P(A \cap B) \neq 0$ , A and B are not mutually exclusive.

**c** If  $P(B) = .3$ , then  $P(B) = P(B|A)$  which means that A and B are independent.

**4.4.18** Refer to the solution to Exercises 1-6 (Section 4.1) where the six simple events in the experiment are given, with  $P(E_i) = 1/6$ .

**a**  $S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$  and  $P(S) = 6/6 = 1$

**b**  $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/3}{1/3} = 1$

**c**  $B = \{E_1, E_2\}$  and  $P(B) = 2/6 = 1/3$

**d**  $A \cap B \cap C$  contains no simple events, and  $P(A \cap B \cap C) = 0$

**e**  $P(A \cap B) = P(A|B)P(B) = 1(1/3) = 1/3$

**f**  $A \cap C$  contains no simple events, and  $P(A \cap C) = 0$

**g**  $B \cap C$  contains no simple events, and  $P(B \cap C) = 0$

**h**  $A \cup C = S$  and  $P(A \cup C) = 1$

**i**  $B \cup C = \{E_1, E_2, E_4, E_5, E_6\}$  and  $P(B \cup C) = 5/6$

**4.4.19** From Exercise 18,  $P(A \cap B) = 1/3$ ,  $P(A|B) = 1$ ,  $P(A) = 1/2$ ,  $P(A|B) \neq P(A)$ , so that A and B are not independent.  $P(A \cap B) \neq 0$ , so that A and B are not mutually exclusive.

**4.4.20** From Exercise 18,  $P(A|C) = P(A \cap C)/P(C) = 0$ ,  $P(A) = 1/2$ ,  $P(A \cap C) = 0$ . Since  $P(A|C) \neq P(A)$ , A and C are dependent. Since  $P(A \cap C) = 0$ , A and C are mutually exclusive.

**4.4.21** In an experiment that consists of tossing two dice, the sample space consists of 36 ordered pairs, the first stage being the number observed on the first die, and the second being the number observed on the second die. The simple events were listed in the solution to Exercise 16 (Section 4.1).

**a** There are 6 pairs that have a sum of 7: (1, 6), (6, 1), (2, 5), (5, 2), (3, 4), and (4, 3). Then

$P(\text{sum is 7}) = \frac{6}{36} = \frac{1}{6}$ . There are 2 pairs that have a sum of 11: (5, 6) and (6, 5), so that

$$P(\text{sum is 11}) = \frac{2}{36} = \frac{1}{18}.$$

**b** There are 6 pairs that have the same upper face: (1, 1), (2, 2), (3, 3), (4, 4), (5, 5), and (6, 6), so that

$$P(\text{doubles}) = \frac{6}{36} = \frac{1}{6}.$$

**c** There are 9 pairs: (1, 1), (1, 3), (1, 5), (3, 1), (3, 3), (3, 5), (5, 1), (5, 3) and (5, 5), so that

$$P(\text{both odd}) = \frac{9}{36} = \frac{1}{4}.$$

**4.4.22** Define the following events:

P: test is positive for drugs

N: test is negative for drugs

D: employee is a drug user

It is given that, on a given test,  $P(P|D) = .98$ ,  $P(N|D^c) = .98$ .

**a** For a given test,  $P(P|D^c) = 1 - .98 = .02$ . Since the tests are independent

$$P(\text{fail both test} | D^c) = (.02)(.02) = .0004$$

**b**  $P(\text{detection} | D) = P(N \cap P | D) + P(P \cap N | D) + P(P \cap P | D)$   
 $= (.98)(.02) + (.02)(.98) + (.98)(.98) = .9996$

**c**  $P(\text{pass both} | D) = P(N \cap N | D) = (.02)(.02) = .0004$

**4.4.23** Define the following events:

A: project is approved for funding

D: project is disapproved for funding

For the first panel,  $P(A_1) = .2$  and  $P(D_1) = .8$ . For the second panel,  $P[\text{same decision as first panel}] = .7$  and  $P[\text{reversal}] = .3$ . That is,

$$P(A_2 | A_1) = P(D_2 | D_1) = .7 \text{ and } P(A_2 | D_1) = P(D_2 | A_1) = .3.$$

**a**  $P(A_1 \cap A_2) = P(A_1)P(A_2 | A_1) = .2(.7) = .14$

**b**  $P(D_1 \cap D_2) = P(D_1)P(D_2 | D_1) = .8(.7) = .56$

**c**  $P(D_1 \cap A_2) + P(A_1 \cap D_2) = P(D_1)P(A_2 | D_1) + P(A_1)P(D_2 | A_1) = .8(.3) + .2(.3) = .30$

**4.4.24** The two-way table in the text gives probabilities for events  $A$ ,  $A^C$ ,  $B$ ,  $B^C$  in the column and row marked “Totals”. The interior of the table contains the four two-way intersections as shown next.

$A \cap B$	$A \cap B^C$
$A^C \cap B$	$A^C \cap B^C$

The necessary probabilities can be found using various rules of probability if not directly from the table.

**a**  $P(A) = .4$

**b**  $P(B) = .37$

**c**  $P(A \cap B) = .10$

**d**  $P(A \cup B) = .4 + .37 - .10 = .67$

**e**  $P(A^C) = 1 - .4 = .6$

**b**  $P(A | B) = P(A \cap B) / P(B) = .1 / .37 = .27$

**g**  $P(B | A) = P(A \cap B) / P(A) = .1 / .4 = .25$

**4.4.25** Refer to Exercise 24.

**a** From the table,  $P(A \cap B) = .1$  while  $P(A)P(B | A) = (.4)(.25) = .1$

**b** From the table,  $P(A \cap B) = .1$  while  $P(B)P(A | B) = (.37)(.1 / .37) = .1$

**c** From the table,  $P(A \cup B) = .1 + .27 + .30 = .67$  while  $P(A) + P(B) - P(A \cap B) = .4 + .37 - .10 = .67$ .

**4.4.26** Define the following events: B: man takes the bus

S: man takes the subway

L: the man is late

It is given that  $P(B) = .3$ ,  $P(S) = .7$ ,  $P(L | B) = .3$ ,  $P(L | S) = .2$ . We need to find

$P(B | L) = P(B \cap L) / P(L)$ . Using the Multiplicative Rule, we can find

$$P(B \cap L) = P(B)P(L | B) = .3(.3) = .09$$

To find  $P(L)$ , remember that the man may be late if he takes *either* the bus or the subway, so that

$$\begin{aligned} P(L) &= P(\text{late and took the bus}) + P(\text{late and took the subway}) \\ &= P(L \cap B) + P(L \cap S) = P(B)P(L | B) + P(S)P(L | S) \\ &= (.3)(.3) + (.7)(.2) = .23. \end{aligned}$$

Finally,  $P(B | L) = P(B \cap L) / P(L) = .09 / .23 = .3913$ .

**NOTE:** This exercise can also be solved using Bayes' Rule, the subject of Section 4.5.

**4.4.27** Define the following events: A: first system fails

B: second system fails

A and B are independent and  $P(A) = P(B) = .001$ . To determine the probability that the combined missile system does not fail, we use the complement of this event; that is,

$$\begin{aligned} P[\text{system does not fail}] &= 1 - P[\text{system fails}] = 1 - P(A \cap B) \\ &= 1 - P(A)P(B) = 1 - (.001)^2 \\ &= .999999 \end{aligned}$$

**4.4.28** An investor can invest in three of five recommended stocks. Unknown to him, only 2 out of 5 will show substantial profit. Let  $P_1$  and  $P_2$  be the two profitable stocks. A typical simple event might be  $(P_1P_2N_3)$ , which represents the selection of two profitable and one nonprofitable stock. The ten simple events are listed next:

$E_1: (P_1P_2N_1)$      $E_2: (P_1P_2N_2)$      $E_3: (P_1P_2N_3)$      $E_4: (P_2N_1N_2)$

$E_5: (P_2N_1N_3)$      $E_6: (P_2N_2N_3)$      $E_7: (N_1N_2N_3)$      $E_8: (P_1N_1N_2)$

$E_9: (P_1N_1N_3)$      $E_{10}: (P_1N_2N_3)$

Then  $P[\text{investor selects two profitable stocks}] = P(E_1) + P(E_2) + P(E_3) = 3/10$  since the simple events are equally likely, with  $P(E_i) = 1/10$ . Similarly,

$$P[\text{investor selects only one of the profitable stocks}] = P(E_4) + P(E_5) + P(E_6) + P(E_8) + P(E_9) + P(E_{10}) = 6/10$$

**4.4.29** Define the following events:

S: student chooses Starbucks

P: student chooses Peet's

C: student orders a café mocha

Then  $P(S) = .7$ ;  $P(P) = .3$ ;  $P(C|S) = P(C|P) = .60$

**a** Using the given probabilities,  $P(S \cap C) = P(S)P(C|S) = .7(.6) = .42$

**b** Since  $P(C) = .6$  regardless of whether the student visits Starbucks or Peet's, the two events are independent.

**c** 
$$P(P|C) = \frac{P(P \cap C)}{P(C)} = \frac{P(P)P(C|P)}{P(C)} = P(P) = .3$$

**d** 
$$P(S \cup C) = P(S) + P(C) - P(S \cap C) = .7 + .6 - (.7)(.6) = .88$$

**4.4.30** Let events A and B be defined as follows:

A: article gets by the first inspector    B: article gets by the second inspector

The event of interest is then the event  $A \cap B$ , that the article gets by both inspectors. It is given that  $P(A) = .1$ , and also that  $P(B|A) = .5$ . Applying the Multiplication Rule,

$$P(A \cap B) = P(A)P(B|A) = (.1)(.5) = .05$$

**4.4.31** Define the events: D: person dies    S: person smokes



It is given that  $P(S) = .2$ ,  $P(D) = .006$ , and  $P(D|S) = 10P(D|S^c)$ . The probability of interest is  $P(D|S)$ . The event  $D$ , whose probability is given, can be written as the union of two mutually exclusive intersections. That is,  $D = (D \cap S) \cup (D \cap S^c)$ .

Then, using the Addition and Multiplication Rules,

$$P(D) = P(D \cap S) + P(D \cap S^c) = P(D|S)P(S) + P(D|S^c)P(S^c) = P(D|S)(.2) + [(1/10)P(D|S)](.8)$$

Since  $P(D) = .006$ , the above equation can be solved for  $P(D|S)$

$$P(D|S)(.2 + .08) = .006$$

$$P(D|S) = .006/.28 = .0214$$

- 4.4.32** Define      A: smoke is detected by device A  
                          B: smoke is detected by device B

If it is given that  $P(A) = .95$ ,  $P(B) = .98$ , and  $P(A \cap B) = .94$ .

**a**       $P(A \cup B) = P(A) + P(B) - P(A \cap B) = .95 + .98 - .94 = .99$

**b**       $P(A^c \cap B^c) = 1 - P(A \cup B) = 1 - .99 = .01$

- 4.4.33 a** Each of the four cell events is equally likely with  $P(E_i) = 1/4$ . Hence, the probability of at least one dominant (R) allele is  $P(rR) + P(Rr) + P(RR) = \frac{3}{4}$ .

**b** Similar to part **a**. The probability of at least one recessive allele is  $P(rR) + P(Rr) + P(rr) = \frac{3}{4}$ .

- c** Define the events:      A: plant has red flowers                      B: plant has one recessive allele  
 Then

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(rR) + P(Rr)}{P(rR) + P(Rr) + P(RR)} = \frac{2/4}{3/4} = \frac{2}{3}$$

- 4.4.34** The completed table is shown next, and each of the possible pairings are equally likely with probability 1/16.

ss yy	Ss yY	ssYy	ss YY
sS yy	sS yY	sS Yy	sS YY
Ss yy	Ss yY	Ss Yy	Ss YY
SS yy	SS yY	SS Yy	SS YY

- a** Smooth yellow peas result from all pairing having at least one S and at least one Y. Hence,  $P(\text{smooth yellow}) = 9/16$ .

- b** Smooth green peas result when the pairing has at least one S and the pair yy. Hence,  $P(\text{smooth green}) = 3/16$ .

- c** Wrinkled yellow peas result when the pairing has at least one Y and the pair ss. Hence,  $P(\text{wrinkled yellow}) = 3/16$

- d** Wrinkled green peas result only when the pairing is ss yy. Hence,  $P(\text{wrinkled green}) = 1/16$ .

- e** Define:      A: offspring has smooth yellow peas

B: offspring has one s allele

C: offspring has one s allele and one y allele

Then  $P(A) = 9/16$ ;  $P(A \cap B) = 6/16$ ;  $P(A \cap C) = 4/16$ . Using the definition of conditional probability,

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{6/16}{9/16} = \frac{2}{3} \quad \text{and} \quad P(C|A) = \frac{P(A \cap C)}{P(A)} = \frac{4/16}{9/16} = \frac{4}{9}$$

**4.4.35** Similar to Exercise 24.

**a**  $P(A) = \frac{154}{256}$

**b**  $P(G) = \frac{155}{256}$

**c**  $P(A \cap G) = \frac{88}{256}$

**d**  $P(G|A) = \frac{P(G \cap A)}{P(A)} = \frac{88/256}{154/256} = \frac{88}{154}$

**e**  $P(G|B) = \frac{P(G \cap B)}{P(B)} = \frac{44/256}{67/256} = \frac{44}{67}$

**f**  $P(G|C) = \frac{P(G \cap C)}{P(C)} = \frac{23/256}{35/256} = \frac{23}{35}$

**g**  $P(C|P) = \frac{P(C \cap P)}{P(P)} = \frac{12/256}{101/256} = \frac{12}{101}$

**h**  $P(B^c) = 1 - P(B) = 1 - \frac{67}{256} = \frac{189}{256}$

**4.4.36** Similar to Exercise 35.

**a**  $P(F) = .35 + .36 = .71$

**b**  $P(G) = .20 + .09 = .29$

**c**  $P(F|M) = \frac{P(F \cap M)}{P(M)} = \frac{.35}{.55} = .63$

**d**  $P(F|W) = \frac{P(F \cap W)}{P(W)} = \frac{.36}{.45} = .80$

**e**  $P(M|F) = \frac{P(M \cap F)}{P(F)} = \frac{.35}{.71} = .49$

**e**  $P(W|G) = \frac{P(W \cap G)}{P(G)} = \frac{.09}{.29} = .31$

**4.4.37** Define the following events:

A: player makes first free throw

B: player makes second free throw

The probabilities of events A and B will depend on which player is shooting.

**a**  $P(A \cap B) = P(A)P(B) = .81(.81) = .6561$ , since the free throws are independent.

**b** The event that Kevin Durant makes exactly one of the two free throws will occur if he makes the first and misses the second, or vice versa. Then

$$\begin{aligned} P(\text{makes exactly one}) &= P(A \cap B^c) + P(A^c \cap B) \\ &= .62(.38) + .38(.62) = .4712 \end{aligned}$$

**c** This probability is the intersection of the individual probabilities for both DeAndre Jordan and Kevin Durant..

$$P(\text{DeAndre makes both and Kevin makes neither}) = [.81(.81)][.38(.38)] = .0947$$

## Section 4.5

**4.5.1** Use the Law of Total Probability, writing

$$P(A) = P(S_1)P(A|S_1) + P(S_2)P(A|S_2) = .7(.2) + .3(.3) = .23$$

**4.5.2-3** Use the results of Exercise 1 in the form of Bayes' Rule:

$$P(S_i | A) = \frac{P(S_i)P(A | S_i)}{P(S_1)P(A | S_1) + P(S_2)P(A | S_2)}$$

$$2. \quad \text{For } i = 1, \quad P(S_1 | A) = \frac{.7(.2)}{.7(.2) + .3(.3)} = \frac{.14}{.23} = .6087$$

$$3. \quad \text{For } i = 2, \quad P(S_2 | A) = \frac{.3(.3)}{.7(.2) + .3(.3)} = \frac{.09}{.23} = .3913$$

**4.5.4-6** Use Bayes' Rule:

$$P(S_i | A) = \frac{P(S_i)P(A | S_i)}{P(S_1)P(A | S_1) + P(S_2)P(A | S_2) + P(S_3)P(A | S_3)}$$

$$4. \quad \text{For } i = 1, \quad P(S_1 | A) = \frac{.2(.2)}{.2(.2) + .5(.1) + .3(.3)} = \frac{.04}{.18} = .2222$$

$$5. \quad \text{For } i = 2, \quad P(S_2 | A) = \frac{.5(.1)}{.2(.2) + .5(.1) + .3(.3)} = \frac{.05}{.18} = .2778$$

$$6. \quad \text{For } i = 3, \quad P(S_3 | A) = \frac{.3(.3)}{.2(.2) + .5(.1) + .3(.3)} = \frac{.09}{.18} = .5000$$

**4.5.7** Use the Law of Total Probability, writing  $P(A) = P(S_1)P(A | S_1) + P(S_2)P(A | S_2) = .6(.3) + .4(.5) = .38$ .

**4.5.8** Define the following events:  $V$ : crime is violent  $R$ : crime is reported

It is given that  $P(V) = .2$ ,  $P(V^c) = .8$ ,  $P(R | V) = .9$ ,  $P(R | V^c) = .7$

**a** The overall reporting rate for crimes is

$$P(R) = P(V)P(R | V) + P(V^c)P(R | V^c) = .2(.9) + .8(.7) = .74$$

$$\text{b} \quad \text{Use Bayes' Rule:} \quad P(V | R) = \frac{P(V)P(R | V)}{P(R)} = \frac{.2(.9)}{.74} = .24$$

$$\text{and} \quad P(V^c | R) = \frac{P(V^c)P(R | V^c)}{P(R)} = \frac{.8(.7)}{.74} = .76$$

**c** Notice that the proportion of non-violent crimes (.8) is much larger than the proportion of violent crimes (.2). Therefore, when a crime is reported, it is more likely to be a non-violent crime.

**4.5.9** Define  $A$ : machine produces a defective item  $B$ : worker follows instructions

Then  $P(A | B) = .01$ ,  $P(B) = .90$ ,  $P(A | B^c) = .03$ ,  $P(B^c) = .10$ . The probability of interest is

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap B^c) \\ &= P(A | B)P(B) + P(A | B^c)P(B^c) \\ &= .01(.90) + .03(.10) = .012 \end{aligned}$$

**4.5.10** Define the following events:

$A$ : passenger uses airport A

$B$ : passenger uses airport B

$C$ : passenger uses airport C

D: a weapon is detected

Suppose that a passenger is carrying a weapon. It is given that

$$P(D|A) = .9 \quad P(A) = .5$$

$$P(D|B) = .8 \quad P(B) = .3$$

$$P(D|C) = .85 \quad P(C) = .2$$

The probability of interest is

$$P(A|D) = \frac{P(A)P(D|A)}{P(A)P(D|A) + P(B)P(D|B) + P(C)P(D|C)} = \frac{.5(.9)}{.5(.9) + .3(.8) + .2(.85)} = .5233$$

Similarly,

$$P(C|D) = \frac{.2(.85)}{.5(.9) + .3(.8) + .2(.85)} = \frac{.17}{.86} = .1977$$

- 4.5.11** Define      L: play goes to the left  
                      R: play goes to the right  
                      S: right guard shifts his stance

**a** It is given that  $P(L) = .3$ ,  $P(R) = .7$ ,  $P(S|R) = .8$ ,  $P(S^C|L) = .9$ ,  $P(S|L) = .1$ , and  $P(S^C|R) = .2$ . Using Bayes' Rule,

$$P(L|S^C) = \frac{P(L)P(S^C|L)}{P(L)P(S^C|L) + P(R)P(S^C|R)} = \frac{.3(.9)}{.3(.9) + .7(.2)} = \frac{.27}{.41} = .6585$$

**b** From part a,  $P(R|S^C) = 1 - P(L|S^C) = 1 - .6585 = .3415$ .

**c** Given that the guard takes a balanced stance, it is more likely (.6585 versus .3415) that the play will go to the left.

- 4.5.12** Define the events:      A: athlete has been disqualified previously  
                                      B: athlete is disqualified for the next 6 weeks

It is given that  $P(B|A^C) = .15$ ,  $P(B|A) = .50$ ,  $P(A) = .30$ . The event of interest is event B, which can be written as the union of two mutually exclusive events:

$$B = (A \cap B) \cup (A^C \cap B)$$

Then, using the Law of Total Probability,

$$P(B) = P(A \cap B) + P(A^C \cap B) = P(B|A)P(A) + P(B|A^C)P(A^C) = .5(.3) + .15(.7) = .255$$

- 4.5.13** The probability of interest is  $P(A|H)$  which can be calculated using Bayes' Rule and the probabilities given in the exercise.

$$\begin{aligned} P(A|H) &= \frac{P(A)P(H|A)}{P(A)P(H|A) + P(B)P(H|B) + P(C)P(H|C)} \\ &= \frac{.01(.90)}{.01(.90) + .005(.95) + .02(.75)} = \frac{.009}{.02875} = .3130 \end{aligned}$$

- 4.5.14** Define the following events, under the assumptions that an incorrect return has been filed.

G<sub>1</sub>: individual guilty of cheating

$G_2$ : individual not guilty (filed incorrectly due to lack of knowledge)

$D$ : individual denies knowledge of error

It is given that  $P(G_1) = .05$ ,  $P(G_2) = .02$ ,  $P(D|G_1) = .80$ . Note that  $P(D|G_2) = 1$  since if the individual has incorrectly filed due to lack of knowledge, he will, with probability 1 deny knowledge of the error. Using Bayes' Rule,

$$P(G_1|D) = \frac{P(G_1)P(D|G_1)}{P(G_1)P(D|G_1) + P(G_2)P(D|G_2)} = \frac{.05(.80)}{.05(.80) + .02(1)} = .6667$$

**4.5.15 a** Using the probability table,

$$P(D) = .08 + .02 = .10$$

$$P(D^c) = 1 - P(D) = 1 - .10 = .90$$

$$P(N|D^c) = \frac{P(N \cap D^c)}{P(D^c)} = \frac{.85}{.90} = .94$$

$$P(N|D) = \frac{P(N \cap D)}{P(D)} = \frac{.02}{.10} = .20$$

**b** Using Bayes' Rule,  $P(D|N) = \frac{P(D)P(N|D)}{P(D)P(N|D) + P(D^c)P(N|D^c)} = \frac{.10(.20)}{.10(.20) + .90(.94)} = .023$

**c** Using the definition of conditional probability,  $P(D|N) = \frac{P(N \cap D)}{P(N)} = \frac{.02}{.87} = .023$

**d**  $P(\text{false positive}) = P(P|D^c) = \frac{P(P \cap D^c)}{P(D^c)} = \frac{.05}{.90} = .056$

**e**  $P(\text{false negative}) = P(N|D) = \frac{P(N \cap D)}{P(D)} = \frac{.02}{.10} = .20$

**f** The probability of a false negative is quite high, and would cause concern about the reliability of the screening method.

## Reviewing What You've Learned

**4.R.1** Define the following events:

A: employee fails to report fraud

B: employee suffers reprisal

It is given that  $P(B|A^c) = .23$  and  $P(A) = .69$ . The probability of interest is

$$P(A^c \cap B) = P(B|A^c)P(A^c) = .23(.31) = .0713$$

**4.R.2** Four customers enter a store which sells two styles of digital video recorders (DVRs). Each customer seeks to purchase one of the two styles. Let the number "1" represent a customer seeking to purchase style 1 and the number "2" represent a customer seeking to purchase style 2. The sample space consists of the following 16 four-tuplets.

1111 2111 1221 2212

1112 1122 2121 2122

1121 1212 2211 1222

1211 2112 2221 2222

The event A, “all four customers seek to purchase the same style of DVR”, consists of the two simple events 1111 and 2222. Since the two styles are assumed to be in equal demand, we assume that each of the 16 simple events is equally likely, and we assign the probability  $1/16$  to each. Hence,  $P(A) = 2/16 = 1/8$ .

- 4.R.3** Two systems are selected from seven, three of which are defective. Denote the seven systems as  $G_1, G_2, G_3, G_4, D_1, D_2, D_3$  according to whether they are good or defective. Each simple event will represent a particular pair of systems chosen for testing, and the sample space, consisting of 21 pairs, is shown next.

$G_1G_2 \quad G_1D_1 \quad G_2D_3 \quad G_4D_2 \quad G_1G_3 \quad G_1G_2 \quad G_3D_1 \quad G_4D_3$   
 $G_1G_4 \quad G_1D_3 \quad G_3D_2 \quad D_1D_2 \quad G_2G_3 \quad G_2D_1 \quad G_3D_3 \quad D_1D_3$   
 $G_2G_4 \quad G_2D_2 \quad G_4D_1 \quad D_2D_3 \quad G_3G_4$

Note that the two systems are drawn simultaneously and that order is unimportant in identifying a simple event. Hence, the pairs  $G_1G_2$  and  $G_2G_1$  are not considered to represent two different simple events. The event A, “no defectives are selected”, consists of the simple events  $G_1G_2, G_1G_3, G_1G_4, G_2G_3, G_2G_4, G_3G_4$ . Since the systems are selected at random, any pair has an equal probability of being selected. Hence, the probability assigned to each simple event is  $1/21$  and  $P(A) = 6/21 = 2/7$ .

- 4.R.4 a** Define the following events:

$B_1$ : client buys on first contact       $B_2$ : client buys on second contact

Since the client may buy on either the first or the second contact, the desired probability is

$$\begin{aligned} P[\text{client will buy}] &= P[\text{client buys on first contact}] + P[\text{client doesn't buy on first, but buys on second}] \\ &= P(B_1) + (1 - P(B_1))P(B_2) = .4 + (1 - .4)(.55) \\ &= .73 \end{aligned}$$

- b** The probability that the client will not buy is one minus the probability that the client will buy, or  $1 - .73 = .27$ .

- 4.R.5 a** If the fourth van tested is the last van with brake problems, then in the first three tests, we must find one van with brake problems and two without. That is, in choosing three from the six vans, we must find one faulty and two that are not faulty. Think of choosing three balls – one white and two red – from a total of six, and the probability can be calculated as

$$P(\text{one faulty and two not}) = \frac{C_1^2 C_2^4}{C_3^6} = \frac{2(6)}{20} = \frac{3}{5}$$

Once this is accomplished, the van with brake problems must be chosen on the fourth test. Using the *mn*

Rule, the probability that the fourth van tested is the last with faulty brakes is  $\frac{3}{5} \left( \frac{1}{3} \right) = \frac{1}{5}$ .

- b** In order that no more than four vans must be tested, you must find one or both of the faulty vans in the first four tests. Proceed as in part **a**, this time choosing four from the six vans, and

$$P(\text{one or two faulty vans}) = \frac{C_0^2 C_4^4 + C_1^2 C_3^4}{C_4^6} = \frac{1+8}{15} = \frac{9}{15}$$

- c** If it is known that the first faulty van is found in the first two tests, there are four vans left from which to select those tested third and fourth. Of these four, only one is faulty. Hence,

$$P(\text{one faulty and one not} \mid \text{one faulty in first two tests}) = \frac{C_1^1 C_3^3 + C_1^1 C_3^4}{C_2^4} = \frac{3}{6} = \frac{1}{2}$$

- 4.R.6** Each ball can be chosen from the set (4, 6) and there are three such balls. Hence, there are a total of  $2(2)(2) = 8$  potential winning numbers.

- 4.R.7 a** Since the Connecticut and Pennsylvania lotteries are independent, and the Connecticut lottery was fair (each ball can be chosen from a set of ten, 0, 1, ..., 9).

$$P(666 \text{ in Connecticut} \mid 666 \text{ in Pennsylvania}) = P(666 \text{ in Connecticut}) = \frac{1}{(10)^3} = \frac{1}{1000}.$$

- b** Since  $P(666 \text{ in Pennsylvania}) = 1/8$ , the probability of a 666 in both lotteries is  $\left(\frac{1}{8}\right)\left(\frac{1}{1000}\right) = \frac{1}{8000}$ .

- 4.R.8 a** Define a birth date by using a three-digit number between 001 and 365. A simple event will be a pair of numbers, each between 001 and 365. A typical simple event might be (264, 023), for two people entering the room.

- b** The probability that a person has a birthday on any specific date is  $1/365$ . Hence, for two people the probability of a specific pair of dates is  $(1/365)^2$ .

- c** The simple events in A are  $\{(001, 001), (002, 002), \dots, (365, 365)\}$ . There are 365 such events.

- d**  $P(A) = \frac{365}{(365)^2} = \frac{1}{365}$ , assuming all simple events are equally likely.

- e**  $P(A^c) = 1 - P(A) = \frac{364}{365}$

- 4.R.9** Fix the birth date of the first person entering the room. Then define the following events:

$A_2$ : second person's birthday differs from the first

$A_3$ : third person's birthday differs from the first and second

$A_4$ : fourth person's birthday differs from all preceding

$\vdots$

$A_n$ :  $n^{\text{th}}$  person's birthday differs from all preceding

Then

$$P(A) = P(A_2)P(A_3)\cdots P(A_n) = \left(\frac{364}{365}\right)\left(\frac{363}{365}\right)\cdots\left(\frac{365-n+1}{365}\right)$$

since at each step, one less birth date is available for selection. Since event B is the complement of event A,

$$P(B) = 1 - P(A)$$

- a** For  $n=3$ ,  $P(A) = \frac{(364)(363)}{(365)^2} = .9918$  and  $P(B) = 1 - .9918 = .0082$

- b** For  $n=4$ ,  $P(A) = \frac{(364)(363)(362)}{(365)^3} = .9836$  and  $P(B) = 1 - .9836 = .0164$

- 4.R.10 a** Use the Law of Total Probability, with  $P(B) = .47$ ,  $P(F) = .53$ ,  $P(MCL \mid B) = .39$ ,  $P(ACL \mid B) = .61$ ,  $P(MCL \mid F) = .33$ , and  $P(ACL \mid F) = .67$ . Then

$$\begin{aligned} P(MCL) &= P(MCL \cap B) + P(MCL \cap F) = P(B)P(MCL \mid B) + P(F)P(MCL \mid F) \\ &= .47(.39) + .53(.33) = .3582 \end{aligned}$$

- b** Use Bayes' Rule,  $P(F \mid MCL) = \frac{P(MCL \cap F)}{P(MCL)} = \frac{P(F)P(MCL \mid F)}{P(MCL)} = \frac{.53(.33)}{.3582} = .4883$ .

$$\text{c} \quad \text{Use Bayes' Rule, } P(B|ACL) = \frac{P(B)P(ACL|B)}{P(B)P(ACL|B) + P(F)P(ACL|F)} = \frac{.47(.61)}{.47(.61) + .53(.67)} = .4467$$

$$\text{4.R.11 a} \quad P(T \cap P) = 27/70 = .386$$

$$\text{b} \quad P(T^c \cap P) = 0$$

$$\text{c} \quad P(T \cap N) = 4/70$$

$$\text{d} \quad P(P|T) = \frac{P(T \cap N)}{P(T)} = \frac{27/70}{31/70} = \frac{27}{31}$$

$$\text{e} \quad P(\text{false negative}) = P(N|T) = \frac{P(T \cap N)}{P(T)} = \frac{4/70}{31/70} = \frac{4}{31}$$

**4.R.12 a** Consider a single trial which consists of tossing two coins. A match occurs when either HH or TT is observed. Hence, the probability of a match on a single trial is  $P(HH) + P(TT) = 1/4 + 1/4 = 1/2$ . Let MMM denote the event “match on trials 1, 2, and 3”.

$$\text{Then} \quad P(MMM) = P(M)P(M)P(M) = (1/2)^3 = 1/8.$$

**b** On a single trial the event A, “two tails are observed” has probability  $P(A) = P(TT) = 1/4$ . Hence, in three trials  $P(AAA) = P(A)P(A)P(A) = (1/4)^3 = 1/64$

**c** This low probability would not suggest collusion, since the probability of three matches is low only if we assume that each student is merely guessing at each answer. If the students have studied together or if they both know the correct answer, the probability of a match on a single trial is no longer 1/2, but is substantially higher. Hence, the occurrence of three matches is not unusual.

**4.R.13** Define  $A$ : union strike fund is adequate to support a strike  
 $C$ : union-management team makes a contract settlement within 2 weeks

It is given that  $P(C) = .5$ ,  $P(A) = .6$ ,  $P(A \cap C) = .3$ . Then

$$P(C|A) = \frac{P(A \cap C)}{P(A)} = \frac{.3}{.6} = .5.$$

Since  $P(C) = .5$  and  $P(C|A) = .5$ , it appears that the settlement of the contract is independent of the ability of the union strike fund to support the strike.

**4.R.14** Define  $R$ : the employee remains 10 years or more

**a** The probability that the man will stay less than 10 years is  $P(R^c) = 1 - P(R) = 1 - 1/6 = 5/6$

**b** The probability that the man and the woman, acting independently, will both work less than 10 years is  $P(R^c R^c) = P(R^c)P(R^c) = (5/6)^2 = 25/36$

**c** The probability that either or both people work 10 years or more is

$$1 - P(R^c R^c) = 1 - (5/6)^2 = 1 - 25/36 = 11/36$$

**4.R.15** Define the events:  $A$ : the man waits five minutes or longer  
 $B$ : the woman waits five minutes or longer

The two events are independent, and  $P(A) = P(B) = .2$ .



**a**  $P(A^c) = 1 - P(A) = .8$

**b**  $P(A^c B^c) = P(A^c)P(B^c) = (.8)(.8) = .64$

**c**  $P[\text{at least one waits five minutes or longer}]$   
 $= 1 - P[\text{neither waits five minutes or longer}] = 1 - P(A^c B^c) = 1 - .64 = .36$

**4.R.16** This exercise provides an example of a lot acceptance sampling plan. Seven items are drawn from a large lot of bearings and we wish to calculate the probability of accepting the lot; that is, the probability of observing no defectives in the sample of seven. In order to obtain  $P[\text{acceptance}]$ , it is necessary to assume that the lot is large enough so that the probability of getting a defective is not noticeably affected by repeated draws. For example, consider a lot which contains 10,000 items, 5000 of which are defective. The probability of getting a defective on the first draw is  $5000/10,000$  or  $1/2$ . Assume that a defective has been drawn. Then the probability of getting a defective on the second draw is  $4999/9999$ , which is not noticeably different from  $1/2$ . Define the following events:

D: draw a defective

G: draw a nondefective, where  $G = D^c$  and  $P(G) = 1 - P(D)$

A: the lot is accepted

In each case, the desired probability is  $P(A) = P(GGGGGGG) = [P(G)]^7$ . If all the items in the lot are nondefective, then  $P(D) = 0$ ,  $P(G) = 1$  and the probability of acceptance is  $P(A) = 1$ . If  $1/10$  are defective, then  $P(D) = .1$ ,  $P(G) = .9$ , and  $P(A) = (.9)^7 = .4783$ . If  $1/2$  are defective, then  $P(D) = .5$ ,  $P(G) = .5$ , and  $P(A) = (.5)^7 = .00781$ .

**4.R.17** It is given that 40% of all people in a community favor the development of a mass transit system. Thus, given a person selected at random, the probability that the person will favor the system is .4. Since the pollings are independent events, when four people are selected at random,

$$P[\text{all 4 favor the system}] = (.4)^4 = .0256$$

Similarly,  $P[\text{none favor the system}] = (1 - .4)^4 = .1296$

**4.R.18** In this exercise, we assume that the presence or absence of disease in any pair of 4 pairs of identical twins represent independent events. If there is no difference in the effect of the drugs, the probability that one drug causes a greater drop in blood pressure than the other is .5. Then

$$P[\text{reading for drug A exceeds reading for drug B for all 4 pairs}] = (.5)^4 = .0625$$

If we do observe the above event, we can reach one of two conclusions:

- 1 The two drugs are equally effective and we have observed a rare event
- 2 Drug A is more effective than drug B

Since the probability of the above event is very small, we would draw the second conclusion.

**4.R.19** Since the first pooled test is positive, we are interested in the probability of requiring five single tests to detect the disease in the single affected person. There are  $(5)(4)(3)(2)(1)$  ways of ordering the five tests, and there are  $4(3)(2)(1)$  ways of ordering the tests so that the diseased person is given the final test. Hence, the desired probability is  $\frac{4!}{5!} = \frac{1}{5} = .2$ .

If two people are diseased, six tests are needed if the last two tests are given to the diseased people. There are  $3(2)(1)$  ways of ordering the tests of the other three people and  $2(1)$  ways of ordering the tests of the two diseased people. Hence, the probability that six tests will be needed is  $\frac{2!3!}{5!} = \frac{1}{10} = .1$ .

- 4.R.20** The objective is to determine how many times a coin must be tossed in order that the following inequality will be true:

$$P[\text{observe at least one head}] \geq .9$$

Using the complement of this event, we have:

$$1 - P[\text{observe no heads}] \geq .9 \quad \text{or} \quad P[\text{observe no heads}] \leq .1$$

Since the probability of observing a tail on a given toss is .5,  $P[\text{observe no heads in } n \text{ tosses}] = (.5)^n$ .

Evaluating this probability for increasing values of  $n$ , we obtain the following table.

$n$	$P[\text{observe no heads}]$
1	.5
2	.25
3	.125
4	.0625

Note that the inequality will be satisfied if and only if  $n$  is greater than or equal to 4. Thus, the coin must be tossed four times.

- 4.R.21** The necessary probabilities can be found by summing the necessary cells in the probability table and dividing by 200, the total number of individuals.

**a**  $P(Y) = \frac{96}{200} = .48$

**b**  $P(M \cap AE) = \frac{20}{200} = .10$

**c**  $P(M | S) = \frac{P(M \cap S)}{P(S)} = \frac{16 / 200}{61 / 200} = \frac{16}{61} = .262$

## Case Study: Probability and Decision Making in the Congo

Let  $p$  be the probability that a single person successfully completes the jump, and assume that the event that any one person successfully completes the jump is independent of the outcome of the other 11 individual jumps. Define the events:

J: all members of the team successfully complete the jump

A<sub>1</sub>: Ross successfully completes the jump

A<sub>2</sub>: Elliot successfully completes the jump

A<sub>3</sub>: Munro successfully completes the jump

⋮

A<sub>12</sub>: Porter 8 successfully completes the jump

Then  $J$  is the intersection of the events  $A_1, A_2, \dots, A_{12}$  and

$$P(J) = P(A_1 \cap A_2 \cap A_3 \dots \cap A_{12}) = P(A_1)P(A_2) \cdots P(A_{12}) = p(p)(p) \cdots p = p^{12}$$

Substituting  $P(J) = .7980$  into this expression and solving for  $p$ , we find  $p = .9814$ . Thus Houston is basing its probability of a successful team jump on a probability equal to .9814 that a single individual will successfully land in the soft volcanic scree.

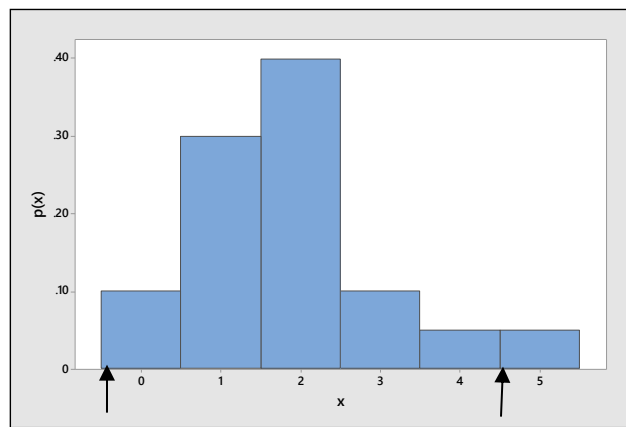
## 5: Discrete Probability Distributions

### Section 5.1

- 5.1.1**  $0 \leq p(x) \leq 1$  and  $\sum p(x) = 1$
- 5.1.2** The number of points scored is a discrete random variable taking the countably infinite number of values, 0, 1, 2, ...
- 5.1.3** Shelf life is a continuous random variable since it can take on any positive real value.
- 5.1.4** Height is a continuous random variable, taking on any positive real value.
- 5.1.5** Length is a continuous random variable, taking on any positive real value.
- 5.1.6** Number of near collisions is a discrete random variable, taking the values 0, 1, 2, ...
- 5.1.7** The increase in length of life achieved by a cancer patient as a result of surgery is a continuous random variable, since an increase in life (measured in units of time) can take on any of an infinite number of values in a particular interval.
- 5.1.8** The tensile strength, in pounds per square inch, of one-inch diameter steel wire cable is a continuous random variable.
- 5.1.9** The number of deer killed per year in a state wildlife preserve is a discrete random variable taking the values 0, 1, 2, ...
- 5.1.10** The number of overdue accounts in a department store at a particular point in time is a discrete random variable, taking the values 0, 1, 2, ...
- 5.1.11** Blood pressure is a continuous random variable.
- 5.1.12** Since one of the requirements of a probability distribution is that  $\sum_x p(x) = 1$ , we need

$$p(4) = 1 - (.1 + .3 + .4 + .1 + .05) = 1 - .95 = .05$$

- 5.1.13** The probability histogram is shown in the figure that follows.



- 5.1.14** For the random variable  $x$  given here,

$$\mu = E(x) = \sum xp(x) = 0(.1) + 1(.3) + \dots + 5(.05) = 1.85$$

The variance of  $x$  is defined as

$$\sigma^2 = E[(x - \mu)^2] = \sum (x - \mu)^2 p(x) = (0 - 1.85)^2(.1) + (1 - 1.85)^2(.3) + \dots + (5 - 1.85)^2(.05) = 1.4275$$

and  $\sigma = \sqrt{1.4275} = 1.19$ .

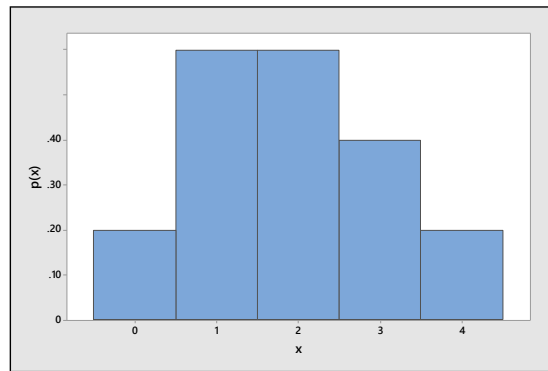
**5.1.15** Refer to Exercise 14. The interval of interest is  $\mu \pm 2\sigma = 1.85 \pm 2.38$  or  $-.53$  to  $4.23$ . This interval is shown on the probability histogram in Exercise 13. Then  $P[-.53 \leq x \leq 4.23] = P[0 \leq x \leq 4] = .95$ .

**5.1.16** Since the probability that  $x$  falls in the interval  $\mu \pm 2\sigma$  is .95 from Exercise 15, we would expect most of the observations to fall in this interval.

**5.1.17** Since one of the requirements of a probability distribution is that  $\sum_x p(x) = 1$ , we need

$$p(3) = 1 - (.1 + .3 + .3 + .1) = 1 - .8 = .2$$

**5.1.18** The probability histogram is shown next.



**5.1.19** For the random variable  $x$  given here,  $\mu = E(x) = \sum xp(x) = 0(.1) + 1(.3) + \dots + 4(.1) = 1.9$

The variance of  $x$  is defined as

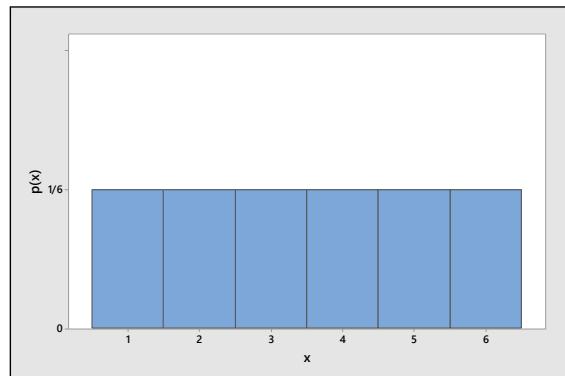
$$\sigma^2 = E[(x - \mu)^2] = \sum (x - \mu)^2 p(x) = (0 - 1.9)^2 (.1) + (1 - 1.9)^2 (.3) + \dots + (4 - 1.9)^2 (.1) = 1.29$$

and  $\sigma = \sqrt{1.29} = 1.136$ .

**5.1.20** Using the table form of the probability distribution,  $P(x > 2) = .2 + .1 = .3$ .

**5.1.21** Using the table form of the probability distribution,  $P(x \leq 3) = 1 - P(x = 4) = 1 - .1 = .9$ .

**5.1.22** Since each of the six possible values,  $x = 1, 2, 3, 4, 5, 6$  is equally likely with  $p(x) = 1/6$  for all values of  $x$ , the graph of the probability distribution has a flat shape, called the *discrete uniform probability distribution*, shown in the figure that follows.



The average value of  $x$  is  $\mu = E(x) = \sum xp(x) = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + \cdots + 6\left(\frac{1}{6}\right) = 3.5$ , the variance of  $x$  is

$$\sigma^2 = E\left[(x - \mu)^2\right] = \sum (x - \mu)^2 p(x) = (1 - 3.5)^2 \left(\frac{1}{6}\right) + (2 - 3.5)^2 \left(\frac{1}{6}\right) + \cdots + (6 - 3.5)^2 \left(\frac{1}{6}\right) = 2.9167$$

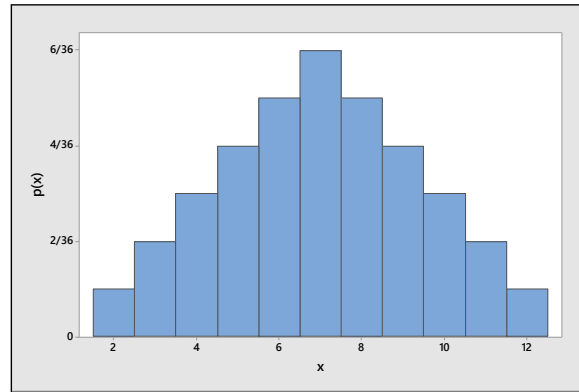
and  $\sigma = \sqrt{2.9167} = 1.71$ .

**5.1.23** There are  $6(6) = 36$  equally likely simple events in this experiment, shown in the table that follows.

	Second Die					
First Die	1	2	3	4	5	6
1	1,1	1,2	1,3	1,4	1,5	1,6
2	2,1	2,2	2,3	2,4	2,5	2,6
3	3,1	3,2	3,3	3,4	3,5	3,6
4	4,1	4,2	4,3	4,4	4,5	4,6
5	5,1	5,2	5,3	5,4	5,5	5,6
6	6,1	6,2	6,3	6,4	6,5	6,6

Each simple event has a particular value of  $x$  associated with it, and by summing the probabilities of all simple events producing a particular value of  $x$ , the probability distribution that follows is obtained. The distribution is mound-shaped.

$x$	2	3	4	5	6	7	8	9	10	11	12
$p(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$



The average value of  $x$  is  $\mu = E(x) = \sum xp(x) = 2\left(\frac{1}{36}\right) + 3\left(\frac{2}{36}\right) + \cdots + 12\left(\frac{1}{36}\right) = 7.0$ , the variance of  $x$  is

$$\sigma^2 = E\left[(x - \mu)^2\right] = \sum (x - \mu)^2 p(x) = (2 - 7)^2 \left(\frac{1}{36}\right) + (3 - 7)^2 \left(\frac{2}{36}\right) + \cdots + (12 - 7)^2 \left(\frac{1}{36}\right) = 5.83333$$

and  $\sigma = \sqrt{5.83333} = 2.415$ .

**5.1.24** Define

T: person admits to texting while driving

N: person does not admit to texting while driving

There are eight simple events in the experiment:

TTT    TTN    TNN    TNT  
NTN    NTT    NNT    NNN

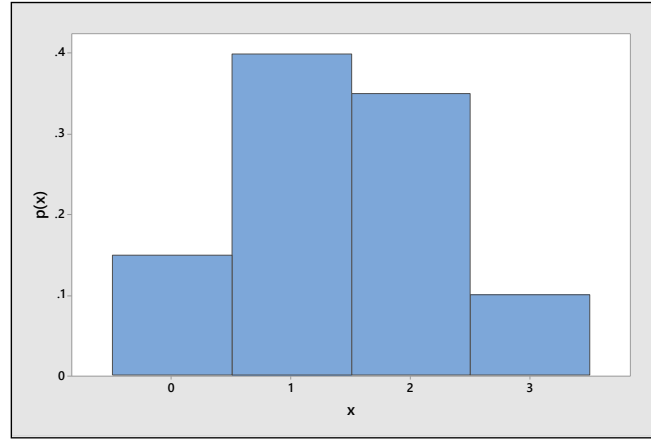
and the probabilities for  $x$  = number who admit to texting while driving = 0, 1, 2, 3 and the probability histogram are shown next.

$$P(x=0) = P(NNN) = (.53)^3 = .148877$$

$$P(x=1) = P(TNN) + P(NTN) + P(NNT) = 3(.47)(.53)^2 = .396069$$

$$P(x=2) = P(NTT) + P(TTN) + P(TNT) = 3(.47)^2(.53) = .351231$$

$$P(x=3) = P(TTT) = (.47)^3 = .103823$$



The average value of  $x$  is  $\mu = E(x) = \sum xp(x) = 0(.148877) + 1(.396069) + 2(.351231) + 3(.103823) = 1.41$

The variance of  $x$  is

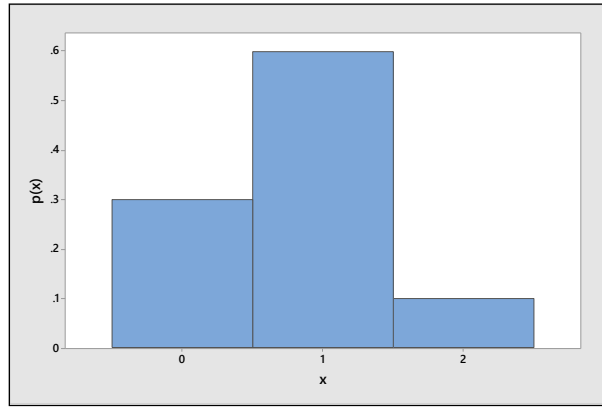
$$\begin{aligned}\sigma^2 &= E[(x - \mu)^2] = \sum (x - \mu)^2 p(x) \\ &= (0 - 1.41)^2 (.148877) + (1 - 1.41)^2 (.396069) + (2 - 1.41)^2 (.351231) + (3 - 1.41)^2 (.103823) \\ &= .7473\end{aligned}$$

and  $\sigma = \sqrt{.7473} = .864$ .

**5.1.25** Let  $W_1$  and  $W_2$  be the two women while  $M_1$ ,  $M_2$  and  $M_3$  are the three men. There are 10 ways to choose the two people to fill the positions. Let  $x$  be the number of women chosen. The 10 equally likely simple events are:

$E_1: W_1W_2 \quad (x=2)$	$E_6: W_2M_2 \quad (x=1)$
$E_2: W_1M_1 \quad (x=1)$	$E_7: W_2M_3 \quad (x=1)$
$E_3: W_1M_2 \quad (x=1)$	$E_8: M_1M_2 \quad (x=0)$
$E_4: W_1M_3 \quad (x=1)$	$E_9: M_1M_3 \quad (x=0)$
$E_5: W_2M_1 \quad (x=1)$	$E_{10}: M_2M_3 \quad (x=0)$

The probability distribution for  $x$  is then  $p(0) = 3/10$ ,  $p(1) = 6/10$ ,  $p(2) = 1/10$  and the probability histogram is shown next.



The average value of  $x$  is  $\mu = E(x) = \sum xp(x) = 0\left(\frac{3}{10}\right) + 1\left(\frac{6}{10}\right) + 2\left(\frac{1}{10}\right) = .8$ , the variance of  $x$  is

$$\begin{aligned}\sigma^2 &= E[(x - \mu)^2] = \sum (x - \mu)^2 p(x) \\ &= (0 - .8)^2 \left(\frac{3}{10}\right) + (1 - .8)^2 \left(\frac{6}{10}\right) + (2 - .8)^2 \left(\frac{1}{10}\right) = .36\end{aligned}\quad \text{and } \sigma = \sqrt{.36} = .6.$$

**5.1.26** Similar to Exercise 25. The random variable  $x$  can take on the values 0, 1, or 2. The associated probabilities can be found by summing probabilities of the simple events for the respective numerical events or by using the laws of probability:

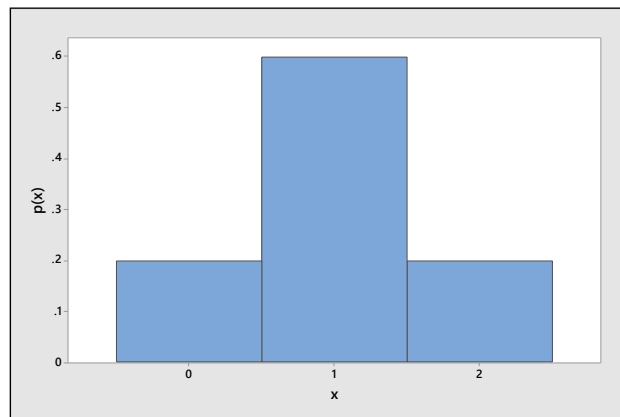
$$\begin{aligned}P[x = 0] &= P[\text{nondefection on first selection}]P[\text{nondefective on second} \mid \text{nondefective on first}] \\ &\quad \times P[\text{nondefective on third} \mid \text{nondefective on first and second}] = \frac{4}{6}\left(\frac{3}{5}\right)\left(\frac{2}{4}\right) = \frac{1}{5}\end{aligned}$$

$$P[x = 1] = P(DNN) + P(NDN) + P(NND) = \frac{2}{6}\left(\frac{4}{5}\right)\left(\frac{3}{4}\right) + \frac{4}{6}\left(\frac{2}{5}\right)\left(\frac{3}{4}\right) + \frac{4}{6}\left(\frac{3}{5}\right)\left(\frac{2}{4}\right) = \frac{3}{5}$$

$$P[x = 2] = P(DDN) + P(DND) + P(NDD) = \frac{2}{6}\left(\frac{1}{5}\right)\left(\frac{4}{4}\right) + \frac{2}{6}\left(\frac{4}{5}\right)\left(\frac{1}{4}\right) + \frac{4}{6}\left(\frac{2}{5}\right)\left(\frac{1}{4}\right) = \frac{1}{5}$$

The probability distribution for  $x$  and the probability histogram follow.

$x$	0	1	2
$p(x)$	1/5	3/5	1/5





The average value of  $x$  is  $\mu = E(x) = \sum xp(x) = 0\left(\frac{1}{5}\right) + 1\left(\frac{3}{5}\right) + 2\left(\frac{1}{5}\right) = 1$ , the variance of  $x$  is

$$\begin{aligned}\sigma^2 &= E\left[(x-\mu)^2\right] = \sum (x-\mu)^2 p(x) \\ &= (0-1)^2\left(\frac{1}{5}\right) + (1-1)^2\left(\frac{3}{5}\right) + (2-1)^2\left(\frac{1}{5}\right) = .4\end{aligned}\quad \text{and } \sigma = \sqrt{.4} = .632.$$

**5.1.27** For the probability distribution given in this exercise,

$$\mu = E(x) = \sum xp(x) = 0(.1) + 1(.4) + 2(.4) + 3(.1) = 1.5.$$

**5.1.28 a-b** On the first try, the probability of selecting the proper key is  $1/4$ . If the key is not found on the first try, the probability changes on the second try. Let F denote a failure to find the key and S denote a success. The random variable is  $x$ , the number of keys tried before the correct key is found. The four associated simple events are shown next.

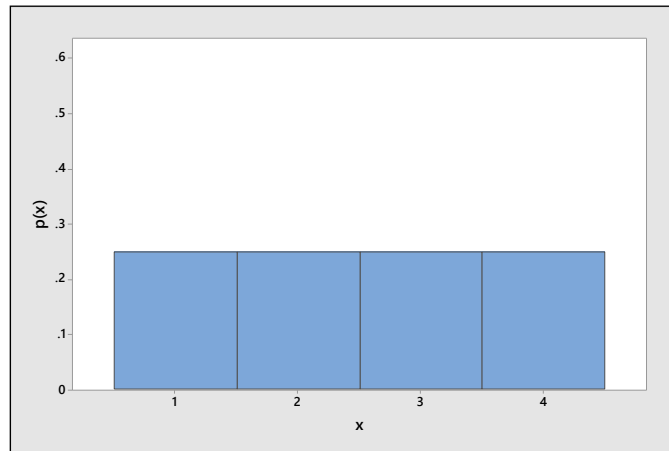
$$\begin{array}{ll} E_1: S & (x=1) \\ E_2: FS & (x=2) \\ E_3: FFS & (x=3) \\ E_4: FFFS & (x=4) \end{array}$$

**c-d** Then

$$\begin{aligned}p(1) &= P(x=1) = P(S) = 1/4 \\ p(2) &= P(x=2) = P(FS) = P(F)P(S) = (3/4)(1/3) = 1/4 \\ p(3) &= P(x=3) = P(FFS) = P(F)P(F)P(S) = (3/4)(2/3)(1/2) = 1/4 \\ p(4) &= P(x=4) = P(FFFS) = P(F)P(F)P(F)P(S) = (3/4)(2/3)(1/2)(1) = 1/4\end{aligned}$$

The probability distribution and probability histogram follow.

$x$	1	2	3	4
$p(x)$	1/4	1/4	1/4	1/4



**5.1.29** Let  $x$  be the number of drillings until the first success (oil is struck). It is given that the probability of striking oil is  $P(O) = .1$ , so that the probability of no oil is  $P(N) = .9$

**a**  $p(1) = P[\text{oil struck on first drilling}] = P(O) = .1$

$p(2) = P[\text{oil struck on second drilling}]$ . This is the probability that oil is not found on the first drilling, but is found on the second drilling. Using the Multiplication Law,

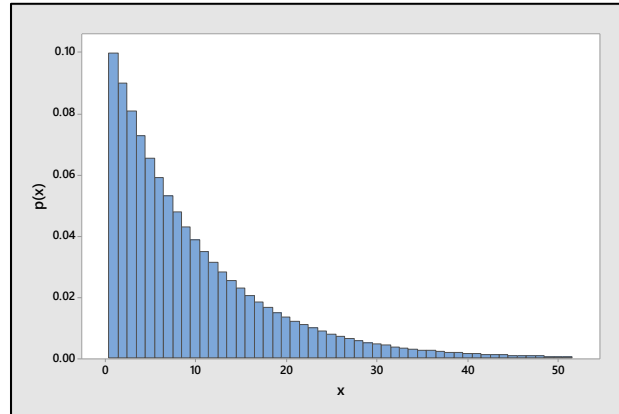
$$p(2) = P(NO) = (.9)(.1) = .09.$$

Finally,  $p(3) = P(NNO) = (.9)(.9)(.1) = .081$ .

**b-c** For the first success to occur on trial  $x$ ,  $(x - 1)$  failures must occur before the first success. Thus,

$$p(x) = P(NNN \dots NNO) = (.9)^{x-1}(.1)$$

since there are  $(x - 1)$   $N$ 's in the sequence. The probability histogram generated by *Minitab* follows.



**5.1.30** The random variable  $x$ , defined as the number of householders insured against fire, can assume the values 0, 1, 2, 3 or 4. The probability that, on any of the four draws, an insured person is found is .6; hence, the probability of finding an uninsured person is .4. Note that each numerical event represents the intersection of the results of four independent draws.

**1**  $P[x = 0] = (.4)(.4)(.4)(.4) = .0256$ , since all four people must be uninsured.

**2**  $P[x = 1] = 4(.6)(.4)(.4)(.4) = .1536$  (Note: the 4 appears in this expression because  $x = 1$  is the union of four mutually exclusive events. These represent the 4 ways to choose the single insured person from the fours.)

**3**  $P[x = 2] = 6(.6)(.6)(.4)(.4) = .3456$ , since the two insured people can be chosen in any of 6 ways.

**4**  $P[x = 3] = 4(.6)^3(.4) = .3456$  and  $P[x = 4] = (.6)^4 = .1296$ .

Then

$$P[\text{at least three insured}] = p(3) + p(4) = .3456 + .1296 = .4752$$

**5.1.31** If a \$5 bet is placed on the number 18, the gambler will either win \$175 ( $35 \times \$5$ ) with probability  $1/38$  or lose \$5 with probability  $37/38$ . Hence, the probability distribution for  $x$ , the gambler's gain is

$x$	$p(x)$
-5	$37/38$
175	$1/38$

The expected gain is  $\mu = E(x) = \sum xp(x) = -5(37/38) + 175(1/38) = \$-0.26$ . The expected gain is in fact negative, a loss of \$0.26.

**5.1.32** In this exercise,  $x$  may take the values 0, 1, 2, or 3, and the probabilities associated with  $x$  are evaluated as in Exercise 30.

$$\begin{aligned}\mathbf{a} \quad P[x=0] &= (.2)^3 = .008 \\ P[x=1] &= 3(.8)(.2)^2 = .096 \\ P[x=2] &= 3(.8)^2(.2) = .384 \\ P[x=3] &= (.8)^3 = .512\end{aligned}$$

The reader may verify that the probabilities sum to one and that  $0 \leq p(x) \leq 1$  for  $x=0, 1, 2$ , and  $3$ . The requirements for a probability distribution have been satisfied.

**b** The alarm will function if  $x=1, 2$ , or  $3$ . Hence,

$$P[\text{alarm functions}] = p(1) + p(2) + p(3) = .096 + .384 + .512 = .992$$

$$\mathbf{c} \quad \mu = E(x) = \sum xp(x) = 0(.008) + 1(.096) + 2(.384) + 3(.512) = 2.4$$

$$\begin{aligned}\sigma^2 &= \sum (x - \mu)^2 p(x) = (0 - 2.4)^2 (.008) + (1 - 2.4)^2 (.096) \\ &\quad + (2 - 2.4)^2 (.384) + (3 - 2.4)^2 (.512) = .48\end{aligned}$$

**5.1.33** The random variable  $G$ , total gain to the insurance company, will be  $D$  if there is no theft, but  $D - 50,000$  if there is a theft during a given year. These two events will occur with probability .99 and .01, respectively. Hence, the probability distribution for  $G$  is given next.

$G$	$p(G)$	The expected gain is
$D$	.99	$E(G) = \sum Gp(G) = .99D + .01(D - 50,000)$
$D - 50,000$	.01	$= D - 50,000$

In order that  $E(G) = 1000$ , it is necessary to have  $1000 = D - 500$  or  $D = \$1500$ .

$$\mathbf{5.1.34} \quad \mathbf{a} \quad \mu = E(x) = \sum xp(x) = 3(.03) + 4(.05) + \cdots + 13(.01) = 7.9$$

$$\begin{aligned}\mathbf{b} \quad \sigma^2 &= \sum (x - \mu)^2 p(x) = (3 - 7.9)^2 (.03) + (4 - 7.9)^2 (.05) + \cdots + (13 - 7.9)^2 (.01) = 4.73 \text{ and} \\ \sigma &= \sqrt{4.73} = 2.1749.\end{aligned}$$

$$\mathbf{c} \quad \text{Calculate } \mu \pm 2\sigma = 7.9 \pm 4.350 \text{ or } 3.55 \text{ to } 12.25. \text{ Then, referring to the probability distribution of } x, \\ P[3.55 \leq x \leq 12.25] = P[4 \leq x \leq 12] = 1 - p(3) - p(13) = 1 - .04 = .96.$$

**5.1.35** **a** Refer to the probability distribution for  $x$  given in this exercise.

$$P(\text{no coffee breaks}) = P(x=0) = .28$$

$$\mathbf{b} \quad P(\text{more than 2 coffee breaks}) = P(x > 2) = .12 + .05 + .01 = .18$$

$$\mathbf{c} \quad \mu = E(x) = \sum xp(x) = 0(.28) + 1(.37) + \cdots + 5(.01) = 1.32$$

$$\begin{aligned}\sigma^2 &= \sum (x - \mu)^2 p(x) = (0 - 1.32)^2 (.28) + (1 - 1.32)^2 (.37) + \cdots + (5 - 1.32)^2 (.01) = 1.4376 \text{ and} \\ \sigma &= \sqrt{1.4376} = 1.199.\end{aligned}$$

$$\mathbf{d} \quad \text{Calculate } \mu \pm 2\sigma = 1.32 \pm 2.398 \text{ or } -1.078 \text{ to } 3.718. \text{ Then, referring to the probability distribution of } x, \\ P[-1.078 \leq x \leq 3.718] = P[0 \leq x \leq 3] = .28 + .37 + .17 + .12 = .94.$$

- 5.1.36** The company will either gain (\$15.50 – 14.80) if the package is delivered on time or will lose \$14.80 if the package is not delivered on time. We assume that, if the package is not delivered within 24 hours, the company does not collect the \$15.50 delivery fee. Then the probability distribution for  $x$ , the company's gain is

$x$	$p(x)$	and
.70	.98	$\mu = E(x) = .70(.98) - 14.80(.02) = .39.$
-14.80	.02	

The expected gain per package is \$0.39.

- 5.1.37** We are asked to find the premium that the insurance company should charge in order to break even. Let  $c$  be the unknown value of the premium and  $x$  be the gain to the insurance company caused by marketing the new product. There are three possible values for  $x$ . If the product is a failure or moderately successful,  $x$  will be negative; if the product is a success, the insurance company will gain the amount of the premium and  $x$  will be positive. The probability distribution for  $x$  follows:

$x$	$p(x)$	
$c$	.94	.
$-800,000 + c$	.01	
$-250,000 + c$	.05	

In order to break even,  $E(x) = \sum xp(x) = 0$ . Therefore,

$$\begin{aligned} .94(c) + .01(-800,000 + c) + (.05)(-250,000 + c) &= 0 \\ -8000 - 12,500 + (.01 + .05 + .94)c &= 0 \\ c &= 20,500 \end{aligned}$$

Hence, the insurance company should charge a premium of \$20,500.

- 5.1.38 a** There are six volunteers, from whom we must choose two people for the committee. The number of choices is then

$$C_2^6 = \frac{6(5)}{2(1)} = 15$$

If the number of women chosen from the two women is  $x$ , and the number of men chosen from the four men must be  $2 - x$ . Then

$$\begin{aligned} P(x=0) &= \frac{C_0^2 C_2^4}{15} = \frac{6}{15} \\ P(x=1) &= \frac{C_1^2 C_1^4}{15} = \frac{8}{15} \\ P(x=2) &= \frac{C_2^2 C_0^4}{15} = \frac{1}{15} \end{aligned}$$

and the probability distribution for  $x$  is shown in the table.

$x$	0	1	2
$p(x)$	6/15	8/15	1/15

- b** The probability that both orchestra representatives will be women is  $P(x=2) = 1/15$ .

$$\text{c} \quad \mu = E(x) = \sum xp(x) = 0\left(\frac{6}{15}\right) + 1\left(\frac{8}{15}\right) + 2\left(\frac{1}{15}\right) = \frac{10}{15} = \frac{2}{3}$$

$$\sigma^2 = \sum (x - \mu)^2 p(x) = \left(0 - \frac{2}{3}\right)^2 \left(\frac{6}{15}\right) + \left(1 - \frac{2}{3}\right)^2 \left(\frac{8}{15}\right) + \left(2 - \frac{2}{3}\right)^2 \left(\frac{1}{15}\right) = \frac{48}{135} = \frac{16}{45}$$

## Section 5.2

**5.2.1** See the definition of a binomial experiment at the beginning of this section (Section 5.2)

**5.2.2-6** These probabilities can be found individually using the binomial formula, or alternatively using the cumulative binomial tables (Table 1 in Appendix I). The values in Table 1 for  $n = 8$  and  $p = .7$  are shown in the table that follows.

$k$	0	1	2	3	4	5	6	7	8
$P(x \leq k)$	.000	.001	.011	.058	.194	.448	.745	.942	1.000

2. Using the formula,

$$P(x \leq 3) = C_0^8 (.7)^0 (.3)^8 + C_1^8 (.7)^1 (.3)^7 + C_2^8 (.7)^2 (.3)^6 + C_3^8 (.7)^3 (.3)^5 = .058$$

or read directly from the table as  $P(x \leq 3) = .058$ .

3. Using the formula,  $P(x \geq 3) = \sum_{x=3}^8 C_x^8 (.7)^x (.3)^{8-x} = .989$  or use Table 1, writing

$$P(x \geq 3) = 1 - P(x \leq 2) = 1 - .011 = .989.$$

4. Using the formula,  $P(x < 3) = C_0^8 (.7)^0 (.3)^8 + C_1^8 (.7)^1 (.3)^7 + C_2^8 (.7)^2 (.3)^6 = .011$  or use Table 1, writing  $P(x < 3) = P(x \leq 2) = .011$ .

5. Using the formula,  $P(x = 3) = C_3^8 (.7)^3 (.3)^5 = 56(.343)(.00243) = .047$  or use Table 1, writing

$$P(x = 3) = P(x \leq 3) - P(x \leq 2) = .058 - .011 = .047.$$

6. Using the formula,  $P(3 \leq x \leq 5) = C_3^8 (.7)^3 (.3)^5 + C_4^8 (.7)^4 (.3)^4 + C_5^8 (.7)^5 (.3)^3 = .437$  or use Table 1, writing  $P(3 \leq x \leq 5) = P(x \leq 5) - P(x \leq 2) = .448 - .011 = .437$ .

**5.2.7-10** Similar to Exercises 2-6. These probabilities can be found individually using the binomial formula, or alternatively using the cumulative binomial tables (Table 1 in Appendix I). The values in Table 1 for  $n = 9$  and  $p = .3$  are shown in the table that follows. We choose to use the table for ease of calculation.

$k$	0	1	2	3	4	5	6	7	8	9
$P(x \leq k)$	.040	.196	.463	.730	.901	.975	.996	1.000	1.000	1.000

7.  $P(x = 2) = P(x \leq 2) - P(x \leq 1) = .463 - .196 = .267$

8.  $P(x < 2) = P(x \leq 1) = .196$

9.  $P(x > 2) = 1 - P(x \leq 2) = 1 - .463 = .537$

10.  $P(2 \leq x \leq 4) = P(x \leq 4) - P(x \leq 1) = .901 - .196 = .705$

**5.2.11-15** Similar to Exercises 2-6. These probabilities can be found individually using the binomial formula, or alternatively using the cumulative binomial tables (Table 1 in Appendix I). The values in Table 1 for  $n = 7$  and  $p = .5$  are shown in the table that follows. We choose to use the table for ease of calculation.

$k$	0	1	2	3	4	5	6	7
$P(x \leq k)$	.008	.062	.227	.500	.773	.938	.992	1.000

$$11. \quad P(x=4) = P(x \leq 4) - P(x \leq 3) = .773 - .500 = .273$$

$$12. \quad P(x \leq 1) = .062$$

$$13. \quad P(x > 1) = 1 - P(x \leq 1) = 1 - .062 = .938$$

$$14. \quad \mu = np = 7(.5) = 3.5$$

$$15. \quad \sigma = \sqrt{npq} = \sqrt{7(.5)(.5)} = 1.323$$

$$5.2.16 \quad C_2^8 (.3)^2 (.7)^6 = \frac{8(7)}{2(1)} (.09)(.117649) = .2965$$

$$5.2.17 \quad C_0^4 (.05)^0 (.95)^4 = (.95)^4 = .8145$$

$$5.2.18 \quad C_3^{10} (.5)^3 (.5)^7 = \frac{10(9)(8)}{3(2)(1)} (.5)^{10} = .1172$$

$$5.2.19 \quad C_1^7 (.2)^1 (.8)^6 = 7(.2)(.8)^6 = .3670$$

$$5.2.20 \quad C_0^8 (.2)^0 (.8)^8 = \frac{8!}{0!8!} (1)(.1678) = .1678$$

$$5.2.21 \quad C_1^8 (.2)^1 (.8)^7 = 8(.2)(.20971) = .3355$$

$$5.2.22 \quad C_2^8 (.2)^2 (.8)^6 = 28(.04)(.262144) = .2936$$

$$5.2.23 \quad P(x \leq 1) = .1678 + .3355 = .5033$$

$$5.2.24 \quad P(x \leq 2) = .1678 + .3355 + .2936 = .7969$$

5.2.25 If  $p = .5$ ,  $p(x) = C_x^n (.5)^x (.5)^{n-x} = C_x^n (.5)^n$ . Since  $C_k^n = \frac{n!}{k!(n-k)!} = C_{n-k}^n$ ,  $C_x^n = C_{n-x}^n$ , you can see that for any value of  $k$ ,

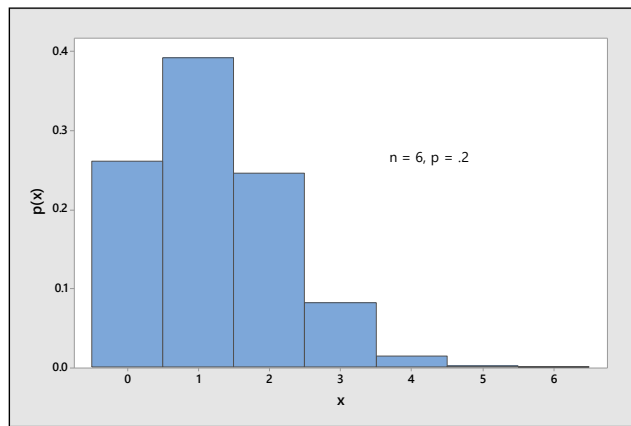
$$P(x=k) = C_k^n (.5)^n = C_{n-k}^n (.5)^n = P(x=n-k)$$

This indicates that the probability distribution is exactly symmetric around the center point  $\mu = np = .5n$ .

5.2.26 With  $n=6$  and  $p=.2$ , the values of  $p(x) = C_x^6 (p)^x (q)^{6-x}$  are calculated for  $x=0,1,2,3,4,5$ , and  $6$ . The values of  $p(x)$  are shown in the table that follows.

$x$	0	1	2	3	4	5	6
$p(x)$	.262	.393	.246	.082	.015	.002	.000

The probability histogram is shown next.

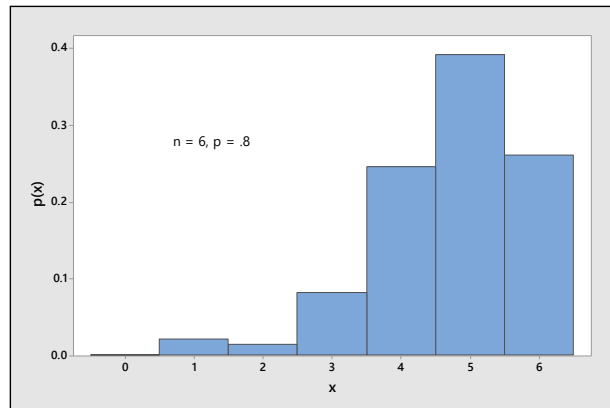


**5.2.27 a** Notice that when  $p = .8$ ,  $p(x) = C_x^6 (.8)^x (.2)^{6-x}$ . In Exercise 26, with  $p = .2$ ,  $p(x) = C_x^6 (.2)^x (.8)^{6-x}$ . The probability that  $x = k$  when  $p = .8$  ---  $C_k^6 (.8)^k (.2)^{6-k}$  --- is the same as the probability that  $x = n - k$  when  $p = .2$  ---  $C_{n-k}^6 (.2)^{n-k} (.8)^k$ . This follows because

$$C_k^n = \frac{n!}{k!(n-k)!} = C_{n-k}^n$$

Therefore, the probabilities  $p(x)$  for a binomial random variable  $x$  when  $n = 6$  and  $p = .8$  are shown in the table. The probability distribution for  $x$  is shown followed by the probability histogram.

$x$	0	1	2	3	4	5	6
$p(x)$	.000	.002	.015	.082	.246	.393	.262



**b** The probabilities found in part a and the probability histogram are the mirror images of those found in Exercise 26.

**5.2.28** Table 1, Appendix I gives the sum,  $p(0) + p(1) + \cdots + p(k)$ , for various values of  $n$  and  $p$ . The three necessary sums can be found directly in the table, by indexing the proper values of  $n$ ,  $p$  and  $k$ .

- a**  $p(0) + p(1) + p(2) + p(3) = .987$  for  $n = 10$  and  $p = .1$
- b**  $p(0) + p(1) + \cdots + p(7) = .213$  for  $n = 15$  and  $p = .6$
- c**  $p(0) + p(1) + \cdots + p(14) = .788$  for  $n = 25$  and  $p = .5$

- 5.2.29** a  $P[x \geq 4] = 1 - P[x \leq 3] = 1 - .099 = .901$   
b  $P[x = 2] = P[x \leq 2] - P[x \leq 1] = .017 - .002 = .015$   
c  $P[x < 2] = P[x \leq 1] = .002$   
d  $P[x > 1] = 1 - P[x \leq 1] = 1 - .002 = .998$

- 5.2.30** a  $P[x < 12] = P[x \leq 11] = .748$   
b  $P[x \leq 6] = .610$   
c  $P[x > 4] = 1 - P[x \leq 4] = 1 - .633 = .367$   
d  $P[x \geq 6] = 1 - P[x \leq 5] = 1 - .034 = .966$   
e  $P[3 < x < 7] = P[x \leq 6] - P[x \leq 3] = .828 - .172 = .656$

- 5.2.31** a  $\mu = 100(.01) = 1; \sigma = \sqrt{100(.01)(.99)} = .99$   
b  $\mu = 100(.9) = 90; \sigma = \sqrt{100(.9)(.1)} = 3$   
c  $\mu = 100(.3) = 30; \sigma = \sqrt{100(.3)(.7)} = 4.58$   
d  $\mu = 100(.7) = 70; \sigma = \sqrt{100(.7)(.3)} = 4.58$   
e  $\mu = 100(.5) = 50; \sigma = \sqrt{100(.5)(.5)} = 5$

- 5.2.32** The table that follows shows the values of  $p$  and  $\sigma$  (which depends on  $p$ ) for a given value of  $n$ . The maximum value of  $p$  occurs when  $p = .5$ .

$p$	.01	.30	.50	.70	.90
$\sigma$	.99	4.58	5.00	4.58	3.00

- 5.2.33** a  $p(0) = C_0^{20}(.1)^0(.9)^{20} = .1215767$   $p(3) = C_3^{20}(.1)^3(.9)^{17} = .1901199$   
 $p(1) = C_1^{20}(.1)^1(.9)^{19} = .2701703$   $p(4) = C_4^{20}(.1)^4(.9)^{16} = .0897788$   
 $p(2) = C_2^{20}(.1)^2(.9)^{18} = .2851798$

so that  $P[x \leq 4] = p(0) + p(1) + p(2) + p(3) + p(4) = .9568255$

- b** Using Table 1, Appendix I,  $P[x \leq 4]$  is read directly as .957.  
**c** Adding the entries for  $x = 0, 1, 2, 3, 4$ , we have  $P[x \leq 4] = .9569$ .  
**d**  $\mu = np = 20(.1) = 2$  and  $\sigma = \sqrt{npq} = \sqrt{1.8} = 1.3416$   
**e** For  $k = 1$ ,  $\mu \pm \sigma = 2 \pm 1.342$  or .658 to 3.342 so that

$$P[.658 \leq x \leq 3.342] = P[1 \leq x \leq 3] = .2702 + .2852 + .1901 = .7455$$

For  $k = 2$ ,  $\mu \pm 2\sigma = 2 \pm 2.683$  or  $-.683$  to 4.683 so that

$$P[-.683 \leq x \leq 4.683] = P[0 \leq x \leq 4] = .9569$$



For  $k = 3$ ,  $\mu \pm 3\sigma = 2 \pm 4.025$  or  $-2.025$  to  $6.025$  so that

$$P[-2.025 \leq x \leq 6.025] = P[0 \leq x \leq 6] = .9977$$

**f** The results are consistent with Tchebysheff's Theorem and the Empirical Rule.

**5.2.34** The random variable  $x$  is not a binomial random variable since the balls are selected without replacement. For this reason, the probability  $p$  of choosing a red ball changes from trial to trial.

**5.2.35** If the sampling in Exercise 34 is conducted with replacement, then  $x$  is a binomial random variable with  $n = 2$  independent trials, and  $p = P[\text{red ball}] = 3/5$ , which remains constant from trial to trial.

**5.2.36** Although there are  $n = 30$  days on which it either rains (S) or does not rain (F), the random variable  $x$  would not be a binomial random variable because the trials (days) are *not independent*. If there is rain on one day, it will probably affect the probability that there will be rain on the next day.

**5.2.37** Although there are trials (telephone calls) which result in either a person who will answer (S) or a person who will not (F), the number of trials,  $n$ , is not fixed in advance. Instead of recording  $x$ , the number of *successes* in  $n$  trials, you record  $x$ , the number of *trials* until the first success. This is *not* a binomial experiment.

**5.2.38** There are  $n = 100$  students which represent the experimental units.

**a** Each student either took (S) or did not take (F) the SAT. Since the population of students is large, the probability  $p = .45$  that a particular student took the SAT will not vary from student to student and the trials will be independent. This is a binomial random variable.

**b** The measurement taken on each student is *score* which can take more than two values. This is not a binomial random variable.

**c** Each student either will (S) or will not (F) score above average. As in part **a**, the trials are independent although the value of  $p$ , the proportion of students in the population who score above average, is unknown. This is a binomial experiment.

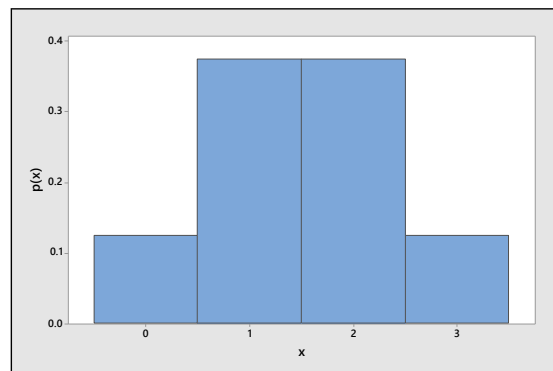
**d** The measurement taken on each student is *amount of time* which can take more than two values. This is not a binomial random variable.

**5.2.39** The random variable  $x$  is defined to be the number of heads observed when a coin is flipped three times. Then  $p = P[\text{success}] = P[\text{head}] = 1/2$ ,  $q = 1 - p = 1/2$  and  $n = 3$ . The binomial formula yields the following results.

$$\mathbf{a} \quad P[x = 0] = p(0) = C_0^3 (1/2)^0 (1/2)^3 = 1/8 \quad P[x = 1] = p(1) = C_1^3 (1/2)^1 (1/2)^2 = 3/8$$

$$P[x = 2] = p(2) = C_2^3 (1/2)^2 (1/2)^1 = 3/8 \quad P[x = 3] = p(3) = C_3^3 (1/2)^3 (1/2)^0 = 1/8$$

**b** The associated probability histogram is shown next.



**c**  $\mu = np = 3(1/2) = 1.5$  and  $\sigma = \sqrt{npq} = \sqrt{3(1/2)(1/2)} = .866$

**d** The desired intervals are

$$\mu \pm \sigma = 1.5 \pm .866 \quad \text{or} \quad .634 \text{ to } 2.366$$

$$\mu \pm 2\sigma = 1.5 \pm 1.732 \quad \text{or} \quad -.232 \text{ to } 3.232$$

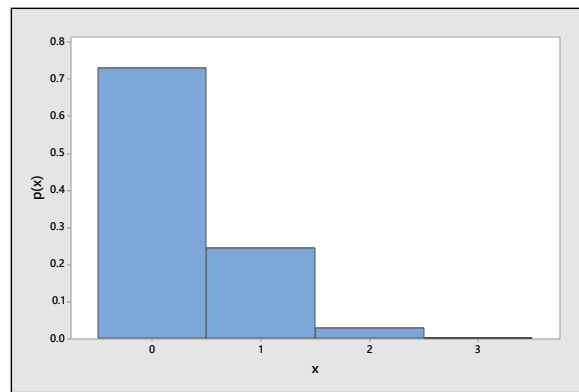
The values of  $x$  which fall in this first interval are  $x = 1$  and  $x = 2$ , and the fraction of measurement in this interval will be  $3/8 + 3/8 = 3/4$ . The second interval encloses all four values of  $x$  and thus the fraction of measurements within 2 standard deviations of the mean will be 1, or 100%. These results are consistent with both Tchebysheff's Theorem and the Empirical Rule.

**5.2.40** Refer to Exercise 39 and assume that  $p = .1$  instead of  $p = .5$ .

**a**  $P[x = 0] = p(0) = C_0^3(.1)^0(.9)^3 = .729$   $P[x = 1] = p(1) = C_1^3(.1)^1(.9)^2 = .243$

$$P[x = 2] = p(2) = C_2^3(.1)^2(.9)^1 = .027 \quad P[x = 3] = p(3) = C_3^3(.1)^3(.9)^0 = .001$$

**b** Note that the probability distribution is no longer symmetric; that is, since the probability of observing a head is so small, the probability of observing a small number of heads on three flips is increased (see the following figure).



**c**  $\mu = np = 3(.1) = .3$  and  $\sigma = \sqrt{npq} = \sqrt{3(.1)(.9)} = .520$

**d** The desired intervals are

$$\mu \pm \sigma = .3 \pm .520 \quad \text{or} \quad -.220 \text{ to } .820$$

$$\mu \pm 2\sigma = .3 \pm 1.04 \quad \text{or} \quad -.740 \text{ to } 1.34$$

The only value of  $x$  which falls in this first interval is  $x = 0$ , and the fraction of measurements in this interval will be .729. The values of  $x = 0$  and  $x = 1$  are enclosed by the second interval, so that  $.729 + .243 = .972$  of the measurements fall within two standard deviations of the mean, consistent with both Tchebysheff's Theorem and the Empirical Rule.

**5.2.41** It is given that  $n = 20$ ,  $p = .5$  and  $x =$  number of patients surviving ten years.

**a**  $P[x \geq 1] = 1 - P[x = 0] = 1 - .000 = 1$

**b**  $P[x \geq 10] = 1 - P[x \leq 9] = 1 - .412 = .588$

**c**  $P[x \geq 15] = 1 - P[x \leq 14] = 1 - .979 = .021$

**5.2.42** Define  $x$  to be the number of successful operations. Then  $p = P[\text{success}] = .8$  and  $n = 5$ . Use the binomial formula or Table 1 in Appendix I to calculate the necessary probabilities.

**a**  $P[x = 5] = C_5^5 (.8)^5 (.2)^0 = (.8)^5 = .3277$

**b**  $P[x = 4] = C_4^5 (.8)^4 (.2)^1 = 5(.8)^4 (.2) = .4096$

**c**  $P[x < 2] = p(0) + p(1) = C_0^5 (.8)^0 (.2)^5 + C_1^5 (.8)^1 (.2)^4 = .00032 + .0064 = .0067$

**d** It is known from part c that the probability of less than two successful operations is .0067. This is a very rare event, and the fact that it has occurred leads to one of two conclusions. Either the success rate is in fact 80% and a very rare occurrence has been observed, or the success rate is less than 80% and we are observing a likely event under the latter assumption. The second conclusion is more likely. We would probably conclude that the success rate for this team is less than 80% and would put little faith in the team.

**5.2.43** Define  $x$  to be the number of failures observed among the four engines. Then  $p = P[\text{engine fails}] = .01$  and  $q = 1 - p = .99$ , with  $n = 4$ .

**a**  $P[\text{no failures}] = P[x = 0] = C_0^4 (.01)^0 (.99)^4 = .9606$

**b**  $P[\text{no more than one failure}] = P[x \leq 1] = p(0) + p(1)$   
 $= C_0^4 (.01)^0 (.99)^4 + C_1^4 (.01)^1 (.99)^3 = .9606 + .0388 = .9994$

**5.2.44** Define  $x$  to be the number of young adults who prefer McDonald's. Then  $x$  has a binomial distribution with  $n = 10$  and  $p = .5$ .

**a**  $P(x > 6) = 1 - P(x \leq 6) = 1 - .828 = .172$ .

**b**  $P(4 \leq x \leq 6) = P(x \leq 6) - P(x \leq 3) = .828 - .172 = .656$ .

**c** If 4 prefer Burger King, then 6 prefer McDonalds, and vice versa. The probability is the same as that calculated in part b, since  $p = .5$ .

**5.2.45** The random variable  $x$ , the number of California homeowners with earthquake insurance, has a binomial distribution with  $n = 15$  and  $p = .1$ .

**a**  $P(x \geq 1) = 1 - P(x = 0) = 1 - .206 = .794$

**b**  $P(x \geq 4) = 1 - P(x \leq 3) = 1 - .944 = .056$

**c** Calculate  $\mu = np = 15(.1) = 1.5$  and  $\sigma = \sqrt{npq} = \sqrt{15(.1)(.9)} = 1.1619$ . Then approximately 95% of the values of  $x$  should lie in the interval

$$\mu \pm 2\sigma \Rightarrow 1.5 \pm 2(1.1619) \Rightarrow -.82 \text{ to } 3.82.$$

or between 0 and 3.

**5.2.46** Let  $x$  be the number of lost calls in a series of  $n = 11$  trials. If the coin is fair, then  $p = \frac{1}{2}$  and  $x$  has a binomial distribution.

**a**  $P[x = 11] = \left(\frac{1}{2}\right)^{11} = \frac{1}{2048}$  which is the same as odds of 1:2047.

- b** If  $n = 13$ ,  $P[x = 13] = \left(\frac{1}{2}\right)^{13} = \frac{1}{8192}$  which is a very unlikely event.
- 5.2.47 a** The random variable  $x$ , the number of plants with red petals, has a binomial distribution with  $n = 10$  and  $p = P[\text{red petals}] = .75$ .
- b** Since the value  $p = .75$  is not given in Table 1, you must use the binomial formula to calculate
- $$P(x \geq 9) = C_9^{10} (.75)^9 (.25)^1 + C_{10}^{10} (.75)^{10} (.25)^0 = .1877 + .0563 = .2440$$
- c**  $P(x \leq 1) = C_0^{10} (.75)^0 (.25)^{10} + C_1^{10} (.75)^1 (.25)^9 = .0000296$ .
- d** Refer to part **c**. The probability of observing  $x = 1$  or something even more unlikely ( $x = 0$ ) is very small  $-.0000296$ . This is a highly unlikely event if in fact  $p = .75$ . Perhaps there has been a nonrandom choice of seeds, or the 75% figure is not correct for this particular genetic cross.
- 5.2.48** The random variable  $x$ , the number of offspring with Tay-Sachs disease, has a binomial distribution with  $n = 3$  and  $p = .25$ . Use the binomial formula.
- a**  $P(x = 3) = C_3^3 (.25)^3 (.75)^0 = (.25)^3 = .015625$
- b**  $P(x = 1) = C_1^3 (.25)^1 (.75)^2 = 3(.25)(.75)^2 = .421875$
- c** Remember that the trials are independent. Hence, the occurrence of Tay-Sachs in the first two children has no effect on the third child, and  $P(\text{third child develops Tay-Sachs}) = .25$ .
- 5.2.49** The random variable  $x$ , the number of subjects who revert to their first learned method under stress, has a binomial distribution with  $n = 6$  and  $p = .8$ . The probability that at least five of the six subjects revert to their first learned method is
- $$P(x \geq 5) = 1 - P(x \leq 4) = 1 - .345 = .655$$
- 5.2.50** Define  $x$  to be the number of people with Rh-positive blood. Then  $p = P[\text{Rh-positive}] = .85$  and  $n = 2$ . Since there is no table available for  $p = .85$ , you will have to calculate the probabilities using the binomial formula.
- $$P(\text{both Rh-negative}) = P(x = 0) = C_0^2 (.85)^0 (.15)^2 = (.15)^2 = .0225$$
- 5.2.51** Define  $x$  to be the number of cars that are black. Then  $p = P[\text{black}] = .1$  and  $n = 25$ . Use Table 1 in Appendix I.
- a**  $P(x \geq 5) = 1 - P(x \leq 4) = 1 - .902 = .098$
- b**  $P(x \leq 6) = .991$
- c**  $P(x > 4) = 1 - P(x \leq 4) = 1 - .902 = .098$
- d**  $P(x = 4) = P(x \leq 4) - P(x \leq 3) = .902 - .764 = .138$
- e**  $P(3 \leq x \leq 5) = P(x \leq 5) - P(x \leq 2) = .967 - .537 = .430$
- f**  $P(\text{more than 20 not black}) = P(\text{less than 5 black}) = P(x \leq 4) = .902$

**5.2.52** Define  $x$  to be the number of players born outside of the United States. Then  $p = P[\text{success}] = .7$  and  $n = 12$ . You can use either the binomial formula or Table 1 to calculate the necessary probabilities.

**a**  $P[x \geq 5] = 1 - P[x \leq 4] = 1 - .009 = .991$

**b**  $P[x = 7] = P[x \leq 7] - P[x \leq 6] = .276 - .118 = .158$

**c**  $P[x < 6] = P[x \leq 5] = .039$

**5.2.53** Define a success to be a patient who fails to pay his bill and is eventually forgiven. Assuming that the trials are independent and that  $p$  is constant from trial to trial, this problem satisfies the requirements for the binomial experiment with  $n = 4$  and  $p = .3$ . You can use either the binomial formula or Table 1.

**a**  $P[x = 4] = p(4) = C_4^4 (.3)^4 (.7)^0 = (.3)^4 = .0081$

**b**  $P[x = 1] = p(1) = C_4^1 (.3)^1 (.7)^3 = 4(.3)(.7)^3 = .4116$

**c**  $P[x = 0] = C_4^0 (.3)^0 (.7)^4 = (.7)^4 = .2401$

**5.2.54** The random variable  $x$  has a binomial distribution with  $n = 2000$  and  $p = .3$ .

**a** The mean number of bills that would have to be forgiven is given by  $\mu = np = 2000(.3) = 600$ .

**b** The variance of  $x$  is  $\sigma^2 = npq = 2000(.3)(.7) = 420$  and the standard deviation is  $\sigma = \sqrt{420} = 20.4939$ .

**c** It is necessary to approximate  $P[x > 700]$ . From Tchebysheff's Theorem we know that at least  $1 - 1/k^2$  of the measurements lie within  $k\sigma$  of the mean. The value  $x = 700$  is 100 units away from the mean  $\mu = 600$ . This distance is equivalent to  $100/\sigma = 4.88$  standard deviations from the mean. For a point  $k = 4.88$  standard deviations from the mean, Tchebysheff's Theorem concludes that at least  $[1 - 1/(4.88)^2] = .96$  of the measurements lie within  $4.88\sigma$  of the mean (i.e.,  $600 \pm 100$ ). Therefore, at most  $1 - .96 = .04$  of the measurements are less than 500 or greater than 700. Since the distribution is fairly mound-shaped and symmetric ( $p = .3$  and  $n$  large) we can say that at most .04 but more likely .02 of the measurements are greater than 700.

**5.2.55** Define  $x$  to be the number of cell phone owners who indicate that they had walked into something or someone while talking on their cell phone. Then,  $n = 8$  and  $p = .23$ .

**a**  $P(x = 1) = C_8^1 (.23)^1 (.77)^7 = 8(.23)(.160485) = .2953$

**b** From the *Minitab Probability Density Function*, read  $P(x = 1) = .295293$ .

**c** Use the *Minitab Cumulative Distribution Function* to find  $P(x \geq 2) = 1 - P(x \leq 1) = 1 - .41887 = .58113$ .

**5.2.56** Define  $x$  to be the number of Americans who are "tasters". Then,  $n = 20$  and  $p = .7$ . Using the binomial tables in Appendix I,

**a**  $P(x \geq 17) = 1 - P(x \leq 16) = 1 - .893 = .107$

**b**  $P(x \leq 15) = .762$

**5.2.57** Define  $x$  to be the number of dog-owning households that have small dogs. Then,  $n = 15$  and  $p = .5$ . Using the binomial tables in Appendix I,

**a**  $P(x = 8) = P(x \leq 8) - P(x \leq 7) = .696 - .500 = .196$

**b**  $P(x \leq 4) = .059$

**c**  $P(x > 10) = 1 - P(x \leq 10) = 1 - .941 = .059$

### Section 5.3

**5.3.1** The Poisson random variable can be used as an approximation when  $n$  is large and  $p$  is small so that  $np < 7$ .

**5.3.2** The Poisson random variable can be used to estimate certain binomial probabilities but can also be used to model the number of events occurring in a specific period of time or space.

**5.3.3** Use the Poisson formula with  $\mu = 2.5$ .

$$P(x = 0) = \frac{2.5^0 e^{-2.5}}{0!} = .082085; \quad P(x = 1) = \frac{2.5^1 e^{-2.5}}{1!} = .205212;$$

$$P(x = 2) = \frac{2.5^2 e^{-2.5}}{2!} = .256516; \quad P(x \leq 2) = .082085 + .205212 + .256516 = .543813$$

**5.3.4** Use the Poisson formula with  $\mu = 3$ .

$$P(x = 0) = \frac{3^0 e^{-3}}{0!} = .049787; \quad P(x = 1) = \frac{3^1 e^{-3}}{1!} = .149361$$

$$P(x > 1) = 1 - P(x \leq 1) = 1 - p(0) - p(1) = 1 - .049787 - .149361 = .800852$$

**5.3.5** Using  $p(x) = \frac{\mu^x e^{-\mu}}{x!} = \frac{2^x e^{-2}}{x!}$ ,

$$P[x = 0] = \frac{2^0 e^{-2}}{0!} = .135335; \quad P[x = 1] = \frac{2^1 e^{-2}}{1!} = .27067$$

$$P[x > 1] = 1 - P[x \leq 1] = 1 - .135335 - .27067 = .593994$$

$$P[x = 5] = \frac{2^5 e^{-2}}{5!} = .036089$$

**5.3.6** These probabilities should be found using the cumulative Poisson tables (Table 2 in Appendix I). The values in Table 2 for  $\mu = 3$  are shown in the table that follows.

$k$	0	1	2	3	4	5	6	7	8	9	10
$P(x \leq k)$	.055	.199	.423	.647	.815	.916	.966	.988	.996	.999	1.000

$$P(x \leq 3) = .647, \text{ directly from the table.}$$

$$P(x > 3) = 1 - P(x \leq 3) = 1 - .647 = .353.$$

$$P(x = 3) = P(x \leq 3) - P(x \leq 2) = .647 - .423 = .224.$$

$$P(3 \leq x \leq 5) = P(x \leq 5) - P(x \leq 2) = .916 - .423 = .493.$$

- 5.3.7** These probabilities should be found using the cumulative Poisson tables (Table 2 in Appendix I). The values in Table 2 for  $\mu = 0.8$  are shown in the table that follows.

$k$	0	1	2	3	4	5
$P(x \leq k)$	.449	.809	.953	.991	.999	1.000

$P(x = 0) = .449$ , directly from the table.

$P(x \leq 2) = .953$ , directly from the table.

$$P(x > 2) = 1 - P(x \leq 2) = 1 - .953 = .047$$

$$P(2 \leq x \leq 4) = P(x \leq 4) - P(x \leq 1) = .999 - .809 = .190.$$

- 5.3.8** Table 2 in Appendix I gives the cumulative probabilities,  $P[x \leq k]$  for various values of  $\mu$  and  $k$ . For  $\mu = 2.5$ ,

$$P[x \geq 5] = 1 - P[x \leq 4] = 1 - .891 = .109$$

$$P[x < 6] = P[x \leq 5] = .958$$

$$P[x = 2] = P[x \leq 2] - P[x \leq 1] = .544 - .287 = .257$$

$$P[1 \leq x \leq 4] = P[x \leq 4] - P[x \leq 0] = .891 - .082 = .809$$

- 5.3.9** Using Table 1, Appendix I,  $P[x \leq 2] = .677$ . Then with  $\mu = np = 20(.1) = 2$ , the Poisson approximation is

$$\begin{aligned} P[x \leq 2] &\approx \frac{2^0 e^{-2}}{0!} + \frac{2^1 e^{-2}}{1!} + \frac{2^2 e^{-2}}{2!} \\ &= .135335 + .27067 + .27067 = .676675 \end{aligned}$$

The approximation is very accurate.

- 5.3.10** For a binomial distribution with  $n = 25$  and  $p = .05$ , the mean is  $\mu = 25(.05) = 1.25$ . Using the Poisson distribution with  $\mu = 1.25$ , we can approximate the binomial probabilities required:

$$\begin{aligned} p(0) &= C_0^{25} (.05)^0 (.95)^{25} \approx \frac{(1.25)^0 e^{-1.25}}{0!} = .2865 \\ p(1) &= C_1^{25} (.05)^1 (.95)^{24} \approx \frac{(1.25)^1 e^{-1.25}}{1!} = .3581 \end{aligned}$$

From Table 1, Appendix I in the text with  $n = 25$ ,  $p = .05$ , we find that the exact probabilities are

$$p(0) = .277 \text{ and } p(1) = .642 - .277 = .365$$

Note the accuracy of the Poisson approximation for  $n$  as small as 25.

- 5.3.11** Using Table 1, Appendix I,  $P[x > 6] = 1 - P(x \leq 6) = 1 - .780 = .220$ . Then with  $\mu = np = 25(.2) = 5$ , the Poisson approximation is found from Table 2 in Appendix I to be

$$P[x > 6] = 1 - P(x \leq 6) = 1 - .762 = .238$$

The approximation is quite accurate.

- 5.3.12** Let  $x$  be the number of bankruptcies filed in the district court per week. Then  $x$  has a Poisson distribution with  $\mu = 5$ . Calculate the probabilities using the Poisson formula or Table 2 in Appendix I.

**a**  $p(0) = e^{-5} = .0067$       **b**  $P(x \geq 1) = 1 - P(x = 0) = 1 - .0067 = .9933$

**c** Use Tchebysheff's Theorem with  $\mu = 5$  and  $\sigma = \sqrt{5} = 2.236$ . At least  $3/4$  of the measurements will be in the interval  $\mu \pm 2\sigma \Rightarrow 5 \pm 2(2.236) \Rightarrow 5 \pm 4.472$ . Since the Poisson random variable is integer valued, this interval,  $(.528, 9.472)$ , is the same as  $(1, 9)$ .

**5.3.13** Let  $x$  be the number of calls to a consumer hotline in a 30-minute period. Then  $x$  has a Poisson distribution with  $\mu = 5$ . Use Table 2 in Appendix I.

**a**  $P(x > 8) = 1 - P(x \leq 8) = 1 - .932 = .068$

**b** Since there is an average of 5 calls per half-hour, there would be an average of 10 calls per hour. The random variable now has a Poisson distribution with  $\mu = 10$ .

**c** Refer to part b and Table 2 with  $\mu = 10$ .  $P(x < 15) = P(x \leq 14) = .917$

**d** There is a "high probability" that  $x$  will lie within 3 standard deviations of the mean. Using Tchebysheff's Theorem with  $k = 3$ , we find

$$\mu \pm 3\sigma \Rightarrow 10 \pm 3\sqrt{10} \Rightarrow 10 \pm 9.49 \Rightarrow (.51, 19.49) \text{ or from } x = 1 \text{ to } x = 19.$$

**5.3.14** Let  $x$  be the number of visits to a website per minute.

**a** The random variable  $x$  has a Poisson distribution with  $\mu = 8$ , so that  $p(x) = \frac{8^x e^{-8}}{x!}$  for  $x = 0, 1, \dots$

**b** Use Table 2 in Appendix I.  $P(x \leq 12) = .936$

**c**  $P(x > 16) = 1 - P(x \leq 16) = 1 - .996 = .004$

**d** Using Tchebysheff's Theorem with  $k = 3$ , we find

$$\mu \pm 3\sigma \Rightarrow 8 \pm 3\sqrt{8} \Rightarrow 8 \pm 8.49 \Rightarrow (-.49, 16.49) \text{ or from } x = 0 \text{ to } x = 16.$$

**5.3.15** Let  $x$  be the number of work-related accidents per week. Then  $x$  has a Poisson distribution with  $\mu = 2$ . Use Table 2 in Appendix I.

**a**  $P(x = 0) = .135$       **b**  $P(x \geq 1) = 1 - P(x = 0) = 1 - .135 = .865$

**c-d** If we assume approximately 4 weeks per month, the number of work-related accidents per month will have a Poisson distribution with  $\mu = 4(2) = 8$  and  $P(x = 0) = \frac{8^0 e^{-8}}{0!} = e^{-8} = .000335$

**5.3.16 a** Let  $x$  be the number of births per day. Then  $x$  has a Poisson distribution with  $\mu = 6$ .

**b** Since there are 24 hours in a day, the number of hourly births will have a Poisson distribution with  $\mu = 6/24 = .25$ .

**c** Use the Poisson formula with  $\mu = .25$ .

$$P(x < 3) = P(x \leq 2) = \frac{(.25)^0 e^{-.25}}{0!} + \frac{(.25)^1 e^{-.25}}{1!} + \frac{(.25)^2 e^{-.25}}{2!} = 1.28125e^{-.25} = .9978$$

**d** Using Tchebysheff's Theorem with  $k = 3$ , we find

$$\mu \pm 3\sigma \Rightarrow .25 \pm 3\sqrt{.25} \Rightarrow .25 \pm 1.5 \Rightarrow (-1.25, 1.75) \text{ or from } x = 0 \text{ to } x = 1.$$



**5.3.17** Since  $np = 200(.01) = 2$ , we can use the Poisson approximation with  $\mu = 2$  to find  
 $P(x \geq 5) = 1 - P(x \leq 4) = 1 - .947 = .053$ .

**5.3.18** Let  $x$  be the number of misses during a given month. Then  $x$  has a Poisson distribution with  $\mu = 5$ .

**a**  $p(0) = e^{-5} = .0067$       **b**  $p(5) = \frac{5^5 e^{-5}}{5!} = .1755$

**c**  $P[x \geq 5] = 1 - P[x \leq 4] = 1 - .440 = .560$  from Table 2.

**5.3.19 a**  $p(2) = \frac{5^2 e^{-5}}{2!} = .08422$  and  $P[x \leq 2] = \frac{5^0 e^{-5}}{0!} + \frac{5^1 e^{-5}}{1!} + \frac{5^2 e^{-5}}{2!} = .12465$

**b** Recall that for the Poisson distribution,  $\mu = 5$  and  $\sigma = \sqrt{5} = 2.236$ . Therefore the value  $x = 10$  lies  $(10 - 5)/2.236 = 2.236$  standard deviations above the mean. It is not a very likely event. Alternatively, from Table 2,  $P[x > 10] = 1 - P[x \leq 10] = 1 - .986 = .014$  which is an unlikely event.

**5.3.20** Let  $x$  be the number of injuries per year, with  $\mu = 2$ .

**a**  $P[x = 2] = P[x \leq 2] - P[x \leq 1] = .677 - .406 = .271$

**b**  $P[x \geq 2] = 1 - P[x \leq 1] = 1 - .406 = .594$

**c**  $P[x \leq 1] = .406$

**5.3.21** Refer to Exercise 20.

**a**  $\mu = 2$  and  $\sigma = \sqrt{\mu} = \sqrt{2} = 1.414$

**b** The value of  $x$  should be in the interval  $\mu \pm 2\sigma \Rightarrow 2 \pm 2.828 \Rightarrow -.828$  to  $4.828$  at least  $3/4$  of the time (and probably more). Hence, most value of  $x$  will lie between 0 and 4.

**5.3.22** The random variable  $x$ , number of bacteria, has a Poisson distribution with  $\mu = 2$ . The probability of interest is

$$P[x \text{ exceeds maximum count}] = P[x > 5]$$

Using the fact that  $\mu = 2$  and  $\sigma = 1.414$ , most of the observations should fall within  $\mu \pm 2\sigma$  or 0 to 4. Hence, it is unlikely that  $x$  will exceed 5. In fact, the exact Poisson probability is  $P[x > 5] = .017$ .

**5.3.23** The random variable  $x$  is the number of cases per 100,000 of *E. Coli*. If you assume that  $x$  has an approximate Poisson distribution with  $\mu = 2.85$ ,

**a**  $P[x \leq 2] = \frac{2.85^0 e^{-2.85}}{0!} + \frac{2.85^1 e^{-2.85}}{1!} + \frac{2.85^2 e^{-2.85}}{2!} = .4576$

**b**  $P(x > 3) = 1 - P(x \leq 3) = 1 - \left[ \frac{2.85^0 e^{-2.85}}{0!} + \frac{2.85^1 e^{-2.85}}{1!} + \frac{2.85^2 e^{-2.85}}{2!} + \frac{2.85^3 e^{-2.85}}{3!} \right] = 1 - .6808 = .3192$

**5.3.24 a** Using the grouped formula given in Chapter 2, calculate

$$\bar{x} = \frac{\sum x_i f_i}{n} = \frac{0(144) + 1(91) + 2(32) + 3(11) + 4(2)}{280} = 0.7$$

**b** Using the value from part a as an estimate for  $\mu$ , calculate  $P(x=1) = \frac{0.7^1 e^{-0.7}}{1!} = .3476$ .

**c** 
$$P[x \leq 2] = \frac{0.7^0 e^{-0.7}}{0!} + \frac{0.7^1 e^{-0.7}}{1!} + \frac{0.7^2 e^{-0.7}}{2!} = .9659$$

**d** From the table, the proportion of times that one soldier was killed is  $91/280 = .3250$  (compared to the Poisson probability of .3476). The proportion of times that at most two soldiers were killed is  $267/280 = .9536$  (compared to the Poisson probability of .9659).

## Section 5.4

**5.4.1** The hypergeometric distribution is appropriate when sampling (without replacement) from a finite rather than an infinite population of successes and failures. In this case, the probability of  $p$  of a success is not constant from trial to trial and the binomial distribution is not appropriate.

**5.4.2** 
$$\frac{C_1^2 C_1^1}{C_2^3} = \frac{2(1)}{3} = \frac{2}{3}$$

**5.4.3** 
$$\frac{C_0^4 C_2^2}{C_2^6} = \frac{1(1)}{15} = \frac{1}{15}$$

**5.4.4** 
$$\frac{C_2^2 C_1^2}{C_3^4} = \frac{1(2)}{4} = \frac{1}{2}$$

**5.4.5** 
$$\frac{C_1^3 C_1^2}{C_2^5} = \frac{3(2)}{10} = .6$$

**5.4.6** 
$$\frac{C_2^4 C_1^3}{C_3^7} = \frac{6(3)}{35} = .5143$$

**5.4.7** 
$$\frac{C_4^5 C_0^3}{C_4^8} = \frac{5(1)}{70} = .0714$$

**5.4.8-10** The formula for  $p(x)$  is  $p(x) = \frac{C_x^5 C_{4-x}^3}{C_4^8}$  for  $x = 1, 2, 3, 4$ . Since there are only 3 “failures” and we are selecting 4 items, we must observe at least one “success”.

**8.** The probability of observing 4 successes is 
$$P(x=4) = \frac{C_4^5 C_0^3}{C_4^8} = \frac{5(1)}{70} = \frac{5}{70}$$

**9.** The probability of exactly one success is 
$$P(x=1) = \frac{C_1^5 C_3^3}{C_4^8} = \frac{5(1)}{70} = \frac{5}{70}$$

**10.** The probability of at most two successes is

$$P(x \leq 2) = p(1) + p(2) = \frac{C_1^5 C_3^3}{70} + \frac{C_2^5 C_2^3}{70} = \frac{5}{70} + \frac{10(3)}{70} = \frac{35}{70} = \frac{1}{2}$$

**5.4.11-13** The formula for  $p(x)$  is  $p(x) = \frac{C_x^6 C_{5-x}^4}{C_5^{10}}$  for  $x = 1, 2, 3, 4, 5$ . Since there are only 4 “failures” and we are selecting 5 items, we must select at least one “success”.

**11.** Since we must select at least one “success”,  $P[x=0] = 0$ .

**12.** 
$$P[x=2] = \frac{C_2^6 C_3^4}{C_5^{10}} = \frac{15(4)}{252} = .2381$$

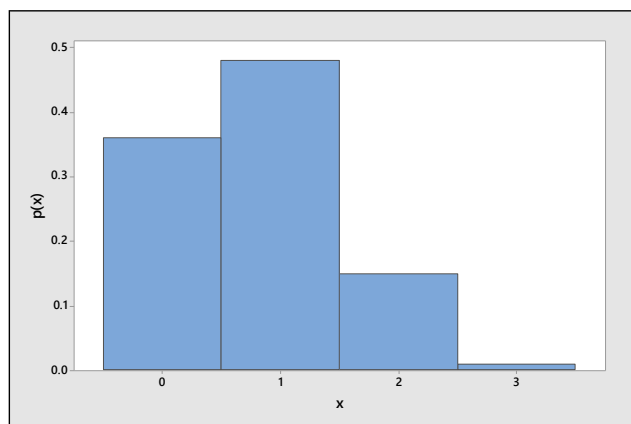
**13.** 
$$P[x \geq 2] = 1 - P[x \leq 1] = 1 - p(1) = 1 - \frac{C_1^6 C_4^4}{C_5^{10}} = 1 - \frac{6}{252} = .9762$$

**5.4.14-17** The formula for  $p(x)$  is  $p(x) = \frac{C_x^4 C_{3-x}^{11}}{C_3^{15}}$  for  $x = 0, 1, 2, 3$

$$14. \quad p(0) = \frac{C_0^4 C_3^{11}}{C_3^{15}} = \frac{165}{455} = .36 \quad p(1) = \frac{C_1^4 C_2^{11}}{C_3^{15}} = \frac{220}{455} = .48$$

$$p(2) = \frac{C_2^4 C_1^{11}}{C_3^{15}} = \frac{66}{455} = .15 \quad p(3) = \frac{C_3^4 C_0^{11}}{C_3^{15}} = \frac{4}{455} = .01$$

15. The probability histogram is shown next.



16. Using the formulas given in Section 5.4.

$$\mu = E(x) = n \left( \frac{M}{N} \right) = 3 \left( \frac{4}{15} \right) = .8$$

$$\sigma^2 = n \left( \frac{M}{N} \right) \left( \frac{N-M}{N} \right) \left( \frac{N-n}{N-1} \right) = 3 \left( \frac{4}{15} \right) \left( \frac{15-4}{15} \right) \left( \frac{15-3}{15-1} \right) = .50286$$

17. Calculate the intervals  $\mu \pm 2\sigma = .8 \pm 2\sqrt{.50286} = .8 \pm 1.418$  or  $-.618$  to  $2.218$   
 $\mu \pm 3\sigma = .8 \pm 3\sqrt{.50286} = .8 \pm 1.418$  or  $-1.327$  to  $2.927$

Then,  $P[-.618 \leq x \leq 2.218] = p(0) + p(1) + p(2) = .99$   
 $P[-1.327 \leq x \leq 2.927] = p(0) + p(1) + p(2) = .99$

These results agree with Tchebysheff's Theorem.

**5.4.18-20** The formula for  $p(x)$  is  $p(x) = \frac{C_x^5 C_{3-x}^3}{C_3^8}$  for  $x = \text{number of brown candies} = 0, 1, 2, 3$ .

$$18. \quad P(x=2) = \frac{C_2^5 C_1^3}{C_3^8} = \frac{10(3)}{56} = .5357$$

$$19. \quad P(x=0) = \frac{C_0^5 C_3^3}{C_3^8} = \frac{1}{56} = .0179$$

$$20. \quad P(x=3) = \frac{C_3^5 C_0^3}{C_3^8} = \frac{10}{56} = .1786$$

**5.4.21** The formula for  $p(x)$  is  $p(x) = \frac{C_x^3 C_{3-x}^5}{C_3^8}$  for  $x$  = number of red candies = 0, 1, 2, 3.

**5.4.22-25** The formula for  $p(x)$  is  $p(x) = \frac{C_x^{26} C_{5-x}^{26}}{C_5^{52}}$  for  $x$  = number of red cards = 0, 1, 2, 3, 4, 5.

$$22. \quad P(x=5) = \frac{C_5^{26} C_0^{26}}{C_5^{52}} = \frac{65780}{2,598,960} = .0253.$$

$$23. \quad P(x=3) = \frac{C_3^{26} C_2^{26}}{C_5^{52}} = \frac{2600(325)}{2,598,960} = .3251.$$

$$24. \quad P(x=0) = \frac{C_0^{26} C_5^{26}}{C_5^{52}} = \frac{65780}{2,598,960} = .0253.$$

$$25. \quad P(x \leq 1) = \frac{C_0^{26} C_5^{26}}{C_5^{52}} + \frac{C_1^{26} C_4^{26}}{C_5^{52}} = \frac{65780}{2,598,960} + \frac{26(14950)}{2,598,960} = .1749$$

**5.4.26-28** The formula for  $p(x)$  is  $p(x) = \frac{C_x^4 C_{3-x}^{48}}{C_3^{52}}$  for  $x$  = number of kings = 0, 1, 2, 3.

$$26. \quad P(x=3) = \frac{C_3^4 C_{3-3}^{48}}{C_3^{52}} = \frac{4}{22100} = .000181$$

$$27. \quad P(x=0) = \frac{C_0^4 C_{3-0}^{48}}{C_3^{52}} = \frac{17296}{22100} = .782624 \quad P(x=1) = \frac{C_1^4 C_{3-1}^{48}}{C_3^{52}} = \frac{4(1128)}{22100} = .204163$$

$$P(x=2) = \frac{C_2^4 C_{3-2}^{48}}{C_3^{52}} = \frac{6(48)}{22100} = .013032 \quad P(x=3) = \frac{C_3^4 C_{3-3}^{48}}{C_3^{52}} = \frac{4}{22100} = .000181$$

**28.** The probabilities in Exercise 27 be the same if  $x$  were defined to be the number of aces.

**5.4.29 a** The formula for  $p(x)$  is  $p(x) = \frac{C_x^{125} C_{10-x}^{75}}{C_{10}^{200}}$  for  $x$  = number of Republicans = 0, 1, 2, ..., 10.

**b.** The probability of selecting 10 Republicans is  $p(10) = \frac{C_{10}^{125} C_{10-10}^{75}}{C_{10}^{200}} = .0079.$

**c** The probability of selecting no Republicans is  $p(0) = \frac{C_0^{125} C_{10-0}^{75}}{C_{10}^{200}} = .0000369.$

**5.4.30** The instructor selects 5 problems from a list of 10, 6 of which the student can solve. The probability that the student can solve all 5 is

$$P(x=5) = \frac{C_5^6 C_0^4}{C_5^{10}} = \frac{6(1)}{252} = .0238$$

**5.4.31** The formula for  $p(x)$  is  $p(x) = \frac{C_x^2 C_{3-x}^4}{C_3^6}$  for  $x$  = number of defectives = 0, 1, 2.

$$p(0) = \frac{C_0^2 C_3^4}{C_3^6} = \frac{4}{20} = .2 \quad p(1) = \frac{C_1^2 C_2^4}{C_3^6} = \frac{12}{20} = .6 \quad p(2) = \frac{C_2^2 C_1^4}{C_3^6} = \frac{4}{20} = .2$$

These results agree with the probabilities calculated in Exercise 26 (Section 5.1).

**5.4.32 a** The random variable  $x$  has a hypergeometric distribution with  $N = 5, M = 2$  and  $n = 2$ . Then

$$p(x) = \frac{C_x^2 C_{2-x}^3}{C_2^5} \text{ for } x = 0, 1, 2$$

**b** Using the formulas given in this section of the text,

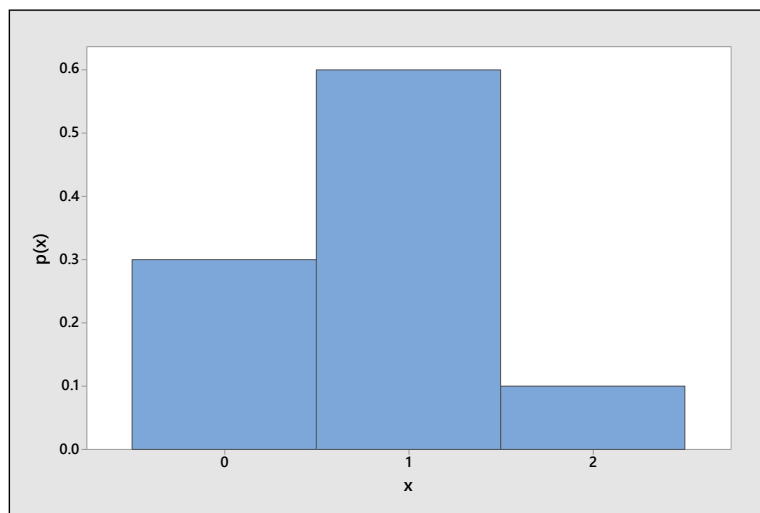
$$\mu = n \left( \frac{M}{N} \right) = 2 \left( \frac{2}{5} \right) = \frac{4}{5} = .8$$

$$\sigma^2 = n \left( \frac{M}{N} \right) \left( \frac{N-M}{N} \right) \left( \frac{N-n}{N-1} \right) = 2 \left( \frac{2}{5} \right) \left( \frac{3}{5} \right) \left( \frac{3}{4} \right) = \frac{9}{25} = .36$$

$$\sigma = \sqrt{.36} = .6$$

**c** The probability distribution and histogram for  $x$  follow.

$x$	0	1	2
$p(x)$	3/10	6/10	1/10



**5.4.33 a** The random variable  $x$  has a hypergeometric distribution with  $N = 8, M = 5$  and  $n = 3$ . Then

$$p(x) = \frac{C_x^5 C_{3-x}^3}{C_3^8} \text{ for } x = 0, 1, 2, 3$$

**b**  $P(x = 3) = \frac{C_3^5 C_0^3}{C_3^8} = \frac{10}{56} = .1786$

**c**  $P(x = 0) = \frac{C_0^5 C_3^3}{C_3^8} = \frac{1}{56} = .01786$

**d**  $P(x \leq 1) = \frac{C_0^5 C_3^3}{C_3^8} + \frac{C_1^5 C_2^3}{C_3^8} = \frac{1+15}{56} = .2857$

**5.4.34** The random variable  $x$  has a hypergeometric distribution with  $N = 10, M = 5$  and  $n = 4$ . Then

$$p(x) = \frac{C_x^5 C_{4-x}^5}{C_4^{10}} \text{ for } x = 0, 1, 2, 3, 4$$

$$\mathbf{a} \quad P(x=4) = \frac{C_4^5 C_0^5}{C_4^{10}} = \frac{5}{210} = \frac{1}{42} \quad \mathbf{b} \quad P(x \leq 3) = 1 - P(x=4) = 1 - \frac{1}{42} = \frac{41}{42}$$

$$\mathbf{c} \quad P(2 \leq x \leq 3) = p(2) + p(3) = \frac{C_2^5 C_2^5 + C_3^5 C_1^5}{C_4^{10}} = \frac{10(10) + 10(5)}{210} = \frac{150}{210} = \frac{5}{7}$$

**5.4.35** The random variable  $x$  has a hypergeometric distribution with  $N = 8, M = 2$  and  $n = 4$ . Then

$$P(x=0) = \frac{C_0^2 C_4^6}{C_4^8} = \frac{15}{70} = .214 \quad \text{and} \quad P(x=2) = \frac{C_2^2 C_2^6}{C_4^8} = \frac{15}{70} = .214$$

## Reviewing What You've Learned

**5.R.1 a** Consider the event  $x=3$ . This will occur only if either A or B wins three sets in a row; that is, if the event AAA or BBB occurs. Then

$$P[x=3] = P(AAA) + P(BBB) = [P(A)]^3 + [P(B)]^3 = (.6)^3 + (.4)^3 = .28$$

Consider the event  $x=4$ . This will occur if 3 A-wins are spread over 4 sets (with the last A-win in the fourth set) or if 3 B-wins are similarly spread over 4 sets. The associated simple events are

ABAA	BABB	AABA
BAAA	ABBB	BBAB

and the probability that  $x=4$  is

$$p(4) = 3(.6)^3(.4) + 3(.6)(.4)^3 = .3744$$

The event  $x=5$  will occur if 4 A-wins are spread over 5 sets (with the last A-win in the fifth set) or similarly for B. The associated simple events are

ABBAA	AABBA	BBAAA	BAABA	ABABA	BABAA
ABBAB	AABBB	BBAAB	BAABB	ABABB	BABAB

and the probability that  $x=5$  is  $p(5) = 6(.6)^3(.4)^2 + 6(.6)^2(.4)^3 = .3456$

Notice that  $p(3) + p(4) + p(5) = .28 + .3744 + .3456 = 1.00$ .

For a general value  $P(A) = p$  and  $P(B) = 1 - p$ , the probability distribution for  $x$  is

$$\begin{aligned} p(3) &= p^3 + (1-p)^3 \\ p(4) &= 3p^3(1-p) + 3p(1-p)^3 \\ p(5) &= 6p^3(1-p)^2 + 6p^2(1-p)^3 \end{aligned}$$

For the three values of  $p$  given in this exercise, the probability distributions and values of  $E(x)$  are given in the tables and parts **b**, **c**, and **d** that follow.

$P(A) = .6$	$P(A) = .5$	$P(A) = .9$
$\frac{x}{3} \quad \frac{p(x)}{.2800}$	$\frac{x}{3} \quad \frac{p(x)}{.25}$	$\frac{x}{3} \quad \frac{p(x)}{.7300}$

4	.3744
5	.3456

4	.375
5	.375

4	.2214
5	.0486

**b**  $E(x) = 4.0656$

**c**  $E(x) = 4.125$

**d**  $E(x) = 3.3186$

**e** Notice that as the probability of winning a single set,  $P(A)$ , increases, the expected number of sets to win the match,  $E(x)$ , decreases.

**5.R.2** Define the random variable  $x$  to be daily sales;  $x$  can take the values \$0, 50,000 or 100,000 depending on the number of customers the salesman contacts. The associated probabilities are shown next:

$$\begin{aligned} P[x = 0] &= P[\text{contact one, fail to sell}] + P[\text{contact two, fail to sell}] \\ &= P[\text{contact one}]P[\text{fail to sell}] + P[\text{contact two}]P[\text{fail with first}]P[\text{fail with second}] \\ &= (1/3)(9/10) + (2/3)(9/10)(9/10) \\ &= 9/30 + 162/300 = 252/300 \end{aligned}$$

Similarly,

$$\begin{aligned} P[x = 50,000] &= P[\text{contact one, sell}] + P[\text{contact two, sell to one}] \\ &= P[\text{contact one, sell}] + P[\text{contact two, sell to first only}] + P[\text{contact two, sell to second only}] \\ &= (1/3)(1/10) + (2/3)(1/10)(9/10) + (2/3)(9/10)(1/10) = 46/300 \end{aligned}$$

$$\begin{aligned} \text{Finally, } P[x = 100,000] &= P[\text{contact two, sell to both}] \\ &= (2/3)(1/10)(1/10) = 2/300 \end{aligned}$$

Then  $E(x) = 0(252/300) + 50,000(46/300) + 100,000(2/300) = 8333.33$ . Thus, the expected value of daily sales is \$8333.33.

**5.R.3 a**  $E(x) = 2(.12) + 3(.80) + 4(.06) + 5(.02) = 2.98$

**b**  $E(x) = 3(.14) + 4(.80) + 5(.04) + 6(.02) = 3.94$

**c**  $E(x) = 4(.04) + 5(.80) + 6(.12) + 7(.04) = 5.16$

**5.R.4** Let  $y$  represent the value of the premium which the insurance company charges and let  $x$  be the insurance company's gain. There are four possible values for  $x$ . If no accident occurs or if an accident results in no damage to the car, the insurance company gains  $y$  dollars. If an accident occurs and the car is damaged, the company will gain either  $y - 22,000$  dollars,  $y - .6(22,000)$  dollars, or  $y - .2(22,000)$  dollars, depending upon whether the damage to the car is total, 60% of market value, or 20% of market value, respectively. The following probabilities are known.

$$P[\text{accident occurs}] = .15$$

$$P[\text{total loss} \mid \text{accident occurs}] = .08$$

$$P[60\% \text{ loss} \mid \text{accident occurs}] = .12$$

$$P[20\% \text{ loss} \mid \text{accident occurs}] = .80$$

$$\text{Hence, } P[x = y - 22,000] = P[\text{accident}]P[\text{total loss} \mid \text{accident}] = .15(.08) = .012$$

$$\text{Similarly, } P[x = y - 13,200] = .15(.12) = .018 \text{ and } P[x = y - 4400] = .15(.80) = .12$$

The gain  $x$  and its associated probability distribution are shown in the table that follows. Note that  $p(y)$  is found by subtraction.

$x$	$p(x)$
$y - 22,000$	.012
$y - 13,200$	.018
$y - 4400$	.12
$y$	.85

Letting the expected gain equal zero, the value of the premium is obtained.

$$E(x) = \sum xp(x) = .012(y - 22,000) + .018(y - 13,200) + .12(y - 4400) + .85y$$

$$E(x) = y - (264 + 237.6 + 528) = y - 1029.6$$

$$y = \$1029.6$$

**5.R.5 a** Define the event R: subject chooses red and N: subject does not choose red. Then

$P(R) = \frac{1}{3}$  and  $P(N) = \frac{2}{3}$ . There are 8 simple events in the experiment:

NNN ( $x = 0$ )	RRN ( $x = 2$ )
RNN ( $x = 1$ )	RNR ( $x = 2$ )
NRN ( $x = 1$ )	NRR ( $x = 2$ )
NNR ( $x = 1$ )	RRR ( $x = 3$ )

Then

$$P(x = 0) = P(NNN) = P(N)P(N)P(N) = \left(\frac{2}{3}\right)^3 = \frac{8}{27}$$

$$P(x = 1) = 3P(N)P(N)P(R) = 3\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right) = \frac{12}{27}$$

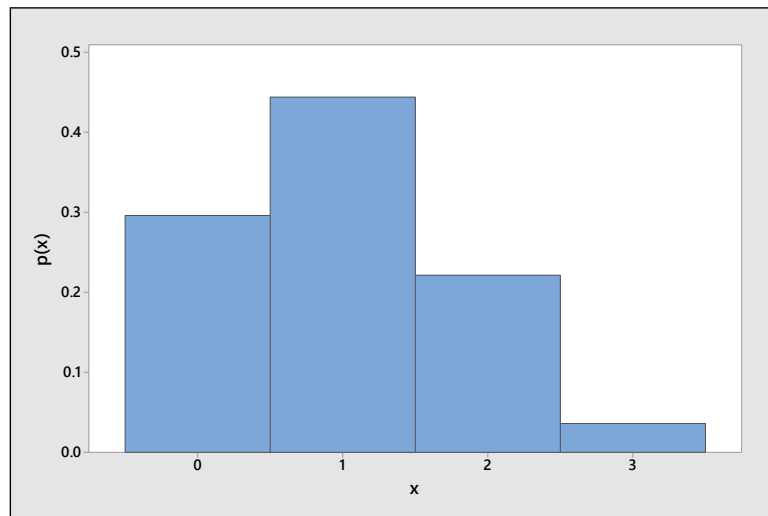
$$P(x = 2) = 3P(N)P(R)P(R) = 3\left(\frac{2}{3}\right)\left(\frac{1}{3}\right)^2 = \frac{6}{27}$$

$$P(x = 3) = P(RRR) = P(R)P(R)P(R) = \left(\frac{1}{3}\right)^3 = \frac{1}{27}$$

The probability distribution for  $x$  is shown in the table.

$x$	0	1	2	3
$p(x)$	8/27	12/27	6/27	1/27

**b** The probability histogram is shown next.



**5.R.6** Define  $x$  to be the number supporting the commissioner's claim, with  $p = .8$  and  $n = 25$ .



- a** Using the binomial tables for  $n = 25$ ,  $P[x \geq 22] = 1 - P[x \leq 21] = 1 - .766 = .234$
- b**  $P[x = 22] = P[x \leq 22] - P[x \leq 21] = .902 - .766 = .136$
- c** The probability of observing an event as extreme as  $x = 22$  (or more extreme) is quite high assuming that  $p = .8$ . Hence, this is not an unlikely event and we would not doubt the claim.
- 5.R.7** **a** Define  $x$  to be the integer between 0 and 9 chosen by a person. If the digits are equally likely to be chosen, then  $p(x) = \frac{1}{10}$  for  $x = 0, 1, 2, \dots, 9$ .
- b**  $P[4, 5 \text{ or } 6 \text{ is chosen}] = p(4) + p(5) + p(6) = \frac{3}{10}$
- c**  $P[\text{not } 4, 5 \text{ or } 6] = 1 - \frac{3}{10} = \frac{7}{10}$
- 5.R.8** Refer to Exercise 7. Redefine  $x$  to be the number of people who choose a 4, 5 or 6 in the sample of  $n = 20$ . Then  $x$  has a binomial distribution with  $p = .3$ .
- a**  $P[x \geq 8] = 1 - P[x \leq 7] = 1 - .772 = .228$
- b** Observing eight or more people choosing a 4, 5 or 6 is not an unlikely event, assuming that the integers are all equally likely. Therefore, there is no evidence to indicate that people are more likely to choose these three numbers than any others.
- 5.R.9** Let  $x$  be the number of Millennials in a sample of  $n = 20$  who say that they check in with the office at least once or twice a week while on vacation. Then  $x$  has a binomial distribution with  $p = .6$ .
- a**  $P(x = 16) = P(x \leq 16) - P(x \leq 15) = .984 - .949 = .035$
- b**  $P(16 \leq x \leq 18) = P(x \leq 18) - P(x \leq 15) = .999 - .949 = .050$
- c**  $P(x \leq 5) = .002$ . This would be an unlikely occurrence, since it occurs only 2 times in 1000.
- 5.R.10** It is given that  $x =$  number of patients with a psychosomatic problem,  $n = 25$ , and  $p = P[\text{patient has psychosomatic problem}]$ . A psychiatrist wishes to determine whether or not  $p = .8$ .
- a** Assuming that the psychiatrist is correct (that is,  $p = .8$ ), the expected value of  $x$  is  $E(x) = np = 25(.8) = 20$ .
- b**  $\sigma^2 = npq = 25(.8)(.2) = 4$
- c** Given that  $p = .8$ ,  $P[x \leq 14] = .006$  from Table 1 in Appendix I.
- d** Assuming that the psychiatrist is correct, the probability of observing  $x = 14$  or the more unlikely values,  $x = 0, 1, 2, \dots, 13$  is very unlikely. Hence, one of two conclusions can be drawn. Either we have observed a very unlikely event, or the psychiatrist is incorrect and  $p$  is actually less than .8. We would probably conclude that the psychiatrist is incorrect. The probability that we have made an incorrect decision is  $P[x \leq 14 \text{ given } p = .8] = .006$  which is quite small.
- 5.R.11** Define  $x$  to be the number of students favoring the issue, with  $n = 25$  and  $p = P[\text{student favors the issue}]$  assumed to be .8. Using the binomial tables in Appendix I,

$$P[x \leq 15] = .017$$

Thus, the probability of observing  $x = 15$  or the more extreme values,  $x = 0, 1, 2, \dots, 14$  is quite small under the assumption that  $p = .8$ . We probably should conclude that  $p$  is actually smaller than .8.

**5.R.12** Define  $x$  to be the number of students 30 years or older, with  $n = 200$  and  $p = P[\text{student is 30+ years}] = .25$ .

**a** Since  $x$  has a binomial distribution,  $\mu = np = 200(.25) = 50$  and  $\sigma = \sqrt{npq} = \sqrt{200(.25)(.75)} = 6.124$ .

**b** The observed value,  $x = 35$ , lies  $\frac{35-50}{6.124} = -2.45$  standard deviations below the mean. It is unlikely that  $p = .25$ .

**5.R.13 a**  $p = P[\text{rain according to the forecaster}] = .3$

**b** With  $n = 25$  and  $p = .3$ ,  $\mu = np = 25(.3) = 7.5$  and  $\sigma = \sqrt{npq} = \sqrt{5.25} = 2.29129$

**c** The observed value,  $x = 10$ , lies  $z = \frac{10-7.5}{2.29} = 1.09$  standard deviations above the mean.

**d** The observed event is not unlikely under the assumption that  $p = .3$ . We have no reason to doubt the forecaster.

**5.R.14 a** If there is no preference for either design, then  $p = P[\text{choose the second design}] = .5$ .

**b** Using the results of part **a** and  $n = 25$ ,  $\mu = np = 25(.5) = 12.5$  and  $\sigma = \sqrt{npq} = \sqrt{6.25} = 2.5$ .

**c** The observed value,  $x = 20$ , lies  $z = \frac{20-12.5}{2.5} = 3$  standard deviations above the mean. This is an unlikely event, assuming  $p = .5$ . We would probably conclude that there is a preference for the second design and that  $p > .5$ .

**5.R.15** The random variable  $x$ , the number of neighbors per square meter, has a Poisson distribution with  $\mu = 4$ . Use the Poisson formula or Table 2 in Appendix I.

**a**  $P(x = 0) = .018$

**b**  $P(x \leq 3) = .433$

**c**  $P(x \geq 5) = 1 - P(x \leq 4) = 1 - .629 = .371$

**d** With  $\mu = 4$  and  $\sigma = \sqrt{\mu} = 2$ , approximately 95% of the values of  $x$  should lie in the interval

$$\mu \pm 2\sigma \Rightarrow 4 \pm 4 \Rightarrow 0 \text{ to } 8$$

In fact, using Table 2, we can calculate the probability of observing between 0 and 8 neighbors per square meter to be  $P(x \leq 8) = .979$ , which is close to our approximation.

**5.R.16 a** The random variable  $x$ , the number of chickens with blue feathers, has a binomial distribution with  $n = 20$  and  $p = P[\text{blue feathers}] = P(Bb) + P(bB) = .5$ .

**b**  $\mu = np = 20(.5) = 10$ .

**c** From Table 1,  $P(x < 5) = P(x \leq 4) = .006$ .

**d** From Table 1,  $P(10 \leq x \leq 12) = P(x \leq 12) - P(x \leq 9) = .868 - .412 = .456$ .

**5.R.17 a** Since cases of adult onset diabetes is not likely to be contagious, these cases of the disorder occur independently at a rate of 12.5 per 100,000 per year. This random variable can be approximated by the Poisson random variable with  $\mu = 12$ .

**b** Use Table 2 to find  $P(x \leq 10) = .347$ .

**c**  $P(10 < x < 15) = P(x \leq 14) - P(x \leq 10) = .772 - .347 = .425$ .

**d** The probability of observing 19 or more cases per 100,000 in a year is

$$P(x \geq 19) = 1 - P(x \leq 18) = 1 - .963 = .037$$

This is an occurrence which we would not expect to see very often, if in fact  $\mu = 12$ .

**5.R.18 a** The distribution of  $x$  is actually hypergeometric, with  $N = 50$ ,  $n = 10$  and  $M = 12$  phones that will crash when tested. That is,

$$p(x) = \frac{C_x^{12} C_{10-x}^{38}}{C_{10}^{50}} \text{ for } x = 0, 1, 2, \dots, 10$$

**b** If there are 12 defectives in the carton, then  $M = 12$  and the probability that the lot is shipped is

$$P(x = 0) = \frac{C_0^{12} C_{10}^{38}}{C_{10}^{50}} = \frac{38!40!}{28!50!} = .0460$$

**c** All 50 phones will be tested if there are one or more crashes in the 10 phones. That is,

$$P(x \geq 1) = 1 - P(x = 0) = 1 - .0460 = .9540$$

**5.R.19** The random variable  $x$ , the number of adults who prefer milk chocolate to dark chocolate, has a binomial distribution with  $n = 5$  and  $p = .47$ .

**a** Since  $p = .47$  is not in Table 1, you must use the binomial formula to find

$$P(x = 5) = C_5^5 (.47)^5 (.53)^0 = .0229$$

**b** The probability that exactly three of the adults prefer milk chocolate to dark chocolate is

$$P(x = 3) = C_3^5 (.47)^3 (.53)^2 = .5833$$

**c**  $P(x \geq 1) = 1 - P(x = 0) = 1 - C_0^5 (.47)^0 (.53)^5 = 1 - (.53)^5 = .9582$

**5.R.20** The random variable  $x$ , the number of consumers who say they have become conservative spenders, has a binomial distribution with  $n = 15$  and the probability of success approximated as  $p = .6$ . Use Table 1 in Appendix I to find the necessary probabilities.

**a**  $P(x > 6) = 1 - P(x \leq 6) = 1 - .095 = .905$ .

**b**  $P(x \leq 5) = .009$ .

**c**  $P(x = 9) = P(x \leq 9) - P(x \leq 8) = .597 - .390 = .207$ .

**5.R.21 a** The random variable  $x$ , the number of tasters who pick the correct sample, has a binomial distribution with  $n = 5$  and, if there is no difference in the taste of the three samples,

$$p = P(\text{taster picks the correct sample}) = \frac{1}{3}$$

- b** The probability that exactly one of the five tasters chooses the latest batch as different from the others is

$$P(x=1) = C_1^5 \left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^4 = .3292$$

- c** The probability that at least one of the tasters chooses the latest batch as different from the others is

$$P(x \geq 1) = 1 - P(x=0) = 1 - C_0^5 \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^5 = .8683$$

**5.R.22** The random variable  $x$ , the number of questionnaires that are filled out and returned, has a binomial distribution with  $n = 20$  and  $p = .7$ . Use Table 1 in Appendix I to find the necessary probabilities.

- a**  $P(x=10) = P(x \leq 10) - P(x \leq 9) = .048 - .017 = .031$ .
- b**  $P(x \geq 12) = 1 - P(x \leq 11) = 1 - .113 = .887$ .
- c**  $P(x \leq 10) = .048$ .

**5.R.23** Refer to Exercise 22.

- a** The average value of  $x$  is  $\mu = np = 20(.7) = 14$ .
- b** The standard deviation of  $x$  is  $\sigma = \sqrt{npq} = \sqrt{20(.7)(.3)} = \sqrt{4.2} = 2.049$ .
- c** The  $z$ -score corresponding to  $x = 10$  is  $z = \frac{x - \mu}{\sigma} = \frac{10 - 14}{2.049} = -1.95$ . Since this  $z$ -score does not exceed 3 in absolute value, we would not consider the value  $x = 10$  to be an unusual observation.

**5.R.24** The random variable  $x$ , the number of treated birds who are still infected with the parasite, has a binomial distribution with  $n = 25$ . If the diet supplement is ineffective, then the proportion of infected birds should still be  $p = .3$ , even after two weeks of treatment.

- a** Using Table 1 with  $n = 25$  and  $p = .3$ , we find  $P(x \leq 3) = .033$ .
- b** If the treatment is effective, reducing the value of  $p$  to  $p = .1$ , then  $P(x \leq 3) = .764$ .

**5.R.25** The random variable  $x$  has a Poisson distribution with  $\mu = 2$ . Use Table 2 in Appendix I or the Poisson formula to find the following probabilities.

- a**  $P(x=0) = \frac{2^0 e^{-2}}{0!} = e^{-2} = .135335$
- b** 
$$P(x \leq 2) = \frac{2^0 e^{-2}}{0!} + \frac{2^1 e^{-2}}{1!} + \frac{2^2 e^{-2}}{2!}$$

$$= .135335 + .270671 + .270671 = .676676$$

**5.R.26** The random variable  $x$ , the number of salesmen who will be involved in a serious accident during the coming year, has a binomial distribution with  $n = 100$  and  $p = .01$ . To use the Poisson approximation, calculate  $\mu = np = 100(.01) = 1$ . The probability that  $x = 2$  is approximated as

$$P(x=2) = \frac{1^2 e^{-1}}{2!} = .183940$$

- 5.R.27** The random variable  $x$ , the number of applicants who will actually enroll in the freshman class, has a binomial distribution with  $n=1360$  and  $p=.9$ . Calculate  $\mu=np=1360(.9)=1224$  and  $\sigma=\sqrt{npq}=\sqrt{1360(.9)(.1)}=11.06$ . Then approximately 95% of the values of  $x$  should lie in the interval

$$\mu \pm 2\sigma \Rightarrow 1224 \pm 2(11.06) \Rightarrow 1201.87 \text{ to } 1246.12.$$

or between 1202 and 1246.

- 5.R.28** The random variable  $x$ , the number of women who get fast food when they are too busy, has a binomial distribution with  $n=100$  and  $p=.36$ .

**a** The average value of  $x$  is  $\mu=np=100(.36)=36$ .

**b** The standard deviation of  $x$  is  $\sigma=\sqrt{npq}=\sqrt{100(.36)(.64)}=\sqrt{23.04}=4.8$ .

**c** The  $z$ -score corresponding to  $x=49$  is

$$z = \frac{x - \mu}{\sigma} = \frac{49 - 36}{4.8} = 2.71$$

Since this value is not greater than 3 in absolute value, it is not an extremely unusual observation. It is between 2 and 3, however, so it is somewhat unusual.

- 5.R.29 a** The random variable  $x$ , the number of Californians who believe that college is not important, has a binomial distribution with  $n=25$  and  $p=.5$ .

**b**  $P(x \geq 17) = 1 - P(x \leq 16) = 1 - .946 = .054$ .

**c** From Table 1 in Appendix I, the largest value of  $c$  (labeled “ $k$ ” in Table 1) for which  $P(x \leq c) \leq .05$  is the value  $c=7$  for which  $P(x \leq 7) = .022$ .

**d**  $P(x \leq 6) = .007$ . This is a very unlikely event if  $p=.5$ .

## On Your Own

- 5.R.30** Define  $x$  to be the number of fields infested with whitefly. Then  $p=P[\text{infested field}]=.1$  and  $n=100$ .

**a**  $\mu=np=100(.1)=10$

**b** Since  $n$  is large, this binomial distribution should be fairly mound-shaped, even though  $p=.1$ . Hence you would expect approximately 95% of the measurements to lie within two standard deviation of the mean with  $\sigma=\sqrt{npq}=\sqrt{100(.1)(.9)}=3$ . The limits are calculated as

$$\mu \pm 2\sigma \Rightarrow 10 \pm 6 \text{ or from } 4 \text{ to } 16$$

**c** From part **b**, a value of  $x=25$  would be very unlikely, assuming that the characteristics of the binomial experiment are met and that  $p=.1$ . If this value were actually observed, it might be possible that the trials (fields) are not independent. This could easily be the case, since an infestation in one field might quickly spread to a neighboring field. This is evidence of *contagion*.

- 5.R.31** Define  $x$  to be the number of times the mouse chooses the red door. Then, if the mouse actually has no preference for color,  $p=P[\text{red door}]=.5$  and  $n=10$ . Since  $\mu=np=5$  and  $\sigma=\sqrt{npq}=1.58$ , you would expect that, if there is no color preference, the mouse should choose the red door

$$\mu \pm 2\sigma \Rightarrow 5 \pm 3.16 \Rightarrow 1.84 \text{ to } 8.16$$

or between 2 and 8 times. If the mouse chooses the red door more than 8 or less than 2 times, the unusual results might suggest a color preference.

**5.R.32** Define  $x$  to be the number of Americans who look for services close to the highway. Then,  $n = 25$  and  $p = .4$ .

**a**  $\mu = np = 25(.4) = 10$  and  $\sigma = \sqrt{npq} = \sqrt{25(.4)(.6)} = 2.449$

**b**  $\mu \pm 2\sigma \Rightarrow 10 \pm 4.898 \Rightarrow 5.102$  to  $14.898$ . Since  $x$  can take only integer values from 0 to 25, this interval consists of the values of  $x$  in the range  $6 \leq x \leq 14$ .

**c** Using Table 1 in Appendix I,  $P(6 \leq x \leq 14) = P(x \leq 14) - P(x \leq 5) = .966 - .029 = .937$ . This value agrees with Tchebysheff's Theorem (at least  $3/4$  of the measurements in this interval) and also with the Empirical Rule (approximately 95% of the measurements in this interval).

## Case Study: A Mystery: Cancers Near a Reactor

**1** Let  $x$  be the number of cancer cases in the state and let  $\mu$  be the average number of cancer cases statewide. If 52 is 50% more than the average, then

$$52 = .5\mu + \mu \text{ or } \mu = 52/1.5 = 34.67$$

**2** Since  $x$  has a Poisson distribution with  $\mu = 34.67$ , the standard deviation of  $x$  is

$$\sigma = \sqrt{\mu} = \sqrt{34.67} = 5.89.$$

**c** The  $z$ -score for the observed value of  $x$  is

$$z = \frac{x - \mu}{\sigma} = \frac{52 - 34.67}{5.89} = 2.94$$

which is somewhat unlikely, under the assumption that the average is still  $\mu = 34.67$  for this area. It is more likely that there is a difference in the cancer rates for the area close to the reactor.

## 6: The Normal Probability Distribution

### Section 6.1

**6.1.1-8** For a uniform random variable  $x$  over the interval from  $a$  to  $b$ , the probability density function is given as  $f(x) = 1/(b-a)$  for  $a \leq x \leq b$ . The probability that  $x$  lies in any particular interval is calculated as the area under the probability density function (a rectangle) over that interval.

$$1. \quad P(x < 5) = \frac{5-0}{10} = \frac{1}{2}$$

$$2. \quad P(3 < x < 7) = \frac{7-3}{10} = \frac{4}{10} = \frac{2}{5}$$

$$3. \quad P(x > 8) = \frac{10-8}{10} = \frac{1}{5}$$

$$4. \quad P(2.5 < x < 8.3) = \frac{8.3-2.5}{10} = \frac{5.8}{10} = 0.58$$

$$5. \quad P(x < 0) = \frac{0-(-1)}{1-(-1)} = \frac{1}{2}$$

$$6. \quad P(x > .7) = \frac{1-.7}{1-(-1)} = \frac{.3}{2} = .15$$

$$7. \quad P(-.5 < x < .5) = \frac{.5-(-.5)}{1-(-1)} = \frac{1}{2}$$

$$8. \quad P(-.7 < x < .2) = \frac{.2-(-.7)}{1-(-1)} = \frac{.9}{2} = .45$$

**6.1.9-16** For an exponential random variable  $x$  with parameter  $\lambda$ , the probability density function is given as  $f(x) = \lambda e^{-\lambda x}$  for  $x \geq 0$  and  $\lambda > 0$ . The probability that  $x$  lies in any particular interval is calculated as the area under the probability density function over that interval. Use the fact that  $P(x > a) = e^{-\lambda a}$  and  $P(x \leq a) = 1 - e^{-\lambda a}$ .

$$9. \quad P(x > 1) = e^{-1} = .3679$$

$$10. \quad P(1 < x < 5) = P(x < 5) - P(x < 1) = 1 - e^{-5} - (1 - e^{-1}) = e^{-1} - e^{-5} = .3611$$

$$11. \quad P(x < 1.5) = 1 - e^{-1.5} = .7769$$

$$12. \quad P(2 < x < 4) = P(x < 4) - P(x < 2) = 1 - e^{-4} - (1 - e^{-2}) = e^{-2} - e^{-4} = .1170$$

$$13. \quad P(x > 6) = e^{-2(6)} = .3012$$

$$14. \quad P(4 < x < 6) = P(x < 6) - P(x < 4) = 1 - e^{-2(6)} - (1 - e^{-2(4)}) = e^{-8} - e^{-1.2} = .1481$$

$$15. \quad P(x < 5) = 1 - e^{-2(5)} = .6321$$

$$16. \quad P(3 < x < 8) = P(x < 8) - P(x < 3) = 1 - e^{-2(8)} - (1 - e^{-2(3)}) = e^{-6} - e^{-1.6} = .3469$$

**6.1.17** It is given that  $x$  has a uniform distribution from  $a = 0$  to  $b = 30$ , so that  $f(x) = \frac{1}{30}$  for  $0 \leq x \leq 30$ .

$$\text{a.} \quad P(x > 20) = \frac{30-20}{30} = \frac{1}{3}$$

$$\text{b.} \quad P(x < 10) = \frac{10-0}{30} = \frac{1}{3}$$

$$\text{c.} \quad P(10 < x < 20) = \frac{20-10}{30} = \frac{1}{3}$$

**6.1.18** It is given that  $x$  has a uniform distribution from  $a = 25$  to  $b = 50$ , so that

$$f(x) = \frac{1}{50-25} = \frac{1}{25} \text{ for } 25 \leq x \leq 50.$$

$$\text{a. } P(x > 45) = \frac{50-45}{25} = \frac{5}{25} = \frac{1}{5} \quad \text{b. } P(35 < x < 45) = \frac{45-35}{25} = \frac{10}{25} = \frac{2}{5}$$

$$\text{c. } P(x < 40) = \frac{40-25}{25} = \frac{15}{25} = \frac{3}{5}$$

**6.1.19** It is given that  $x$  has an exponential distribution with  $\lambda = 1/365$ .

$$\text{a. } P(x < 180) = 1 - e^{-\frac{180}{365}} = .3893 \quad \text{b. } P(x > 365) = e^{-\frac{365}{365}} = .3679$$

c. If there are two batteries required and the batteries operate independently,

$$P(\text{both batteries last more than one year}) = (.3679)(.3679) = .1353$$

**6.1.20** It is given that  $x$  has an exponential distribution with  $\lambda = 1/40 = .025$ .

$$\text{a. } P(x < 5) = 1 - e^{-.025(5)} = .1175 \quad \text{b. } P(x > 50) = e^{-.025(50)} = .2865$$

$$\text{c. } P(20 < x < 30) = P(x < 30) - P(x < 20) = 1 - e^{-.025(30)} - (1 - e^{-.025(20)}) = .1342$$

d. The interval  $40 \pm 2(40)$  or  $-40$  to  $120$  includes values of  $x$  between  $0$  and  $120$ . Hence,

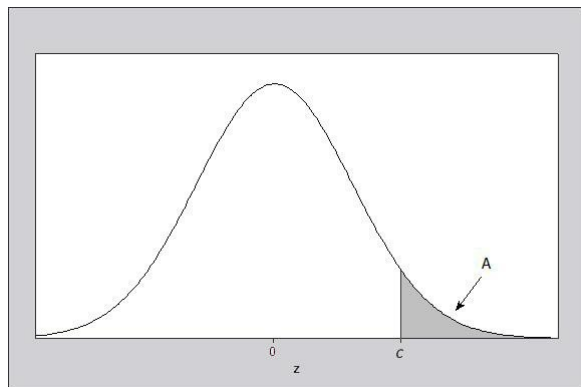
$$P(0 < x < 120) = 1 - e^{-.025(120)} = .9502 \text{ which agrees with the minimum of 75\% given by Tchebysheff's Theorem.}$$

## Section 6.2

**6.2.1-24** The first exercises are designed to provide practice for the student in evaluating areas under the normal curve. The following notes may be of some assistance.

**I** Table 3, Appendix I tabulates the cumulative area under a standard normal curve to the left of a specified value of  $z$ .

**II** Since the total area under the curve is one, the total area lying to the right of a specified value of  $z$  and the total area to its left must add to 1. Thus, in order to calculate a “tail area”, such as the one shown in Figure 6.1, the value of  $z = c$  will be indexed in Table 3, and the area that is obtained will be subtracted from 1. Denote the area obtained by indexing  $z = c$  in Table 3 by  $A(c)$  and the desired area by  $A$ . Then, in the above example,  $A = 1 - A(c)$ .





**III** To find the area under the standard normal curve between two values,  $c_1$  and  $c_2$ , calculate the difference in their cumulative areas,  $A = A(c_2) - A(c_1)$ .

**IV** Note that  $z$ , similar to  $x$ , is actually a random variable which may take on an infinite number of values, both positive and negative. Negative values of  $z$  lie to the left of the mean,  $z = 0$ , and positive values lie to the right.

1.  $P(z < 2) = P(z < 2.00) = A(2.00) = .9772$
2.  $P(z > 1.16) = 1 - P(z \leq 1.16) = 1 - A(1.16) = 1 - .8770 = .1230$
3.  $P(-2.33 < z < 2.33) = A(2.33) - A(-2.33) = .9901 - .0099 = .9802$
4.  $P(z < 1.88) = A(1.88) = .9699$
5.  $P(z > 5) < P(z > 3.49) = 1 - .9998 = .0002$  or “approximately zero”. Notice that the values in Table 3 approach 1 as the value of  $z$  increases. When the value of  $z$  is larger than  $z = 3.49$  (the largest value in the table), we can assume that the area to its left is approximately 1, and the area to its right is approximately zero.
6.  $P(-3 < z < 3) = A(3.00) - A(-3.00) = .9987 - .0013 = .9974$
7.  $P(z < 2.81) = A(2.81) = .9975$
8.  $P(z > 2.81) = 1 - P(z \leq 2.81) = 1 - A(2.81) = 1 - .9975 = .0025$
9.  $P(z < 2.33) = A(2.33) = .9901$
10.  $P(z < 1.645) = A(1.645)$ . However, the value for  $z = 1.645$  is not given in Table 3, but falls halfway between two tabulated values,  $z = 1.64$  and  $z = 1.65$ . One solution is to choose an area  $A(1.645)$  which lies halfway between the two tabulated areas,  $A(1.64) = .9495$  and  $A(1.65) = .9505$ . Then

$$A(1.645) = .9500 \quad \text{and} \quad P(z < 1.645) = A(.9500).$$

This method of evaluation is called “linear interpolation” and is performed as follows:

- 1 The difference between two entries in the table is called a “tabular difference”. Interpolation is accomplished by taking appropriate portions of this difference.
- 2 Let  $P_c$  be the probability associated with the  $z$ -value of interest, call it  $c$  (i.e.  $P_c = A(c)$ ) and let  $P_1$  and  $P_2$  be the two tabulated probabilities with corresponding  $z$  values,  $z_1$  and  $z_2$ . Consider  $\frac{c - z_1}{z_2 - z_1}$  which is the proportion of the distance from  $z_1$  to  $c$ .

$$3 \quad \text{Multiply} \quad \frac{c - z_1}{z_2 - z_1} (P_2 - P_1)$$

to obtain a corresponding proportion for the probabilities and add this value to  $P_1$ . This value is the desired  $P_c = A(c)$ . Thus, in this case,

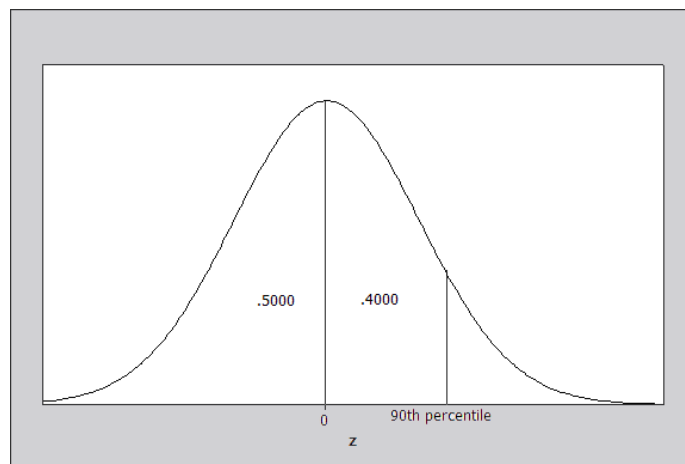
$$\frac{c - z_1}{z_2 - z_1} = \frac{1.645 - 1.64}{1.65 - 1.64} = \frac{.005}{.010} = \frac{1}{2}$$

and

$$P_c = A(c) = P_1 + \frac{c - z_1}{z_2 - z_1} (P_2 - P_1) = .9495 + \left(\frac{1}{2}\right)[.9505 - .9495] = .9495 + .0005 = .9500$$

11.  $P(z > 1.96) = 1 - A(1.96) = 1 - .9750 = .0250$
12.  $P(-2.58 < z < 2.58) = A(2.58) - A(-2.58) = .9951 - .0049 = .9902$
13. It is necessary to find the area to the left of  $z = 1.6$ . That is,  $A = A(1.6) = .9452$ .
14. The area to the left of  $z = 1.83$  is  $A = A(1.83) = .9664$ .
15. This probability is the mirror image of the probability in Exercise 14. The area to the right of  $z = -1.83$  is  $1 - A(-1.83) = 1 - .0336 = .9664$
16.  $A = A(4.18) \approx 1$ . Notice that the values in Table 3 approach 1 as the value of  $z$  increases. When the value of  $z$  is larger than  $z = 3.49$  (the largest value in the table), we can assume that the area to its left is approximately 1.
17. The area to the right of  $-1.96$  is  $1 - A(-1.96) = 1 - .0250 = .9750$
18.  $A = A(1.4) - A(-1.4) = .9192 - .0808 = .8384$
19.  $P(-1.43 < z < .68) = A(.68) - A(-1.43) = .7517 - .0764 = .6753$
20.  $P(-1.55 < z < -.44) = A(-.44) - A(-1.55) = .3300 - .0606 = .2694$
21.  $P(z > 1.34) = 1 - A(1.34) = 1 - .9099 = .0901$
22. Since the value of  $z = -4.32$  is not recorded in Table 3, you can assume that the area to the left of  $z = -4.32$  is very close to 0. Then  $P(z < -4.32) \approx 0$ .
23.  $P(.58 < z < 1.74) = A(1.74) - A(.58) = .9591 - .7190 = .2401$
24.  $A = A(1.96) - A(-1.96) = .9750 - .0250 = .9500$

**6.2.25-28** The  $p$ th percentile of the standard normal distribution is a value of  $z$  which has area  $p/100$  to its left. Since the four percentiles in these exercises are greater than the 50<sup>th</sup> percentile, the value of  $z$  will all lie to the right of  $z = 0$ , as shown for the 90<sup>th</sup> percentile in the figure that follows.

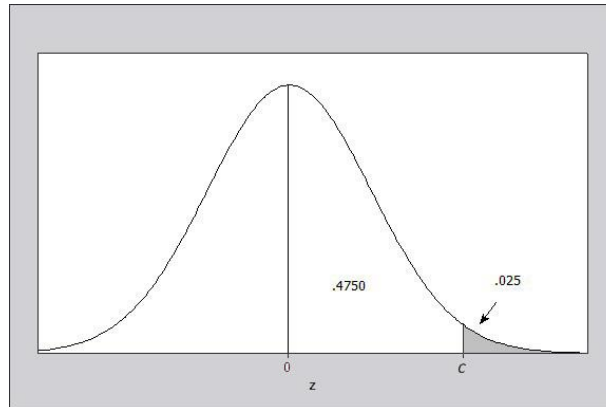


25. From the figure, the area to the left of the 90<sup>th</sup> percentile is .9000. From Table 3, the appropriate value of  $z$  is closest to  $z = 1.28$  with area .8997. Hence the 90<sup>th</sup> percentile is approximately  $z = 1.28$ .
26. As in Exercise 25, the area to the left of the 95<sup>th</sup> percentile is .9500. From Table 3, the appropriate value of  $z$  is found using linear interpolation (see Exercise 10) as  $z = 1.645$ . Hence the 95<sup>th</sup> percentile is  $z = 1.645$ .

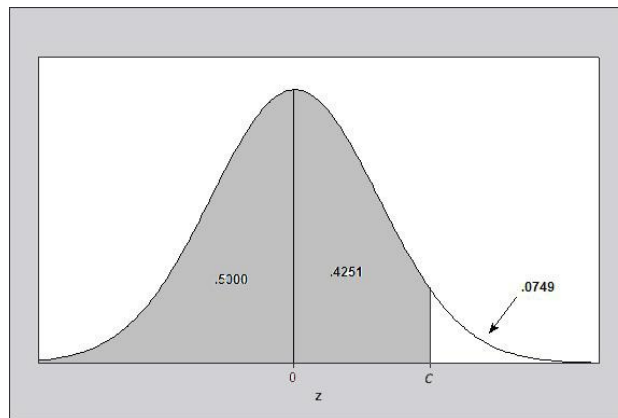
**27.** The area to the left of the 98<sup>th</sup> percentile is .9800. From Table 3, the appropriate value of  $z$  is closest to  $z = 2.05$  with area .9798. Hence the 98<sup>th</sup> percentile is approximately  $z = 2.05$ .

**28.** The area to the left of the 99<sup>th</sup> percentile is .9900. From Table 3, the appropriate value of  $z$  is closest to  $z = 2.33$  with area .9901. Hence the 99<sup>th</sup> percentile is approximately  $z = 2.33$ .

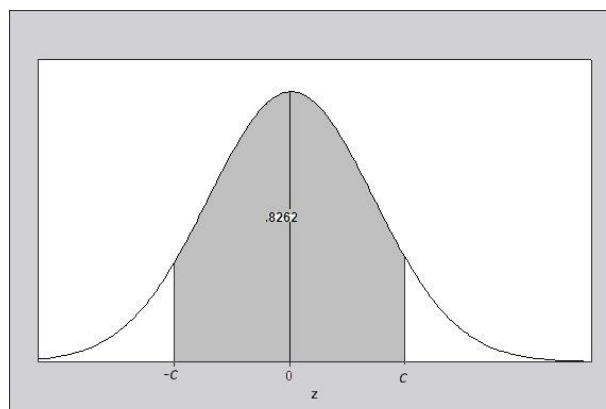
**6.2.29** We need to find a  $c$  such that  $P(z > c) = .025$ . This is equivalent to finding an indexed area of  $1 - .025 = .975$ . Search the interior of Table 3 until you find the four-digit number **.9750**. The corresponding  $z$ -value is **1.96**; that is,  $A(1.96) = .9750$ . Therefore,  $c = 1.96$  is the desired  $z$ -value (see the figure that follows).



**6.2.30** We need to find a  $c$  such that  $P(z < c) = .9251$  (see the figure that follows). Using Table 3, we find a value such that the indexed area is .9251. The corresponding  $z$ -value is  $c = 1.44$ .



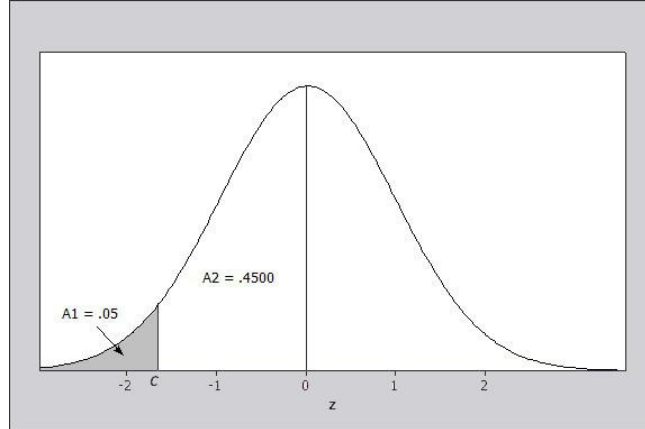
**6.2.31** We want to find a  $z$ -value such that  $P(-c < z < c) = .8262$  (see the next figure.)



Since  $A(c) - A(-c) = .8262$ , the total area in the two tails of the distribution must be  $1 - .8262 = .1738$  so that the lower tail area must be  $A(-c) = .1738/2 = .0869$ . From Table 3,  $-c = -1.36$  and  $c = 1.36$ .

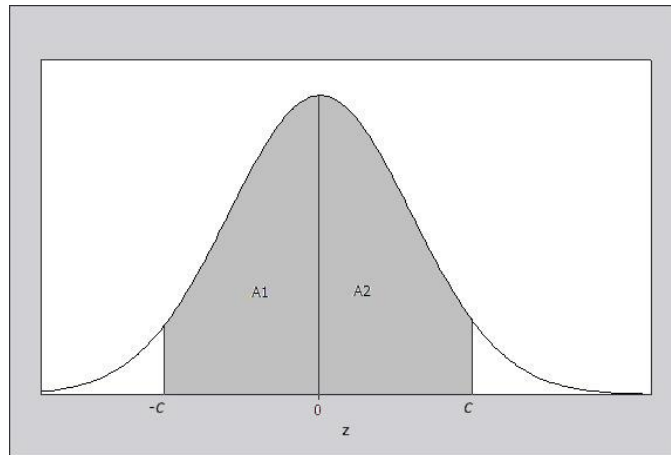
**6.2.32** Similar to Exercise 30. The value of  $c$  must be positive and  $A(c) = .9505$ . Hence,  $c = 1.65$ .

**6.2.33** It is given that the area to the left of  $c$  is .05, shown as  $A_1$  in the figure that follows. The desired value is not tabulated in Table 3 but falls between two tabulated values, .0505 and .0495. Hence, using linear interpolation (as you did in Exercise 10)  $c$  will lie halfway between  $-1.64$  and  $-1.65$ , or  $c = -1.645$ .



**6.2.34** Refer to the figure that follows. It is given that  $P(-c < z < c) = .9000$ . That is,

$$A(c) + A(-c) = .9000 \Rightarrow A(-c) = \frac{1}{2}(1 - .9000) \Rightarrow A(-c) = .0500$$



From Exercise 10,  $z_0 = 1.645$ .

**6.2.35** Refer to the figure above and consider  $P(-c < z < c) = A_1 + A_2 = .9900$ . Then

$A(-c) = \frac{1}{2}(1 - .9900) = .0050$ . Linear interpolation must now be used to determine the value of  $-c$ , which will lie between  $z_1 = -2.57$  and  $z_2 = -2.58$ . Hence, using a method similar to that in Exercise 10, we find

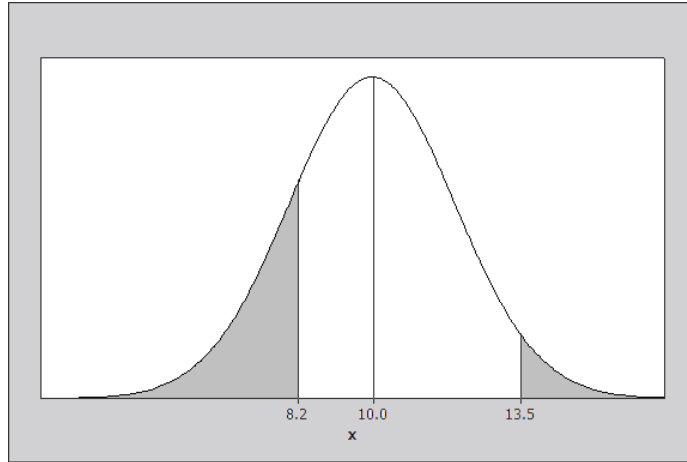
$$\begin{aligned} c &= z_1 + \frac{P_c - P_1}{P_2 - P_1}(z_2 - z_1) = -2.57 + \left( \frac{.4950 - .4949}{.4951 - .4949} \right)(-2.58 + 2.57) \\ &= -2.57 - \left( \frac{1}{2} \right)(.01) = -2.575 \end{aligned}$$

If Table 3 were correct to more than 4 decimal places, you would find that the actual value of  $c$  is  $c = 2.576$ ; many texts chose to round up and use the value  $c = 2.58$ .

- 6.2.36** Since  $z = (x - \mu)/\sigma$  measures the number of standard deviations an observation lies from its mean, it can be used to standardize any normal random variable  $x$  so that Table 3 can be used. Calculate

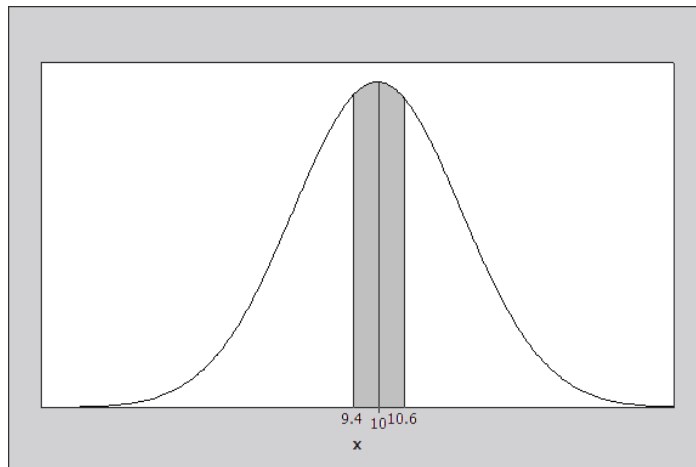
$$z = \frac{x - \mu}{\sigma} = \frac{13.5 - 10}{2} = 1.75. \text{ Then } P(x > 13.5) = P(z > 1.75) = 1 - .9599 = .0401. \text{ This probability is the}$$

shaded area in the right tail of the normal distribution shown next.



- 6.2.37** Calculate  $z = \frac{x - \mu}{\sigma} = \frac{8.2 - 10}{2} = -0.9$ . Then  $P(x < 8.2) = P(z < -0.9) = A(-.9) = .1841$ . This probability is the shaded area in the left tail of the normal distribution above.

- 6.2.38** Refer to the figure that follows.



Calculate  $z_1 = \frac{9.4 - 10}{2} = -0.3$  and  $z_2 = \frac{10.6 - 10}{2} = 0.3$ . Then

$$P(9.4 < x < 10.6) = P(-0.3 < z < 0.3) = .6179 - .3821 = .2358$$

- 6.2.39** Similar to Exercise 38. Calculate  $z_1 = \frac{1.00 - 1.20}{.15} = -1.33$  and  $z_2 = \frac{1.10 - 1.20}{.15} = -.67$ . Then

$$P(1.00 < x < 1.10) = P(-1.33 < z < -.67) = .2514 - .0918 = .1596$$

**6.2.40** Calculate  $z = \frac{x - \mu}{\sigma} = \frac{1.38 - 1.20}{.15} = 1.2$ . Then  $P(x > 1.38) = P(z > 1.2) = 1 - .8849 = .1151$ .

**6.2.41** Calculate  $z_1 = \frac{1.35 - 1.20}{.15} = 1$  and  $z_2 = \frac{1.50 - 1.20}{.15} = 2$ . Then

$$P(1.35 < x < 1.50) = P(1 < z < 2) = .9772 - .8413 = .1359.$$

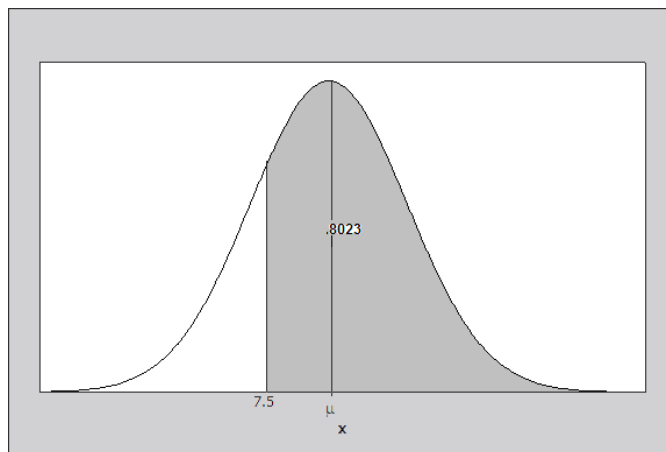
**6.2.42** The 99<sup>th</sup> percentile of the standard normal distribution was found in Exercise 28 to be  $z = 2.33$ . Since the relationship between the general normal random variable  $x$  and the standard normal  $z$  is  $z = \frac{x - \mu}{\sigma}$ , the corresponding percentile for this general normal random variable is found by solving for  $x = \mu + z\sigma$ ;

$$2.33 = \frac{x - 35}{10}$$

$$x - 35 = 23.3 \text{ or } x = 58.3$$

**6.2.43** The  $z$ -value corresponding to  $x = 0$  is  $z = \frac{x - \mu}{\sigma} = \frac{0 - 50}{15} = -3.33$ . Since the value  $x = 0$  lies more than three standard deviations away from the mean, it is considered an unusual observation. The probability of observing a value of  $z$  as large or larger than  $z = -3.33$  is  $A(-3.33) = .0004$ .

**6.2.44** It is given that  $x$  is normally distributed with  $\sigma = 2$  but with unknown mean  $\mu$ , and that  $P(x > 7.5) = .8023$ . This normal distribution is shown next.



Consider the probability  $P(x > 7.5)$ . In terms of the standard normal random variable  $z$ , we can write the  $z$ -value corresponding to  $x = 7.5$  as

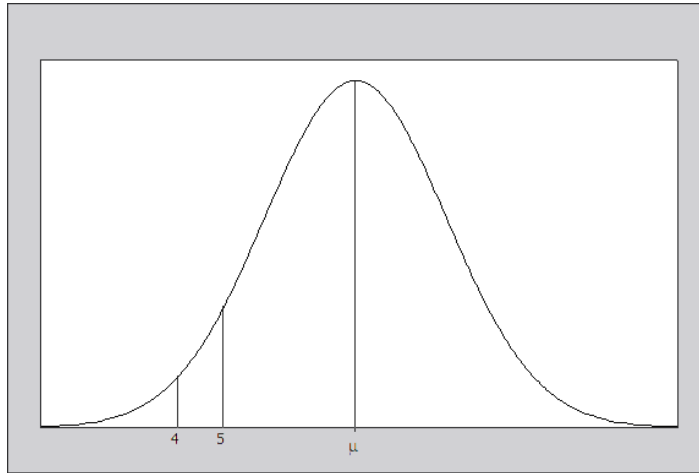
$$z = \frac{7.5 - \mu}{\sigma} = \frac{7.5 - \mu}{2}$$

And  $P(x > 7.5) = P\left(z > \frac{7.5 - \mu}{2}\right) = .8023$ . From Table 3, the value  $\frac{7.5 - \mu}{2}$  must be negative, with

$$A\left(\frac{7.5 - \mu}{2}\right) = 1 - .8023 = .1977 \text{ or } \frac{7.5 - \mu}{2} = -.85. \text{ Solving for } \mu, \mu = 7.5 + 2(.85) = 9.2.$$

**6.2.45** The random variable  $x$  is normal with unknown  $\mu$  and  $\sigma$ . However, it is given that

$P(x > 4) = P\left(z > \frac{4 - \mu}{\sigma}\right) = .9772$  and  $P(x > 5) = P\left(z > \frac{5 - \mu}{\sigma}\right) = .9332$ . These probabilities are shown in the figure that follows.



The value  $\frac{4 - \mu}{\sigma}$  is negative, with  $A\left(\frac{4 - \mu}{\sigma}\right) = 1 - .9772 = .0228$  or  $\frac{4 - \mu}{\sigma} = -2$  (i)

The value  $\frac{5 - \mu}{\sigma}$  is also negative, with  $A\left(\frac{5 - \mu}{\sigma}\right) = 1 - .9332 = .0668$  or  $\frac{5 - \mu}{\sigma} = -1.5$  (ii)

Equations (i) and (ii) provide two equations in two unknowns which can be solved simultaneously for  $\mu$

and  $\sigma$ . From (i),  $\sigma = \frac{\mu - 4}{2}$  which, when substituted into (ii) yields

$$\begin{aligned} 5 - \mu &= -1.5\left(\frac{\mu - 4}{2}\right) \\ 10 - 2\mu &= -1.5\mu + 6 \\ \mu &= 8 \end{aligned}$$

and from (i),  $\sigma = \frac{8 - 4}{2} = 2$ .

**6.2.46** Calculate  $z_1 = \frac{12 - 12.9}{1.6} = -.56$  and  $z_2 = \frac{13 - 12.9}{1.6} = .06$ . Then

$$P(12 < x < 13) = P(-.56 < z < .06) = .5239 - .2877 = .2362.$$

**6.2.47** Calculate  $z = \frac{x - \mu}{\sigma} = \frac{16 - 14.1}{1.6} = 1.19$ . Then  $P(x > 16) = P(z > 1.19) = 1 - .8830 = .1170$ .

**6.2.48** The  $z$ -value corresponding to  $x = 10$  is  $z = \frac{x - \mu}{\sigma} = \frac{10 - 14.1}{1.6} = -2.56$ . The probability of observing a value of  $x$  as small or smaller than  $x = 10$  is  $A(-2.56) = .0052$  which is a highly unlikely event.

**6.2.49** The random variable  $x$ , the height of a male human, has a normal distribution with  $\mu = 69.5$  and  $\sigma = 3.5$ . A height of 6'0" represents  $6(12) = 72$  inches, so that

$$P(x > 72) = P\left(z > \frac{72 - 69.5}{3.5}\right) = P(z > .71) = 1 - .7611 = .2389$$

- 6.2.50** The random variable  $x$ , the height of a male human, has a normal distribution with  $\mu = 69.5$  and  $\sigma = 3.5$ . Heights of 5'8" and 6'1" represent  $5(12) + 8 = 68$  and  $6(12) + 1 = 73$  inches, respectively. Then

$$P(68 < x < 73) = P\left(\frac{68-69.5}{3.5} < z < \frac{73-69.5}{3.5}\right) = P(-.43 < z < 1.00) = .8413 - .3336 = .5077$$

- 6.2.51** The random variable  $x$ , the height of a male human, has a normal distribution with  $\mu = 69.5$  and  $\sigma = 3.5$ . A height of 6'2" represents  $6(12) + 2 = 74$  inches, which has a  $z$ -value of

$$z = \frac{74-69.5}{3.5} = 1.29$$

The probability of observing a value of  $x$  as large or larger than  $x = 74$  is  $1 - A(1.29) = 1 - .9015 = .0985$ . This is not an unlikely event.

- 6.2.52** The random variable  $x$ , the height of a male human, has a normal distribution with  $\mu = 69.5$  and  $\sigma = 3.5$ . The probability that a man is 6'0" or taller was found in part **a** to be .2389, which is not an unusual occurrence. However, if you define  $y$  to be the number of men in a random sample of size  $n = 44$  who are 6'0" or taller, then  $y$  has a binomial distribution with mean  $\mu = np = 44(.2389) = 10.5116$  and standard deviation  $\sigma = \sqrt{npq} = \sqrt{44(.2389)(.7611)} = 2.828$ . The value  $y = 19$  lies

$$\frac{y-\mu}{\sigma} = \frac{19-10.5116}{2.828} = 3.002$$

standard deviations from the mean and would be considered an unusual occurrence for the general population of male humans. Perhaps our presidents do not represent a *random* sample from this population.

- 6.2.53** The random variable  $x$ , cerebral blood flow, has a normal distribution with  $\mu = 74$  and  $\sigma = 16$ .

$$P(60 < x < 80) = P\left(\frac{60-74}{16} < z < \frac{80-74}{16}\right) = P(-.88 < z < .38) = .6480 - .1894 = .4586$$

- 6.2.54** The random variable  $x$ , cerebral blood flow, has a normal distribution with  $\mu = 74$  and  $\sigma = 16$ .

$$P(x > 100) = P\left(z > \frac{100-74}{16}\right) = P(z > 1.62) = 1 - .9474 = .0526$$

- 6.2.55** The random variable  $x$ , cerebral blood flow, has a normal distribution with  $\mu = 74$  and  $\sigma = 16$ .

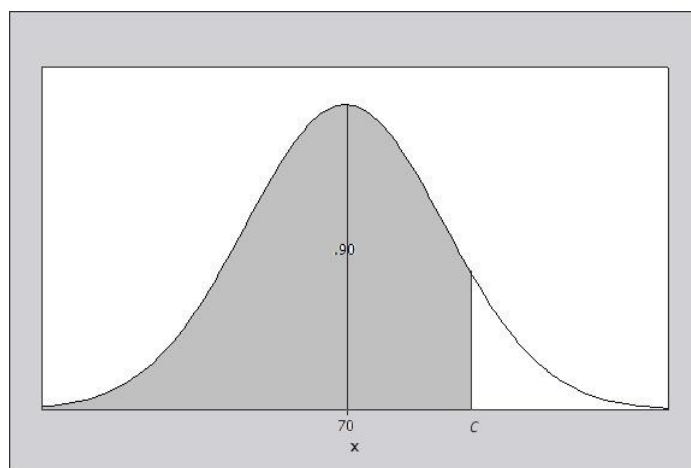
$$P(x < 40) = P\left(z < \frac{40-74}{16}\right) = P(z < -2.12) = .0170$$

- 6.2.56** It is given that  $\mu = 10.5$  and  $\sigma = 3.0$ . A washer must be repaired or replaced if its lifetime is less than five years, so that the desired fraction is  $A = P(x < 5)$ . The corresponding  $z$ -value is

$$z = \frac{x-\mu}{\sigma} = \frac{5-10.5}{3.0} = -1.83 \text{ and } P(x < 5) = P(z < -1.83) = .0336.$$

- 6.2.57** For this exercise  $\mu = 70$  and  $\sigma = 12$ . The object is to determine a value,  $c$ , for the random variable  $x$  so that  $P(x < c) = .90$  (that is, 90% of the students will finish the examination before the set time limit). Refer to the figure that follows.





We must have

$$P(x < c) = P\left(z \leq \frac{c-70}{12}\right) = .90$$

$$A\left(\frac{c-70}{12}\right) = .90$$

Consider  $z = \frac{c-70}{12}$ . Without interpolating, the approximate value for  $z$  is

$$z = \frac{c-70}{12} = 1.28 \quad \text{or} \quad c = 85.36$$

- 6.2.58** It is given that the random variable  $x$  (ounces of fill) is normally distributed with mean  $\mu$  and standard deviation  $\sigma = .3$ . It is necessary to find a value of  $\mu$  so that  $P(x > 8) = .01$ . That is, an 8-ounce cup will overflow when  $x > 8$ , and this should happen only 1% of the time. Then

$$P(x > 8) = P\left(z > \frac{8-\mu}{.3}\right) = .01.$$

From Table 3, the value of  $z$  corresponding to an area (in the upper tail of the distribution) of .01 is  $z_0 = 2.33$ . Hence, the value of  $\mu$  can be obtained by solving for  $\mu$  in the following equation:

$$2.33 = \frac{8-\mu}{.3} \quad \text{or} \quad \mu = 7.301$$

- 6.2.59** The random variable  $x$ , the gestation time for a human baby is normally distributed with  $\mu = 278$  and  $\sigma = 12$ .

**a.** To find the upper and lower quartiles, first find these quartiles for the standard normal distribution. We need a value  $c$  such that  $P(z > c) = 1 - A(c) = .2500$ . Hence,  $A(c) = .7500$ . The desired value,  $c$ , will be between  $z_1 = .67$  and  $z_2 = .68$  with associated probabilities  $P_1 = .2514$  and  $P_2 = .2483$ . Since the desired tail area, .2500, is closer to  $P_1 = .2514$ , we approximate  $c$  as  $c = .67$ . The values  $z = -.67$  and  $z = .67$  represent the 25<sup>th</sup> and 75<sup>th</sup> percentiles of the standard normal distribution. Converting these values to their equivalents for the general random variable  $x$  using the relationship  $x = \mu + z\sigma$ , you have:

The lower quartile:  $x = -.67(12) + 278 = 269.96$  and

The upper quartile:  $x = .67(12) + 278 = 286.04$

**b** If you consider a month to be approximately 30 days, the value  $x = 6(30) = 180$  is unusual, since it lies

$$z = \frac{x - \mu}{\sigma} = \frac{180 - 278}{12} = -8.167 \quad \text{standard deviations below the mean gestation time.}$$

**6.2.60** The scores are approximately normal with mean  $\mu = 75$  and standard deviation  $\sigma = 12$ . We need to find a value of  $x$ , say  $x = c$ , such that  $P(x > c) = .15$ . The  $z$ -value corresponding to  $x = c$  is

$$z = \frac{x - \mu}{\sigma} = \frac{c - 75}{12}$$

From Table 3, the  $z$ -value corresponding to an area of .15 in the right tail of the normal distribution is a value that has area .8500 to its left. The closest value given in the table is  $z = 1.04$ . Then,

$$\frac{c - 75}{12} = 1.04 \Rightarrow c = 75 + 1.04(12) = 87.48$$

The proper score to designate “extroverts” would be any score higher than 87.48.

**6.2.61** The random variable  $x$ , the weight of a package of ground beef, has a normal distribution with  $\mu = 1$  and  $\sigma = .15$ .

**a**  $P(x > 1) = P\left(z > \frac{1 - 1}{.15}\right) = P(z > 0) = 1 - .5 = .5$

**b**  $P(.95 < x < 1.05) = P\left(\frac{.95 - 1}{.15} < z < \frac{1.05 - 1}{.15}\right) = P(-.33 < z < .33) = .6293 - .3707 = .2586$

**c**  $P(x < .80) = P\left(z < \frac{.80 - 1}{.15}\right) = P(z < -1.33) = .0918$

**d** The  $z$ -value corresponding to  $x = 1.45$  is  $z = \frac{x - \mu}{\sigma} = \frac{1.45 - 1}{.15} = 3$ , which would be considered an unusual observation. Perhaps the setting on the scale was accidentally changed to 1.5 pounds!

**6.2.62** The random variable  $x$ , the diameter of a Douglas fir, has a normal distribution with  $\mu = 4$  and  $\sigma = 1.5$ .

**a**  $P(3 < x < 5) = P\left(\frac{3 - 4}{1.5} < z < \frac{5 - 4}{1.5}\right) = P(-.67 < z < .67) = .7486 - .2514 = .4972$

**b**  $P(x < 3) = P\left(z < \frac{3 - 4}{1.5}\right) = P(z < -.67) = .2514$

**c**  $P(x > 6) = P\left(z > \frac{6 - 4}{1.5}\right) = P(z > 1.33) = 1 - .9082 = .0918$

**6.2.63** It is given that  $x$  is normally distributed with  $\mu = 140$  and  $\sigma = 16$ .

**a**  $P(x \leq 120) = P\left(z \leq \frac{120 - 140}{16}\right) = P(z \leq -1.25) = .1056$

$$P(x \leq 140) = P\left(z \leq \frac{140 - 140}{16}\right) = P(z \leq 0) = .5$$

**b** In order to avoid a collision, you must brake within 160 feet or less. Hence,

$$P(x \leq 160) = P\left(z \leq \frac{160 - 140}{16}\right) = P(z \leq 1.25) = .8944$$

- 6.2.64** The random variable  $x$ , daily discharge, has mean  $\mu = 27$  and a variance  $\sigma^2 = 196$ . If  $x$  is normally distributed, find  $P(x > 50)$ . The  $z$ -value corresponding to  $x = 50$  is

$$z = \frac{x - \mu}{\sigma} = \frac{50 - 27}{14} = 1.64$$

$$\text{and } P(x > 50) = P(z > 1.64) = 1 - A(1.64) = 1 - .9495 = .0505.$$

- 6.2.65** It is given that  $x$ , the unsupported stem diameter of a sunflower plant, is normally distributed with  $\mu = 35$  and  $\sigma = 3$ .

**a**  $P(x > 40) = P\left(z > \frac{40 - 35}{3}\right) = P(z > 1.67) = 1 - .9525 = .0475$

- b** From part **a**, the probability that one plant has stem diameter of more than 40 mm is .0475. Since the two plants are independent, the probability that two plants both have diameters of more than 40 mm is

$$(.0475)(.0475) = .00226$$

- c** Since 95% of all measurements for a normal random variable lie within 1.96 standard deviations of the mean, the necessary interval is

$$\mu \pm 1.96\sigma \Rightarrow 35 \pm 1.96(3) \Rightarrow 35 \pm 5.88$$

or in the interval 29.12 to 40.88.

- d** The 90<sup>th</sup> percentile of the standard normal distribution was found in Exercise 25 to be  $z = 1.28$ . Since the relationship between the general normal random variable  $x$  and the standard normal  $z$  is  $z = \frac{x - \mu}{\sigma}$ , the corresponding percentile for this general normal random variable is found by solving for  $x = \mu + z\sigma$ .

$$x = 35 + 1.28(3) \text{ or } x = 38.84$$

- 6.2.66 a** It is given that the prime interest rate forecasts,  $x$ , are approximately normal with mean  $\mu = 4.25$  and standard deviation  $\sigma = 0.1$ . It is necessary to determine the probability that  $x$  exceeds 4.00. Calculate

$$z = \frac{x - \mu}{\sigma} = \frac{4.00 - 4.25}{0.1} = -2.5. \text{ Then } P(x > 4.00) = P(z > -2.5) = 1 - .0062 = .9938.$$

- b** Calculate  $z = \frac{x - \mu}{\sigma} = \frac{4.30 - 4.25}{0.1} = 0.5$ . Then  $P(x < 4.30) = P(z < 0.5) = .6915$ .

- 6.2.67** It is given that the counts of the number of bacteria are normally distributed with  $\mu = 85$  and  $\sigma = 9$ . The  $z$ -value corresponding to  $x = 100$  is  $z = \frac{x - \mu}{\sigma} = \frac{100 - 85}{9} = 1.67$  and

$$P(x > 100) = P(z > 1.67) = 1 - .9525 = .0475$$

- 6.2.68** The amount of money spent between 4 and 6 pm on Sundays is normally distributed with  $\mu = 185$  and  $\sigma = 20$ .

- a** The  $z$ -value corresponding to  $x = 195$  is  $z = \frac{x - \mu}{\sigma} = \frac{195 - 185}{20} = 0.5$ . Then

$$P(x > 195) = P(z > 0.5) = 1 - .6915 = .3085$$

- b** The  $z$ -value corresponding to  $x = 215$  is  $z = \frac{x - \mu}{\sigma} = \frac{215 - 185}{20} = 1.5$ . Then

$$P(195 < x < 215) = P(0.5 < z < 1.5) = .9332 - .6915 = .2417$$

**c** First, find  $P(x > 215) = P(z > 1.5) = 1 - .9332 = .0668$  for a single shopper. For two shoppers, use the Multiplication Rule.

$$P(\text{both shoppers spend more than \$215}) = P(\text{1st spends more than \$215}) \times P(\text{2nd spends more than \$215}) \\ = (.0668)(.0668) = .0045$$

**6.2.69** The pulse rates are normally distributed with  $\mu = 78$  and  $\sigma = 12$ .

**a** The  $z$ -values corresponding to  $x = 60$  and  $x = 100$  are  $z = \frac{x - \mu}{\sigma} = \frac{60 - 78}{12} = -1.5$  and

$$z = \frac{x - \mu}{\sigma} = \frac{100 - 78}{12} = 1.83 \quad \text{Then} \quad P(60 < x < 100) = P(-1.5 < z < 1.83) = .9664 - .0668 = .8996.$$

**b** From Exercise 26, we found that the 95<sup>th</sup> percentile of the *standard* normal ( $z$ ) distribution is  $z = 1.645$ .

Since  $z = \frac{x - \mu}{\sigma} = \frac{x - 78}{12}$ , solve for  $x$  to find the 95<sup>th</sup> percentile for the pulse rates:

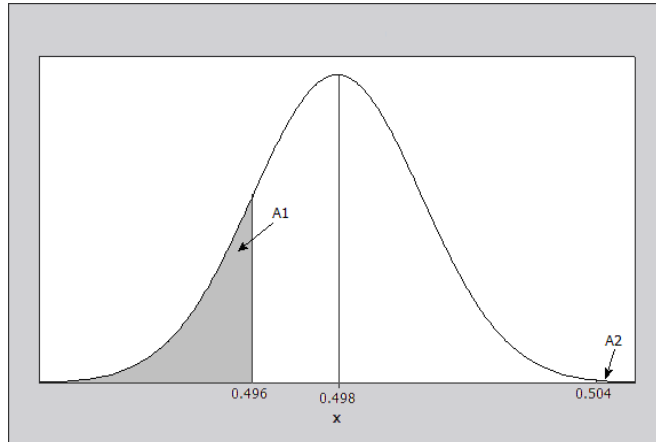
$$1.645 = \frac{x - 78}{12} \Rightarrow x = 78 + 1.645(12) = 97.74$$

**c** The  $z$ -score for  $x = 110$  is  $z = \frac{x - \mu}{\sigma} = \frac{110 - 78}{12} = 2.67$  and

$$P(x > 110) = P(z > 2.67) = 1 - .9962 = .0038.$$

The  $z$ -score is between 2 and 3; the probability of observing a value this large or larger is quite small. This pulse rate would be somewhat unusual.

**6.2.70** For this exercise, it is given that the population of bolt diameters is normally distributed with  $\mu = .498$  and  $\sigma = .002$ . Thus, no correction for continuity is necessary. The fraction of acceptable bolts will be those which lie in the interval from .496 to .504. All others are unacceptable. The desired fraction of acceptable bolts is calculated, and the fraction of unacceptable bolts (shaded in the figure that follows) is obtained by subtracting from the total probability, which is 1.



The fraction of acceptable bolts is then

$$P(.496 \leq x \leq .504) = P\left(\frac{.496 - .498}{.002} \leq z \leq \frac{.504 - .498}{.002}\right) \\ = P(-1 \leq z \leq 3) = .9987 - .1587 = .8400$$

and the fraction of unacceptable bolts is  $1 - .84 = .16$ .

### Section 6.3

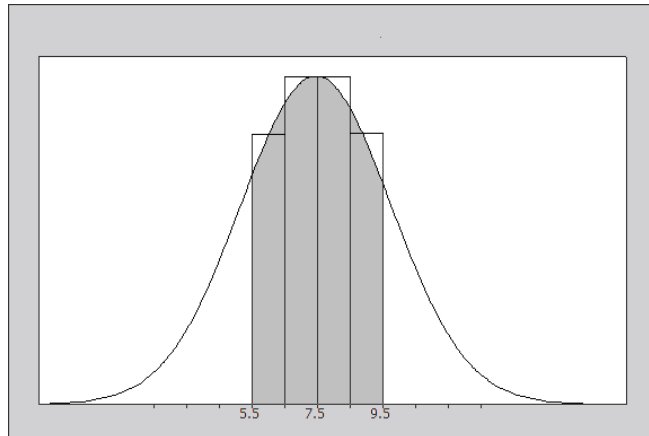
**6.3.1-4** The normal distribution can be used to approximate binomial probabilities when both  $np$  and  $nq$  are greater than 5. If  $n$  is large and  $np$  is small (preferably less than 7), the Poisson distribution can be used to approximate binomial probabilities.

1. Calculate  $np = 25(.6) = 15$  and  $nq = 25(.4) = 10$ . The normal approximation is appropriate.
2. Calculate  $np = 45(.05) = 2.25$  and  $nq = 45(.95) = 42.75$ . The normal approximation is not appropriate, but the Poisson approximation can be used.
3. Calculate  $np = 25(.3) = 7.5$  and  $nq = 25(.7) = 17.5$ . The normal approximation is appropriate.
4. Calculate  $np = 15(.5) = 7.5$  and  $nq = 15(.5) = 7.5$ . The normal approximation is appropriate.

**6.3.5** Calculate  $\mu = np = 15$  and  $\sigma = \sqrt{npq} = \sqrt{25(.6)(.4)} = 2.449$ . To find the probability of more than 9 successes, we need to include the values  $x = 10, 11, \dots, 25$ . To include the entire block of probability for the first value of  $x = 10$ , we need to start at 9.5. Then the probability of more than 9 successes is approximated as

$$P(x > 9.5) = P\left(z > \frac{9.5 - 15}{2.449}\right) = P(z > -2.25) = 1 - .0122 = .9878.$$

**6.3.6** For the binomial random variable,  $\mu = np = 7.5$  and  $\sigma = \sqrt{npq} = \sqrt{25(.3)(.7)} = 2.291$ . The probability of interest is the area under the binomial probability histogram corresponding to the rectangles  $x = 6, 7, 8$  and  $9$  in the figure that follows.



To approximate this area, use the “correction for continuity” and find the area under a normal curve with mean  $\mu = 7.5$  and  $\sigma = 2.291$  between  $x_1 = 5.5$  and  $x_2 = 9.5$ . The  $z$ -values corresponding to the two values of  $x$  are

$$z_1 = \frac{5.5 - 7.5}{2.291} = -.87 \quad \text{and} \quad z_2 = \frac{9.5 - 7.5}{2.291} = .87$$

The approximating probability is  $P(5.5 < x < 9.5) = P(-.87 < z < .87) = .8078 - .1922 = .6156$ .

**6.3.7** For this binomial random variable,  $\mu = np = 100(.2) = 20$  and  $\sigma = \sqrt{npq} = 100\sqrt{25(.2)(.8)} = 4$ . To include the entire rectangles for  $x = 21$  and  $x = 24$ , the approximating probability is

$$P(20.5 < x < 24.5) = P(.12 < z < 1.12) = .8686 - .5478 = .3208$$

**6.3.8** For this binomial random variable,  $\mu = np = 100(.2) = 20$  and  $\sigma = \sqrt{npq} = 100\sqrt{25(.2)(.8)} = 4$ . The approximating probability will be  $P(x > 22.5)$  where  $x$  has a normal distribution with  $\mu = 100(.2) = 20$  and  $\sigma = \sqrt{100(.2)(.8)} = 4$ . Then

$$P(x > 22.5) = P\left(z > \frac{22.5 - 20}{4}\right) = P(z > .62) = 1 - .7324 = .2676$$

**6.3.9** For this binomial random variable,  $\mu = np = 100(.2) = 20$  and  $\sigma = \sqrt{npq} = 100\sqrt{25(.2)(.8)} = 4$ . The approximating probability is  $P(x > 21.5)$  since the entire rectangle corresponding to  $x = 22$  must be included.

$$P(x > 21.5) = P\left(z > \frac{21.5 - 20}{4}\right) = P(z > .38) = 1 - .6480 = .3520$$

**6.3.10** For this binomial random variable,  $\mu = np = 100(.2) = 20$  and  $\sigma = \sqrt{npq} = 100\sqrt{25(.2)(.8)} = 4$ . To include the entire rectangle for  $x = 25$ , the approximating probability is

$$P(x < 25.5) = P(z < 1.38) = .9162$$

**6.3.11** The normal approximation with “correction for continuity” is  $P(354.5 < x < 360.5)$  where  $x$  is normally distributed with mean and standard deviation given by

$$\mu = np = 400(.9) = 360 \text{ and } \sigma = \sqrt{npq} = \sqrt{400(.9)(.1)} = 6$$

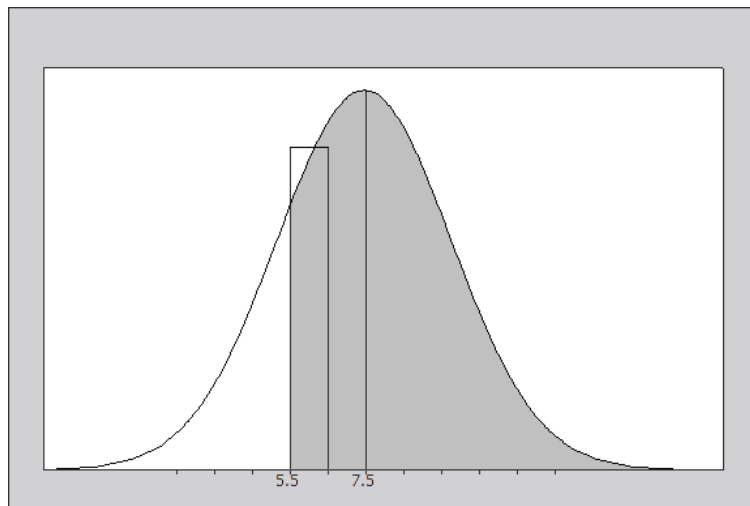
$$\begin{aligned} \text{Then } P(354.5 < x < 360.5) &= P\left(\frac{354.5 - 360}{6} < z < \frac{360.5 - 360}{6}\right) \\ &= P(-.92 < z < .08) = .5319 - .1788 = .3531 \end{aligned}$$

**6.3.12** From Table 1 in Appendix I, the exact binomial probabilities are:

$$P(x \geq 6) = 1 - P(x \leq 5) = 1 - .151 = .849$$

$$P(x > 6) = 1 - P(x \leq 6) = 1 - .304 = .696$$

To approximate these probabilities, first calculate  $\mu = np = 7.5$  and  $\sigma = \sqrt{npq} = \sqrt{15(.5)(.5)} = 1.9365$ . The first probability of interest is the area under the binomial probability histogram corresponding to the rectangles  $x = 6, 7, \dots, 15$  in the figure that follows



$$\text{Then } P(x > 5.5) = P\left(z > \frac{5.5 - 7.5}{1.9365}\right) = P(z > -1.03) = 1 - .1515 = .8485.$$

To approximate  $P(x > 6)$ , the area corresponding to the rectangles  $x = 7, 8, \dots, 15$  you use

$$P(x > 6.5) = P\left(z > \frac{6.5 - 7.5}{1.9365}\right) = P(z > -.52) = 1 - .3015 = .6985$$

These probabilities are not far from the actual probabilities, .849 and .696, respectively.

**6.3.13** Given a binomial random variable  $x$  with  $n = 25$  and  $p = .2$ , use Table 1 to calculate

$$P(4 \leq x \leq 6) = P(x \leq 6) - P(x \leq 3) = .780 - .234 = .546$$

To approximate this probability, we find the area under the binomial distribution for  $x = 4, 5, 6$ , using the “correction for continuity” and find the area under the normal curve between 3.5 and 6.5. This is done in order to include the entire area under the rectangles associated with the different values of  $x$ . First find the mean and standard deviation of the binomial random variable  $x$ :

$$\mu = np = 25(.2) = 5 \text{ and } \sigma = \sqrt{npq} = \sqrt{25(.2)(.8)} = 2. \text{ Then}$$

$$P(3.5 < x < 6.5) = P\left(\frac{3.5 - 5}{2} < z < \frac{6.5 - 5}{2}\right) = P(-.75 < z < .75) = .7734 - .2266 = .5468$$

The actual and approximated probabilities are very close.

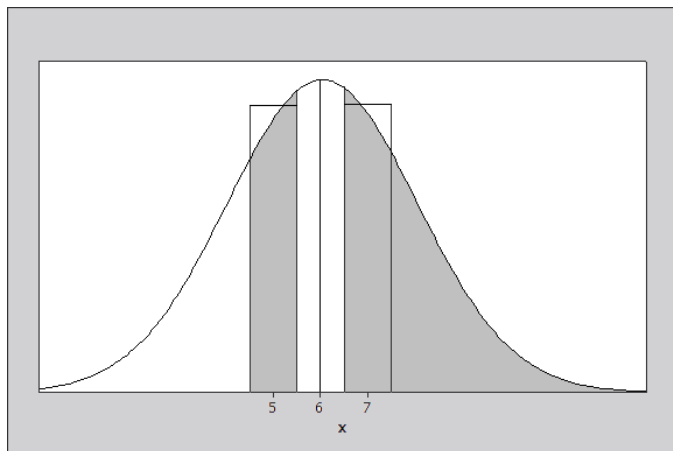
**6.3.14** Using the binomial tables for  $n = 20$  and  $p = .3$ , you can verify that

$$P(x \geq 7) = 1 - P(x \leq 6) = 1 - .608 = .392$$

$$P(x = 5) = P(x \leq 5) - P(x \leq 4) = .416 - .238 = .178$$

To approximate these probabilities, first calculate the mean and standard deviation of this binomial distribution:

$$\mu = np = 20(.3) = 6 \text{ and } \sigma = \sqrt{npq} = \sqrt{20(.3)(.7)} = \sqrt{4.2} = 2.049$$



The appropriate correction for continuity (see figure above) yields  $P(x \geq 7)$  is

$$P(x \geq 6.5) = P(z \geq .24) = 1 - .5948 = .4052$$

Similarly,  $P(x = 5) \approx P(4.5 \leq x \leq 5.5)$ . The  $z$ -values corresponding to  $x_1 = 4.5$  and  $x_2 = 5.5$  are

$$z_1 = \frac{4.5-6}{2.049} = -.73 \quad \text{and} \quad z_2 = \frac{5.5-6}{2.049} = -.24$$

so that  $P(4.5 \leq x \leq 5.5) = P(-.73 \leq z \leq -.24) = .4052 - .2327 = .1725$ . The actual and approximated probabilities are very close.

**6.3.15** Similar to previous exercises. With  $n = 20$  and  $p = .4$ , use Table 1 to find

$$P(x \geq 10) = 1 - P(x \leq 9) = 1 - .755 = .245.$$

To use the normal approximation, find the mean and standard deviation of this binomial random variable:

$$\mu = np = 20(.4) = 8 \quad \text{and} \quad \sigma = \sqrt{npq} = \sqrt{20(.4)(.6)} = \sqrt{4.8} = 2.191.$$

Using the continuity correction, it is necessary to find the area to the right of 9.5. The  $z$ -value

$$\text{corresponding to } x = 9.5 \text{ is} \quad z = \frac{9.5-8}{2.191} = .68 \quad \text{and}$$

$$P(x \geq 10) \approx P(z > .68) = 1 - .7517 = .2483.$$

Note that the normal approximation is very close to the exact binomial probability.

**6.3.16 a** The approximating probability will be  $P(x < 29.5)$  where  $x$  has a normal distribution with

$$\mu = 50(.47) = 23.5 \quad \text{and} \quad \sigma = \sqrt{50(.47)(.53)} = 3.529.$$
 Then

$$P(x < 29.5) = P\left(z < \frac{29.5-23.5}{3.529}\right) = P(z < 1.70) = .9554$$

**b** The approximating probability is

$$P(x > 20.5) = P\left(z > \frac{20.5-23.5}{3.529}\right) = P(z > -.85) = 1 - .1977 = .8023$$

**c** If more than 10 individuals *do not* associate “recycling” with Earth Day, then less than  $50 - 10 = 40$  *do* associate it with Earth Day. The approximating probability is

$$P(x < 39.5) = P\left(z < \frac{39.5-23.5}{3.529}\right) = P(z < 4.53) \approx 1$$

**6.3.17** The random variable  $x$  has a binomial distribution with  $n = 100$  and  $p = .51$ . Calculate

$\mu = np = 100(.51) = 51$  and  $\sigma = \sqrt{npq} = \sqrt{100(.51)(.49)} = 4.999$ . The probability that  $x$  is 60 or more is approximated as

$$P(x \geq 59.5) = P\left(z \geq \frac{59.5-51}{4.999}\right) = P(z \geq 1.70) = 1 - .9554 = .0446$$

**6.3.18** The following information is available:

$x$  = number of relays from supplier A

$p = P[\text{relay comes from supplier A}] = 2/3$

$n = 75$

Calculate  $\mu = np = 75(2/3) = 50$  and  $\sigma = \sqrt{npq} = \sqrt{16.667} = 4.082$ . The probability of interest is  $P(x \leq 48)$ , which is approximated by  $P(x \leq 48.5)$  using the normal approximation to the binomial

distribution. The  $z$ -value corresponding to  $x = 48.5$  is  $z = \frac{x-\mu}{\sigma} = \frac{48.5-50}{\sqrt{16.667}} = -.37$  and

$$P(x \leq 48) \approx P(z \leq -.37) = .3557.$$



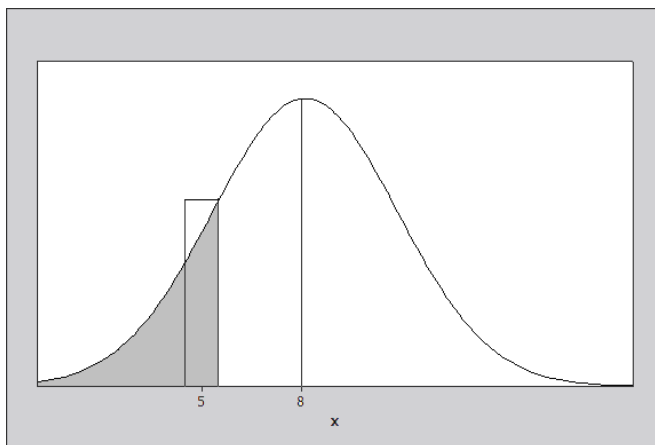
- 6.3.19** The random variable of interest is  $x$ , the number of persons not showing up for a given flight. This is a binomial random variable with  $n = 160$  and  $p = P[\text{person does not show up}] = .05$ . If there is to be a seat available for every person planning to fly, then there must be at least five persons not showing up. Hence, the probability of interest is  $P(x \geq 5)$ . Calculate

$$\mu = np = 160(.05) = 8 \quad \text{and} \quad \sigma = \sqrt{npq} = \sqrt{160(.05)(.95)} = \sqrt{7.6} = 2.7$$

Referring to the figure that follows, a correction for continuity is made to include the entire area under the rectangle associated with the value  $x = 5$ , and the approximation becomes  $P(x \geq 4.5)$ . The  $z$ -value

corresponding to  $x = 4.5$  is  $z = \frac{x - \mu}{\sigma} = \frac{4.5 - 8}{\sqrt{7.6}} = -1.27$  so that

$$P(x \geq 4.5) = P(z \geq -1.27) = 1 - .1020 = .8980$$



- 6.3.20** Define  $x$  to be the number of children with defects. Then  $p = .001$  and  $n = 50,000$ . Calculate

$$\mu = np = 50,000(.001) = 50 \quad \text{and} \quad \sigma = \sqrt{npq} = \sqrt{50,000(.001)(.999)} = 7.068.$$

Since  $n$  is large, with  $np = 50$  and  $nq = 49,950$  both greater than five, the normal approximation to the binomial can be used.

- a** The probability  $P(x \geq 60)$  is approximated by the area under the appropriate normal curve to the right of 59.5. The  $z$ -value corresponding to  $x = 59.5$  is  $z = \frac{59.5 - 50}{7.068} = 1.34$  and

$$P(x \geq 60) \approx P(z > 1.34) = 1 - .9099 = .0901$$

- b** Since this probability is not too small (it is greater than .05), we would not say that observing 60 or more defects is rare event.

- 6.3.21** Define  $x$  to be the number of guests claiming a reservation at the motel. Then  $p = P[\text{guest claims reservation}] = 1 - .1 = .9$  and  $n = 215$ . The motel has only 200 rooms. Hence, if  $x > 200$ , a guest will not receive a room. The probability of interest is then  $P(x \leq 200)$ . Using the normal approximation, calculate

$$\mu = np = 215(.9) = 193.5 \quad \text{and} \quad \sigma = \sqrt{npq} = \sqrt{215(.9)(.1)} = \sqrt{19.35} = 4.399$$

The probability  $P(x \leq 200)$  is approximated by the area under the appropriate normal curve to the left of 200.5. The  $z$ -value corresponding to  $x = 200.5$  is  $z = \frac{200.5 - 193.5}{\sqrt{19.35}} = 1.59$  and

$$P(x \leq 200) \approx P(z < 1.59) = .9441$$

- 6.3.22** Define  $x$  to be the number of workers with identifiable lung cancer. If the rate of lung cancer in the population of workers in the air-polluted environment is the same as the population in general, then  $p = 1/40 = .025$  and the random variable  $x$  has a binomial distribution with  $n = 400$  and  $p = .025$ . Calculate

$$\mu = np = 400(.025) = 10 \text{ and } \sigma = \sqrt{400(.025)(.975)} = \sqrt{9.75} = 3.122$$

If we want to show that the rate of lung cancer in the polluted environment is greater than the general population rate, we need to show that the number of cancer victims in the sample from the polluted environment is unusually high. Hence, we calculate the  $z$ -score associated with  $x = 19$  as

$$z = \frac{x - \mu}{\sigma} = \frac{19 - 10}{3.122} = 2.882$$

which is quite large. This would imply that the value  $x = 19$  is unusually large and would indicate that in fact  $p$  is greater than  $1/40$  for the workers in the polluted environment.

- 6.3.23** Define  $x$  to be the number of elections in which the taller candidate won. If Americans are not biased by height, then the random variable  $x$  has a binomial distribution with  $n = 53$  and  $p = .5$ . Calculate

$$\mu = np = 53(.5) = 26.5 \text{ and } \sigma = \sqrt{53(.5)(.5)} = \sqrt{13.25} = 3.640$$

- a** Using the normal approximation with correction for continuity, we find the area to the right of  $x = 27.5$  :

$$P(x > 27.5) = P\left(z > \frac{27.5 - 26.5}{3.640}\right) = P(z > 0.27) = 1 - .6064 = .3936$$

- b** Since the occurrence of 28 out of 53 taller choices is not unusual, based on the results of part **a**, there is insufficient evidence to conclude that Americans consider height when casting a vote for a candidate.

- 6.3.24** Define  $x$  to be the number of people with Rh negative blood. Then the random variable  $x$  has a binomial distribution with  $n = 92$  and  $p = .15$ . Calculate

$$\mu = np = 92(.15) = 13.8 \text{ and } \sigma = \sqrt{92(.15)(.85)} = \sqrt{11.73} = 3.4249$$

- a** Using the normal approximation with correction for continuity, we find the area to the left of  $x = 10.5$  :

$$P(x < 10.5) = P\left(z > \frac{10.5 - 13.8}{3.4249}\right) = P(z < -.96) = .1685$$

- b** Similar to part **a**, finding the area between  $x = 14.5$  and  $x = 20.5$  :

$$P(14.5 < x < 20.5) = P(.20 < z < 1.96) = .9750 - .5793 = .3957$$

- c** The probability that more than 80 donors are Rh positive is the same as the probability that less than 12 donors are Rh negative, approximated with the area to the left of  $x = 11.5$  :

$$P(x < 11.5) = P(z < -.67) = .2514$$

- 6.3.25** Define  $x$  to be the number of consumers who preferred a *Coke* product. Then the random variable  $x$  has a binomial distribution with  $n = 500$  and  $p = .36$ , if *Coke*'s market share is indeed 36%. Calculate

$$\mu = np = 500(.36) = 180 \text{ and } \sigma = \sqrt{500(.36)(.64)} = \sqrt{115.2} = 10.733$$

- a** Using the normal approximation with correction for continuity, we find the area between  $x = 199.5$  and  $x = 200.5$  :

$$P(199.5 < x < 200.5) = P\left(\frac{199.5-180}{10.733} < z < \frac{200.5-180}{10.733}\right) = P(1.82 < z < 1.91) = .9719 - .9656 = .0063$$

**b** Find the area between  $x = 174.5$  and  $x = 200.5$ :

$$P(174.5 < x < 200.5) = P\left(\frac{174.5-180}{10.733} < z < \frac{200.5-180}{10.733}\right) = P(-.51 < z < 1.91) = .9719 - .3050 = .6669$$

**c** Find the area to the left of  $x = 199.5$ :

$$P(x < 199.5) = P\left(z < \frac{199.5-180}{10.733}\right) = P(z < 1.82) = .9656$$

**d** The value  $x = 232$  lies  $z = \frac{232-180}{10.733} = 4.84$  standard deviations above the mean, if *Coke's* market share is indeed 36%. This is such an unusual occurrence that we would conclude that *Coke's* market share is higher than claimed.

**6.3.26 a** Let  $x$  be the number of plants with red petals. Then  $x$  has a binomial distribution with  $n = 100$  and  $p = .75$ .

**b** Since  $np = 100(.75) = 75$  and  $nq = 100(.25)$  are both greater than 5, the normal approximation is appropriate.

**c** Calculate  $\mu = np = 100(.75) = 75$  and  $\sigma = \sqrt{npq} = \sqrt{100(.75)(.25)} = \sqrt{18.75} = 4.33$ . A correction for continuity is made to include the entire area under the rectangles corresponding to  $x = 70$  and  $x = 80$ . Hence the approximation will be

$$P(69.5 < x < 80.5) = P\left(\frac{69.5-75}{4.33} < z < \frac{80.5-75}{4.33}\right) = P(-1.27 < z < 1.27) = .8980 - .1020 = .7960$$

**d** The probability that 53 or fewer plants have red flower is approximated as

$$P(x < 53.5) = P\left(z < \frac{53.5-75}{4.33}\right) = P(z < -4.97) \approx 0$$

This would be considered an unusual event.

**e** If the value  $p = .75$  is correct, the only explanation for the unusual occurrence in part **d** is that the  $n = 100$  seeds do not represent a random sample from the population of peony plants. Perhaps the sample became contaminated in some way; some other uncontrolled variable is affecting the flower color.

**6.3.27** It is given that the probability of a successful single transplant from the early gastrula stage is .65. In a sample of 100 transplants, the mean and standard deviation of the binomial distribution are

$$\mu = np = 100(.65) = 65 \quad \text{and} \quad \sigma = \sqrt{npq} = \sqrt{100(.65)(.35)} = 4.770$$

It is necessary to find the probability that more than 70 transplants will be successful. This is approximated by the area under a normal curve with  $\mu = 65$  and  $\sigma = 4.77$  to the right of 70.5. The  $z$ -value corresponding

$$\text{to } x = 70.5 \text{ is } z = \frac{x - \mu}{\sigma} = \frac{70.5 - 65}{4.77} = 1.15 \quad \text{and} \quad P(x > 70) \approx P(z > 1.15) = 1 - .8749 = .1251$$

**6.3.28** Define  $x$  to be the number of Americans who feel pressure to work overtime. Then the random variable  $x$  has a binomial distribution with  $n = 25$  and  $p = .6$ . Calculate

$$\mu = np = 25(.6) = 15 \quad \text{and} \quad \sigma = \sqrt{npq} = \sqrt{25(.6)(.4)} = \sqrt{6} = 2.449$$

**a** From Table 1 in Appendix I, the exact probability that  $x$  is more than 20 is

$$P(x > 20) = 1 - P(x \leq 20) = 1 - .991 = .009.$$

**b** Using the normal approximation with correction for continuity, we find the area to the right of  $x = 20.5$  :

$$P(x > 20.5) = P\left(z > \frac{20.5 - 15}{2.449}\right) = P(z > 2.25) = 1 - .9878 = .0122$$

**6.3.29** Define  $x$  to be the number of American shoppers who use their smartphones to search for coupons. Then the random variable  $x$  has a binomial distribution with  $n = 50$  and  $p = .49$  .

**a-b** Calculate  $\mu = np = 50(.49) = 24.5$  and  $\sigma = \sqrt{50(.49)(.51)} = \sqrt{12.495} = 3.535$

**c** Using the normal approximation with correction for continuity to approximate the probability that  $x$  is less than or equal to 15, we find the area to the right of  $x = 15.5$  :

$$P(x < 15.5) = P\left(z < \frac{15.5 - 24.5}{3.535}\right) = P(z < -2.55) = .0054$$

This would be considered a very unusual occurrence.

## Reviewing What You've Learned

**6.R.1** It is given that  $x$  has an exponential distribution with  $\lambda = 1/5 = .2$ .

$$\mathbf{a} \quad P(x < 4) = 1 - e^{-.2(4)} = .5507 \quad \mathbf{b} \quad P(x > 6) = e^{-.2(6)} = .3012$$

**c** If there are two such switches that operate independently,

$$P(\text{neither switch fails before year 8}) = [P(x > 8)]^2 = [e^{-.2(8)}]^2 = .0408$$

**6.R.2** **a** It is given that your arrival time  $x$  has a uniform distribution from  $a = 0$  to  $b = 30$ , so that

$$f(x) = \frac{1}{30} \text{ for } 0 \leq x \leq 30.$$

**b** You will wait less than 10 minutes as long as you arrive within 10 minutes of the next showing—that is, 10 minutes before 7:15 ( $5 < x < 15$ ) or 10 minutes before 7:30 ( $20 < x < 30$ ). The probability is calculated as

$$P(5 < x < 15) + P(20 < x < 30) = \frac{15-5}{30} + \frac{30-20}{30} = \frac{2}{3}$$

**c** The probability that you wait more than 10 minutes until a movie starts is the complement of the event in part **b** or

$$P(0 < x < 5) + P(15 < x < 20) = \frac{5-0}{30} + \frac{20-15}{30} = \frac{1}{3}$$

**6.R.3** It is given that  $x$  is approximately normally distributed with  $\mu = 75$  and  $\sigma = 12$  .

$$\mathbf{a} \quad \text{Calculate } z = \frac{x - \mu}{\sigma} = \frac{60 - 75}{12} = -1.25. \text{ Then } P(x < 60) = P(z < -1.25) = .1056$$

$$\mathbf{b} \quad P(x > 60) = 1 - P(x < 60) = 1 - .1056 = .8944$$

- c If the bit is replaced after more than 90 hours, then  $x > 90$ . Calculate  $z = \frac{x - \mu}{\sigma} = \frac{90 - 75}{12} = 1.25$ . Then  $P(x > 90) = P(z > 1.25) = 1 - .8944 = .1056$ .

**6.R.4** The random variable  $x$  is approximately normally distributed with  $\mu = 1230$  and  $\sigma = 120$ .

- a The  $z$ -value corresponding to  $x = 1400$  is  $z = \frac{x - \mu}{\sigma} = \frac{1400 - 1230}{120} = 1.42$  and

$$P(x > 1400) = P(z > 1.42) = 1 - .9222 = .0778.$$

- b The  $z$ -value corresponding to  $x = 1000$  is  $z = \frac{x - \mu}{\sigma} = \frac{1000 - 1230}{120} = -1.92$  and

$$P[\text{restaurant does not break even}] = P(x < 300) = P(z < -1.92) = .0274.$$

**6.R.5** It is given that  $x$  is normally distributed with  $\mu = 30$  and  $\sigma = 11$ . The probability of interest is

$$P(x > 50) = P\left(z > \frac{50 - 30}{11}\right) = P(z > 1.82) = 1 - .9656 = .0344$$

**6.R.6** It is given that  $\mu = 1.4$  and  $\sigma = .7$ . If we assume that  $x$ , the service time for one vehicle, is normally distributed, the probability of interest is

$$P(x > 1.6) = P\left(z > \frac{1.6 - 1.4}{.7}\right) = P(z > .29) = 1 - .6141 = .3859$$

**6.R.7** For the binomial random variable  $x$ , the mean and standard deviation are calculated under the assumption that the advertiser's claim is correct and  $p = .2$ . Then

$$\mu = np = 1000(.2) = 200 \text{ and } \sigma = \sqrt{npq} = \sqrt{1000(.2)(.8)} = 12.649$$

If the advertiser's claim is correct, the  $z$ -score for the observed value of  $x$ ,  $x = 184$ , is

$$z = \frac{x - \mu}{\sigma} = \frac{184 - 200}{12.649} = -1.26$$

That is, the observed value lies 1.26 standard deviations below the mean. This is not an unlikely occurrence. Hence, we would have no reason to doubt the advertiser's claim.

**6.R.8** For the binomial random variable  $x$ , the mean and standard deviation are calculated under the assumption that there is no difference between the effect of TV and reading on calorie intake, and hence that  $p = .5$ .

Then  $\mu = np = 30(.5) = 15$  and  $\sigma = \sqrt{npq} = \sqrt{30(.5)(.5)} = 2.7386$ . If there is no difference between TV and reading, the  $z$ -score for the observed value of  $x$ ,  $x = 19$ , is

$$z = \frac{x - \mu}{\sigma} = \frac{19 - 15}{2.7386} = 1.461$$

That is, the observed value lies 1.461 standard deviations above the mean. This is not an unlikely occurrence. Hence, we would have no reason to believe that there is a difference between calorie intake for TV watchers versus readers.

**6.R.9** Define  $x$  to be the percentage of returns audited for a particular state. It is given that  $x$  is normally distributed with  $\mu = 1.55$  and  $\sigma = .45$ .

- a  $P(x > 2.5) = P\left(z > \frac{2.5 - 1.55}{.45}\right) = P(z > 2.11) = 1 - .9826 = .0174$

$$\mathbf{b} \quad P(x < 1) = P\left(z < \frac{1-1.55}{.45}\right) = P(z < -1.22) = .1112$$

**6.R.10** The random variable  $y$ , the percentage of tax returns audited, has a normal distribution with  $\mu = 1.55$  and  $\sigma = .45$ .

$$\mathbf{a} \quad P(y > 2) = P\left(z > \frac{2-1.55}{.45}\right) = P(z > 1) = 1 - .8413 = .1587$$

**b** Define  $x$  to be the number of states in which more than 2% of its returns were audited. Then  $x$  has a binomial distribution with  $\mu = np = 50(.1587) = 7.935$  and  $\sigma = \sqrt{npq} = \sqrt{50(.1587)(.8413)} = 2.583$ .

The expected value of  $x$  is  $E(x) = \mu = 7.935$ .

**c** The value  $x = 15$  has  $z$ -score  $z = \frac{x - \mu}{\sigma} = \frac{15 - 7.935}{2.583} = 2.73$ . It is unlikely that as many as 15 states will have more than 2% of its returns audited.

**6.R.11** Define  $x$  to be the number of men who have fished in the last year. Then  $x$  has a binomial distribution with  $n = 180$  and  $p = .41$ . Calculate

$$\mu = np = 180(.41) = 73.8 \quad \text{and} \quad \sigma = \sqrt{npq} = \sqrt{180(.41)(.59)} = 6.5986$$

**a** Using the normal approximation with correction for continuity,

$$P(x < 50) \approx P\left(z < \frac{49.5 - 73.8}{6.5986}\right) = P(z < -3.68) \approx 0$$

$$\mathbf{b} \quad P(50 \leq x \leq 75) \approx P\left(\frac{49.5 - 73.8}{6.5986} \leq z \leq \frac{75.5 - 73.8}{6.5986}\right) = P(-3.68 \leq z \leq .26) = .6026 - .0000 = .6026$$

**c** The sample is not random, since mailing lists for a sporting goods company will probably contain more fishermen than the population in general. Since the sampling was not random, the survey results are not reliable.

**6.R.12** Assume that the temperatures of healthy humans is approximately normal with  $\mu = 98.6$  and  $\sigma = 0.8$ .

$$\mathbf{a} \quad P(x > 99.0) = P\left(z > \frac{99 - 98.6}{0.8}\right) = P(z > .5) = 1 - .6915 = .3085$$

**b** From Exercise 26 (Section 6.2), we found that the 95<sup>th</sup> percentile of the *standard* normal ( $z$ ) distribution is  $z = 1.645$ . Since  $z = \frac{x - \mu}{\sigma} = \frac{x - 98.6}{0.8}$ , solve for  $x$  to find the 95<sup>th</sup> percentile for the temperatures:

$$1.645 = \frac{x - 98.6}{0.8} \Rightarrow x = 98.6 + 1.645(0.8) = 99.92 \text{ degrees.}$$

**6.R.13 a** It is given that the scores on a national achievement test were approximately normally distributed with a mean of 540 and standard deviation of 110. It is necessary to determine how far, in standard deviations, a score of 680 departs from the mean of 540. Calculate

$$z = \frac{x - \mu}{\sigma} = \frac{680 - 540}{110} = 1.27.$$

**b** To find the percentage of people who scored higher than 680, we find the area under the standardized normal curve greater than 1.27. Using Table 3, this area is equal to

$$P(x > 680) = P(z > 1.27) = 1 - .8980 = .1020$$

Thus, approximately 10.2% of the people who took the test scored higher than 680.

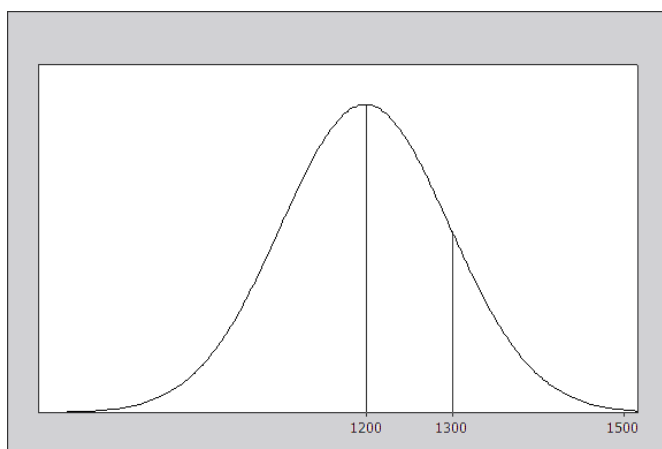
**6.R.14** Define  $x$  to be the salaries of assistant professors at public 4-year colleges during the 2016 academic year. It is given that  $x$  is normally distributed with  $\mu = 83,398$  and  $\sigma = 4000$ .

$$\text{a } P(x < 75,000) = P\left(z < \frac{75,000 - 83,398}{4000}\right) = P(z < -2.10) = .0179$$

$$\begin{aligned} \text{b } P(75,000 < x < 85,000) &= P\left(\frac{75,000 - 83,398}{4000} < z < \frac{85,000 - 83,398}{4000}\right) \\ &= P(-2.10 < z < .40) = .6554 - .0179 = .6375 \end{aligned}$$

### On Your Own

**6.R.15** The random variable  $x$ , total weight of 8 people, has a mean of  $\mu = 1200$  and a standard deviation  $\sigma = 99$ . It is necessary to find  $P(x > 1300)$  and  $P(x > 1500)$  if the distribution of  $x$  is approximately normal. Refer to the figure that follows.



The  $z$ -value corresponding to  $x_1 = 1300$  is  $z_1 = \frac{x_1 - \mu}{\sigma} = \frac{1300 - 1200}{99} = 1.01$ . Hence,

$$P(x > 1300) = P(z > 1.01) = 1 - A(1.01) = 1 - .8438 = .1562.$$

Similarly, the  $z$ -value corresponding to  $x_2 = 1500$  is  $z_2 = \frac{x_2 - \mu}{\sigma} = \frac{1500 - 1200}{99} = 3.03$ .

$$\text{and } P(x > 1500) = P(z > 3.03) = 1 - A(3.03) = 1 - .9988 = .0012.$$

**6.R.16** It is given that  $x$  is normally distributed with  $\mu$  unknown and  $\sigma = 25.7$ . It is necessary to have  $P(x > 2000) = .01$ . Calculate

$$z = \frac{2000 - \mu}{25.7} \quad \text{and} \quad P(x > 2000) = P\left(z > \frac{2000 - \mu}{25.7}\right) = .01$$

Since the value  $(2000 - \mu)/25.7$  is a constant (although its value is unknown), it can be treated as such in Table 3. It is necessary to find  $z_0$  such that  $P(z > z_0) = .01$  or  $A(z_0) = .9900$ . From Table 3,  $z_0 = 2.33$ , so that

$$\frac{2000 - \mu}{25.7} = 2.33 \quad \text{and} \quad \mu = 1940.119$$

**6.R.17** Let  $w$  be the number of words specified in the contract. Then  $x$ , the number of words in the manuscript, is normally distributed with  $\mu = w + 20,000$  and  $\sigma = 10,000$ . The publisher would like to specify  $w$  so that

$$P(x < 100,000) = .95.$$

As in Exercise 16, calculate 
$$z = \frac{100,000 - (w + 20,000)}{10,000} = \frac{80,000 - w}{10,000}.$$

Then  $P(x < 100,000) = P\left(z < \frac{80,000 - w}{10,000}\right) = .95$ . It is necessary that  $z_0 = (80,000 - w)/10,000$  be such that

$$P(z < z_0) = .95 \Rightarrow A(z_0) = .9500 \text{ or } z_0 = 1.645.$$

Hence,

$$\frac{80,000 - w}{10,000} = 1.645 \text{ or } w = 63,550.$$

**6.R.18 a** No. The errors in forecasting would tend to be skewed, with more underestimates than overestimates.

**b** Let  $x$  be the number of estimates that are in error by more than 15%. Then  $x$  has a binomial distribution with  $n = 100$  and  $p = P[\text{estimate is in error by more than 15\%}] = .5$ . Using the normal approximation to the binomial distribution,

$$P(x > 60) \approx P\left(z > \frac{60.5 - 100(.5)}{\sqrt{100(.5)(.5)}}\right) = P(z > 2.1) = 1 - .9821 = .0179$$

**6.R.19** The random variable  $x$  is the size of the freshman class. That is, the admissions office will send letters of acceptance to (or accept deposits from) a certain number of qualified students. Of these students, a certain number will actually enter the freshman class. Since the experiment results in one of two outcomes (enter or not enter), the random variable  $x$ , the number of students entering the freshman class, has a binomial distribution with

$n$  = number of deposits accepted and

$$p = P[\text{student, having been accepted, enters freshman class}] = .8$$

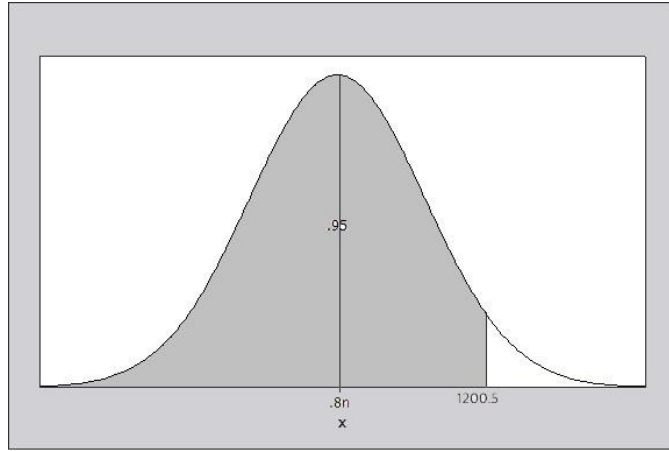
**a** It is necessary to find a value for  $n$  such that  $P(x \leq 1200) = .95$ . Note that,

$$\mu = np = .8n \text{ and } \sigma = \sqrt{npq} = \sqrt{.16n}$$

Using the normal approximation, we need to find a value of  $n$  such that  $P(x \leq 1200) = .95$ , as shown in the next figure. The  $z$ -value corresponding to  $x = 1200.5$  is

$$z = \frac{x - \mu}{\sigma} = \frac{1200.5 - .8n}{\sqrt{.16n}}$$





From Table 3, the  $z$ -value corresponding to an area of .05 in the right tail of the normal distribution is 1.645. Then,

$$\frac{1200.5 - .8n}{\sqrt{.16n}} = 1.645$$

Solving for  $n$  in the above equation, we obtain the following quadratic equation:

$$.8n + .658\sqrt{n} - 1200.5 = 0$$

Let  $x = \sqrt{n}$ . Then the equation takes the form  $ax^2 + bx + c = 0$  which can be solved using the quadratic

formula,  $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ , or

$$x = \frac{-.658 \pm \sqrt{.432964 + 4(960.4)}}{1.6} = \frac{-.658 \pm 61.9841}{2}$$

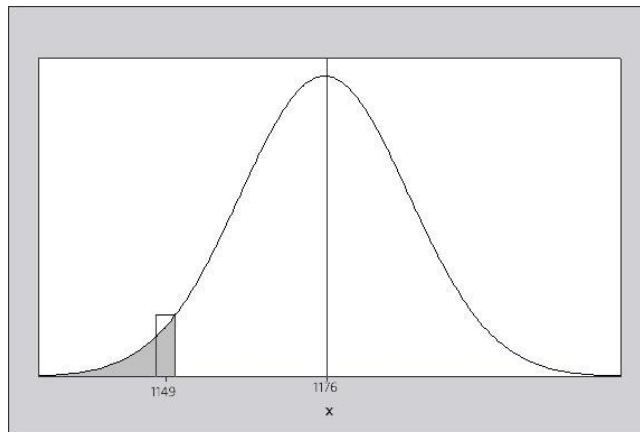
Since  $x$  must be positive, the desired root is

$$x = \sqrt{n} = \frac{61.3261}{1.6} = 38.329 \text{ or } n = (38.329)^2 = 1469.1.$$

Thus, 1470 deposits should be accepted.

**b** Once  $n = 1470$  has been determined, the mean and standard deviation of the distribution are

$$\mu = np = 1470(.8) = 1176 \quad \text{and} \quad \sigma = \sqrt{npq} = \sqrt{235.2} = 15.3362.$$



Then the approximation for  $P(x < 1150) = P(x \leq 1149)$ , shown above, is

$$P(x \leq 1149.5) = P\left(z \leq \frac{1149.5 - 1176}{15.3362}\right) = P(z \leq -1.73) = .0418$$

- 6.R.20 a** The measurements are approximately normal with mean  $\mu = 39.83$  and standard deviation  $\sigma = 2.05$ . To find the 95<sup>th</sup> percentile, we need to find a value of  $x$ , say  $x = c$ , such that  $P(x > c) = .05$ . The  $z$ -value corresponding to  $x = c$  is

$$z = \frac{x - \mu}{\sigma} = \frac{c - 39.83}{2.05}$$

From Table 3, the  $z$ -value corresponding to an area of .05 in the right tail of the normal distribution is a value that has area .9500 to its left. Using linear interpolation in the table, we choose  $z = 1.645$ . Then,

$$\frac{c - 39.83}{2.05} = 1.645 \Rightarrow c = 39.83 + 1.645(2.05) = 43.202$$

The 95<sup>th</sup> percentile of the chest measurements is 43.202.

- b** The empirical estimate of the 95<sup>th</sup> percentile, based on the 5738 chest measurements, would be a value of  $x$  that is greater than approximately 95% of the measurements or less than approximately 5% of the measurements. To find the position of the 95<sup>th</sup> percentile, use a formula similar to the quartile positions in Chapter 2. Calculate  $.95(n + 1) = .95(5739) = 5452.05$  and look in the frequency distribution for a value of  $x$  which is in position 5452.05 from the lower tail of the distribution. Since there are 5738 measurements, this would be equivalent to looking in the frequency distribution for a value of  $x$  which is in position  $5738 - 5452.05 = 285.95$  from the upper tail of the distribution. The table that follows shows the cumulative counts in the upper tail of the distribution. Position “285.95” falls in the group of soldiers having chest size 43, close to the value calculated in part a.

Number of measurements $\geq x$	1	5	26	76	168	538	1196	2130
Chest size, $x$	48	47	46	45	44	43	42	41

- c** To find the 90<sup>th</sup> percentile, we need to find a value of  $x$ , say  $x = c$ , such that  $P(x > c) = .10$ . The  $z$ -value corresponding to  $x = c$  is

$$z = \frac{x - \mu}{\sigma} = \frac{c - 39.83}{2.05}$$

From Table 3, the  $z$ -value corresponding to an area of .10 in the right tail of the normal distribution is a value that has area .9000 to its left. The closest value in the table is  $z = 1.28$ . Then,

$$\frac{c - 39.83}{2.05} = 1.28 \Rightarrow c = 39.83 + 1.28(2.05) = 42.454$$

The 90<sup>th</sup> percentile of the chest measurements is 42.454.

The empirical estimate of the 90<sup>th</sup> percentile, based on the 5738 chest measurements, would be a value of  $x$  that is greater than approximately 90% of the measurements or less than approximately 10% of the measurements. As in part b, we find the position of the 90<sup>th</sup> percentile by calculating  $.90(n + 1) = .90(5739) = 5165.1$ , so that the 90<sup>th</sup> percentile is a value of  $x$  which is in position 5165.1 from the lower tail of the distribution. Since there are 5738 measurements, this would be equivalent to looking in the frequency distribution for a value of  $x$  which is in position  $5738 - 5165.1 = 572.9$  from the upper tail of the distribution. Using the table given in part a, position “572.9” falls in the group of soldiers having chest size 42, close to the value calculated above.

- 6.R.21** The measurements are approximately normal with mean  $\mu = 39.83$  and standard deviation  $\sigma = 2.05$ .

a

$$P(36.5 < x < 43.5) = P\left(\frac{36.5 - 39.83}{2.05} < z < \frac{43.5 - 39.83}{2.05}\right) = P(-1.62 < z < 1.79) = .9633 - .0526 = .9107$$

b From Exercise 24 (Section 6.2), we know that  $P(-1.96 < z < 1.96) = .95$  for the standard normal  $z$ . Substituting for  $z$  in the probability statement, we write

$$P(-1.96 < \frac{x - \mu}{\sigma} < 1.96) \Rightarrow P(\mu - 1.96\sigma < x < \mu + 1.96\sigma) = .95$$

That is, 95% of the chest measurements,  $x$ , will lie between  $\mu - 1.96\sigma$  and  $\mu + 1.96\sigma$ , or

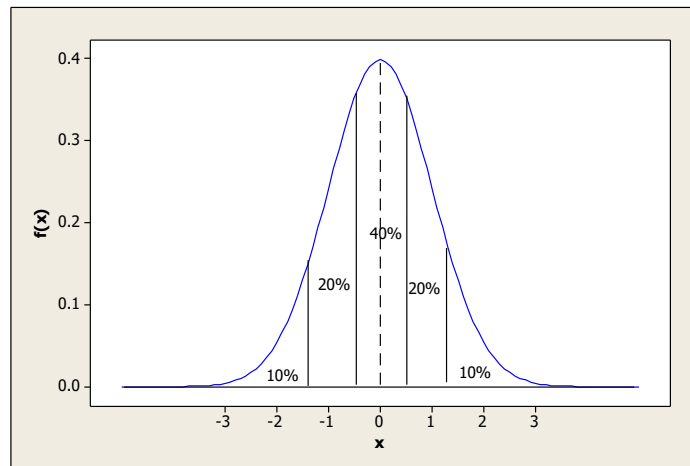
$$39.83 \pm 1.96(2.05) \text{ which is } 35.812 \text{ to } 43.848.$$

c Refer to the data distribution given in Exercise 20, and count the number of measurements falling in each of the two intervals.

- Between 36.5 and 43.5:  $\frac{420 + 749 + \dots + 370}{5738} = \frac{5283}{5738} = .921$  (compared to .9107)
- Between 35.812 and 43.848:  $\frac{185 + 420 + \dots + 370}{5738} = \frac{5468}{5738} = .953$  (compared to .95)

## Case Study: “Are You Going to Curve the Grades?”

In order to implement the traditional interpretation of “curving the grades”, the proportions shown in the table need to be applied to the normal curve, as shown in the figure that follows.



- The C grades constitute the middle 40%, that is, 20% on either side of the mean. The lower boundary has area .3000 to its left. From Table 3, we need to find a value of  $z$  such that  $A(z) = .3000$ . The closest value in the table is .3015 with  $z = -.52$ . The upper boundary is then  $z = +.52$ .
- The cutoff for the lowest D and highest B grades constitute the lower and upper boundaries of the middle 80%, that is, 40% on either side of the mean. The lower boundary has area .1000 to its left, so we need to find a value of  $z$  such that  $A(z) = .1000$ . The closest value in the table is .1003 with  $z = -1.28$ . The upper boundary is then  $z = 1.28$ .
- If the grades of this large class have a mean of 79.972 and a standard deviation of 12.271, the appropriate cutoffs are  $\pm 0.5$  standard deviations and  $\pm 1.5$  standard deviations from the mean. Calculate

$$79.972 \pm 0.5(12.271) \Rightarrow 79.972 \pm 6.1355 \text{ or } 73.8365 \text{ and } 86.1075$$

$$79.972 \pm 1.5(12.271) \Rightarrow 79.972 \pm 18.4065 \text{ or } 61.5655 \text{ and } 98.3785$$

The cutoff points are shown in the table that follows.

Letter grade	A	B	C	D	F
Score	98.38 and above	86.11 to 98.38	73.84 to 86.11	61.57 to 73.84	below 61.57

4. In the “curving” scheme, the student with a score of 92 would receive a B rather than an A on an absolute standard scale. He or she would definitely not benefit by “curving the grades”. Curving grades will only serve its designed purpose when the data is approximately mound-shaped, in order to achieve the desired percentages. Since this data is clearly skewed to the left, the “curving” procedure does not have the desired effect for most students; only a few students (those who score between 60 and 61) will benefit when the grades are “curved”.

## 7: Sampling Distributions

### Section 7.1

**7.1.1** The experimental units are each assigned two random numbers (001 and 501 to unit 1, 002 and 502 to unit 2, ..., 499 and 999 to unit 499, and 000 and 500 to unit 500). You can select a simple random sample of size  $n = 20$  using Table 10 in Appendix I. First choose a starting point and consider the first three digits in each number. The three digits OR the (three digits – 500) will identify the proper experimental unit. For example, if the three digits are 742, you should select the experimental unit numbered  $742 - 500 = 242$ . The probability that any three-digit number is selected is  $2/1000 = 1/500$ . One possible selection for the sample size  $n = 20$  is

242	134	173	128	399
056	412	188	255	388
469	244	332	439	101
399	156	028	238	231

**7.1.2** Students may choose different sampling schemes, and will obtain different samples, using Table 10 in Appendix I.

**7.1.3** Students may choose different sampling schemes, and will obtain different samples, using Table 10 in Appendix I.

**7.1.4** Students may choose different sampling schemes, and will obtain different samples, using Table 10 in Appendix I.

**7.1.5** Each home represents a cluster within which each occupant is surveyed. The ten homes represent a *cluster sample*.

**7.1.6** This is a *systematic* 1-in- $k$  sample with  $k = 100$ .

**7.1.7** A random sample is selected from each of six departments. This is a *stratified sample* with the departments representing the strata.

**7.1.8** The doctor is judging that his “at risk” patients will benefit the most by participating in the clinical trial. This is a *judgement sample*.

**7.1.9** Since the student procrastinated until the last moment to complete the survey, he chooses the most convenient people (his fraternity brothers) to participate. This is a *convenience sample*.

**7.1.10** The researcher is specifically choosing men and women in the same proportions as they appear in the population. This is a *quota sample*.

**7.1.11** The professor would like to be certain that he has a high return rate on his questionnaire, so he chooses his twenty most reliable students. This is a *convenience sample*. Some students may decide that this is a *judgement sample*, since the professor is using his best judgement in the selection of the students to receive the questionnaire.

**7.1.12** Since each cluster (a city block) is censused, this is an example of cluster sampling.

**7.1.13** This is a 1-in-10 systematic sample.

**7.1.14** The wards are the strata, and the sample is a stratified sample.

**7.1.15** This is a 1-in-10 systematic sample.

**7.1.16** This is a simple random sample from the population of all tax returns filed in the city of San Bernardino, California.

**7.1.17** The survey responses do not represent a representative sample from the total number of surveys that were mailed out. Since only 30% of the surveys were returned, there is a problem of *nonresponse* and the 90%

figure may indicate that only those people who were in favor of the proposed zoning changes bothered to return their survey.

- 7.1.18** The question is worded in such a way to be biased in favor of a “yes” response. There is *wording bias* in the question.
- 7.1.19** Only readers of your Facebook page will respond to the survey. There is a problem with *undercoverage* and the responses will not be representative of the entire population.
- 7.1.20** If all of the town citizenry is likely to pass this corner, a sample obtained by selecting every tenth person is probably a fairly random sample.
- 7.1.21** The questionnaires that were returned do not constitute a representative sample from the 1000 questionnaires that were randomly sent out. It may be that the voters who chose to return the questionnaire were particularly adamant about the Parks and Recreation surcharge, while the others had no strong feelings one way or the other. The nonresponse of half the voters in the sample will undoubtedly bias the resulting statistics.
- 7.1.22** Voter registration lists and DMV records will systematically exclude a large segment of the general population – nondrivers, people who do not own cars, nonvoters and so on.
- 7.1.23** The wording of the question is biased to suggest that a “yes” response is the correct one. A more unbiased way to phrase the question might be: “Is there too much sex and violence during prime TV viewing hours?”
- 7.1.24** Use a randomization scheme similar to that used in Exercise 1. Number each of the 100 individuals from 00 to 99. To choose the 50 individuals who will receive the experimental treatment, select 50 two-digit random numbers from Table 10. Each two-digit number will identify the proper experimental unit. The other 50 individuals will be placed in the control group.
- 7.1.25** **a** Since the question is particularly sensitive to people of different ethnic origins, you may find that the answers may not always be truthful, depending on the ethnicity of the interviewer and the person being interviewed.
- b** Notice that the percentage in favor of affirmative action increases as the ethnic origin of the interviewer changes from Caucasian to Asian to African-American. The people being interviewed may be changing their response to match what they perceive to be the response which the interviewer wants to hear.
- 7.1.26** Answers will vary from student to student. All three questions involve some sort of wording bias.
- 7.1.27** **a** Since each subject must be randomly assigned to either a tai chi class or a wellness education class with equal probability, assign the digits 0-4 to the tai chi treatment, and the digits 5-9 to the wellness treatment. As each subject enters the study, choose a random digit using Table 10 and assign the appropriate treatment.
- b** The randomization scheme in part **a** does not guarantee an equal number of subjects in each group.
- 7.1.28** **a** The first question is more unbiased.
- b** Notice that the percentage favoring the new space program drops dramatically when the phrase “spending billions of dollars” is added to the question.
- 7.1.29** Answers will vary from student to student. Paying cash for opinions will not necessarily produce a random sample of opinions of all Pepsi and Coke drinkers.
- 7.1.30** **a** To divide a group of 20 people into two groups of 10, use Table 10 in Appendix I. Assign an identification number from 01 to 20 to each person. Then select ten two-digit numbers from the random number table to identify the ten people in the first group. (If the number is greater than 20, subtract multiples of 20 from the random number until you obtain a number between 01 and 20.)
- b** Although it is not possible to select an actual random sample from this hypothetical population, the researcher must obtain a sample that *behaves like* a random sample. A large database of some sort should be used to ensure a fairly representative sample.

c The researcher has actually selected a *convenience sample*; however, it will probably behave like a simple random sample, since a person's enthusiasm for a paid job should not affect his response to this psychological experiment.

## Section 7.2

**7.2.1-2**  $C_4^6 = \frac{6!}{4!2!} = 15$  samples are possible. The 15 samples along with the sample means and medians for each are shown in the table that follows.

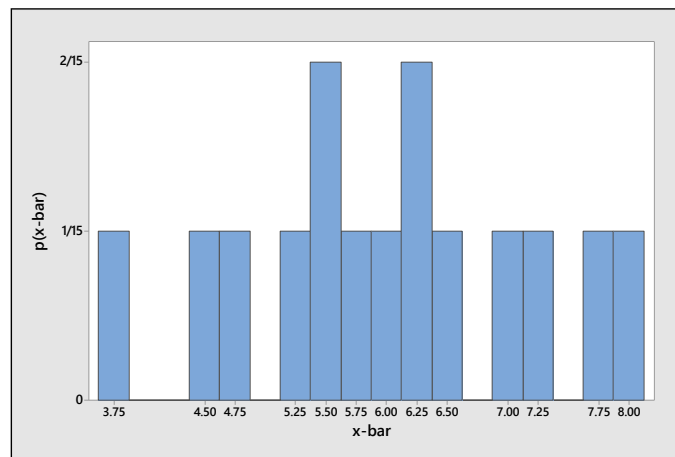
Sample	Observations	$\bar{x}$	$m$	Sample	Observations	$\bar{x}$	$m$
1	1, 3, 4, 7	3.75	3.5	9	1, 4, 10, 11	6.5	7
2	1, 3, 4, 10	4.5	3.5	10	3, 4, 7, 10	6	5.5
3	1, 3, 4, 11	4.75	3.5	11	3, 4, 7, 11	6.25	5.5
4	1, 3, 10, 11	6.25	6.5	12	3, 4, 10, 11	7	7
5	1, 3, 7, 10	5.25	5	13	3, 7, 10, 11	7.75	8.5
6	1, 3, 7, 11	5.5	5	14	4, 7, 10, 11	8	8.5
7	1, 4, 7, 10	5.5	5.5	15	1, 7, 10, 11	7.25	8.5
8	1, 4, 7, 11	5.75	5.5				

1. There are 15 possible samples which are equally likely (due to random sampling), each with probability  $1/15$ . However, there are only 13 distinct values of  $\bar{x}$ , since two values of  $\bar{x}$  (6.25 and 5.5) occur twice. Therefore, the sampling distribution of  $\bar{x}$  is given as

$$p(\bar{x}) = \frac{1}{15} \text{ for } \bar{x} = 3.75, 4.5, 4.75, 5.25, 5.75, 6, 6.5, 7, 7.25, 7.75, 8$$

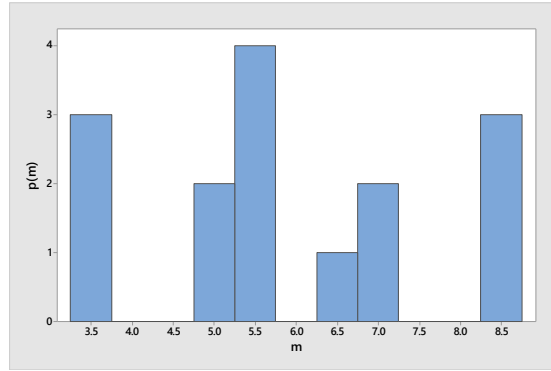
$$p(\bar{x}) = \frac{2}{15} \text{ for } \bar{x} = 5.5, 6.25$$

The sampling distribution is shown next.



2. There are 15 possible samples which are equally likely (due to random sampling), each with probability  $1/15$ . However, there are only 6 distinct values of the sample median  $m$ . The sampling distribution of the sample median is shown next.

$m$	$f(m)$
3.5	3/15
5	2/15
5.5	4/15
6.5	1/15
7	2/15
8.5	3/15



**7.2.3-6**  $C_3^5 = \frac{5!}{3!2!} = 10$  samples are possible. The 10 samples along with the sample mean, median, range and sample variance for each are shown next.

Sample	Observations	$\bar{x}$	$m$	$R$	$s^2$
1	11, 12, 15	12.67	12	4	4.33
2	11, 12, 18	13.67	12	7	14.33
3	11, 12, 20	14.33	12	9	24.33
4	11, 15, 18	14.67	15	7	12.33
5	11, 15, 20	15.33	15	9	20.33
6	11, 18, 20	16.33	18	9	22.33
7	12, 15, 18	15.00	15	6	9.00
8	12, 15, 20	15.67	15	8	16.33
9	12, 18, 20	16.67	18	8	17.33
10	15, 18, 20	17.67	18	5	6.33

**3.** There are 10 unique values for  $\bar{x}$ , each with probability 1/10 because of the random sampling. The sampling distribution of  $\bar{x}$  is:

$$p(\bar{x}) = \frac{1}{10} \text{ for } \bar{x} = 12.67, 13.67, 14.33, 14.67, 15, 15.33, 15.67, 16.33, 16.67, 17.67$$

**4.** There are 10 values for the median  $m$ , each with probability 1/10, but only three unique values. The sampling distribution of  $m$  is:

$m$	12	15	18
$p(m)$	3/10	4/10	3/10

**5.** There are 10 values for the range, each with probability 1/10, but only 6 unique values. The sampling distribution of  $R$  is

$R$	4	5	6	7	8	9
$p(R)$	1/10	1/10	1/10	2/10	2/10	3/10

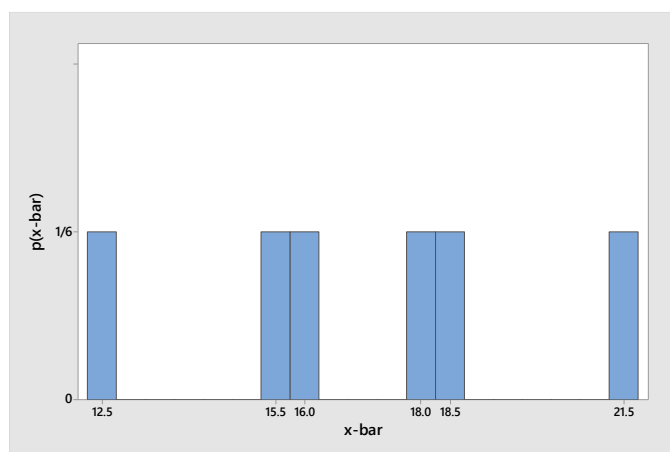
**6.** There are 10 values for the sample variance, all of which are unique, and each with probability 1/10. The sampling distribution of  $s^2$  is

$$p(s^2) = \frac{1}{10} \text{ for } s^2 = 4.33, 6.33, 9, 12.33, 14.33, 16.33, 17.33, 20.33, 22.33, 24.33$$

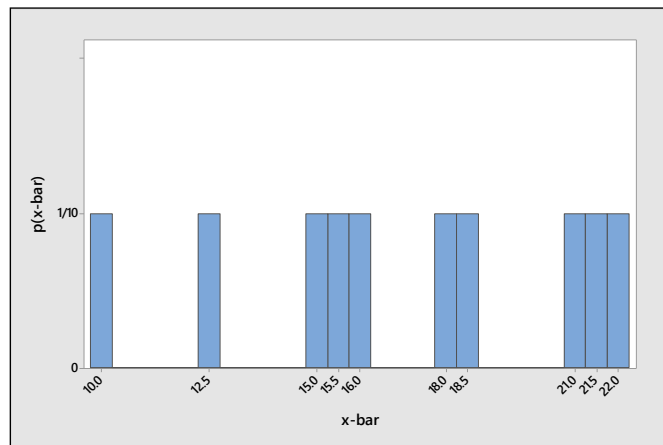
**7.2.7** There are  $C_2^4 = \frac{4!}{2!2!} = 6$  possible samples without replacement, shown as follows along with the value of  $\bar{x}$  for each. Since each value of  $\bar{x}$  is unique, each has probability 1/6.



Sample	Observations	$\bar{x}$
1	10, 15	12.5
2	10, 21	15.5
3	10, 22	16
4	15, 21	18
5	15, 22	18.5
6	21, 22	21.5



**7.2.8** Refer to Exercise 7. When sampling *with replacement*, there are four more possible samples—(10, 10), (15, 15), (21, 21) and (22, 22)—with means 10, 15, 21 and 22, respectively. Each value of  $\bar{x}$  now has probability 1/10, and the sampling distribution is shown next.



**7.2.9** Refer to Exercises 7 and 8. For Exercise 7, the mean of the 6 values of  $\bar{x}$  is

$$\mu_{\bar{x}} = \frac{12.5 + 15.5 + \dots + 21.5}{6} = \frac{102}{6} = 17 \quad \text{while for Exercise 8, the mean of the 10 values of } \bar{x} \text{ is}$$

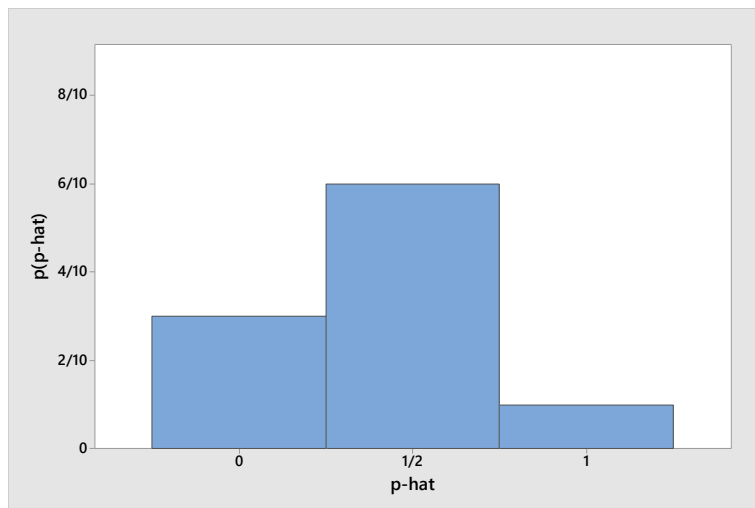
$$\mu_{\bar{x}} = \frac{12.5 + 15.5 + \dots + 21.5 + 10 + 15 + 21 + 22}{10} = \frac{170}{10} = 17. \quad \text{In both cases, the mean of the sampling distribution of } \bar{x} \text{ is the same as the mean } \mu = 17 \text{ of the original distribution.}$$

**7.2.10-11** The distribution is an example of a hypergeometric distribution with  $N = 5$  and  $n = 2$ . When sampling without replacement, there are  $C_2^5 = 10$  possible samples, listed next along with the proportion of successes in the sample.

Sample	Outcome	Proportion of Successes	Sample	Outcome	Proportion of Successes
1	$S_1 S_2$	1	6	$S_2 F_2$	1/2
2	$S_1 F_1$	1/2	7	$S_2 F_3$	1/2
3	$S_1 F_2$	1/2	8	$F_1 F_2$	0
4	$S_1 F_3$	1/2	9	$F_1 F_3$	0
5	$S_2 F_1$	1/2	10	$F_2 F_3$	0

**7.2.12** Refer to Exercise 10-11. The sampling distribution for the sample proportion  $\hat{p}$  is shown in the table and as a probability histogram that follow:

$\hat{p}$	$p(\hat{p})$
0	3/10
1/2	6/10
1	1/10



**7.2.13** Both of the sampling distributions are centered at the mean  $\mu = 5.5$ , but distribution 2 is much less variable. Distribution 2 would be the better plan to use in estimating the population mean.

### Section 7.3

**7.3.1-3** If the sampled populations are normal, the distribution of  $\bar{x}$  is also normal *for all values of  $n$* . The sampling distribution of the sample mean will have a mean  $\mu$  equal to the mean of the population from which we are sampling, and a standard deviation equal to  $\sigma/\sqrt{n}$ .

1.  $\mu = 10$ ;  $\sigma/\sqrt{n} = 3/\sqrt{36} = .5$
2.  $\mu = 5$ ;  $\sigma/\sqrt{n} = 2/\sqrt{100} = .2$
3.  $\mu = 120$ ;  $\sigma/\sqrt{n} = 1/\sqrt{8} = .3536$

**7.3.4-6** The Central Limit Theorem states that for sample sizes as small as  $n = 30$ , the sampling distribution of  $\bar{x}$  will be approximately normal. But regardless of the shape of the population from which we are sampling, the sampling distribution of the sample mean will have a mean  $\mu$  equal to the mean of the population from which we are sampling, and a standard deviation equal to  $\sigma/\sqrt{n}$ .

4. The sample size,  $n = 8$ , is too small to assume that the distribution of  $\bar{x}$  is approximately normal. The mean and standard error are  $\mu = 120$ ;  $\sigma/\sqrt{n} = 1/\sqrt{8} = .3536$ .
5. The sample size,  $n = 10$ , is too small to assume that the distribution of  $\bar{x}$  is approximately normal. The mean and standard error are  $\mu = 15$ ;  $\sigma/\sqrt{n} = \sqrt{4}/\sqrt{10} = .6325$ .
6. Since the sample size is large, the distribution of  $\bar{x}$  is approximately normal. The mean and standard error are  $\mu = 36$ ;  $\sigma/\sqrt{n} = \sqrt{6}/\sqrt{80} = .2739$ .

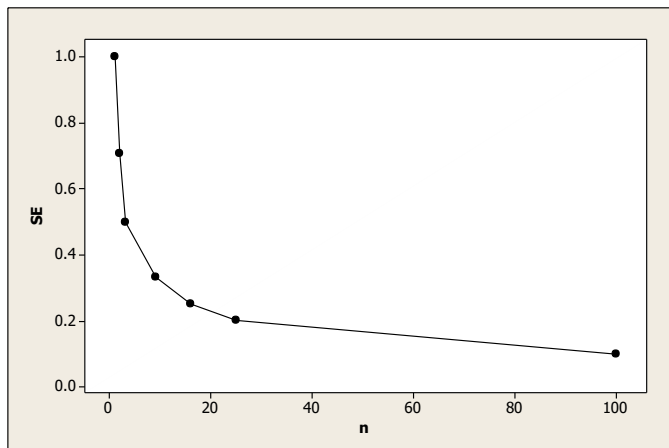
**7.3.7-13** For a population with  $\sigma = 1$ , the standard error of the mean is

$$\sigma/\sqrt{n} = 1/\sqrt{n}$$

The values of  $\sigma/\sqrt{n}$  for various values of  $n$  are tabulated in the following table.

Exercise Number	7	8	9	10	11	12	13
$n$	1	2	4	9	16	25	100
$SE(\bar{x}) = \sigma/\sqrt{n}$	1.00	.707	.500	.333	.250	.200	.100

**7.3.14** Refer to Exercises 7-13. The plot is shown next. Notice that the standard error *decreases* as the sample size *increases*.



**7.3.15** Since  $n = 49$ , the Central Limit Theorem is used. The sampling distribution of  $\bar{x}$  will be approximately normal. The mean of the sampling distribution of  $\bar{x}$  is  $\mu = 53$  and the standard deviation (or standard error) is  $\sigma/\sqrt{n} = 21/\sqrt{49} = 3$ .

**7.3.16** Since  $n = 40$ , the Central Limit Theorem is used. The sampling distribution of  $\bar{x}$  will be approximately normal. The mean of the sampling distribution of  $\bar{x}$  is  $\mu = 100$  and the standard deviation (or standard error) is  $\sigma/\sqrt{n} = 20/\sqrt{40} = 3.16$ .

**7.3.17** If the sample population is normal, the sampling distribution of  $\bar{x}$  will also be normal (regardless of the sample size) with mean  $\mu = 106$  and standard deviation (or *standard error*) given as

$$\sigma/\sqrt{n} = 12/\sqrt{25} = 2.4$$

**7.3.18** Refer to Exercise 15. The probability of interest is  $P(\bar{x} > 55)$ . When  $\bar{x} = 55$ ,  $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{55 - 53}{3} = .67$ .  
Then,  $P(\bar{x} > 55) = P(z > .67) = 1 - .7486 = .2514$ .

**7.3.19** Refer to Exercise 16. The probability of interest is  $P(105 < \bar{x} < 110)$ . When  $\bar{x} = 105$  and  $\bar{x} = 110$ ,

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{105 - 100}{3.16} = 1.58 \text{ and } z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{110 - 100}{3.16} = 3.16$$

Then,  $P(105 < \bar{x} < 110) = P(1.58 < z < 3.16) = .9992 - .9429 = .0563$ .

**7.3.20** Refer to Exercise 17.

**a** Calculate  $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{110 - 106}{2.4} = 1.67$ , so that

$$P(\bar{x} > 110) = P(z > 1.67) = 1 - .9525 = .0475$$

**b**  $P(102 < \bar{x} < 110) = P(-1.67 < z < 1.67) = .9525 - .0475 = .9050$

**7.3.21** Since the weights are normally distributed, the sampling distribution for the average weight for a sample of  $n = 16$  baby girls will also be normally distributed with mean  $\mu = 12.9$  and standard error  $\sigma/\sqrt{n} = 1.6/\sqrt{16} = 0.4$ .

**a** Calculate  $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{14 - 12.9}{0.4} = 2.75$ , so that  $P(\bar{x} > 14) = P(z > 2.75) = 1 - .9970 = .0030$ .

**b** Calculate  $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{11 - 12.9}{0.4} = -4.75$ , so that  $P(\bar{x} < 11) = P(z < -4.75) \approx 0$ .

**c** Since the probability in part b is close to zero, this would be a very unlikely event, assuming that the average weight of babies at this facility is 12.9 as it is in the general population. Perhaps babies in this area are underweight.

**7.3.22** Assume that the CBF in the population of stroke patients is normally distributed with mean  $\mu = 74$  and standard deviation  $\sigma = 16$ . For a random sample of size  $n = 16$ ,

$$P(\bar{x} \leq 69.7) = P(z \leq \frac{69.7 - 74}{16/\sqrt{25}}) = P(z \leq -1.34) = .0901$$

**7.3.23** Since  $n = 400$ , the Central Limit Theorem ensures that the sampling distribution of  $\bar{x}$  will be approximately normal with mean  $\mu = 1110$  and standard deviation  $\sigma/\sqrt{n} = 80/\sqrt{400} = 4$ .

**a**  $P(1100 < \bar{x} < 1110) = P(\frac{1100 - 1110}{4} < z < \frac{1110 - 1110}{4}) = P(-2.5 < z < 0) = .5000 - .0062 = .4938$ .

**b**  $P(\bar{x} > 1120) = P(z > \frac{1120 - 1110}{4}) = P(z > 2.5) = 1 - .9938 = .0062$ .

**c**  $P(\bar{x} < 900) = P(z < \frac{900 - 1110}{4}) = P(z < -52.5) = .0000$ .

**7.3.24** Since the package weights are normal with mean 16 and standard deviation 0.6, the total weight for a box of 24 packages will also have a normal distribution with mean  $n\mu = 24(16) = 384$  and standard deviation  $\sigma\sqrt{n} = 0.6\sqrt{24} = 2.939$ . Let  $T$  be the total weight for the box. Then

$$\begin{aligned} P(T > 392) &= P(z > \frac{392 - 384}{2.939}) \\ &= P(z > 2.72) = 1 - .9967 = .0033 \end{aligned}$$

**7.3.25 a** Since the sample size is large, the sampling distribution of  $\bar{x}$  will be approximately normal with mean  $\mu = 75,878$  and standard deviation  $\sigma/\sqrt{n} = 4000/\sqrt{60} = 516.3978$ .

**b** From the Empirical Rule (and the general properties of the normal distribution), approximately 95% of the measurements will lie within 2 standard deviations of the mean:

$$\begin{aligned} \mu \pm 2SE &\Rightarrow 75,878 \pm 2(516.3978) \\ &75,878 \pm 1032.80 \text{ or } 74,845.20 \text{ to } 76,910.80 \end{aligned}$$

**c** Use the mean and standard deviation for the distribution of  $\bar{x}$  given in part a.

$$\begin{aligned} P(\bar{x} > 78,000) &= P\left(z > \frac{78,000 - 75,878}{516.3978}\right) \\ &= P(z > 4.11) \approx 0 \end{aligned}$$

**d** Refer to part **c**. You have observed a very unlikely occurrence (one that almost never happens), assuming that  $\mu = 75,878$ . Perhaps your sample was not a random sample, or perhaps the average salary of \$75,878 is no longer correct.

**7.3.26 a** Age of equipment, technician error, technician fatigue, equipment failure, difference in chemical purity, contamination from outside sources, and so on.

**b** The variability in the average measurement is measured by the standard error,  $\sigma/\sqrt{n}$ . In order to decrease this variability, you should increase the sample size  $n$ .

**7.3.27** The weight of a package of 12 tomatoes is the sum of the 12 individual tomato weights. Hence, since weights and heights are in general normally distributed, so will be the sum of the 12 weights, according to the Central Limit Theorem.

**7.3.28** The number of bacteria in one cubic foot of water can be thought of as the sum of 1728 random variables, each of which is the number of bacteria in a particular cubic inch of water. Hence, the Central Limit Theorem insures the approximate normality of the sum.

**7.3.29 a** Since the original population is normally distributed, the sample mean  $\bar{x}$  is also normally distributed (for any sample size) with mean  $\mu$  and standard deviation

$$\sigma/\sqrt{n} \approx 2/\sqrt{10}$$

**b** If  $\mu = 21$  and  $n = 10$ , the  $z$ -value corresponding to  $\bar{x} = 20$  is

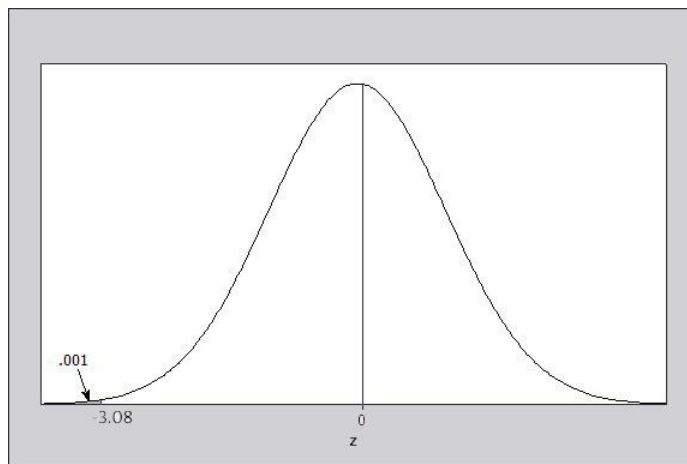
$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{20 - 21}{2/\sqrt{10}} = -1.58 \quad \text{and} \quad P(\bar{x} < 20) = P(z < -1.58) = .0571$$

**c** Refer to part **b** and assume that  $\mu$  is unknown. It is necessary to find a value  $c$  so that  $P(\bar{x} < 20) = .001$ . Recall that if  $\mu = c$ , the  $z$ -value corresponding to  $\bar{x} = 20$  is

$$z = \frac{\bar{x} - c}{\sigma/\sqrt{n}} = \frac{20 - c}{2/\sqrt{10}} = \frac{20 - c}{.632}.$$

It is necessary then to have  $P(\bar{x} < 20) = P\left(z < \frac{20 - c}{.632}\right) = A\left(\frac{20 - c}{.632}\right) = .001$ .

Refer to the figure that follows.



From Table 3, the value of  $z$  that cuts off .001 in the left-hand tail of the normal distribution is  $z = -3.08$ . Hence, the two values,  $(20 - c)/.632$  and  $z = -3.08$ , must be the same. Solving for  $c$ ,

$$\frac{20-c}{.632} = -3.08 \quad \text{or} \quad c = 21.948.$$

- 7.3.30 a** The random variable  $T = \sum x_i$ , where  $x_i$  is normally distributed with mean  $\mu = 422$  and standard deviation  $\sigma = 13$  for  $i = 1, 2, 3$ . The Central Limit Theorem states that  $T$  is normally distributed with mean  $n\mu = 3(422) = 1266$  and standard deviation  $\sigma\sqrt{n} = 13\sqrt{3} = 22.5167$ .

**b** Calculate  $z = \frac{T - 1266}{22.5167} = \frac{1300 - 1266}{22.5167} = 1.51$ . Then

$$P(T > 1300) = P(z > 1.51) = 1 - .9345 = .0655.$$

- 7.3.31 a** Since the total daily sales is the sum of the sales made by a fixed number of customers on a given day, it is a sum of random variables, which, according to the Central Limit Theorem, will have an approximate normal distribution.

**b** Let  $x_i$  be the total sales for a single customer, with  $i = 1, 2, \dots, 30$ . Then  $x_i$  has a probability distribution with  $\mu = 8.50$  and  $\sigma = 2.5$ . The total daily sales can now be written as  $x = \sum x_i$ . If  $n = 30$ , the mean and standard deviation of the sampling distribution of  $x$  are given as

$$n\mu = 30(8.5) = 255 \quad \text{and} \quad \sigma\sqrt{n} = 2.5\sqrt{30} = 13.693$$

- 7.3.32 a** Since the original population is normally distributed, the sample mean  $\bar{x}$  is also normally distributed (for any sample size) with mean  $\mu$  and standard deviation  $\sigma/\sqrt{n} = 0.8/\sqrt{130} = .07016$ .

The  $z$ -value corresponding to  $\bar{x} = 98.25$  is

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{98.25 - 98.6}{0.8/\sqrt{130}} = -4.99$$

and  $P(\bar{x} < 98.25) = P(z < -4.99) \approx 0$ .

**b** Since the probability is extremely small, the average temperature of 98.25 degrees is very unlikely.

- 7.3.33** The sampled population has a mean of 5.97 with a standard deviation of 1.95.

**a** With  $n = 31$ , calculate  $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{6.5 - 5.97}{1.95/\sqrt{31}} = 1.51$ , so that  $P(\bar{x} \leq 6.5) = P(z \leq 1.51) = .9345$ .

**b** Calculate  $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{9.80 - 5.97}{1.95/\sqrt{31}} = 10.94$ , so that  $P(\bar{x} \geq 9.80) = P(z \geq 10.94) \approx 1 - 1 = 0$ .

**c** The probability of observing an average diameter of 9.80 or higher is extremely unlikely, if indeed the average diameter in the population of injured tendons was no different from that of healthy tendons (5.97). We would conclude that the average diameter in the population of patients with injured tendons is higher than 5.97.

## Section 7.5

**7.5.1**  $p = .3$ ;  $SE(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{.3(.7)}{100}} = .0458$

**7.5.2**  $p = .1$ ;  $SE(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{.1(.9)}{400}} = .015$

**7.5.3**  $p = .6; SE(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{.6(.4)}{250}} = .0310$

**7.5.4** For this binomial distribution, calculate  $np = 2.5$  and  $nq = 47.5$ . Since  $np$  is less than 5, the normal approximation is not appropriate.

**7.5.5** For this binomial distribution, calculate  $np = 30$  and  $nq = 45$ . Since both values are greater than 5, the normal approximation is appropriate.

**7.5.6** For this binomial distribution, calculate  $np = 7.5$  and  $nq = 67.5$ . Since both values are greater than 5, the normal approximation is appropriate.

**7.5.7** For this binomial distribution, calculate  $np = 50$  and  $nq = 450$ . Since both values are greater than 5, the normal approximation is appropriate.

**7.5.8** For this binomial distribution, calculate  $np = 247.5$  and  $nq = 2.5$ . Since  $nq$  is less than 5, the normal approximation is not appropriate.

**7.5.9** Since  $\hat{p}$  is approximately normal, with standard deviation  $SE(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{.4(.6)}{75}} = .0566$ , the probability of interest is  $P(\hat{p} \leq .43) = P(z \leq \frac{.43 - .4}{.0566}) = P(z \leq .53) = .7019$ .

**7.5.10** Since  $\hat{p}$  is approximately normal, with standard deviation  $SE(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{.4(.6)}{75}} = .0566$ , the probability of interest is  $P(\hat{p} > .38) = P(z > \frac{.38 - .4}{.0566}) = P(z > -.35) = 1 - .3632 = .6368$ .

**7.5.11** Since  $\hat{p}$  is approximately normal, with standard deviation  $SE(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{.4(.6)}{75}} = .0566$ , the probability of interest is approximated as

$$\begin{aligned} P(.35 \leq \hat{p} \leq .43) &= P\left[\frac{.35 - .4}{.0566} \leq z \leq \frac{.43 - .4}{.0566}\right] \\ &= P(-.88 \leq z \leq .53) = .7019 - .1894 = .5125 \end{aligned}$$

**7.5.12** Since  $\hat{p}$  is approximately normal, with standard deviation  $SE(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{.4(.6)}{75}} = .0566$ , the probability of interest is  $P(\hat{p} < .30) = P(z < \frac{.30 - .4}{.0566}) = P(z < -1.77) = .0384$ .

**7.5.13-15** For  $n = 500$  and  $p = .1$ ,  $np = 50$  and  $nq = 450$  are both greater than 5. Therefore, the normal approximation will be appropriate.

**13.**  $P(\hat{p} > .12) = P\left(z > \frac{.12 - .1}{.0134}\right) = P(z > 1.49) = 1 - .9319 = .0681$

**14.**  $P(\hat{p} < .10) = P\left(z < \frac{.10 - .1}{.0134}\right) = P(z < 0) = .5$

**15.**  $P(-.02 \leq (\hat{p} - p) \leq .02) = P(-1.49 \leq z \leq 1.49) = .9319 - .0681 = .8638$

**7.5.16-22** The values  $SE = \sqrt{pq/n}$  for  $n = 100$  and various values of  $p$  are tabulated next.